Speed and Quality of Collective Decision Making: Incentives for Information Provision*

Hans Peter GRÜNER
University of Mannheim and CEPR, London

Elisabeth SCHULTE
University of Mannheim

February 2006

*We thank Enno Mammen, Thomas Gall, Nicola Persico, Jörg Oechssler, Frank Riedel, Benny Moldovanu, Michel LeBreton and participants of workshops and seminars at Erasmus University Rotterdam, Gerzensee, Bonn University, Mannheim University, UCLA, at ZEI Bonn, Université Toulouse and Aachen University, and participants of the 2004 ENTER Jamboree in Barcelona and of the Econometric Society World Congress 2005 in London for helpful comments and suggestions. Elisabeth Schulte acknowledges financial support through the DFG research group on the institutionalization of international negotiation systems (IINS).
Correspondence address: University of Mannheim, Department of Economics, L7, 3-5, D-68131 Mannheim, Germany; email: hgruener@rumms.uni-mannheim.de; elisabeth.schulte@vwl.uni-mannheim.de.

Speed and Quality of Collective Decision Making: Incentives for Information Provision

Abstract This paper provides a game-theoretic extension of Radner's (1993) model of hierarchical information aggregation. It studies the role of the hierarchy design for the speed and quality of a collective decision process. We consider a programmed network of managers. The network's task is to aggregate information in form of a set of n data items and to make an informed decision. The programme describes how information is processed within the network. Each manager benefits from reaching an accurate decision but suffers from an individual cost of unobservable effort. Effort has to be provided in order to understand the information contained in a data item properly. Complementarity of effort provision arises endogenously in our model. Given these complementarities, decentralized information processing increases incentives for information provision. Under certain conditions, the delay-minimizing reduced tree proposed by Radner (1993) is the only efficient organization along the dimensions speed and quality of decision making.

Keywords: Information processing, hierarchies, incentives for information provision.

JEL D23, D70, D83, L22, P51.

1 Introduction

In complex decision procedures, the aggregation of decision-relevant information plays an important role. In order to make a good decision, the relevant information should be processed accurately. In order to reach a decision quickly, the information aggregation procedure should be fast. In general the design of an institution which governs a decision affects both dimensions of the outcome, the quality of the decision as well as the speed at which the decision is reached. While each of these dimensions has been studied in recent strands of the theoretical literature, very little is known about the nature of the trade-off between speed and quality of decision making. This paper provides a game theoretic extension of Radner's (1993) model of decentralized information aggregation and analyzes the impact of the institutional design on the speed and quality of collective decision processes in this extended framework.

The aggregation of information is a key task of almost all complex economic organizations (Radner, 1993). A trade-off in institution design between speed and quality of the information aggregation process should obtain if the complexity of individual tasks is high. Consider a financial institution which has to identify the best investment opportunity for newly available funds. The evaluation of an investment opportunity is a complex and time consuming task and it is natural to delegate this task to several individuals (experts) in order to save on processing time. Under such a decentralized arrangement incentive problems naturally arise. These incentive problems are likely to affect the quality of the overall outcome. Many other large organizations (firms or government agencies) which decide upon the proper allocation of a given budget face similar problems. Incentive problems of the sort studied in this paper may also arise in firms which operate in several regional markets when headquarters have to rely on aggregates of regional customer data. The organization of information processing will determine both the delay of a decision and the individuals' incentives for information provision.

In this paper, an institution is modeled as a programmed network of agents.[1] The programme describes how information is processed within the network. It specifies the assignment of information to the various decision makers and the sequence of information processing and message stages. Thereby, it underlies certain restrictions related to the time that some operations consume. The members of the network derive state-dependent utility from the decision. The state of the world can be learned by aggregating information. The decision quality may suffer from a simple moral hazard problem. We assume that agents

[1] See Radner (1993) or Radner and Van Zandt (1992) for a detailed introduction into this concept.

can understand information properly only if they provide the necessary amount of effort. Individual effort is unobservable for other agents. A programmed network therefore induces a game among its members. We characterize the equilibria of this game in terms of (i) the speed at which the decision is reached and (ii) the cost and quality of the decision. Our objective is to study incentives for information provision in such networks and to obtain characteristics of efficient institutions.

In a first step we study the case of a single decision maker whose task is to work on n objects. It turns out that for some particular distributions of these n random variables, the marginal gain from effort provision increases in the number of objects the decision maker already read. This yields a new rationale for decentralized information processing: If the information processing task is spilt into subtasks, and the number of tasks assigned to each agent is sufficiently small, every agent is willing to provide effort while a single decision maker is not. Parallel information processing is not only faster than full centralization – it also provides better incentives for the participating managers. The reason is the complementarity of effort of different managers which arises endogenously in our model.

