The decentralization of information processing in the presence of interactions

Dimitri Vayanos

This is an electronic version of an Article published in the Review of economic studies, 70 (3). pp. 667-695 © 2003 Blackwell Publishing.
http://www.blackwellpublishing.com/journals/ROES

The Decentralization of Information Processing in the Presence of Interactions

Dimitri Vayanos *

February 4, 2003

Abstract

We propose a model of organizational decision making, in which information processing is decentralized. Our model incorporates two features of many actual organizations: aggregation entails a loss of useful information, and the decision problems of different agents interact. We assume that an organization forms a portfolio of risky assets, following a hierarchical procedure. Agents' decision rules and the organization's hierarchical structure are derived endogenously. Typically, in the optimal hierarchical structure, all agents have one subordinate, and returns to ability are at least as high at the bottom as at the top. However, these results can be reversed in the presence of returns to specialization.

*MIT and NBER. Address: MIT Sloan School of Management, 50 Memorial Drive E52-437, Cambridge MA 02142-1347, tel: 617-2532956, e-mail: dimitriv@mit.edu. I thank Patrick Bolton, Jacques Crémer, James Dow (the editor), Lu Hong, Bruno Julien, Jim Malcomson, David Martimort, Roy Radner, Stefan Reichelstein, Patrick Rey, Sherwin Rosen, José Scheinkman, Jean Tirole, Tim Van Zandt, two anonymous referees, seminar participants at Berkeley, Chicago, LSE, MIT, Stanford, Toulouse, and participants at the CEPR Information Processing Organizations, Econometric Society, Gerzensee Economic Theory, and Northwestern Microeconomic Theory conferences, for very useful comments. I am especially grateful to Denis Gromb for his detailed and very insightful comments on multiple versions of this paper. I thank Hakan Orbay for excellent research assistance.

1 Introduction

Organizational decisions are usually based on large quantities of information. Such information cannot be processed by a single agent. Therefore, information processing has to be decentralized among many agents. Consider, for example, the decision to expand a production plant. This decision depends on information about the plant's cost, known mainly by the plant manager. It also depends on information about the demand for the plant's output, known by marketing managers. Finally, it can depend on broader information about the organization's strategy, known by top-level managers.

This paper proposes a model of organizational decision making, in which information processing is decentralized, and information is communicated along hierarchical lines. The model has two novel elements, relative to previous literature. First, aggregation entails a loss of useful information, in the sense that when agents summarize their information to their hierarchical superiors, information which is useful to the superiors is lost. Second, agents' decision problems interact, in the sense that an agent's optimal decision should depend on information held by agents in other parts of the organization. We use our model to examine issues of organization design. In particular, we determine what hierarchical structure an organization should adopt, and in what hierarchical level the returns to employing agents who are better able to process information are the highest.

Our model is motivated from how investment firms form their portfolios. Because of the large number of securities involved, investment firms rarely perform a full-scale portfolio optimization. Rather, they follow a multi-stage hierarchical procedure. An example of this procedure, for the case of three stages, is described in Sharpe (1985, p.657):

> In the first stage (security selection), combinations of securities in each of the several stock groups and in each of the several bond groups are selected. The second stage (group selection) involves the determination of an appropriate combination of the stock group portfolios and an appropriate combination of the bond group portfolios. The final stage is devoted to asset allocation, using the bond and stock portfolios as asset-class portfolios. In every stage but the last, decisions are made myopically, considering only a subset of the available securities. In every stage but the first, groups of securities are "locked together" in fixed proportions determined in prior stages.

In this hierarchical procedure, agents' decision problems obviously interact. Consider, for example, an analyst forming a portfolio of stocks within a given group. The analyst's optimal portfolio should depend on information about all securities, including those outside the group. If, for example, the securities outside the group have high systematic risk, the analyst should favor those securities within the group that have low systematic risk. Analysts seem, however, to form their portfolios myopically, without using information on the securities outside their groups, presumably because they have little such information.

In addition to interactions, the hierarchical procedure involves aggregation loss. A stock analyst, for example, summarizes his information on the stocks within his group, through his choice of a group portfolio. This summary, however, entails a loss of useful information for the manager in charge of the overall stock portfolio. Indeed, if the manager knew the analyst's detailed information, he could add it to his own information on the other stock groups, and improve on the analyst's portfolio. It is worth noting that asset-class managers seem not to change the composition of analysts' portfolios, presumably because they lack the detailed information to do so.

Our model is as follows. An organization can invest in one riskless and multiple risky assets. Investing in a risky asset involves both systematic and idiosyncratic risk. Systematic risk is represented by one aggregate factor (e.g., the business cycle). Assets differ only in how sensitive their returns are to the factor, and we refer to these return sensitivities as factor loadings.

The organization's portfolio formation process is subject to three constraints. The *communication* constraint is that the organization must have a hierarchical structure, and communication must take place along hierarchical lines, as follows. An agent at the bottom of the hierarchy examines some assets, and observes their factor loadings. He then forms a portfolio of these assets, and communicates the portfolio's factor loading to his (direct) superior. The superior forms a portfolio of the assets included in his subordinates' portfolios, and of any additional assets he examines directly. He then communicates this portfolio's factor loading to his own superior, and so on. The *scaling* constraint is that agents cannot change the composition of their subordinates' portfolios, but can only scale portfolios up or down, i.e., multiply the investment in each asset by the same scalar. The *processing* constraint is that agents can form a portfolio of at most $K < \infty$ inputs, where an input can be either an asset examined directly, or a subordinate's portfolio. The processing constraint captures agents' limitations in processing information.

The organization is designed optimally, subject to the three constraints above. The design of the organization takes place ex-ante, before factor loadings are realized. There are three design parameters: the hierarchical structure, the assets each agent examines, and agents' decision rules, i.e., the way agents map their information into portfolio weights. The optimal set of parameters must implement a decision rule for the organization that maximizes the expected utility of asset payoffs. Intuitively, this decision rule must select a portfolio that is the closest possible to the first-best portfolio, selected in the absence of agents' information processing limitations.

While our model is motivated from portfolio formation in investment firms, it is also applicable to other organizational settings. A direct application is to risk management, by both financial and non-financial firms. Indeed, risk management can be viewed as a hierarchical portfolio formation procedure, where each unit in a firm determines its portfolio of risky activities, and then risk managers control the overall level of risk.[1] A more indirect application is to capital budgeting, i.e., firms' choice of physical investments. Consider, for example, the decision to build a production plant, and suppose that the plant's design (which can be viewed as a "portfolio" of attributes) should depend significantly on the specific mix of products that will be manufactured in the plant. A design proposed to headquarters by the organization's manufacturing department might then be suboptimal, because it might be missing important marketing information. Such interactions between manufacturing and marketing, or between different manufacturing departments, seem quite common in practice.[2]

To solve the organization design problem, we must make an assumption on the probability distribution of factor loadings as of the design stage. We first consider the simple case where factor loadings are i.i.d. across assets, with mean zero. Agents' optimal decision rules are then myopic as described in Sharpe (1985), i.e., agents form their portfolios ignoring interactions with assets outside the portfolios. Under these decision rules, the organization's investment in a particular asset differs from the first-best investment, as long as the agent at the top of the hierarchy does not examine the asset directly. This is because the top agent needs to adjust the investment in the asset to take into account interactions ignored

[1] For a general presentation of risk management and value at risk, see, for example, Litterman (1996), Alexander (1998), and Jorion (2000). Ch.9 in Alexander and ch.16 in Jorion emphasize, in particular, that while risk managers set limits for the level of risk that each unit in a firm can take, they do not dictate the portfolio of a unit's risky activities. For the notion that the portfolios formed by different units in a firm may fail to be globally optimal, see also Naik and Yadav (2001). These authors find that dealers in the London Stock Exchange control their inventories without fully taking into account the covariance with the inventories of other dealers in the same firm.

[2] Bower (1970) presents four detailed case studies of capital budgeting in a large corporation. Closest to our discussion is the case on Specialty Plastics, where a proposal for a new plant was formulated by engineers in one manufacturing department. The proposal was not very successful, and one reason was that although the engineers had consulted with divisional headquarters, they had not consulted sufficiently with their marketing department, or with other manufacturing departments that could benefit from using the plant.

by his subordinates. Since he has only aggregated information, however, his adjustment is imperfect.

Given the optimal decision rules, we can evaluate the performance of different hierarchical structures. Quite surprisingly, in the optimal hierarchical structure, all agents have one subordinate. The intuition is that for independent factor loadings, the average factor loading of even a small set of assets differs significantly from the factor loading of each asset in the set. Therefore, aggregation results in a loss of useful information, even when it concerns only a few assets. As a result, it is optimal for the top agent to examine directly as many assets as possible.

In the optimal hierarchical structure, all agents work at full capacity, handling exactly K assets or portfolios. Only one agent may work below capacity, due to integer constraints. Interestingly, this agent can be at any hierarchical level. Interpreting this agent as a low ability agent, who can handle fewer than K assets or portfolios, our result implies that returns to ability are independent of the hierarchical level. Intuitively, the benefit of the top agent working at full capacity is that he can process more disaggregated information, while the corresponding benefit for the bottom agent is that he can take more interactions into account. For independent factor loadings, these turn out to be equal.

When factor loadings are not i.i.d., the organization design problem becomes more complicated, and we solve it only in some special cases. We first assume that factor loadings are the sum of a component common to all assets, and an i.i.d. component. We show that the one-subordinate result of the i.i.d. case still holds, but returns to ability are highest at the bottom of the hierarchy. We next consider an example where assets are partitioned into groups, and factor loadings are the sum of a group and an i.i.d. component. The one-subordinate result holds again, which is perhaps more surprising than in the i.i.d. and common component cases. Indeed, one would expect the loss from information aggregation to decrease if the top agent has multiple subordinates, each examining assets in one group.

One feature of actual organizations which is not captured in our model, and which might explain why agents have multiple subordinates, is that there are returns to specialization. In investment firms, for example, it is efficient to assign all stocks within an industry sector to a single analyst, so that the analyst can develop expertise on that sector. To capture returns to specialization, we consider our group component example, with the modification that agents observe factor loadings imperfectly. This ensures that knowing an asset's factor loading is useful when observing factor loadings of other assets in the same group. We then show that it is optimal for the top agent to have multiple subordinates, and that returns to ability are highest at the top of the hierarchy. These results are, of course, derived in the context of an example, but they suggest an interesting direction to extend the model.

This paper belongs to a large literature that studies organizations with boundedly rational agents, and abstracts from incentive issues.[3] Crémer (1980), Aoki (1986), and Geanakoplos and Milgrom (1991), study resource allocation in organizations, using the team-theoretic approach of Marschak and Radner (1972). We also use this approach in the derivation of agents' optimal decision rules. The main difference with these papers is that we consider hierarchical communication, where an agent's information comes from his subordinates.

Radner (1993) considers a model with hierarchical communication. He assumes that an organization performs an associative operation involving many items. An agent needs one unit of time to perform the operation on two items. Decentralizing the operation to a hierarchy of agents, where agents communicate partial results to their superiors, is valuable because it reduces the time it takes to process all the items. Van Zandt (1999b) extends Radner's "batch processing" framework to "real-time processing", where items arrive in each period, and the organization has to select which items to process.[4] Beckmann (1960, 1983) and Keren and Levhari (1979, 1983) are precursor papers that restrict attention to balanced hierarchies, where all agents at a given level have the same number of subordinates. Bolton and Dewatripont (1994) assume that decentralization is valuable not because it reduces delay, but because it allows agents to specialize by processing the same type of items more frequently. These papers, however, consider associative operations, where there is no aggregation loss and no interactions.[5]

Garicano (2000) and Beggs (2001) consider hierarchical communication in organizations that handle heterogeneous tasks. Tasks that cannot be handled by agents at a given level, are sent one level up, where agents can either handle a larger set of tasks (Beggs), or specialize in handling less frequent tasks (Garicano). In these papers, however, tasks can be handled independently, and thus there are no interactions.

Harris and Raviv (2001) assume interactions between activities, and examine how these determine the structure (matrix, functional, or divisional) that an organization should

[3]For a survey of this literature, see Van Zandt (1999a).

[4]The batch and real-time processing models have been applied to a number of organizational issues such as returns to scale (Radner and Van Zandt (1992), Van Zandt and Radner (2001)), resource allocation (Van Zandt (2000a, 2000b)), returns to ability (Prat (1997)), and internal structure (Orbay (2002)).

[5]Consider, for example, the selection of the best project out of a pool, i.e., the maximum operation. There is no aggregation loss because the best project out of a subset is a "sufficient statistic" for all the projects in the subset. There are no interactions because the best project out of a subset does not depend on the quality of the projects outside the subset.

adopt. Hart and Moore (2000) also emphasize interactions in their theory of allocation of decision rights within a firm. These papers, however, adopt a reduced-form approach in modelling interactions.

The rest of this paper is organized as follows. Section 2 presents the model. Section 3 considers the case where factor loadings are i.i.d., and Section 4 considers the general case. Section 5 concludes, and all proofs are in two Appendices.