In a second step we study incentives for information processing in programmed networks. First, we investigate a simple two-tier hierarchy and show that there are endogenous limits to decentralization. Then, we address the issue of efficient hierarchy design in more general terms. In his seminal paper Radner (1993) studies the design of hierarchical structures when information processing takes time. This leads him to propose a hierarchical structure – which he calls a reduced tree – within which information is processed at maximum speed. The virtue of the reduced tree is that processors on all levels simultaneously process information right from the beginning. This guarantees maximal parallel processing and minimizes the delay of the entire information processing procedure. We study the properties of the reduced tree in our modified setup. It turns out that it has virtues that lie beyond the efficient treatment of information characterized in Radner's paper.

The most interesting result of our analysis is that the speed and the quality of decision procedures may go hand in hand. Under certain conditions, an organization with a given number of managers maximizes decision quality if and only if it minimizes decision delay. The reason is that an "equal" sharing of processing tasks is a precondition for both objectives. We show that any programmed network N in which all agents provide effort can – under certain conditions – be replaced with a reduced tree $R(N)$ with the same number of processors that also has a full effort equilibrium. Therefore $R(N)$ generates the same decision quality and a faster decision. Consequently, for these decision problems the efficiency frontier is constituted of reduced trees.

This paper is most closely related to a recent literature on organization design which draws on insights from computer science, starting with Radner (1992, 1993), Radner and Van Zandt (1992). The reduced tree is designed for one-shot problems (to which we restrict attention). These are problems in which there is only one set of data to be processed, or the processing of the data is finished before another calculation task occurs. Van Zandt (1997, 1998) and Meagher, Orbay and Van Zandt (2001) study the case when new data comes in before the processing of the old set is finished. Orbay (2002) adds the frequency with which new data arrives as a new dimension to the analysis of efficient hierarchies. Prat (1997) studies hierarchies in which some managers are able to work faster than others, and the wage a manager is paid is a function of his ability. It turns out that with these modifications – except for the one made by Prat (1997) – the reduced tree is still (close to) efficient. Bolton and Dewatripont (1994) allow for specialization, which reduces the time an agent needs to understand information he handles frequently. In their model, the trade-off between specialization and communication costs determines the extent and the form of decentralization.

In Radner's model – and in most of the information processing literature which followed – individuals are thought of as machines, perfectly doing what they are programmed to do. There are at least two important features of information processing by human agents which may require modifications of this basic model. First, an individual's calculation ability might not be perfect, i.e. occasionally individuals may make mistakes. Second, when delegating tasks, one has to make sure that the agent has an incentive to perform them. Thus, there emerges another dimension for the evaluation of hierarchies: the quality of the decision.

The joint analysis of speed and quality of hierarchical decision processes has previously been studied in Jehiel (1999) and in Schulte and Grüner (2004). Jehiel considers the case where some signals get lost in the hierarchy with an exogenous probability, depending on the size of the groups of which the hierarchy consists. Schulte and Grüner study the role of the hierarchy design when individuals make mistakes with an exogenously given probability. In the present paper, the quality of collected information is endogenously detemined by the actions of self-interested agents. The result that the reduced tree provides a (weakly) better decision quality than other organizations is the same.

Our paper is also related to a huge game theoretic literature that studies incentives in hierarchies, in particular Aghion and Tirole (1997), Mookherjee and Reichelstein (1997), and Melhumad, Mookherjee and Reichelstein (1995). These papers consider problems in which certain tasks as well as authority have to be delegated (and sub-delegated) to (and by) agents whose interests diverge from that of the principal. Delegation involves a loss of control for the principal, but strengthens the incentives for the agent. In our model, all agents have the same objective: The available information shall be processed as accurately as possible in order to make an informed decision.

Winter (2004) studies incentive provision in a hierarchy via a transfer scheme. In his paper, the n tasks are assigned to n agents right from the beginning. He does not allow for the possibility to assign tasks differently, nor that one agent performs all of them. A hierarchical position is identified by a payment to the agent. Payments to agents in higher positions are high to induce them to provide effort independently of what the other agents are doing. Increasing returns to effort provision imply that the next agent, knowing that there is someone who provides effort, can be paid less to ensure his effort provision. The optimal mechanism is symmetric if and only if the technology is such that the last effort provision is the least productive one. It is fully disrcriminating (i.e. assigning different transfers to every agent), if and only if the technology has increasing returns to scale. Unlike Winter (2004), we are not interested in a transfer scheme that ensures effort provision, but rather in the effect of task assignment on effort provision. A related moral-hazard-in-team problem is studied in Dewatripont and Tirole (2004). They consider a sender-receiver game, where the sender has payoff-relevant information for the receiver, and both must invest costly and unobservable effort for the receiver to understand the information. In our model, communication is costless for the sender, but he has to provide effort to acquire the information in the first place.

Other papers derive decentralized (hierarchical) organizations from technology. Crémer (1980) considers a problem of resource allocation under constraints on managerial time and finds that hierarchical organizations increase the amount of information that can be applied to a particular decision. Garicano (2000) deals with the optimal organization of knowledge. Knowledge is costly to acquire and, once acquired, may be used to help other agents. This yields a hierarchical structure in which the higher level agents are knowledgeable about how to solve the most difficult or least frequent problems. Rosen (1982) has a paper in which a hierarchical structure emerges due to the need to supervise production (and supervision).

The paper is organized as follows: In the next section, we present the model of the decision problem and the organization of decentralized information processing. Next, in Section 3 we provide some examples for our framework. In Section 4, we study the incentives for information processing in a single decision maker organization and in a two-tier hierarchy, the centralized tree. In Section 5, we turn to the analysis of the reduced tree. We present our conclusions concerning the design of efficient organizations in Section 6. In the final section we summarize our main results and discuss some possible extensions of our model.

2 The Model

2.1 The Decision Problem

In this section, we introduce the general model framework. Examples are presented in Section 3. There is a collective decision $d \in D$ to be made which generates a state-dependent payoff $u(d, x)$ for the members of society. The state of the world $x \in X$ is the aggregate of n information variables[2] x_i, where $x_i \in X$, $i = 1, ..., n$. The vector of information variables $\mathbf{x} = (x_1, ..., x_n)$, where $\mathbf{x} \in X^n$, is transformed into the aggregate x via a function $f_n : X^n \rightarrow X$.

A simple example for such an aggregation is an addition, where the values of the information variables are aggregated to the sum of the values. Another example is a maximization, where the aggregation yields the object with the highest value.

The information variables x_i are drawn from commonly known distribution functions $\phi(x_i)$, $i = 1, ..., n$. The realization can be learned only if costly effort is provided. In order to learn the aggregate state of the world x, the information variables have to be processed, that is $f_n(\mathbf{x})$ has to be calculated. The information processing task consists of $n-1$ binary operations on the information variables. The binary operation is a function that transforms two information variables x_i and x_j into a partial result, a new information variable $f_2(x_i, x_j)$. The operation is commutative and associative, that is $f_n(\mathbf{x}) = f_{n-1}(x_1, ...x_{i-1}, f_2(x_i, x_{i+1}), x_{i+2,}...x_n)$. Hence information processing can be decentralized, i.e. raw data can be aggregated to partial results[3] by some agents which can be further aggregated by other agents. The information variables are independent random variables. Since there is common knowledge about the distributions $\phi(x_i)$ and the processing task, the distributions of partial aggregates and the distribution of the aggregate x are common knowledge as well.

An agent can learn the realization of an information variable by investing unobservable effort. Effort comes at a cost c. If an agent does not provide effort when reading an

[2] We use the terms "information variables", "data items" and "objects" interchangebly.

[3] We use the term "partial result" for an information variable that is the outcome of an aggregation operation and "raw data (item)" for an information variable that has not yet been combined with another one.

information variable, he does not receive any information about its realization. However, if the realization is not learned, it may be guessed. This is the moral hazard problem we study in this paper.

Since agents' preferences concerning the decision are identical, we consider only equilibria which have the following property. If a random variable is guessed, then the "best guess" is formulated in the sense that it maximizes expected utility when used as input for subsequent aggregation operations. In case information variable x_i is guessed, the input to the operation $f_2(.)$ is the best guess. We will be more explicit about the best guess in subsection 2.4, in which we define agents' preferences.[4]

When information processing is done in a decentralized manner, the production of the final result requires that the partial results are submitted to other agents within the organization. That is, an agent j who is assigned a subtask of the information processing problem is also asked to send a report $r_j \in R_j$ to another agent. R_j is the set of possible results of j's calculations. From the point of view of the receiver, the report produced by another agent is a random variable. We assume that reading a report sent by another agent also requires effort at a cost c.

2.2 Decentralized Information Processing

Since the aggregation operations are commutative and associative, information processing can be decentralized in a programmed network.[5] The program describes how information processing tasks are distributed within the network. There are n data items that have to be aggregated by P agents called managers. Managers are endowed with an inbox, a processing unit and a memory.

The program consists of assignments $a : \{1, ..., n\} \to \{1, ..., P\}$, and $p : \{1, ..., P\} \to \{1, ..., P\}$, and specifies R_j, $j = 1, ..., P$. That is, it assigns the raw data to the managers' inboxes, and specifies for each manager to whom to report and the reporting space. Moreover, the program gives instructions to each agent when to read an item, which operation to perform, and when to send a report. We assume that reading an object takes one unit of time, and that the other tasks do not cause any delay. Hence, in one unit of time, a manager is able to perform the following tasks: (i) reading an object from his inbox into his processing unit, (ii) aggregating it to what is stored in his memory, and (iii) sending a report.

[4]Note that the best guess need not be unique, e.g. when guessing which of two objects has a higher value, both of them are best guesses, if the objects' values are identically distributed.

[5]See Radner (1993) for a more detailed description of a programmed network.