2 The Model

We consider an organization that forms an asset portfolio, over three periods, 0, 1, and 2. In period 0, the organization is designed. In period 1, the assets' factor loadings are randomly drawn, and the organization forms its portfolio in a way that depends on its design. Finally, in period 2, the assets pay off. We first describe the assets in which the organization can invest, and then the organization designer's objective, constraints, and choice variables.

2.1 Assets

There is one riskless and N risky assets. The return on the riskless asset is zero. The returns on the risky assets follow a simple factor structure. Asset n, $n = 1, .., N$, returns

$$r_n = \mu_n + \lambda_n \eta + \epsilon_n, \tag{1}$$

where μ_n is a constant, η an aggregate factor (e.g., the business cycle), λ_n asset n's factor loading (the sensitivity to the factor), and ϵ_n asset n's idiosyncratic risk. The idiosyncratic risk is independent across assets, and is independent of the aggregate factor. Both the aggregate factor and the idiosyncratic risk have mean zero, and thus μ_n is asset n's expected return.[6]

In addition to assuming a simple factor structure, we assume the following. First, η and $\{\epsilon_n\}_{n=1,..,N}$ are normal, and thus portfolio choice can be reduced to a mean-variance problem. Second, all assets have the same expected return and idiosyncratic variance, i.e., $\mu_n = \mu$ and $E(\epsilon_n^2) = \sigma^2$ for all n. Assets thus differ only in their factor loadings. Finally, $E(\eta^2) = \sigma^2$, which is without loss of generality, since we can redefine the factor loadings.

[6] The assumption of only one aggregate factor is for notational simplicity. The analysis can easily be generalized to multiple factors.

2.2 Organization Designer's Objective

We assume that in period 0, the organization designer knows μ and σ^2. (He does not know, however, the factor loadings $\{\lambda_n\}_{n=1,..,N}$, which are randomly drawn according to some probability distribution in period 1.) The organization designer maximizes the expected utility of the organization's period 2 asset payoffs where, for simplicity, utility is exponential with CARA a. Since the organization designer evaluates expected utility in period 0, the expectation is computed w.r.t. η, $\{\epsilon_n\}_{n=1,..,N}$, and $\{\lambda_n\}_{n=1,..,N}$.

Asset payoffs are

$$\sum_{n=1}^{N} x_n(1+r_n) + (W - \sum_{n=1}^{N} x_n),$$

where x_n denotes the organization's investment in risky asset n, and W the organization's initial wealth. Setting, for simplicity, $W = 0$, and using equation (1), we can write the expected utility of asset payoffs as

$$\begin{aligned} & -E\exp\left[-a\sum_{n=1}^{N} x_n r_n\right] \\ = \quad & -E\exp\left[-a\left(\sum_{n=1}^{N} x_n\mu + \left(\sum_{n=1}^{N} x_n\lambda_n\right)\eta + \sum_{n=1}^{N} x_n\epsilon_n\right)\right]. \end{aligned}$$

Taking expectations w.r.t. η and $\{\epsilon_n\}_{n=1,..,N}$, which are independent and normal, we can write expected utility as

$$-E\exp\left[-a\left(\sum_{n=1}^{N} x_n\mu - \frac{a}{2}\left(\left(\sum_{n=1}^{N} x_n\lambda_n\right)^2\sigma^2 + \sum_{n=1}^{N} x_n^2\sigma^2\right)\right)\right], \tag{2}$$

where the expectation is now w.r.t. $\{\lambda_n\}_{n=1,..,N}$ only. Equation (2) represents the organization designer's objective. This objective takes an intuitive form. For a given set of factor loadings, the organization designer takes into consideration the mean and the variance of the organization's portfolio. The mean is equal to $\sum_{n=1}^{N} x_n\mu$, while the variance is the sum of two terms. First, a term which corresponds to aggregate risk, and depends on the factor loading of the organization's portfolio ($\sum_{n=1}^{N} x_n\lambda_n$), and second, a term which corresponds to idiosyncratic risk.

2.3 Organization Designer's Constraints

The organization designer is subject to three constraints.

Communication Constraint: We assume that (i) the organization must have a hierarchical structure, and (ii) communication must take place along hierarchical lines, from the bottom to the top of the hierarchy. We represent the agents in a hierarchy by sequences of positive integers. The agent at the top corresponds to the null sequence, "·". His first subordinate, starting from the left, corresponds to the sequence 1, his second subordinate to the sequence 2, and so on. Similarly, the first subordinate of the first subordinate corresponds to the sequence 1,1, the second subordinate to the sequence 1,2, and so on. We denote a sequence by j, and refer to the corresponding agent as agent j. We denote by J the set of all agents. Finally, we denote by $S(j)$ the number of (direct) subordinates of agent j. Figure 1 illustrates the notation for a simple hierarchy.

Each agent in the hierarchy may examine some risky assets, i.e., observe their factor loadings. We denote by $M(j)$ the number of assets that agent j examines, and by $A_M(j)$ their set. We assume that each asset is examined by only one agent, i.e., the sets $\{A_M(j)\}_{j \in J}$ form a partition of $\{1, .., N\}$. We refer to the assets examined by agent j, or by his direct or indirect subordinates, as the assets under j's control. We denote by $N(j)$ the number of these assets, and by $A_N(j)$ their set.

Communication takes place as follows. An agent at the bottom of the hierarchy observes the factor loadings of the assets he examines. He then forms a portfolio of these assets, and communicates the portfolio's factor loading to his superior. The superior observes the factor loadings of his subordinates' portfolios, and of the assets he (directly) examines. He then forms a portfolio of all the assets under his control, and communicates the portfolio's factor loading to his own superior. This process continues until the top of the hierarchy, and the portfolio formed by the top agent is also that of the organization. We denote by $x_n(j)$ the investment of agent j in an asset n under his control, and by

$$\lambda(j) \equiv \sum_{n \in A_N(j)} x_n(j) \lambda_n,$$

the factor loading of agent j's portfolio.

Scaling Constraint: We assume that agents cannot change the composition of their subordinates' portfolios, but can only scale portfolios up or down, i.e., multiply the investment in each asset by the same scalar. We denote by $y_i(j)$ the scaling factor that agent j applies to the portfolio of his ith subordinate, $i = 1, .., S(j)$. The investment of agent j in an asset $n \in A_N(j, i)$ is thus $x_n(j) = y_i(j) x_n(j, i)$.

Processing Constraint: We assume that agents can form a portfolio of at most $K < \infty$ inputs, where an input can be either an asset agents (directly) examine, or a subordinate's portfolio.[7] The processing constraint for agent j is

$$S(j) + M(j) \leq K. \tag{3}$$

The processing constraint is fundamental to the organization design problem, since in its absence (i.e., if $K = \infty$) the organization would trivially consist of only one agent. This constraint might be capturing agents' limitations in collecting information, i.e., observing assets' factor loadings, and learning portfolios' factor loadings from subordinates. Alternatively, it might be capturing agents' limitations in processing the information. Both types of limitations seem quite important in practice.[8] The communication and scaling constraints capture some features of the portfolio formation process in investment firms. Needless to say, it would be desirable to derive these features endogenously. In Section 4, we show that in some cases, the scaling constraint can indeed be endogenized. We also discuss generalizations of the communication constraint.

2.4 Organization Designer's Choice Variables

The organization designer can choose three aspects of the organization, all of which can influence the portfolio the organization forms in period 1.

Hierarchical Structure: The organization designer can choose any hierarchical structure satisfying the processing constraint (3).

Assignment of Assets to Agents: The organization designer can choose which assets each agent examines. Of course, this choice is trivial when the assets are identical ex-ante, i.e., as of period 0.

Agents' Decision Rules: The organization designer can choose the decision rules that agents use when forming their portfolios. A decision rule for agent j is a mapping from j's information to j's controls. Agent j's information is

$$\gamma(j) \equiv \left(\{\lambda_n\}_{n \in A_M(j)}, \{\lambda(j,i)\}_{i=1,..,S(j)} \right),$$

[7]Note that treating a subordinate's portfolio as a single asset is consistent with the communication and scaling constraints. Indeed, under the communication constraint, an agent's information on a subordinate's portfolio is only the factor loading, as in the case of a single asset. Moreover, under the scaling constraint, the agent's control on the portfolio is only the scaling factor, i.e., is one-dimensional as for a single asset.

[8]For example, in his presentation of risk management, Litterman (1996) states that: "The choice of how finely to disaggregate the data is a compromise between accuracy of risk management, which comes from disaggregation, and the clarity that comes from aggregation. In addition, lack of availability of data or computing resources may limit the degree to which positions can be analyzed on a disaggregated basis."

i.e., the factor loadings of the assets j examines, and the factor loadings of the portfolios of j's subordinates. Agent j's controls are

$$\left(\{x_n(j)\}_{n\in A_M(j)}, \{y_i(j)\}_{i=1,..,S(j)}\right),$$

i.e., investments in the assets j examines, and scaling factors for the portfolios of j's subordinates.

A given hierarchical structure, assignment of assets to agents, and decision rule for each agent, implement a decision rule for the organization, i.e., a mapping from factor loadings to investments by the organization. A decision rule for the organization determines, in turn, the organization designer's expected utility, from equation (2). The organization designer's optimization problem is to implement a decision rule for the organization that maximizes expected utility.

2.5 First-Best

To analyze the organization designer's problem, it is useful to consider the first-best decision rule. This rule selects the investments $\{x_n^*\}_{n=1,..,N}$ that maximize expected utility conditional on factor loadings, i.e., the investments that solve

$$\max_{\{x_n\}_{n=1,..,N}} \left[\sum_{n=1}^{N} x_n\mu - \frac{a}{2}\left(\left(\sum_{n=1}^{N} x_n\lambda_n\right)^2 \sigma^2 + \sum_{n=1}^{N} x_n^2\sigma^2\right)\right]. \tag{4}$$

Obviously, the first-best rule maximizes expected utility. Therefore, if it is implementable (through a choice of hierarchical structure, assignment of assets to agents, and agents' decision rules) it is the solution to the organization designer's problem.

To gain intuition on whether the first-best rule is implementable, we consider the simple case where factor loadings are small. We assume, in particular, that $\lambda_n = \lambda\ell_n$, where λ goes to zero and the probability distribution of $\{\ell_n\}_{n=1,..,N}$ is held constant. For small factor loadings, x_n^* becomes

$$x_n^* = \frac{\mu}{a\sigma^2}\left[1 - \lambda_n \sum_{n'=1}^{N} \lambda_{n'} + o(\lambda^2)\right], \tag{5}$$

where $o(\lambda^2)/\lambda^2$ goes to zero as λ goes to zero.[9] The first-best investment in asset n is equal to $\mu/a\sigma^2$, minus an adjustment for aggregate risk, obtained by multiplying the factor

[9] The first-order condition of the problem (4) is

$$x_n^* = \frac{\mu}{a\sigma^2} - \lambda_n\left(\sum_{n'=1}^{N} x_{n'}^*\lambda_{n'}\right).$$

Plugging $x_n^* = \mu/a\sigma^2 + o(1)$ in the RHS, we obtain equation (5).

loading of asset n times the sum of all assets' factor loadings. Intuitively, asset n is penalized if it is risky, and the penalty increases with the riskiness of the rest of the asset portfolio.

When factor loadings are deterministic, the first-best rule is easily implementable. The organization designer can, for example, instruct agents to choose investments equal to x_n^* for each asset n they examine, and scaling factors equal to 1. This implements the first-best rule, for any hierarchical structure and assignment of assets to agents.

When factor loadings are stochastic, the first-best rule is generally not implementable. The organization designer cannot, for example, instruct an agent to choose x_n^*, since x_n^* depends on the factor loadings of all assets, which are generally not all known by any single agent. For example, an agent at the bottom of the hierarchy, who examines asset n, knows λ_n, but generally not the factor loading of the rest of the asset portfolio. The agent's hierarchical superiors have better aggregate information on the portfolio's factor loading, but do not have the disaggregated information on λ_n.

When the first-best rule is not implementable, the organization designer's problem consists in implementing a second-best rule. This problem is potentially very complicated, since it involves optimization over a large, discrete set of hierarchical structures and assignments of assets to agents, and also over an infinite set of agents' decision rules. In Sections 3 and 4, we solve this problem in a number of special cases.

3 I.I.D. Factor Loadings

In this section, we solve the organization designer's problem in the special case where factor loadings are small, and i.i.d. with mean zero. The small factor loadings assumption ($\lambda_n = \lambda \ell_n$, where λ goes to zero and the probability distribution of $\{\ell_n\}_{n=1,..,N}$ is held constant) greatly simplifies the problem, while preserving many of the economic insights. The independence assumption also simplifies the problem, and provides a useful benchmark for the more general analysis in Section 4. Finally, the assumption that factor loadings have mean zero is only for notational simplicity and does not affect the results.

Since factor loadings are i.i.d., the assets are identical ex-ante. Therefore, the assignment of assets to agents does not matter, and the organization designer's only choice variables are the hierarchical structure and agents' decision rules. We first determine agents' optimal decision rules for a general hierarchical structure, and then determine the optimal hierarchical structure.