At the beginning of the procedure, the memory is empty. After having finished processing the items in his inbox (which may contain raw data as well as partial results provided by subordinates), a manager sends a report to his direct superior. The top manager's report is the final result.

All individuals in the network are interested in the quality of the overall result but they also suffer from an individual cost of providing effort. Moreover, they suffer from delay in decision making.

2.3 The Reduced Tree

By decentralizing information processing, one can economize on time. Radner (1993) derived the optimal (fastest) organization for any given number of managers P and any number of objects n. He proposes the following hierarchy design: Number the managers subsequently from 1 to P and assign $\frac{n}{P}$ objects to each manager. (If $\frac{n}{P}$ is no integer, assign the largest integer smaller than $\frac{n}{P}$ to each manager and another one to the first $n \bmod P$ ones.) When finished with the initially assigned objects, let manager P report to $P-1$. Manager $(P-1)$ therewith becomes P's immediate superior. Let $P-2$ report to $P-3$, etc. Note that if P is an odd number, manager 1 remains unconnected. Renumber the managers who did not yet report to somebody in an appropriate manner and repeat the procedure until a single manager remains (the top manager).[6] The n objects are aggregated in $\lfloor \frac{n}{P} \rfloor + \lceil \log_2 (P + n \bmod P) \rceil$ units of time.[7] We call a hierarchy which is constructed in the way described above a *reduced tree.* Figure 1 depicts the construction of a reduced tree for $P = 8$ managers. A circle represents a manager, an arrow from manager i to manager j means that i reports to j. The figure also illustrates how the managers are renumbered.

Note that the reduced tree is not the unique organization that implements the decision with minimum delay (see Footnote 6). However, if $n = bP$ and $P = 2^a$, where $b \in \mathbb{N}^+$, $a \in \mathbb{N}^+$, all delay minimizing organizations are isomorph and look like those depicted in Figure 2. This figure illustrates a reduced tree of 8 managers processing 24 data items. A circle represents a manager, and a link from a manager to a manager further above means that the upper level manager reads the result provided by the lower level manager.

For b units of time, all managers are busy with their initial objects. Thereafter, initial objects have been aggregated to P partial results, so only $\frac{P}{2}$ managers are needed for further

[6] Note that this construction of a reduced tree slightly deviates from the one in Radner (1993). However, it yields the same delay in information processing. Moreover, if P is a power of 2, it yields the same organizational form. The main difference with respect to the organizational form is that the top manager's

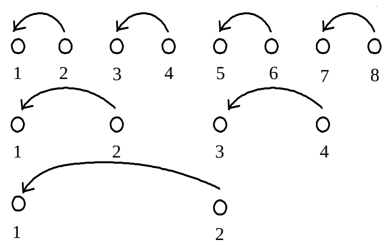

Figure 1: Construction of a reduced tree.

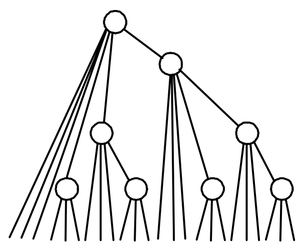

Figure 2: Reduced tree.

aggregation. In the next unit of time, there is still work for only $\frac{P}{4}$ managers, etc.. Note that in this organization, in any moment of time, information processing is done by either all or the sufficient number of managers. Partial results are produced "just in time", that is a report read in time unit t was produced in $t-1$. Moreover, the sender as well as the receiver worked in every time unit till $t-1$.

To see that for the case $n = bP$, $P = 2^a$, all delay minimizing organizations must be isomorph to a reduced tree, first note that given the reporting structure, any other assignment of raw data would strictly increase delay and vice versa. Second, consider an organization with a different assignment of raw data and a different reporting structure. At least one agent has to aggregate $b^+ > b$ raw data items. After b^+ units of time, there are at least $1 + \left\lceil \frac{2^a-1}{2^{b^+-b}} \right\rceil$ partial results left for further aggregation. The minimum delay for these tasks is at least $\left\lceil \log_2 \left(1 + \frac{2^a-1}{2^{b^+-b}}\right) \right\rceil$, and hence total delay for this organization is at least $b^+ + \left\lceil \log_2 \left(1 + \frac{2^a-1}{2^{b^+-b}}\right) \right\rceil = b^+ + \left\lceil \log_2 \left(2^a + 2^{b^+-b} - 1\right) + b - b^+ \right\rceil > a + b$, the delay achieved by the reduced tree.

2.4 Preferences and the Moral Hazard Problem

All agents have identical preferences, which can be represented by the following utility function:

$$U_j = u(d, x) - e_j c,$$

where e_j measures how many times individual j provides effort.

Hence, utility is additively separable in the utility derived from the decision and the cost of effort provision. Agents who are not involved in information processing do not contribute any effort. We assume that they are plenty enough such that utility from better decisions in the aggregate outweigh information processing cost.

For each x, there exists a $d^*(x) \in D$ maximizing $u(d, x)$. We assume that $u(d^*(x), x) = u^*$ for all x. Alternative to processing the whole set of information, agents may guess part or all of the information. If information processing takes place in a decentralized manner, that is if agent j reports r_j to another agent, then the information contained in the report is learned by that agent only if he provides effort. We call an object *processed properly* if every agent who is supposed to read an item in which the object is contained provides effort and

workload is (weakly) lower in our definition of the reduced tree than in Radner's original definition.

[7] See Radner (1993) for a proof that this is indeed the minimum delay for processing n data items with P managers.

truthfully reports to his superior.

Let m denote the number of information variables which are not processed properly, and let $f_{m'}\left(x^{m'}\right)$ denote the aggregate of $m' = n - m$ properly processed objects. We assume that agents are able to formulate the best guess, i.e. to make a decision that maximizes expected utility given the limited information they have, i.e. there exists $d^{m'}$ maximizing $E\left[u\left(d,x\right)\left|f_{m'}\left(x^{m'}\right)\right.\right]$, where $0 \leq m' \leq n$. In our analysis, we restrict attention to problems for which the best guess for the aggregate coincides with the aggregate of best guesses. This could easily be generalized, but would require richer reporting spaces.[8] Let $f_m\left(x^m\right)$ denote the aggregate of the best guesses for the m guessed information variables. We have that $d^{m'} = d^*\left(f_2\left(f_{m'}\left(x^{m'}\right), f_m\left(x^m\right)\right)\right)$.

Definition 1 *The cost of guessing m out of n information variables is denoted*

$$k(m) = u^* - E[u(d^{m'}, x)|f_{m'}(x^{m'})]. \tag{1}$$

The cost of guessing an additional item,

$$k\left(m+1\right) - k\left(m\right) = E\left[u\left(d^{m'+1}, x\right)\left|f_{m'+1}\left(x^{m'+1}\right)\right.\right] - E\left[u\left(d^{m'}, x\right)\left|f_{m'}\left(x^{m'}\right)\right.\right],$$

is the loss of expected utility due to the fact that the decision is based on less information and due to a higher variance of the distribution of the aggregate.

The specification of preferences, the moral hazard problem and the underlying programmed network fully describe a game. In this game a strategy is a plan that fixes (i) when to provide effort (possibly history dependent) and (ii) what to report (again possibly history dependent). The equilibrium concept we use is Nash equilibrium. We restrict attention to equilibria in which agents report either the truth or their best guess. Moreover, we use Pareto-dominance as an equilibrium selection criterion where applicable.

In our analysis, we represent preferences in terms of the cost function $k\left(m\right)$. In the next subsection, we state properties of this function that are important for our analysis. In the following section we present examples for which the cost of guessing has these properties.

[8] In the next section, we present examples for which our assumptions hold. An example for a decision problem in which this requirement is not met is one that requires Bayesian updating: Instead of processing agents' guesses about the realizations of imperfect signals, the decision should be based on more complex reports on processed information.

2.5 Properties of k(m)

Our results hold for decision problems with general distributions $\phi(.)$ and individual preferences $u(.)$ which define the cost function $k(.)$ such that it satisfies two particular concavity properties. These properties are:

Property P1: The function $k(m)$ is positive, weakly increasing and weakly concave.

Property P2: The function $k(m)$ has the properties that

$$k\left(2^{j}m\right)-k\left(2^{j-1}m\right)\leq k\left(2^{j+1}m\right)-k\left(2^{j}m\right),j\in\mathbb{N}^{+}$$

and

$$\frac{k(m)}{m}\leq k(2m)-k(m),m\in\mathbb{N}^{+}.$$

For $m=0$, the state of the world is known with certainty, so the best decision can be made and utility u^{*} obtained. However, if an information variable is guessed, this introduces uncertainty about the state of the world and decreases expected utility. The property that $k(m)$ is weakly increasing means that a decision based on more information is weakly better than a decision based on less information. The function is concave if the more information is guessed already, the less additional guessing harms. This is the case if the mistakes which are caused by guessing partly outweigh each other. Property P2 requires that this effect is not too strong.

It is useful to define the function $c(m)=\frac{k(m)}{m}$, which measures the cost of guessing m items per item guessed. Given P1, and the fact that $k(0)=0$, $c(m)$ is non-increasing in m.