3.1 Agents' Decision Rules

To simplify the derivation of agents' optimal decision rules, we assume that these are smooth, in the sense of having a Taylor expansion of order two around $\lambda = 0$. For $\lambda = 0$, the first-best investment in each asset is $\mu/a\sigma^2$, from equation (5). This can be implemented by decision rules where agents invest $\mu/a\sigma^2$ in each asset they examine, and set scaling factors equal to 1.

Assumption 1 *There exists a selection of optimal decision rules, parametrized by λ, such that for each agent j,*

$$x_n(j) = \frac{\mu}{a\sigma^2}\left[1 + f_n(j) + o(\lambda^2)\right],$$

for $n \in A_M(j)$, and

$$y_i(j) = 1 + g_i(j) + o(\lambda^2),$$

for $i = 1, .., S(j)$, where $f_n(j)$ and $g_i(j)$ contain first- and second-degree terms in $\gamma(j)$ and λ.[10]

In Proposition 1, we determine $f_n(j)$ and $g_i(j)$. To state the proposition, we set

$$\Lambda(j) \equiv \sum_{n \in A_N(j)} \lambda_n.$$

Intuitively, $\Lambda(j)$ represents the factor loading of agent j's portfolio for small λ. Indeed, since for small λ all assets in the portfolio receive approximately the same investment $\mu/a\sigma^2$, the portfolio's factor loading is

$$\lambda(j) = \frac{\mu}{a\sigma^2}\sum_{n \in A_n(j)} \lambda_n + o(1) = \frac{\mu}{a\sigma^2}\Lambda(j) + o(1).$$

Proposition 1 *Suppose that factor loadings are i.i.d. with mean zero. Then, there exists a selection of optimal decision rules such that for each agent j, $x_n(j)$ and $y_i(j)$ are as in Assumption 1 with*

$$f_n(j) = -\lambda_n\Lambda(j) + o(\lambda^2),$$

and

$$g_i(j) = -\frac{\Lambda(j,i)}{N(j,i)}\left[\Lambda(j) - \Lambda(j,i)\right] + o(\lambda^2).$$

[10] We consider a *selection* of optimal decision rules, rather than *the* optimal rules, because optimal rules are not unique. The non-uniqueness is in the trivial sense that we can multiply an agent's investments by a scalar, and divide the scaling factor of the agent's superior by the same scalar.

To illustrate Proposition 1, we consider the hierarchical structure in Figure 1. Equation (5) implies that the first-best investment in asset 1 is

$$x_1^* = \frac{\mu}{a\sigma^2}\left[1 - \lambda_1(\lambda_1 + \lambda_2 + \lambda_3 + \lambda_4) + o(\lambda^2)\right].$$

Consider now the organization's investment in asset 1. The investment of agent 1,1, who is at the bottom of the hierarchy and examines asset 1, is

$$\begin{aligned} x_1(1,1) &= \frac{\mu}{a\sigma^2}\left[1 + f_1(1,1) + o(\lambda^2)\right] \\ &= \frac{\mu}{a\sigma^2}\left[1 - \lambda_1(\lambda_1 + \lambda_2) + o(\lambda^2)\right]. \end{aligned} \tag{6}$$

Intuitively, agent 1,1 attempts to replicate the first-best investment, but does not know the factor loadings λ_3 and λ_4. He replaces these with his best guess, which is zero since factor loadings are i.i.d. with mean zero. Agent 1 scales agent 1,1's investment by

$$\begin{aligned} y_1(1) &= 1 + g_1(1) + o(\lambda^2) \\ &= 1 - \frac{\lambda_1 + \lambda_2}{2}\lambda_3 + o(\lambda^2). \end{aligned} \tag{7}$$

Therefore, the investment of agent 1 is

$$\begin{aligned} x_1(1) &= y_1(1)x_1(1,1) \\ &= \frac{\mu}{a\sigma^2}\left[1 - \lambda_1(\lambda_1 + \lambda_2) - \frac{\lambda_1 + \lambda_2}{2}\lambda_3 + o(\lambda^2)\right]. \end{aligned}$$

Intuitively, agent 1 adjusts agent 1,1's investment to take into account the factor loading λ_3, that he knows but agent 1,1 does not. Unlike agent 1,1, however, agent 1 does not know λ_1, and replaces it by his best guess, which is $(\lambda_1 + \lambda_2)/2$. (Agent 1's only information on λ_1 is the factor loading of agent 1,1's portfolio, which for small λ is $\lambda_1 + \lambda_2$. Since factor loadings are i.i.d., agent 1's expectation of λ_1 conditional on $\lambda_1 + \lambda_2$ is $(\lambda_1 + \lambda_2)/2$.) Proceeding as for agent 1, we find that the investment of the top agent is

$$x_1 = \frac{\mu}{a\sigma^2}\left[1 - \lambda_1(\lambda_1 + \lambda_2) - \frac{\lambda_1 + \lambda_2}{2}\lambda_3 - \frac{\lambda_1 + \lambda_2 + \lambda_3}{3}\lambda_4 + o(\lambda^2)\right].$$

The top agent adjusts agent 1's investment to take into account λ_4. He has, however, to replace λ_1 by his best guess, which is $(\lambda_1 + \lambda_2 + \lambda_3)/3$.

The difference between the organization's investment in an asset n, x_n, and the first-best investment, x_n^*, reflects the organization's decision-making error. Defining the error, e_n, for asset n as the lowest order term in $(x_n - x_n^*)/(\frac{\mu}{a\sigma^2})$, we have in our example

$$e_1 = \left(\lambda_1 - \frac{\lambda_1 + \lambda_2}{2}\right)\lambda_3 + \left(\lambda_1 - \frac{\lambda_1 + \lambda_2 + \lambda_3}{3}\right)\lambda_4. \tag{8}$$

The error for asset 1 is the sum of two errors, one associated with agent 1 and one with the top agent. Each of these errors takes an intuitive form as the product of two terms, an *interaction term* and an *aggregation loss term.* The interaction term reflects the fact that an agent adjusts a subordinate's investment to take into account interactions with assets whose factor loadings the subordinate does not know. For example, the interaction term for agent 1 is λ_3, because this agent takes into account the interaction with asset 3. The aggregation loss term reflects the fact that when adjusting a subordinate's investment, an agent has only aggregate information, which is imperfect. For example, the aggregation loss term for agent 1 is

$$\lambda_1 - \frac{\lambda_1 + \lambda_2}{2}.$$

This is because when adjusting his subordinate's investment, agent 1 does not know the factor loading of asset 1, λ_1, and replaces it by the average factor loading of assets 1 and 2, $(\lambda_1 + \lambda_2)/2$.

It is worth emphasizing that Proposition 1 determines agents' optimal decision rules for *any* given hierarchical structure. Furthermore, these rules are myopic as described in Sharpe (1985), i.e., agents form their portfolios ignoring interactions with assets outside the portfolios. Using Proposition 1, we now proceed to evaluate the performance of different hierarchical structures, and determine the optimal structure.

3.2 Hierarchical Structure

An optimal hierarchical structure must maximize the organization designer's expected utility. For small factor loadings, expected utility takes a simple form.

Lemma 1 *A hierarchical stucture that maximizes expected utility for small λ, must minimize*

$$\sum_{n=1}^{N} E(e_n^2). \tag{9}$$

Equation (9) is a sum over assets of the expected squared error for each asset (where expectation is w.r.t. the factor loadings). This measures the distance between the organization's decision rule and the first-best decision rule. An optimal hierarchical structure must minimize this distance. We should emphasize that, unlike all other results in Section 3, this result does not require i.i.d. factor loadings.

In Lemma 2, we compute the sum of expected squared errors for a general hierarchical structure, and i.i.d. factor loadings.

Lemma 2 *Suppose that factor loadings are i.i.d. with mean zero, and agents follow the optimal decision rules of Proposition 1. Then,*

$$\sum_{n=1}^{N} E(e_n^2) = \sigma_\lambda^4 \sum_{j\in J} \sum_{i=1}^{S(j)} [N(j,i)-1][N(j)-N(j,i)], \tag{10}$$

where $\sigma_\lambda^2 \equiv E(\lambda_n^2)$.

The intuition behind equation (10) is as follows. Consider an agent j, and an asset n in the portfolio of j's ith subordinate (agent j,i). The component of the error for asset n, that is associated with agent j, is the product of an interaction and an aggregation loss term. The interaction term is the sum of the factor loadings of the $N(j)-N(j,i)$ assets that are under the control of j but not of j,i. Since factor loadings are i.i.d., the expected squared interaction term is

$$\sigma_\lambda^2[N(j)-N(j,i)].$$

The aggregation loss term is the factor loading of asset n minus the average factor loading of the $N(j,i)$ assets that are under the control of j,i. Since for i.i.d. random variables $\{\lambda_n\}_{n=1,..,N}$, we have

$$E\left(\lambda_1 - \frac{\sum_{n=1}^{N}\lambda_n}{N}\right)^2 = \sigma_\lambda^2 \frac{N-1}{N},$$

the expected squared aggregation loss term is

$$a[N(j,i)] \equiv \sigma_\lambda^2 \frac{N(j,i)-1}{N(j,i)}. \tag{11}$$

Multiplying the expected squared interaction and aggregation loss terms by the $N(j,i)$ assets in j,i's portfolio, and summing over i and j, we obtain equation (10).

In Proposition 2, we characterize optimal hierarchical structures.

Proposition 2 *Suppose that factor loadings are i.i.d. with mean zero. In any optimal hierarchical structure:*

i. agents must have one subordinate (except the bottom agent),

ii. agents must handle exactly K assets or portfolios, except at most one agent.

The first, and somewhat surprising, result of Proposition 2 is that in an optimal hierarchical structure, agents have only one subordinate. To explain the intuition, we consider the hierarchical structures H_1 and H_2 in Figure 2. Both hierarchical structures are for $N=4$ and $K=2$, and according to Proposition 2, H_1 is optimal.

The main advantage of H_1 over H_2 is that it selects a better investment in asset 4. Indeed, in both H_1 and H_2, it is the top agent who adjusts the investment in asset 4 to take into account the interaction with assets 1 and 2. In H_1, however, the top agent knows the exact factor loading of asset 4, while in H_2 he only knows the average factor loading of assets 3 and 4. Therefore, the aggregation loss term (associated with asset 4 and the top agent) is zero in H_1, but non-zero in H_2. Conversely, the disadvantage of H_1 over H_2 is that it selects worse investments in assets 1 and 2. Indeed, in both H_1 and H_2, it is the top agent who adjusts the investments in assets 1 and 2 to take into account the interaction with asset 4. In H_2, however, the top agent knows the average factor loading of assets 1 and 2, while in H_1 he only knows the average factor loading of assets 1, 2, and 3. Therefore, the aggregation loss term (associated with assets 1 or 2, and the top agent) is larger in H_1 than in H_2.

The reason why H_1 dominates H_2 is that for independent factor loadings, the benefit of knowing an exact factor loading instead of an average (as is the case for asset 4) exceeds the cost of knowing an average that is less precise (as is the case for assets 1 and 2). In other words, the average of two independent factor loadings is not much closer to one factor loading, than is the average of three or more, i.e., the aggregation loss term is not much larger for three than for two assets. Indeed, the expected squared aggregation loss term is $a(N) = [(N-1)/N]\sigma_\lambda^2$. The function $a(N)$ is maximum for $N = \infty$, and achieves half of its maximum value for $N = 2$.

In Table 1, we compute the expected squared errors for H_1 and H_2 (omitting σ_λ^4, for notational simplicity). Table 1 confirms that H_1 dominates H_2 because the better investment in asset 4 outweighs the worse investments in assets 1 and 2.

	Asset 1	Asset 2	Asset 3	Asset 4	Sum
H_1	7/6	7/6	2/3	0	3
H_2	1	1	1	1	4

Table 1: Expected squared errors for H_1 and H_2.

The second result of Proposition 2 is that in an optimal hierarchical structure, all agents except perhaps one work at full capacity, handling exactly K assets or portfolios. The intuition is simply that the decentralization of information processing, which arises from the processing constraint, reduces decision quality. Therefore, it is optimal to minimize decentralization, by making the processing constraint binding.

An agent working below capacity may exist only because of integer constraints. When, for example, $N = 4$ and $K = 3$, the optimal hierarchical structures (H_a and H_b, shown in Figure 3) consist of two agents, only one of whom works at full capacity. By contrast, when $N = 5$ and $K = 3$, both agents work at full capacity.

The existence of an agent working below capacity offers one way to get at the question of how returns to ability depend on the hierarchical level. Indeed, an agent working below capacity can be interpreted as a low ability agent, who can handle fewer than K assets or portfolios. Such an agent should be at the hierarchical level where returns to ability are the smallest.

Interestingly, Proposition 2 shows that an agent working below capacity can be at any hierarchical level. When, for example, $N = 4$ and $K = 3$, such an agent can be either at the bottom of the hierarchy (in H_a) or at the top (in H_b). This means that returns to ability are independent of the hierarchical level. In other words, the benefit of placing a high ability agent at the top of the hierarchy (the agent can process more disagreggated information) is equal to the benefit of placing the agent at the bottom (the agent can form a better portfolio, taking interactions with more assets into account).