2.6 Classical surplus and efficiency

We evaluate an organization along two dimensions: (i) the delay within which a decision is reached and (ii) classical surplus. Recall that an object is called processed properly if every agent who reads the information in which this object is contained provides effort and truthfully reports to his superior. Call the number of objects that are not processed properly m, where $m\in\{0,1,..n\}$. The *classical surplus* of the economy (the sum of utilities) is defined by

$$v:=s\cdot\left(u^{*}-k(m)\right)-\sum_{j=1}^{P}e_{j}c, \tag{2}$$

where s denotes population size, and P denotes the number of agents involved in information processing.[9] In what follows, we assume that processing the whole set of data maximizes the classical surplus of the economy. Note that generally adding agents to the hierarchy generates a cost: If every hierarchy member provides effort, the total cost of effort provision is $(n + P - 1)\, c$. Hence there are two possible reasons for decentralizing information processing: (i) incentives or (ii) the speed of decision making.[10]

Definition 2 *A hierarchy design is efficient for a processing task* (n, c) *if it is not dominated on one of the following dimensions by another design that performs at least equally well on the other dimension: (i) speed (ii) classical surplus.*

3 Examples

In this section, we present concrete decision problems which fit into the general framework introduced above. We derive the cost functions $k(m)$ for these problems and show that they have the Properties P1 and P2.

3.1 Addition of normally distributed random variables

Consider a problem in which the utility-maximizing decision is to match the aggregate state of the world, $d = x$. Agents dislike deviations from the optimal decision. The aggregate state of the world is given by the sum of n i.i.d. normally distributed random variables. We have: $X = D = \mathbb{R}$. Moreover, we may restrict the reporting spaces R_j, where $j = 1, ..., P$, to $\mathbb{R}$ as well. The aggregation operation is given by $f_2(x_i, x_j) = x_i + x_j$, and $f_n(\mathbf{x}) = \sum_{i=1}^{n} x_i$. A sequence of instructions in a programmed network solving the aggregation task could look like this: "Agent 1: in time unit 1, take object 1, read x_1, and store x_1. In time unit 2, take object 2, read x_2, calculate $x_1 + x_2$. Send the result of this calculation to agent 2." A strategy for agent 1 specifies an effort decision in time unit 1, another effort decision in time unit 2 (conditional on effort provision in time unit 1) and a report to be send to agent 2, conditional on effort provision and information learned.

[9]We consider the total cost of effort provision as a measure of cost. See Meagher and Van Zandt (1998) for a discussion of appropriate cost measures for decentralized information processing tasks.

[10]Instead of incorporating the time dimension explicitly into the utility function, we evaluate decision procedures on the dimensions speed and quality separately. We do so in order to avoid a bias in our results concerning the quality dimension towards faster structures built in by assumption.

Agents' preferences are represented by the following utility function:

$$u(d,x) = -|d-x|.$$

The aggregate x is the sum of n normally distributed random variables, $x_i \sim N\left(0, \sigma_{x_i}^2\right)$, $i = 1...n$. We have $x \sim N(0, \sigma^2)$, where $\sigma^2 = n\sigma_{x_i}^2$. The best guess of the total (or a partial) sum is the median of its distribution.[11]

Hence, the cost of guessing m numbers is:

$$\begin{aligned} k(m) &= \int_{-\infty}^{\infty} |x| \cdot \frac{1}{\sigma (2\pi)^{\frac{1}{2}}} e^{-\frac{1}{2\sigma^2}x^2} dx = \\ 2\int_0^{\infty} |x| \cdot \frac{1}{\sigma (2\pi)^{\frac{1}{2}}} e^{-\frac{1}{2\sigma^2}x^2} dx &= \\ 2\frac{1}{\sigma (2\pi)^{\frac{1}{2}}} \int_0^{\infty} x \cdot e^{-\frac{1}{2\sigma^2}x^2} dx &= \\ 2\frac{1}{\sigma (2\pi)^{\frac{1}{2}}} \left(-\sigma^2 e^{-\frac{1}{2\sigma^2}x^2}\right)\Big|_0^{\infty} &= \\ 2\frac{\sigma}{(2\pi)^{\frac{1}{2}}} \cdot \left(-e^{-\frac{1}{2\sigma^2}x^2}\right)\Big|_0^{\infty} &= \\ & 2\frac{\sigma}{(2\pi)^{\frac{1}{2}}}. \end{aligned} \tag{3}$$

Using $\sigma^2 = m\sigma_{x_i}^2$, we get

$$k(m) = \sqrt{\frac{2m}{\pi}}\sigma_{x_i}, \tag{4}$$

which is a strictly increasing, concave function in m. Hence, Property P1 holds. Moreover, P2 holds for $m \geq 3$:

$$\begin{aligned} 2k(2m) &\leq k(4m) + k(m) \\ &\Leftrightarrow 2\sqrt{\frac{4m}{\pi}}\sigma_{x_i} \leq \sqrt{\frac{8m}{\pi}}\sigma_{x_i} + \sqrt{\frac{2m}{\pi}}\sigma_{x_i} \\ &\Leftrightarrow 4 \leq 3\sqrt{2}, \end{aligned}$$

and

$$\begin{aligned} \frac{k(m)}{m} &\leq k(2m) - k(m) \Leftrightarrow \sqrt{\frac{2}{\pi m}}\sigma_{x_i} \leq \sqrt{\frac{4m}{\pi}}\sigma_{x_i} - \sqrt{\frac{2m}{\pi}}\sigma_{x_i} \\ &\Leftrightarrow \frac{1}{\sqrt{2}-1} \leq m. \end{aligned}$$

[11] See DeGroot (1970), p. 232.

3.2 Quadratic utility

Consider next the above problem of adding random variables but with quadratic utilities,

$$u(d,x) = -(d-x)^2 .$$

In a setup with quadratic losses the best guess is the expected value and the loss equals the variance of the error. Given that the x_i are i.i.d. $k(m)$ is linear in m:

$$k(m) = m\sigma_{x_i}^2 . \tag{5}$$

It is easy to verify that this specification also satisfies Properties P1 and P2.

3.3 Binary distribution

Now consider a task of adding n information variables $x_i \in \{0,1\}$, where the information variables are independent random variables, and each outcome is equally likely. Again, let the utility function assume the form

$$u(d,x) = -|d-x| .$$

The expected utility-maximizing guess of the sum of m numbers is the median of its distribution, which coincides with its expected value $\frac{m}{2}$ here. The cost of guessing m numbers are:

$$k(m) = E\left(\left|\frac{m}{2} - \sum_{i=1}^{m} x_i\right|\right) . \tag{6}$$

Note that the true result assumes the value $i < \frac{m}{2}$ with probability $\frac{1}{2^m}\binom{m}{i}$. The loss of utility in this case is $\frac{m}{2} - i$. Note that deviations are symmetric. Hence:

$$\begin{aligned} k(m) &= 2\sum_{i=0}^{\lfloor \frac{m}{2} \rfloor} \frac{1}{2^m}\binom{m}{i}\left(\frac{m}{2} - i\right) \\ &= \frac{1}{2^{m-1}} \sum_{i=0}^{\lfloor \frac{m}{2} \rfloor} \binom{m}{i}\left(\frac{m}{2} - i\right) . \end{aligned} \tag{7}$$

It can be shown that this function can be represented by:

$$k(m) = \left\lceil \frac{m}{2} \right\rceil \frac{1}{2^{2\lceil \frac{m}{2} \rceil}} \binom{2\lceil \frac{m}{2} \rceil}{\lceil \frac{m}{2} \rceil},$$

which is a weakly increasing step-function with diminishing increments. This function satisfies a modified version of Property P1 (note that $k(m)$ is not concave), and satisfies P2.[12]

3.4 Project selection I

Our next example concerns a different aggregation task: the "max" operation. The task is to identify the best option out of a set of n alternatives. We have $X = I \times \Theta$, where $I = \{1, ..., n\}$ is the set of objects, and $\Theta = \mathbb{R}$ is the set of realizations of an object's quality. The decision concerns picking one object, i.e. $D = I$. The aggregation operation $f_2(x_i, x_j)$ yields x_i if $\theta_i \geq \theta_j$ and x_j else. The reporting spaces R_j can be restricted to the set of objects handled by j, $J \subseteq I$, for all j. A report means that j passes on one of his objects to his superior. We represent preferences by the following utility function:

$$u(d, x) = \begin{cases} 1, & \text{if } d = i : \theta_i \geq \theta_j, \forall j \in D \backslash \{i\} \\ 0, & \text{else.} \end{cases}$$

We assume that the distribution of quality is unknown. Hence ex ante, any object is the best one with equal probability. Guessing in this case means randomly choosing one of the objects, which yields expected utility of $\frac{1}{n}$. When reading an object i, θ_i is learned. However, if only one object's quality is learned, nothing is gained, because the probability that this object is the best one still is $\frac{1}{n}$. To make a better decision, the quality of at least two objects have to be learned, such that they can be compared. The better one of the two is the best object with probability $\frac{2}{n}$. Hence, the cost function for this example is given by:

$$k(m) = \begin{cases} \frac{m}{n} & \text{for } m = 0...n-1 \\ \frac{n-1}{n} & \text{for } m = n. \end{cases}$$

It is easy to see that this function has properties P1 and P2.

3.5 Project selection II

Another more complex project selection example relates to situations in which a project has to satisfy several criteria in order to be useful. Consider the task of selecting an object out of a set of objects O, where $|O| = o$. Each object is either useful or not. An object is useful if and only if it has n properties, $n < o$. There are n information items available, each of

[12]The proof that they are satisfied can be found in the appendix of the working paper version of this paper, which is available from the authors upon request.

which identifies $x_i \subset O$, the set of objects which have property i. The reporting spaces are given by the powerset of O. The aggregation task is to exclude objects which are not useful, i.e. $f_2(x_i, x_j) = x_i \cap x_j$. The aggregate state of the world x is given by the set of useful objects, $x = \bigcap_{i=1...n} x_i$. The utility from selecting object j is:

$$u(j) = \begin{cases} 1, & \text{if } j \in x \\ 0, & \text{else.} \end{cases}$$

If m out of n information variables are not processed properly, the best guess is a random choice out of the set $x_1 \cap x_2 .. \cap x_{n-m}$. It is easy to specify the information environment such that the cost of guessing has Properties P1 and P2. For example, assume that each object j has at least $n-1$ properties, and that each property is satisfied by all but one object. The probability that the randomly chosen object is useful is $(o-n)/(o-n+m)$. Hence, the cost of guessing is

$$k(m) = \frac{m}{o-n+m}.$$

Simple calculations verify that $k(m)$ satisfies Properties P1 and P2, if o is sufficiently larger than n.