4 Generalizations and Extensions

In this section, we study the organization designer's problem when factor loadings are not i.i.d., and we examine how the results of the i.i.d. case can generalize. We first characterize agents' optimal decision rules for a general hierarchical structure and probability distribution of factor loadings. We next solve the organization designer's problem for two specific probability distributions, and also in an extension of the model that incorporates returns to specialization. Finally, we discuss ways to relax the communication and scaling constraints.

4.1 Agents' Decision Rules

Before presenting the general characterization of agents' optimal decision rules, we show that these can be computed explicitly when the hierarchical structure and probability distribution of factor loadings satisfy a "sufficient statistic" condition. To state this condition, we denote by $H(j)$ agent j's subordinate hierarchy (consisting of j, and j's direct and indirect subordinates), and by $J(j)$ the set of agents in $H(j)$. We also set

$$\Lambda^c(j) \equiv \sum_{n' \in \{1,..,N\} \setminus A_N(j)} \lambda_{n'},$$

to represent the factor loading of the portfolio of assets outside $H(j)$, and

$$\Gamma(j) \equiv \left(\{\lambda_n\}_{n \in A_M(j)}, \{\Lambda(j,i)\}_{i=1,..,S(j)} \right),$$

to represent agent j's information, both for small λ. Finally, we denote by $I(j)$ the information set generated by $\Gamma(j)$, and by $\bar{I}(j)$ that generated by $\{\Gamma(j')\}_{j' \in J(j)}$.

Condition 1 *For each agent j,*

$$E\left[\Lambda^c(j) | \, \bar{I}(j)\right] = E\left[\Lambda^c(j) | \, I(j)\right]. \tag{12}$$

Condition 1 requires that the agents in j's subordinate hierarchy have no incremental information, relative to agent j, on the factor loading of the portfolio of assets outside the subordinate hierarchy. In other words, agent j knows aggregate information concerning the rest of organization, at least as well as his subordinates. Condition 1 is satisfied for independent factor loadings, since the factor loading of an asset outside a subordinate hierarchy is independent of the information of the agents in that hierarchy. Condition 1 is also satisfied for two-level hierarchies, since the only agent with a non-trivial subordinate hierarchy is the top agent. We will show that it is satisfied in other cases of interest as well.

Proposition 3 *Suppose that Condition 1 holds. Then, there exists a selection of optimal decision rules such that for each agent j, $x_n(j)$ and $y_i(j)$ are as in Assumption 1 with*

$$f_n(j) = -\lambda_n \Lambda(j) - \left[\lambda_n - \frac{\Lambda(j)}{N(j)}\right] E\left[\Lambda^c(j) | \, I(j)\right] + o(\lambda^2),$$

and

$$g_i(j) = -\frac{\Lambda(j,i)}{N(j,i)}\left[\Lambda(j) - \Lambda(j,i)\right] - \left[\frac{\Lambda(j,i)}{N(j,i)} - \frac{\Lambda(j)}{N(j)}\right] E\left[\Lambda^c(j) | \, I(j)\right] + o(\lambda^2).$$

To illustrate Proposition 3, we consider again the hierarchical structure in Figure 1. The investment of agent 1,1 in asset 1 is

$$x_1(1,1) = \frac{\mu}{a\sigma^2}\left[1 - \lambda_1(\lambda_1 + \lambda_2) - \left(\lambda_1 - \frac{\lambda_1 + \lambda_2}{2}\right) E\left[\lambda_3 + \lambda_4 | I(1,1)\right] + o(\lambda^2)\right].$$

Compared to the case where factor loadings are i.i.d. with mean zero (equation (6)), there is a new term, involving agent 1,1's conditional expectation of $\lambda_3 + \lambda_4$. This term reflects the fact that agent 1,1 has information on assets 3 and 4, and can adjust his investment to take into account the interaction with these assets. The adjustment consists of two components. The first is simply

$$-\lambda_1 E\left[\lambda_3 + \lambda_4 | I(1,1)\right],$$

agent 1,1's conditional expectation of the optimal, first-best adjustment, $-\lambda_1(\lambda_3+\lambda_4)$. The second is

$$\frac{\lambda_1+\lambda_2}{2}E\left[\lambda_3+\lambda_4|I(1,1)\right],$$

and can be interpreted as a scaling factor for agent 1,1's portfolio, since it is the same for asset 2 as for asset 1. Agent 1,1 chooses this scaling factor so that his portfolio contains no overall adjustment for the interaction with assets 3 and 4 (i.e., the adjustments in assets 1 and 2 sum to zero). Intuitively, agent 1,1's portfolio can be better adjusted for the interaction with assets 3 and 4 by that agent's superior, agent 1, whose information on $\lambda_3+\lambda_4$ is more accurate.

The scaling factor of agent 1 is

$$y_1(1)=1-\frac{\lambda_1+\lambda_2}{2}\lambda_3-\left(\frac{\lambda_1+\lambda_2}{2}-\frac{\lambda_1+\lambda_2+\lambda_3}{3}\right)E\left[\lambda_4|I(1)\right]+o(\lambda^2).$$

Compared to the i.i.d. case (equation (7)), there is again a new term, reflecting an adjustment for the interaction with asset 4.

The error for asset 1 is equal to that in the i.i.d. case (equation (8)), plus the new terms for agents 1,1 and 1. Rearranging these terms, we can write the error as

$$\begin{aligned} e_1 &= \left(\lambda_1-\frac{\lambda_1+\lambda_2}{2}\right)\left[\lambda_3+E\left[\lambda_4|I(1)\right]-E\left[\lambda_3+\lambda_4|I(1,1)\right]\right] \\ &\quad+\left(\lambda_1-\frac{\lambda_1+\lambda_2+\lambda_3}{3}\right)\left[\lambda_4-E\left[\lambda_4|I(1)\right]\right]. \end{aligned}$$

This equation is similar to that in the i.i.d. case, and it generalizes nicely the analysis of that case. As in the i.i.d. case, the error for asset 1 is the sum of errors associated to agent 1 and the top agent, and each error is the product of an interaction and an aggregation loss term. The aggregation loss term is exactly as in the i.i.d. case. The interaction term takes a more general form, and reflects the incremental information that an agent has relative to a subordinate, on the assets that are not under the subordinate's control. For example, the interaction term for agent 1 is

$$\lambda_3+E\left[\lambda_4|I(1)\right]-E\left[\lambda_3+\lambda_4|I(1,1)\right],$$

and reflects the incremental information agent 1 has relative to agent 1,1, on assets 3 and 4.

It is worth emphasizing how Condition 1 simplifies the analysis. In our example, Condition 1 requires that agent 1,1 has no incremental information on asset 4, relative to agent 1. This ensures that agent 1,1 performs no overall adjustment of his portfolio for the interaction with asset 4, but leaves this to agent 1. If, by contrast, agent 1,1 had incremental information, he would perform part of the adjustment. In that case, however, each agent would have to base his adjustment on that of the other agent. This would involve a complicated analysis of what an agent knows, what he knows the other knows, etc.

While Condition 1 is useful for simplifying the optimal decision rules, it is not needed for characterizing them. In Theorem 1 (stated in the Appendix) we show that optimal decision rules can be characterized through a set of first-order conditions for each agent. The first-order conditions for agent j are

$$E[e_n|I(j)] = 0, \tag{13}$$

for $n \in A_M(j)$, and

$$E\left[\sum_{n' \in A_N(j,i)} e_{n'} \middle| I(j)\right] = 0, \tag{14}$$

for $i = 1, .., S(j)$, where e_n is the error for asset n. The intuition is that agent j must not be able to reduce the sum of expected squared errors, by changing the investment in an asset he examines, or the scaling factor of a subordinate's portfolio. This general characterization does not require Condition 1. Rather, Condition 1 is useful because it allows us to find a simple solution to the first-order conditions.

4.2 The Common Component Case

One way to introduce correlation in factor loadings is to assume that they are the sum of a component which is common to all assets, and an i.i.d. component. This captures the notion that the determinants of an asset's factor loading are either economy-wide, or asset-specific. We set $\lambda_n = \zeta + \xi_n$, where ζ denotes the common component, ξ_n the i.i.d. component, and ζ and $\{\xi_n\}_{n=1,..,N}$ are independent. For notational simplicity, we assume that ζ and $\{\xi_n\}_{n=1,..,N}$ have mean zero. We also assume that they are normal, so that Condition 1 holds. Indeed, under normality, the sum of factor loadings over a set A is a sufficient statistic, relative to all factor loadings in A, for the common component ζ. Therefore, it is also a sufficient statistic for the sum of factor loadings over the complement of A.

Since assets are ex-ante identical, the assignment of assets to agents does not matter, and we only need to determine the optimal hierarchical structure. We first compute the sum of expected squared errors for a general hierarchical structure.

Lemma 3 *Suppose that $\lambda_n = \zeta + \xi_n$, where $\zeta \sim N(0, \sigma_\zeta^2)$, $\xi_n \sim N(0, \sigma_\xi^2)$, and all variables are independent. Suppose also that agents follow the optimal decision rules of Proposition 3. Then,*

$$\sum_{n=1}^{N} E(e_n^2) = \sigma_\xi^4 (1 + rN)^2 \sum_{j \in J} \sum_{i=1}^{S(j)} \frac{[N(j,i) - 1][N(j) - N(j,i)]}{[1 + rN(j,i)][1 + rN(j)]}, \tag{15}$$

where $r \equiv \sigma_\zeta^2 / \sigma_\xi^2$.

Equation (15) generalizes equation (10) of the i.i.d. case. Using equation (15), we can determine the optimal hierarchical structure.

Proposition 4 *Suppose that $\lambda_n = \zeta + \xi_n$, where $\zeta \sim N(0, \sigma_\zeta^2)$, $\xi_n \sim N(0, \sigma_\xi^2)$, and all variables are independent. In any optimal hierarchical structure:*

i. agents must have one subordinate (except the bottom agent),

ii. agents must handle exactly K assets or portfolios, except at most the top agent.

The optimal hierarchical structure is as in in the i.i.d. case: agents have only one subordinate, and work at full capacity except perhaps one. As in the i.i.d. case, the intuition for the one subordinate result is that aggregation loss is large even for two assets. The common component does not affect the aggregation loss term, since this term is the difference between one factor loading and an average of factor loadings.

The only difference with the i.i.d. case concerns the position of an agent working below capacity. While in the i.i.d. case such an agent can be at any hierarchical level, in the common component case he can be only at the top. This means that returns to ability are highest at the bottom of the hierarchy. Intuitively, the benefit of placing a high ability agent at the top is that he can process more disagreggated information. This benefit is the same as in the i.i.d. case, since the aggregation loss term is the same. By contrast, the benefit of placing a high ability agent at the bottom is larger than in the i.i.d. case. Indeed, a high ability agent can examine more assets, and so obtain more accurate information on the common component. Therefore, he can better take into account the interactions with the assets he does not examine, and so form a better portfolio.

4.3 The Group Component Case

An alternative way to introduce correlation in factor loadings is to assume that they are the sum of a component which is common to all assets in a given group, and an i.i.d.

component. Groups may, for example, correspond to industries or countries. When assets are sorted into groups, they are not identical ex-ante (in particular, not in relation to other assets). Therefore, the assignment of assets to agents matters, and the organization designer's problem becomes more complicated. To gain some insights on the solution, we consider a simple example where there are $N = 4$ assets, each agent can process $K = 2$ assets, and there are two groups, one consisting of assets 1 and 2, and the other of assets 3 and 4. We set $\lambda_n = \zeta_{12} + \xi_n$, for $n = 1, 2$, and $\lambda_n = \zeta_{34} + \xi_n$, for $n = 3, 4$, where the group components ζ_{12} and ζ_{34} are i.i.d., and are independent of the i.i.d. components $\{\xi_n\}_{n=1,..,4}$. We also assume that all variables are mean zero and normal.

There are four possible combinations of hierarchical structures and assignments of assets to agents, illustated in Figure 4. First, the *one-subordinate homogeneous* hierarchy, H_1, in which each agent has one subordinate, and the bottom agent examines assets in the same group. Second, the *one-subordinate heterogeneous* hierarchy, $\hat{H}_1$, which differs from H_1 in that the bottom agent examines one asset in each group. Third, the *two-subordinate homogeneous* hierarchy, H_2, in which the top agent has two subordinates, each of whom examines assets in the same group. Finally, the *two-subordinate heterogeneous* hierarchy, $\hat{H}_2$, which differs from H_2 in that subordinates examine one asset in each group.

Proposition 5 *Suppose that $N = 4$, $K = 2$, $\lambda_n = \zeta_{12} + \xi_n$ for $n = 1, 2$, and $\lambda_n = \zeta_{34} + \xi_n$ for $n = 3, 4$, where $\zeta_{12}, \zeta_{34} \sim N(0, \sigma_\zeta^2)$, $\xi_n \sim N(0, \sigma_\xi^2)$, and all variables are independent. Then, H_2 is equivalent to $\hat{H}_2$, and both are dominated by H_1.*

Proposition 5 implies that an one-subordinate hierarchy is optimal. This result is the same as in the independent and common component cases, but it is even more surprising in the group component case. Consider, for example, the hierarchies H_1 and H_2. In the absence of a group component, H_1 dominates H_2 because it is preferable that the top agent knows an exact factor loading and an aggregate of three, rather than two aggregates of two. In the presence of a group component, however, one might expect H_2 to dominate. This is because the aggregates in H_2 are formed by two assets in the same group, and thus involve little loss of information, while the aggregate in H_1 is formed by three assets, not all in the same group. The reason why H_1 still dominates is that the aggregate of three assets is not as relevant, because some of the top agent's tasks are performed by agent 1. Indeed, since assets 3 and 4 are in the same group, agent 1 has some information on λ_4, which he can use to adjust his investments for the interaction with asset 4.[11]

[11]The intuition why H_2 and $\hat{H}_2$ are equivalent goes along similar lines. The advantage of H_2 is that the

4.4 Returns to Specialization

The result that one-subordinate hierarchies are optimal seems contrary to the way many actual organizations are structured. In investment firms, for example, many industry sector analysts report to the same boss, who is in charge of the overall stock portfolio. One feature of actual organizations which is not captured in our model is that there are returns to specialization. Sector analysts, for example, have expertise in their sector, enabling them to analyze the sector stocks better than other analysts. In this section, we show that our model can be extended to incorporate returns to specialization. Furthermore, in the presence of such returns, one-subordinate hierarchies are no longer optimal, and returns to ability can be highest at the top of the hierarchy.

To model returns to specialization, we consider the example of Section 4.3, with one modification. We assume that when an agent examines an asset, he has only the time to learn either the group component, or the i.i.d. component, but not both. The reason why this implies returns to specialization is as follows. Suppose that an agent examines an asset, and chooses to learn the group component. Then, if this agent examines a second asset in the same group, he does not need to learn the group component again. Therefore, he can learn the i.i.d. component, and so be fully informed about the second asset.[12]

To obtain the returns to specialization, it is important to endogenize agents' choice of which information to learn on each asset they examine. Consequently, we treat this choice as part of the solution to the organization designer's problem. We solve this problem in Proposition 6.

Proposition 6 *Suppose that $N = 4$, $K = 2$, $\lambda_n = \zeta_{12} + \xi_n$ for $n = 1, 2$, and $\lambda_n = \zeta_{34} + \xi_n$ for $n = 3, 4$, where $\zeta_{12}, \zeta_{34} \sim N(0, \sigma_\zeta^2)$, $\xi_n \sim N(0, \sigma_\xi^2)$, and all variables are independent. Then, if $\sigma_\zeta^2/\sigma_\xi^2$ is sufficiently large, it is optimal that agents know the group components for the assets they examine. Furthermore, H_2 dominates H_1, $\hat{H}_1$, and $\hat{H}_2$.*

Proposition 6 implies that if returns to specialization are important (in the sense that the variance of the group component is sufficiently large), then the top agent should have two subordinates, each of whom examines two assets in the same group. The intuition

aggregates are formed by assets in the same group, and thus involve little loss of information. The advantage of $\hat{H}_2$, on the other hand, is that the bottom agents can perform some of the top agent's tasks because they have some information on the assets not under their control.

[12] An alternative way to model returns to specialization is to assume that agents can only observe noisy signals of the factor loadings of the assets they examine. Then, an agent examining two assets in the same group has more information on the group component, and so can estimate more accurately each asset's factor loading. We do not model returns to specialization through noisy signals, however, because this is less tractable.

is as follows. Because the group component has a large variance, it represents valuable information, and thus agents should know the group components for the assets they examine. Given this, the only way for the organization to obtain some information on the i.i.d. components is that assets in each group are examined by the same agent.

Since returns to specialization affect the optimal hierarchical structure, it is interesting to know whether they also affect the returns to ability. To examine this, we modify the previous example by introducing a third group of two assets (assets 5 and 6), and a high ability agent who can process four assets, instead of two as the other agents. We can then show[13] that if returns to specialization are important, the high ability agent should be at the top of the hierarchy, examine two assets in the same group, and have two subordinates also examining two assets in the same group. (Hierarchy H_A, illustrated in figure 5.) The intuition why the high ability agent should be at the top rather the bottom (as, for example, in hierarchy H_B) is that he can perform the high-level task of adjusting investment in each group for the interaction with the other groups, leaving to his subordinates (part of) the low-level task of selecting within-group investments.

4.5 Communication and Scaling Constraints

In our model, portfolio formation is subject to the communication constraint, namely, (i) agents form their portfolios by combining the portfolios of their hierarchical subordinates and the assets they examine, and (ii) agents' information concerns only the above portfolios and assets. One way to generalize this constraint is to relax (ii), and allow agents to obtain information on portfolios or assets elsewhere in the hierarchy. If agents can obtain information only on the portfolios of those at lower hierarchical levels (i.e., their subordinates or the subordinates of their hierarchical peers), then the organization's portfolio is still formed in a single iteration, going from the bottom to the top of the hierarchy. If, by contrast, agents can also obtain information on the portfolios of their peers or superiors, then portfolio formation generally involves multiple iterations. For example, the organization can form a trial portfolio, then the top agent can communicate information on this portfolio to the bottom agents, so that a second trial portfolio can be formed, etc. We expect the analysis and the results of the single iteration case to be similar to the ones in this paper. However, the analysis of the multiple iteration case might be substantially more complicated.

An additional constraint that we impose on portfolio formation is the scaling constraint,

[13] The proof is available upon request.

namely, agents cannot change the composition of their subordinates' portfolios, but can only scale the portfolios up or down. This constraint follows, in some cases, from agents' optimal behavior. Indeed, when the assets in the portfolio of an agent's subordinate are identical conditional on the agent's information (as in the independent and common component cases), then it is optimal for the agent not to change the composition of the subordinate's portfolio.[14]

5 Conclusion

This paper proposes a model of organizational decision making, in which information processing is decentralized. The model's novel elements are that aggregation entails a loss of useful information, and agents' decision problems interact. These elements seem important in many organizational settings. One such setting, from which our model is motivated, is portfolio selection in investment firms.

In our model, an organization forms a portfolio of risky assets. Portfolio formation is hiearchical: agents combine their subordinates' portfolios and any assets they examine directly into larger portfolios, and communicate information on these portfolios to their superiors. Agents' ability to process information is limited, in that they can combine only a limited number of inputs into a portfolio.

We determine the hierarchical structure, assignment of assets to agents, and agents' decision rules that maximize the quality of organizational decisions. In the cases we examine, the optimal hierarchical structure has a chain form, where all agents have one subordinate. Furthermore, returns to ability are at least as high at the bottom as at the top of the hierarchy.

One interesting extension is to introduce returns to specialization. This extension is quite realistic, and in the example we examine in this paper, it reverses the one-subordinate result, as well as the result on returns to ability. Other possible extensions are to allow assets to differ in their expected returns, and to consider alternative probability distributions for the factor loadings. In particular, it might be interesting to introduce multiple group components (corresponding, for example, to an asset's industry sector and country). This would raise the question of along which criterion assets should be grouped, a question that has received attention by investment management practitioners. (See, for example,

[14]Formally, in the absence of the scaling constraint, the first-order condition (13) must hold for all assets in $A_N(j)$, instead of $A_M(j)$. When assets in $A_N(j,i)$ are identical conditonal on j's information, $E[e_n|I(j)]$ is the same for all $n \in A_N(j,i)$. Therefore, equation (14) implies (13) for all $n \in A_N(j,i)$.

Gunn (2000).) At a more abstract level, studying optimal grouping might be relevant for understanding whether organizations should be structured according to functions (U-form), products (M-form), or in a hybrid fashion (matrix form).

Appendix

Due to space limitations, this Appendix does not contain the proofs of Propositions 4-6. These proofs are in a separate Appendix, available on the *Review of Economic Studies* website.

We first state and prove Theorem 1 (referred to at the end of Section 4.1). To state the theorem, we introduce some notation. For an asset n, we denote by j_n the agent who examines the asset, i.e., who satisfies $n \in A_M(j_n)$, and by J_n the set of agents who can influence the investment in the asset, i.e., who satisfy $n \in A_N(j)$. For an agent $j \in J_n \backslash \{j_n\}$, we denote by $i_n(j)$ the index of j's subordinate who is in J_n. The organization's investment in asset n can be expressed as

$$x_n = x_n(j_n) \prod_{j \in J_n \backslash \{j_n\}} y_{i_n(j)}(j). \tag{16}$$

Given functions $\{F_n(j)\}_{n \in A_M(j)}$ and $\{G_i(j)\}_{i=1,..,S(j)}$, for each agent j, we set

$$F_n \equiv F_n(j_n) + \sum_{j \in J_n \backslash \{j_n\}} G_{i_n(j)}(j). \tag{17}$$

Intuitively, $F_n(j)$ and $G_i(j)$ will represent the lowest order terms (after the order zero term) in $x_n(j)$ and $y_i(j)$, respectively, and F_n will represent the lowest order term in x_n. We denote by F_n^* the lowest order term in x_n^*, i.e.,

$$F_n^* \equiv -\lambda_n \sum_{n'=1}^{N} \lambda_{n'}.$$

Finally, we recall the definitions of $H(j)$, $J(j)$, $\Lambda^c(j)$, $\Gamma(j)$, $I(j)$, and $\bar{I}(j)$, given at the beginning of Section 4.1.

Theorem 1 *Suppose that for each agent j, there exist functions $\{F_n(j)\}_{n \in A_M(j)}$ and $\{G_i(j)\}_{i=1,..,S(j)}$, which contain second-degree terms in $\Gamma(j)$ and λ, and satisfy the first-order conditions*

$$E[F_n - F_n^* | I(j)] = 0 \tag{18}$$

for $n \in A_M(j)$, and

$$E\left[\sum_{n' \in A_N(j,i)} (F_{n'} - F_{n'}^*) | I(j)\right] = 0 \tag{19}$$

for $i = 1, .., S(j)$. Then, there exists a selection of optimal decision rules such that for each agent j,

$$x_n(j) = \frac{\mu}{a\sigma^2}\left[1 + F_n(j) + o(\lambda^2)\right], \tag{20}$$

for $n \in A_M(j)$*, and*

$$y_i(j) = 1 + G_i(j) + o(\lambda^2), \tag{21}$$

for $i = 1, .., S(j)$*.*

Proof: From Assumption 1, there exists a selection of optimal decision rules such that for each agent j,

$$x_n(j) = \frac{\mu}{a\sigma^2}\left[1 + \hat{f}_n(j) + o(\lambda^2)\right], \tag{22}$$

for $n \in A_M(j)$, and

$$y_i(j) = 1 + \hat{g}_i(j) + o(\lambda^2), \tag{23}$$

for $i = 1, .., S(j)$, where $\hat{f}_n(j)$ and $\hat{g}_i(j)$ contain first- and second-degree terms in $\gamma(j)$ and λ. We assume initially that $\hat{f}_n(j)$ and $\hat{g}_i(j)$ contain only second-degree terms, and prove the theorem in three steps. In a fourth step, we extend the proof to the case where $\hat{f}_n(j)$ and $\hat{g}_i(j)$ contain also first-degree terms.

Step 1: Definition of $\hat{F}_n$ and first-order conditions. Since in the limit when λ goes to zero, the organization's investment in each asset converges to $\mu/a\sigma^2$, we have

$$\lambda(j,i) = \frac{\mu}{a\sigma^2}\Lambda(j,i) + o(\lambda). \tag{24}$$

Therefore,

$$\hat{f}_n(j) = \hat{F}_n(j) + o(\lambda^2), \tag{25}$$

and

$$\hat{g}_i(j) = \hat{G}_i(j) + o(\lambda^2), \tag{26}$$

where $\hat{F}_n(j)$ and $\hat{G}_i(j)$ contain second-degree terms in $\Gamma(j)$ and λ. Equations (16), (22), (23), (25), and (26), imply that the organization's investment in asset n is

$$x_n = \frac{\mu}{a\sigma^2}\left[1 + \hat{F}_n + o(\lambda^2)\right], \tag{27}$$

where $\hat{F}_n$ is defined from $\hat{F}_n(j)$ and $\hat{G}_i(j)$, as in equation (17).

We next show that the first-order conditions (18) and (19) hold if we substitute $\{\hat{F}_n\}_{n=1,..,N}$ for $\{F_n\}_{n=1,..,N}$. Consider the organization designer's objective, given in equation (2):

$$-E\exp\left[-aQ(x)\right],$$

where $x \equiv \{x_n\}_{n=1,..,N}$,

$$Q(x) \equiv \sum_{n=1}^{N} x_n r - \frac{R(x)}{a},$$

and

$$R(x) \equiv \frac{a^2}{2}\left(\left(\sum_{n=1}^{N} x_n\lambda_n\right)^2 \sigma^2 + \sum_{n=1}^{N} x_n^2\sigma^2\right).$$

Since Q is quadratic, and maximum at $x^* \equiv \{x_n^*\}_{n=1,..,N}$, we have

$$Q(x) = Q(x^*) - \frac{R(x-x^*)}{a}.$$

Therefore, we can write the organization designer's objective as

$$-E\left[\exp\left[-aQ(x^*)\right]\exp\left[R(x-x^*)\right]\right]. \tag{28}$$

Consider $n \in A_M(j)$, and suppose that agent j's investment in asset n is perturbed to

$$x_n(j)(1+\alpha h),$$

where α is a scalar, and h a function of $\gamma(j)$. This changes the organization's investment in asset n to

$$x_n(1+\alpha h).$$

The derivative of the organization designer's objective w.r.t. α, for $\alpha = 0$, is

$$-E\left[\exp\left[-aQ(x^*)\right]\exp\left[R(x-x^*)\right]\frac{\partial R}{\partial x_n}(x-x^*)h\right]. \tag{29}$$

This has to equal zero, by the optimality of agent j's decision rule. To derive equation (18), we will determine the lowest order term in equation (29), for small factor loadings. Equation (5) implies that

$$\exp\left[-aQ(x^*)\right] = \exp\left[-N\frac{\mu^2}{2\sigma^2} + o(\lambda^2)\right]. \tag{30}$$

Equations (5) and (27) imply that

$$\exp\left[R(x-x^*)\right] = \exp\left[o(\lambda^2)\right],$$

and

$$\frac{\partial R}{\partial x_n}(x-x^*) = a^2\sigma^2\left[\lambda_n\left(\sum_{n'=1}^{N}(x_{n'}-x_{n'}^*)\lambda_{n'}\right) + (x_n - x_n^*)\right] = a^2\sigma^2(\hat{F}_n - F_n^*) + o(\lambda^2).$$

To determine the lowest order term in h, we set

$$h(\gamma(j)) \equiv H\left(\frac{\gamma(j)}{\lambda}\right),$$

and assume that the function H stays fixed as λ goes to zero. Equation (24) implies that

$$H\left(\frac{\gamma(j)}{\lambda}\right) = K\left(\frac{\Gamma(j)}{\lambda} + o(1)\right),$$

for some function K. Therefore, the lowest order term in $h(\gamma(j))$ is $K(\Gamma_0(j))$, where $\Gamma_0(j) \equiv \Gamma(j)/\lambda$. The lowest order term in equation (29) is

$$-E\left[\exp\left[-N\frac{r^2}{2\sigma^2}\right] a^2\sigma^2(\hat{F}_n - F_n^*)K(\Gamma_0(j))\right].$$

Since this has to equal zero, we must have

$$E\left[(\hat{F}_n - F_n^*)K(\Gamma_0(j))\right] = 0,$$

for all functions $K(\Gamma_0(j))$, i.e., for all $I(j)$-measurable functions. This implies equation (18).

To derive equation (19), we proceed similarly. We suppose that agent j's scaling factor for the portfolio of his ith subordinate is perturbed to

$$y_i(j)(1 + \alpha h).$$

This changes the organization's investment in any asset $n \in A_N(j,i)$ to

$$x_n(1 + \alpha h).$$

The derivative of the organization designer's objective w.r.t. α, for $\alpha = 0$, is

$$-E\left[\exp\left[-aQ(x^*)\right]\exp\left[R(x - x^*)\right]\sum_{n\in A_N(j,i)}\left(\frac{\partial R}{\partial x_n}(x - x^*)\right)h\right]. \tag{31}$$

Writing that the lowest order term in this equation is zero, we find that

$$E\left[\left[\sum_{n'\in A_N(j,i)}(F_{n'} - F_{n'}^*)\right]K(\Gamma_0(j))\right] = 0,$$

which implies equation (19).

Step 2: $F_n = \hat{F}_n$**, for all** $n = 1, .., N$**.** This means that investments by the organization resulting from (i) decision rules of the form (20) and (21), and (ii) the optimal decision rules (22) and (23), are identical up to second-degree terms.

We have

$$E\sum_{n=1}^{N}(F_n - \hat{F}_n)^2 = S - \hat{S},$$

where

$$S = E\sum_{n=1}^{N}\left[(F_n - \hat{F}_n)(F_n - F_n^*)\right],$$

and

$$\hat{S} = E\sum_{n=1}^{N}\left[(F_n - \hat{F}_n)(\hat{F}_n - F_n^*)\right].$$

To show that $F_n = \hat{F}_n$, we will show that $S = \hat{S} = 0$. Equation (17) implies that

$$\begin{aligned} F_n - \hat{F}_n &= \left[F_n(j_n) - \hat{F}_n(j_n)\right] + \sum_{j\in J_n\backslash\{j_n\}}\left[G_{i_n(j)}(j) - \hat{G}_{i_n(j)}(j)\right] \\ &\equiv \Delta_n^F(j_n) + \sum_{j\in J_n\backslash\{j_n\}}\Delta_{i_n(j)}^G(j). \end{aligned}$$

Using this equation, and summing across agents instead of across assets, we can write S as

$$S = E\sum_{j\in J}\left[\sum_{n\in A_M(j)}\Delta_n^F(j)(F_n - F_n^*) + \sum_{i=1}^{S(j)}\Delta_i^G(j)\sum_{n'\in A_N(j,i)}(F_{n'} - F_{n'}^*)\right].$$

Since $\Delta_n^F(j)$ is $I(j)$-measurable, we have

$$E\left[\Delta_n^F(j)(F_n - F_n^*)\right] = E\left[\Delta_n^F(j)E[F_n - F_n^*|I(j)]\right],$$

which is equal to zero from equation (18). Similarly,

$$E\left[\Delta_i^G(j)\sum_{n'\in A_N(j,i)}(F_{n'} - F_{n'}^*)\right] = 0.$$

Therefore, $S = 0$, and by an identical argument, $\hat{S} = 0$.

Step 3: There exist decision rules of the form (20) and (21), which are optimal. To show this, we will start from the optimal decision rules (22) and (23), and construct decision rules of the form (20) and (21), without modifying the organization's investments. Consider an agent j at the bottom of the hierarchy, and an asset $n \in A_M(j)$. Since $F_n = \hat{F}_n$, we have

$$F_n(j) - \hat{F}_n(j) = \sum_{j'\in J_n\backslash\{j\}}\left[\hat{G}_{i_n(j')}(j') - G_{i_n(j')}(j')\right].$$

The LHS of this equation is $I(j)$-measurable, i.e., is a function of $\Gamma(j) = \{\lambda_n\}_{n\in A_M(j)}$. The RHS is a function of $\Gamma(j')$ for $j' \in J_n\backslash\{j\}$, and can thus depend on $\Gamma(j)$ only through $\Lambda(j)$. Therefore, both the LHS and the RHS are equal to a function K of $\Lambda(j)$ (and λ), which, in addition, contains second-degree terms. Equation (24) implies that there exists a function k, which contains second-degree terms in $\lambda(j)$ and λ, such that

$$k = K + o(\lambda^2).$$

Suppose that the decision rule for agent j is modified from (22) to $x_n(j)[1+k]$ for all $n \in A_M(j)$, and the decision rule for j's superior, j^b, is modified to $y_i(j^b)/[1+k]$, only for the index i corresponding to j. The new decision rules are measurable w.r.t. both agents' information, since k is a function of $\lambda(j)$. They also leave the organization's investments unchanged, since the change in j's investment is undone by j^b. Finally, j's decision rule becomes of the form (20), since

$$\begin{aligned} x_n(j)[1+k] &= \frac{\mu}{a\sigma^2}\left[1 + \hat{f}_n(j) + o(\lambda^2)\right][1+k] \\ &= \frac{\mu}{a\sigma^2}\left[1 + \hat{F}_n(j) + o(\lambda^2)\right][1+K+o(\lambda^2)] \\ &= \frac{\mu}{a\sigma^2}\left[1 + \hat{F}_n(j) + K + o(\lambda^2)\right] \\ &= \frac{\mu}{a\sigma^2}\left[1 + F_n(j) + o(\lambda^2)\right]. \end{aligned}$$

Proceeding inductively to the top of the hierarchy, we can similarly modify all agents' decision rules so that they take the form (20) and (21).

Step 4: $\hat{f}_n(j)$ and $\hat{g}_i(j)$ contain first-degree terms. Suppose that $\hat{f}_n(j)$ and $\hat{g}_i(j)$ contain first-degree terms in $\gamma(j)$ and λ. Then

$$\hat{f}_n(j) = \hat{F}_n^1(j) + o(\lambda),$$

and

$$\hat{g}_i(j) = \hat{G}_i^1(j) + o(\lambda),$$

where $\hat{F}_n^1(j)$ and $\hat{G}_i^1(j)$ contain first-degree terms in $\Gamma(j)$ and λ. Moreover,

$$x_n = \frac{\mu}{a\sigma^2}\left[1 + \hat{F}_n^1 + o(\lambda)\right],$$

where $\hat{F}_n^1$ is defined from $\hat{F}_n^1(j)$ and $\hat{G}_i^1(j)$, as in equation (17). Proceeding as in Step 1, and noting that the first-degree term in x_n^* is zero, we can show that the first-order conditions (18) and (19) hold if we substitute $\{\hat{F}_n^1\}_{n=1,..,N}$ for $\{F_n\}_{n=1,..,N}$, and zero for $\{F_n^*\}_{n=1,..,N}$. Since (18) and (19) obviously hold if we substitute zero for both $\{F_n\}_{n=1,..,N}$ and $\{F_n^*\}_{n=1,..,N}$, we can proceed as in Step 2, and show that $\hat{F}_n^1 = 0$ for all $n = 1,..,N$. We can then modify the decision rules (22) and (23), so that they do not contain first-degree terms, and yet imply the same investments by the organization. To do this, we proceed as in Step 3. We start with an agent j at the bottom of the hierarchy and show that $\hat{F}_n^1(j)$ must be a function of $\Lambda(j)$ only. We then adjust j's decision rule so that the first-degree terms cancel, and proceed inductively to the top of the hierarchy. $\|$

Proof of Proposition 1: The proposition follows from Proposition 3, by noting that since factor loadings are i.i.d. with mean zero, we have $E\left[\Lambda^c(j)|\, I(j)\right] = 0$. $\|$

Proof of Lemma 1: Consider a hierarchical structure H, that maximizes expected utility. Since expected utility can be expressed as in equation (28), and x^* does not depend on H, H must also maximize

$$-E\left[\exp\left[-aQ(x^*)\right]\left[\exp\left[R(x-x^*)\right]-1\right]\right]. \tag{32}$$

Since

$$x_n - x_n^* = \frac{\mu}{a\sigma^2}\left[F_n - F_n^* + o(\lambda^2)\right] = \frac{\mu}{a\sigma^2}\left[e_n + o(\lambda^2)\right],$$

we have

$$\exp\left[R(x-x^*)\right] - 1 = \frac{\mu^2}{2\sigma^2}\sum_{n=1}^{N} e_n^2 + o(\lambda^4).$$

Using this equation and equation (30), we can write equation (32) as

$$-\exp\left[-N\frac{\mu^2}{2\sigma^2}\right]\frac{\mu^2}{2\sigma^2}\sum_{n=1}^{N} E(e_n^2) + o(\lambda^4).$$

If H maximizes this equation for λ small, then it must also maximize the lowest order term, i.e., it must minimize $\sum_{n=1}^{N} E(e_n^2)$. $\|$

Proof of Lemma 2: The lemma follows from Lemma 3, by setting $r = 0$ and $\sigma_\xi^2 = \sigma_\lambda^2$. (Lemma 3 requires normality, but when $\zeta = 0$, normality is not required in the proof.) $\|$

Proof of Proposition 2: We proceed in two steps.

Step 1: Properties of a hierarchical structure that satisfies conditions (i) and (ii). Consider such a hierarchical structure, and denote by X the number of agents, by m_x, for $x = 1, \cdots, X-1$, the number of assets examined by the agent in level x (the top is level 1), and by $m_X + 1$ the number of assets examined by the bottom agent. Condition (ii) implies that there exists $x_p \in \{1, .., X\}$ (corresponding to the level of the agent who may be working below capacity) such that $m_x = K - 1$ for all $x \in \{1, .., X\}\backslash\{x_p\}$. We have

$$\sum_{x=1}^{X} m_x + 1 = N \Rightarrow m_{x_p} = N - 1 - (X-1)(K-1),$$

and

$$0 < m_{x_p} \leq K-1 \Rightarrow X-1 < \frac{N-1}{K-1} \leq X \Rightarrow X = \left\lceil \frac{N-1}{K-1} \right\rceil,$$

where $\lceil y \rceil$ denotes the smallest integer that is greater or equal than y. Therefore, m_{x_p} and X are the same for any hierarchical structure that satisfies conditions (i) and (ii), and we denote them by $m(N)$ and $X(N)$, respectively.

Lemma 2 implies that the sum of expected squared errors (which we normalize by σ_λ^4, and refer to as the cost from now on) is

$$\sum_{x=1}^{X(N)-1} \left[\left(\sum_{x'=x+1}^{X(N)} m_{x'} \right) m_x \right]. \tag{33}$$

Equation (33) is invariant to a permutation of $\{m_x\}_{x=1,..,X(N)}$, and thus the cost is the same for any hierarchical structure that satisfies conditions (i) and (ii). We denote the cost by $c(N)$, and show two properties of $c(N)$. First,

$$c(N) - c(N-1) = [X(N)-1](K-1). \tag{34}$$

To show equation (34), we use equation (33), and consider hierarchical structures where the agent who may be working below capacity is at the top. If, with $N-1$ assets, the top agent does work below capacity, then by adding one asset, m_1 increases by 1, and equation (33) changes by

$$\sum_{x'=2}^{X(N)} m_{x'} = [X(N)-1](K-1).$$

If, with $N-1$ assets, the top agent works at full capacity, then by adding one asset, we obtain a new top agent with $m_1 = 1$. The term in equation (33) corresponding to this new agent is

$$\sum_{x'=2}^{X(N)} m_{x'} = [X(N)-1](K-1),$$

and thus equation (34) holds. A second property of $c(N)$ is that for $n_1, n_2 > 1$,

$$c(n_1) + c(n_2) - c(n_1+n_2-1) + 2(n_1-1)(n_2-1) > 0. \tag{35}$$

To show equation (35), we assume without loss of generality that $n_1 \geq n_2$, and denote the LHS by $f(n_2)$. Since $c(1) = 0$, we have $f(1) = 0$. Therefore, $f(n_2)$ is positive if it is increasing. Using equation (34), and the definition of $m(N)$, we have

$$\begin{aligned} f(n_2) - f(n_2-1) &= [X(n_2) - X(n_1+n_2-1)]\,(K-1) + 2(n_1-1) \\ &= n_1 - 1 + m(n_1+n_2-1) - m(n_2). \end{aligned} \tag{36}$$

Since

$$m(n_1 + n_2 - 1) = m(n_1) + m(n_2) - \epsilon(K - 1),$$

where $\epsilon = 0$ if $m(n_1) + m(n_2) \leq K - 1$, and $\epsilon = 1$ if $m(n_1) + m(n_2) > K - 1$, we can write equation (36) as

$$n_1 - 1 + m(n_1) - \epsilon(K - 1). \tag{37}$$

If $n_1 > K$ or $\epsilon = 0$, equation (37) is obviously positive. If $n_1 \leq K$ and $\epsilon = 1$, equation (37) is also positive, since

$$n_1 - 1 + m(n_1) = 2m(n_1) \geq m(n_1) + m(n_2) > K - 1.$$

Step 2: A cost-minimizing hierarchical structure must satisfy conditions (i) and (ii). The key to the proof is the following dynamic programming observation. If a hierarchical structure H is cost-minimizing, then for each agent j, the subordinate hierarchy $H(j)$ must also be cost-minimizing among hierarchies that invest in $N(j)$ assets. Indeed, using Lemma 2, we can write the cost of H as

$$\sum_{j' \in J(j)} \sum_{i=1}^{S(j')} [N(j', i) - 1][N(j') - N(j', i)] + \sum_{j' \in J \setminus J(j)} \sum_{i=1}^{S(j')} [N(j', i) - 1][N(j') - N(j', i)],$$

i.e., as the cost of $H(j)$, plus a term that depends on $H(j)$ only through $N(j)$. Therefore, if $H(j)$ is not cost-minimizing, we can replace it by a cost-minimizing hierarchical structure, and reduce the cost of H.

To show that a cost-minimizing hierarchical structure must satisfy conditions (i) and (ii), we use the dynamic programming observation, and proceed by induction on the number N of assets. Suppose that the result is true for all $N' < N$ (N can be 1), and consider a cost-minimizing hierarchical structure H, in which the top agent has at least two subordinates, agents 1 and 2. We will show that H is dominated by a hierarchical structure H', in which $H(1)$ is replaced by a cost-minimizing hierarchical structure $H'(1)$ for $N(1) + N(2) - 1$ assets, and $H(2)$ is replaced by one asset. Using Lemma 2, and noting that $H(1)$, $H(2)$, and $H'(1)$, must satisfy conditions (i) and (ii) (from the dynamic programming observation and the induction hypothesis), we can write the cost of H and H' as

$$c(H) = c[N(1)] + c[N(2)] + [N(1) - 1][N - N(1)] + [N(2) - 1][N - N(2)] + \hat{c},$$

and

$$c(H') = c[N(1) + N(2) - 1] + [N(1) + N(2) - 2][N - [N(1) + N(2) - 1]] + \hat{c},$$

respectively, where

$$\hat{c} = \sum_{i=3}^{S(\cdot)} c[H(i)] + \sum_{i=3}^{S(\cdot)} [N(i) - 1][N - N(i)].$$

Therefore,

$$c(H) - c(H') = c[N(1)] + c[N(2)] - c[N(1) + N(2) - 1] + 2[N(1) - 1][N(2) - 1].$$

Equation (34) implies that this is positive, contradicting the optimality of H. Therefore, the top agent in H has only one subordinate, agent 1. Since the hierarchical structure $H(1)$ satisfies conditions (i) and (ii), H satisfies condition (i). To show that H satisfies condition (ii), we need to show that the top agent in H and the agent in $H(1)$ who may be working below capacity (and for whom $m_{x_p} = m(N - m_1)$) cannot both be working below capacity. Without loss of generality (since equation (33) is invariant to a permutation of the m_x's), we assume that $m_1 \leq m(N - m_1)$. We will show that if $m(N - m_1) < K - 1$, then H is dominated by the hierarchical structure H' in which one asset is shifted from the top agent to the agent working below capacity in $H(1)$. Using equation (33), we can write the cost of H as

$$c(H) = c[H(1)] + (N - m_1 - 1)m_1 = c(N - m_1) + (N - m_1 - 1)m_1,$$

and the cost of H' as

$$c(H') = c(N - m_1 + 1) + (N - m_1)(m_1 - 1).$$

Using equation (34), and noting that $X(N - m_1) = X(N - m_1 + 1)$, we have

$$\begin{aligned} c(H) - c(H') &= -[X(N - m_1) - 1](K - 1) + (N - 2m_1) \\ &= (m(N - m_1) - m_1 + 1) > 0. \end{aligned}$$

Therefore, H satisfies condition (ii). $\|$

Proof of Proposition 3: From Theorem 1, it suffices to show that the functions

$$F_n(j) \equiv -\lambda_n \Lambda(j) - \left[\lambda_n - \frac{\Lambda(j)}{N(j)}\right] E\left[\Lambda^c(j) | I(j)\right],$$

and

$$G_i(j) \equiv -\frac{\Lambda(j,i)}{N(j,i)} [\Lambda(j) - \Lambda(j,i)] - \left[\frac{\Lambda(j,i)}{N(j,i)} - \frac{\Lambda(j)}{N(j)}\right] E\left[\Lambda^c(j) | I(j)\right],$$

for $j \in J$, $n \in A_M(j)$, and $i = 1, .., S(j)$, satisfy the first-order conditions (18) and (19). Setting $E(j) \equiv E\left[\Lambda^c(j) | I(j)\right]$, we have

$$F_n(j_n) - F_n^* = \lambda_n \sum_{n'=1}^{N} \lambda_{n'} - \lambda_n \Lambda(j_n) - \left[\lambda_n - \frac{\Lambda(j_n)}{N(j_n)}\right] E(j_n)$$

$$= \sum_{j \in J_n \setminus \{j_n\}} \lambda_n \left[\Lambda(j) - \Lambda(j, i_n(j))\right] - \left[\lambda_n - \frac{\Lambda(j_n)}{N(j_n)}\right] E(j_n). \qquad (38)$$

Noting that

$$G_i(j) = -\frac{\Lambda(j,i)}{N(j,i)} \left[\Lambda(j) - \Lambda(j,i)\right] + \left[\lambda_n - \frac{\Lambda(j,i)}{N(j,i)}\right] E(j) - \left[\lambda_n - \frac{\Lambda(j)}{N(j)}\right] E(j),$$

we have

$$\begin{aligned}
\sum_{j \in J_n \setminus \{j_n\}} G_{i_n(j)}(j) &= -\sum_{j \in J_n \setminus \{j_n\}} \frac{\Lambda(j, i_n(j))}{N(j, i_n(j))} \left[\Lambda(j) - \Lambda(j, i_n(j))\right] \\
&\quad + \sum_{j \in J_n \setminus \{j_n\}} \left[\lambda_n - \frac{\Lambda(j, i_n(j))}{N(j, i_n(j))}\right] \left[E(j) - E(j, i_n(j))\right] \\
&\quad - \left[\lambda_n - \frac{\Lambda(\cdot)}{N(\cdot)}\right] E(\cdot) + \left[\lambda_n - \frac{\Lambda(j_n)}{N(j_n)}\right] E(j_n). \qquad (39)
\end{aligned}$$

Combining equations (38) and (39), and noting that $E(\cdot) = 0$, we have

$$e_n = F_n - F_n^* = F_n(j_n) + \sum_{j \in J_n \setminus \{j_n\}} G_{i_n(j)}(j) - F_n^* = \sum_{j \in J_n \setminus \{j_n\}} e_n(j), \qquad (40)$$

where

$$e_n(j) \equiv \left[\lambda_n - \frac{\Lambda(j, i_n(j))}{N(j, i_n(j))}\right] \left[\Lambda(j) - \Lambda(j, i_n(j)) + E(j) - E(j, i_n(j))\right],$$

denotes the component of the error e_n for asset n, that is associated with agent j.

Consider now an agent j, and denote by $B(j)$ the set of j's direct and indirect superiors. We will show that for $j^b \in B(j)$ and $n \in A_N(j)$, we have $E[e_n(j^b)|I(j)] = 0$. Since $I(j) \subset \bar{I}(j^b, i_n(j^b))$, it suffices to show that

$$E\left[e_n(j^b) \middle| \bar{I}(j^b, i_n(j^b))\right] = 0, \qquad (41)$$

and apply the law of iterative expectations. Since the aggregation loss term,

$$\lambda_n - \frac{\Lambda(j^b, i_n(j^b))}{N(j^b, i_n(j^b))},$$

is measurable w.r.t. $\bar{I}(j^b, i_n(j^b))$, it suffices to show that the interaction term,

$$\Lambda(j^b) - \Lambda(j^b, i_n(j^b)) + E(j^b) - E(j^b, i_n(j^b)),$$

has zero expectation, conditional on $\bar{I}(j^b, i_n(j^b))$. Using Condition 1, we have

$$\begin{aligned}
\Lambda(j^b) - \Lambda(j^b, i_n(j^b)) + E(j^b) &= \Lambda(j^b) - \Lambda(j^b, i_n(j^b)) + E\left[\Lambda^c(j^b) \middle| I(j^b)\right] \\
&= \Lambda(j^b) - \Lambda(j^b, i_n(j^b)) + E\left[\Lambda^c(j^b) \middle| \bar{I}(j^b)\right] \\
&= E\left[\Lambda(j^b) - \Lambda(j^b, i_n(j^b)) + \Lambda^c(j^b) \middle| \bar{I}(j^b)\right] \\
&= E\left[\Lambda^c(j^b, i_n(j^b)) \middle| \bar{I}(j^b)\right], \qquad (42)
\end{aligned}$$

and

$$\begin{aligned} E(j^b, i_n(j^b)) &= E\left[\Lambda^c(j^b, i_n(j^b))\middle| I(j^b, i_n(j^b))\right] \\ &= E\left[\Lambda^c(j^b, i_n(j^b))\middle| \bar{I}(j^b, i_n(j^b))\right]. \end{aligned}$$

Therefore, the interaction term is

$$E\left[\Lambda^c(j^b, i_n(j^b))\middle| \bar{I}(j^b)\right] - E\left[\Lambda^c(j^b, i_n(j^b))\middle| \bar{I}(j^b, i_n(j^b))\right], \tag{43}$$

and it has zero expectation, conditional on $\bar{I}(j^b, i_n(j^b))$, since $\bar{I}(j^b, i_n(j^b)) \subset \bar{I}(j^b)$.

Equations (40) and $E[e_n(j^b)|I(j)] = 0$, imply immediately equation (18). They also imply that to show equation (19), it suffices to show that

$$\sum_{n' \in A_N(j,i)} \left[\sum_{j' \in J_{n'}(j) \setminus \{j_{n'}\}} e_{n'}(j')\right] = 0.$$

Summing across agents instead of across assets, we can write the LHS as

$$\sum_{n' \in A_N(j,i)} e_{n'}(j) + \sum_{j' \in J(j,i)} \sum_{i'=1}^{S(j')} \sum_{n' \in A_N(j',i')} e_{n'}(j').$$

This is equal to zero, since for all j and i, we have $\sum_{n' \in A_N(j,i)} e_{n'}(j) = 0$. ||

Proof of Lemma 3: Consider $j, j^b \in J_n \setminus \{j_n\}$, and suppose that $j^b \in B(j)$. Noting that $e_n(j)$ is measurable w.r.t. $\bar{I}(j^b, i_n(j^b))$, and using equation (41), we have

$$\begin{aligned} E[e_n(j)e_n(j^b)] &= E\left[E\left[e_n(j)e_n(j^b)|\bar{I}(j^b, i_n(j^b))\right]\right] \\ &= E\left[e_n(j)E\left[e_n(j^b)|\bar{I}(j^b, i_n(j^b))\right]\right] \\ &= 0. \end{aligned}$$

Equation (40) then implies that

$$E(e_n^2) = \sum_{j \in J_n \setminus \{j_n\}} E\left[e_n(j)^2\right]. \tag{44}$$

Since

$$E[e_n(j)] = E\left[E\left[e_n(j)|\, I(j_n)\right]\right] = 0,$$

the aggregation loss and interaction terms are uncorrelated. Normality then implies that these terms are independent. Therefore, $E\left[e_n(j)^2\right]$ is equal to the expected squared aggregation loss term, times the expected squared interaction term. The former is

$$E\left[\lambda_n - \frac{\sum_{n' \in A_N(j,i_n(j))} \lambda_{n'}}{N(j, i_n(j))}\right]^2 = E\left[\xi_n - \frac{\sum_{n' \in A_N(j,i_n(j))} \xi_{n'}}{N(j, i_n(j))}\right]^2 = \sigma_\xi^2 \frac{N(j, i_n(j)) - 1}{N(j, i_n(j))}.$$

Equation (43) implies that the latter is

$$E\left[V\left[E\left[\Lambda^c(j,i_n(j))\middle|\,\bar{I}(j)\right]\middle|\,\bar{I}(j,i_n(j))\right]\right]. \tag{45}$$

Reversing the steps of equation (42), and using normality, we have

$$\begin{aligned}
E\left[\Lambda^c(j,i_n(j))\middle|\,\bar{I}(j)\right] &= \Lambda(j)-\Lambda(j,i_n(j))+E\left[\Lambda^c(j)\middle|\,\bar{I}(j)\right]\\
&= \Lambda(j)-\Lambda(j,i_n(j))+[N-N(j)]\frac{\sigma_\zeta^2}{\sigma_\zeta^2+\frac{\sigma_\xi^2}{N(j)}}\frac{\Lambda(j)}{N(j)}\\
&= \frac{N\sigma_\zeta^2+\sigma_\xi^2}{N(j)\sigma_\zeta^2+\sigma_\xi^2}\left[\Lambda(j)-\Lambda(j,i_n(j))\right]+\frac{[N-N(j)]\sigma_\zeta^2}{N(j)\sigma_\zeta^2+\sigma_\xi^2}\Lambda(j,i_n(j)),
\end{aligned}$$

and thus,

$$V\left[E\left[\Lambda^c(j,i_n(j))\middle|\,\bar{I}(j)\right]\middle|\,\bar{I}(j,i_n(j))\right]=\left[\frac{N\sigma_\zeta^2+\sigma_\xi^2}{N(j)\sigma_\zeta^2+\sigma_\xi^2}\right]^2 V\left[\Lambda(j)-\Lambda(j,i_n(j))\middle|\,\bar{I}(j,i_n(j))\right].$$

Noting that,

$$\begin{aligned}
V\left[\Lambda(j)-\Lambda(j,i_n(j))\middle|\,\bar{I}(j,i_n(j))\right] &= V\left[\sum_{\substack{n'\in A_{N(j)}\setminus\\ A_{N(j,i_n(j))}}}\xi_{n'}+[N(j)-N(j,i_n(j))]\zeta\,\middle|\,\bar{I}(j,i_n(j))\right]\\
&= [N(j)-N(j,i_n(j))]\sigma_\xi^2+\frac{[N(j)-N(j,i_n(j))]^2\sigma_\zeta^2\sigma_\xi^2}{N(j,i_n(j))\sigma_\zeta^2+\sigma_\xi^2}\\
&= [N(j)-N(j,i_n(j))]\sigma_\xi^2\frac{N(j)\sigma_\zeta^2+\sigma_\xi^2}{N(j,i_n(j))\sigma_\zeta^2+\sigma_\xi^2},
\end{aligned}$$

we can write the expected squared interaction term as

$$\sigma_\xi^2\frac{(1+rN)^2[N(j)-N(j,i_n(j))]}{[1+rN(j)][1+rN(j,i_n(j))]}.$$

Multiplying with the expected squared aggregation loss term, and summing across agents instead of across assets, we obtain equation (15). ‖

References

ALEXANDER, C. (ed.) (1998), *Risk Management and Analysis, Volume 1: Measuring and Modelling Financial Risk*, John Wiley & Sons.

AOKI, M. (1986), "Horizontal vs. Vertical Information Structure of the Firm", *American Economic Review*, **76**, 971-983.

BECKMANN, M. (1960), "Some Aspects of Returns to Scale in Business Administration", *Quarterly Journal of Economics*, **74**, 464-471.

BECKMANN, M. (1983), *Timbergen Lectures in Organization Theory*, Springer-Verlag.

BEGGS, A. (2001), "Queues and Hierarchies", *Review of Economic Studies*, **68**, 297-322.

BOLTON, P. and DEWATRIPONT, M. (1994), "The Firm as a Communication Network", *Quarterly Journal of Economics*, **109**, 809-839.

BOWER, J. (1970), *Managing the Resource Allocation Process: A Study of Corporate Planning and Investment*, Division of Research, Graduate School of Business Administration, Harvard University, Boston.

CREMER, J. (1980), "A Partial Theory of the Optimal Organization of a Bureaucracy", *Bell Journal of Economics*, **11**, 683-693.

GARICANO, L. (2000), "Hierarchies and the Organization of Knowledge in Production", *Journal of Political Economy*, **108**, 874-904.

GEANAKOPLOS, J. and MILGROM, P. (1991), "A Theory of Hierarchies Based on Limited Managerial Attention", *Journal of the Japanese and International Economies*, **5**, 205-225.

GUNN, C. (2000), "Global Equity Management and Valuation", *AIMR Conference Proceedings*, 65-74.

HARRIS, M. and RAVIV A. (2001), "Organizational Design", working paper, University of Chicago.

HART, O. and MOORE, J. (2000), "On the Design of Hierarchies: Coordination Versus Specialization", working paper, Harvard University.

JORION, P. (2000), *Value at Risk*, McGraw-Hill.

KEREN, M. and LEVHARI, D. (1979), "The Optimum Span of Control in a Pure Hierarchy", *Management Science*, **25**, 1162-1172.

KEREN, M. and LEVHARI, D. (1983), "The Internal Organization of the Firm and the Shape of Average Costs", *Bell Journal of Economics*, **14**, 474-486.

LITTERMAN, R. (1996), "Hot Spots and Hedges", *Journal of Portfolio Management*, Special Issue, December, 52-75.

MARSCHAK, T. and RADNER, R. (1972), *Economic Theory of Teams*, Yale University Press.

NAIK, N. and YADAV, P. (2001), "Do Correlated Exposures Influence Intermediary Decision-making? Evidence from Trading Behavior of Equity Dealers", working paper, London Business School.

ORBAY, H. (2002), "Information Processing Hierarchies", *Journal of Economic Theory*, forthcoming.

PRAT, A. (1997), "Hierarchies of Processors with Endogenous Capacity", *Journal of Economic Theory*, **77**, 214-222.

RADNER, R. (1993), "The Organization of Decentralized Information Processing", *Econometrica*, **62**, 1109-1146.

RADNER, R. and VAN ZANDT, T. (1992), "Information Processing in Firms and Returns to Scale", *Annales d'Economie et de Statistique*, **25/26**, 265-298.

SHARPE, W. (1985), *Investments*, Prentice-Hall.

VAN ZANDT, T. (1999a), "Decentralized Information Processing and the Theory of Organizations", in SERTEL, M. (ed.), *Contemporary Economic Issues, Volume 4: Economic Behavior and Design*, MacMillan Press.

VAN ZANDT, T. (1999b), "Real-Time Decentralized Information Processing as a Model of Organizations with Boundedly Rational Agents", *Review of Economic Studies*, **66**, 633-658.

VAN ZANDT, T. (2000a), “Real-Time Hierarchical Resource Allocation with Quadratic Costs”, working paper, INSEAD.

VAN ZANDT, T. (2000b), “Structure and Returns to Scale of Real-Time Hierarchical Resource Allocation”, working paper, INSEAD.

VAN ZANDT, T. and RADNER, R. (2001), “Real-Time Decentralized Information Processing and Returns to Scale”, *Economic Theory*, **17**, 497-544.

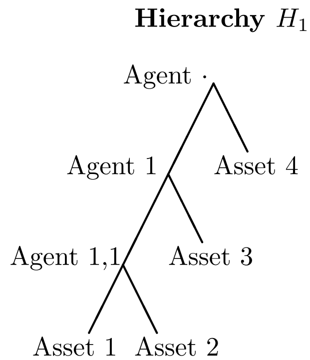

Figure 1: The notation for a simple hierarchy. There are three agents, agent ·, agent 1, and agent 1,1. For agent 1, for example, we have $S(1) = 1$, since this agent has one subordinate, agent 1,1. We also have $M(1) = 1$ and $A_M(1) = \{3\}$, since agent 1 examines asset 3. Finally, we have $N(1) = 3$ and $A_N(1) = \{1, 2, 3\}$, since the assets under agent 1's control are 1, 2, and 3.

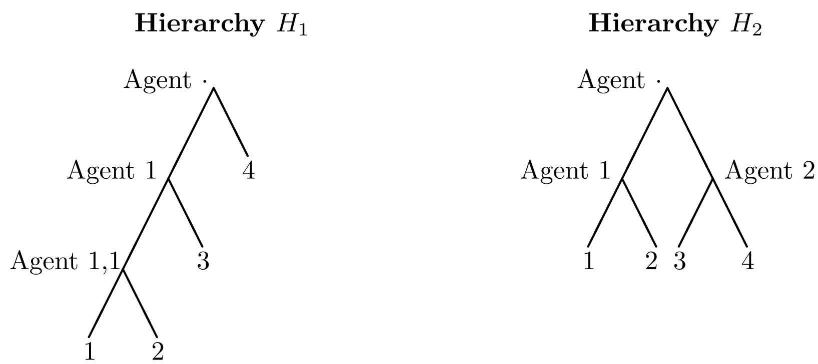

Figure 2: Two hierarchical structures for $N = 4$, $K = 2$.

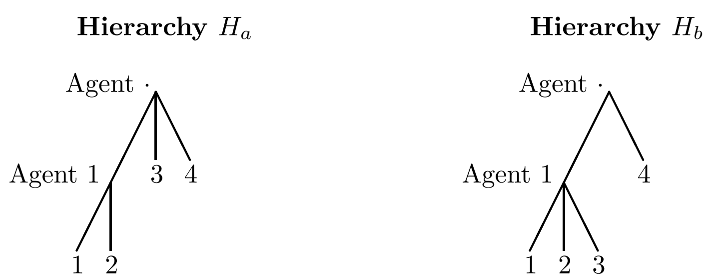

Figure 3: Optimal hierarchical structures for $N = 4$, $K = 3$.

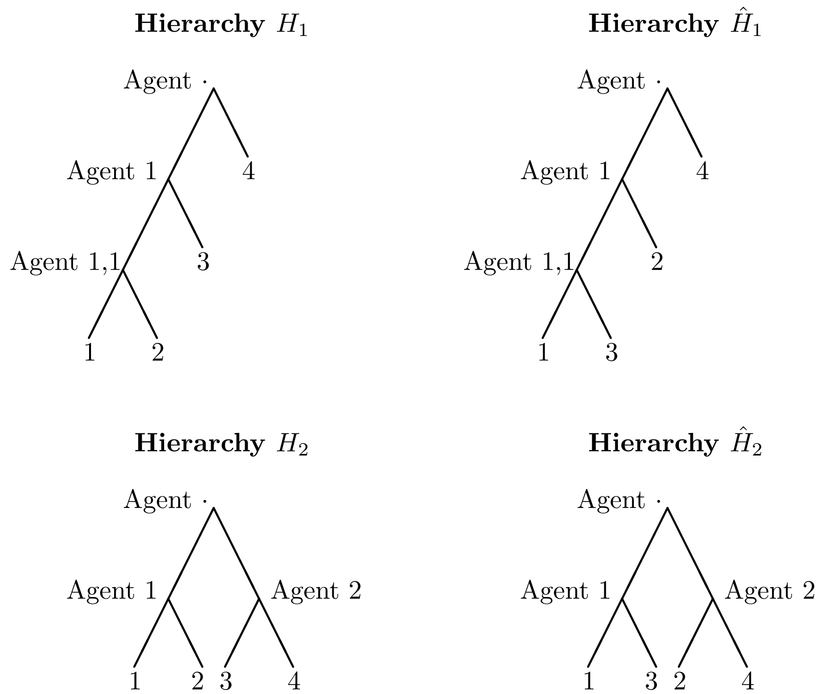

Figure 4: Hierarchical structures and assignments of assets to agents for $N = 4$, $K = 2$, and two groups of assets.

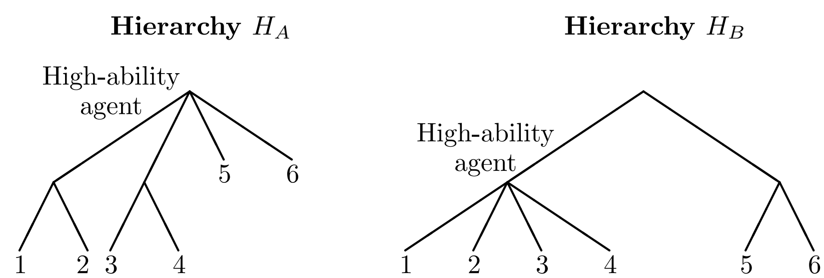

Figure 5: Two hierarchical structures for $N = 6$, $K = 2$, three groups of assets, and a high ability agent.