4 The role of hierarchies

4.1 A single decision maker

It is useful to first consider the case of a single individual which has to aggregate n information items. Suppose that it is optimal for the individual to provide effort on all objects. The concavity of the cost function $k(m)$ guarantees that the initial effort decision is time consistent, i.e. a single decision maker who prefers providing effort n times to providing no effort will continue to provide effort, once he has provided effort the first time. The single decision maker's optimal strategy is stated in Propsition 1.

Proposition 1 *Let $k(m)$ fulfill Property P1. The optimal strategy for the manager in the one-player game is to provide effort on all of the n items if $c \leq c(n)$, and not to provide any effort otherwise.*

This proposition points out a major advantage of decentralized information processing. If $k(m)$ is strictly concave, and hence $c(m)$ strictly decreasing, for $c > c(n)$, a single decision maker would not provide any effort. However, if $c < c\left(\frac{n}{2}\right)$, he would be willing to process half the set, if someone else had already done the other half. Hence, for $c\left(\frac{n}{2}\right) > c > c(n)$, there exists an equilibrium in which the whole data set is processed properly when processing tasks are decentralized, while a single decision maker would guess the realization of the aggregate state of the world.

Whenever a task is divided into several parts which have to be performed by different individuals, the marginal gain from performing the own part must be sufficiently large in order to ensure the existence of a full effort equilibrium. If the function $k(m)$ is concave, the marginal gain of performing a task is the higher, the more tasks are already completed. Hence, distributing the processing tasks to more managers increases the marginal gain of each manager's processing activities. Therefore hierarchies may have the advantage of providing better incentives for information processing. In particular, if $k(m)$ is strictly concave, any hierarchical design provides better incentives for information processing than a single decision maker has.

4.2 The equilibrium concept

We now turn to the analysis of the equilibrium in a hierarchy that consists of more than one single agent. An agent's strategy specifies (i) whether or not the agent provides effort whenever the program asks him to read an object (depending on the history, i.e. the message received before and the data read before) and (ii) the message to be sent to the superior depending on the history. Given that nature first chooses the realizations of all random variables, the game does not have any subgames except the game itself. Therefore we concentrate on Nash equilibria.

We restrict the analysis to equilibria where agents who are not perfectly informed about their items make a best guess when they report to their superior. The superior in turn uses this guess in his further calculations.[13]

4.3 Centralized tree equilibria

We begin our analysis by studying one particular class of hierarchies called centralized trees. The centralized tree has two layers, one layer (lowest level) in which raw data is processed and another one (consisting of one manager, the top manager) in which the reports of the lowest level managers are aggregated. Raw data is distributed equally to the $P-1$ lowest

[13]One can easily exclude signalling equilibria where (i) the sender maps his information one to one into signals and (ii) where these signals are not properly used by the receiver.

level managers.

Thus, in a centralized tree, the $P-1$ lowest level managers process $\frac{n}{P-1}$ items each (on average) and send their partial results to the top manager's inbox. This procedure takes $\left\lceil \frac{n}{P-1} \right\rceil$ units of time. The top manager can start as soon as the first subordinate has finished his tasks, and needs again $P-1$ units of time to come to the final decision. The decision procedure takes $\left\lfloor \frac{n}{P-1} \right\rfloor + P - 1$ units of time.

One obvious disadvantage of any centralized tree is that an equilibrium where all players choose not to provide effort exists independently of the parameter values and functional forms. It remains to be studied whether there are additional equilibria where some or all players provide effort. In deed we find that several such equilibria may exist. Moreover, these equilibria may be ranked according to the Pareto-criterion. First it is useful to characterize strategy profiles as follows:

Definition 3 *Consider a centralized tree. A strategy profile with the following properties is called an l-effort strategy profile, $P-1 \geq l \geq 1$, if*
(i) l lowest level players provide effort,
(ii) $P-1-l$ lowest level players do not provide effort,
(iii) the top player provides effort when reading the report of a low level player who belongs to the first group but not when he reads the report of a player who belongs to the second group.

Definition 4 *An l-effort strategy profile is an l-effort equilibrium if it is a Nash Equilibrium of the game played by the members of the centralized tree.*

In order to avoid case differentiations that add few further insights, we assume in this section that n is a multiple of $P-1$.[14] Proposition 2 describes the set of equilibria when all agents on the lowest level process the same number of items.

Proposition 2 *Let $k(m)$ fulfill Property P1.*
(i) Every centralized tree has an equilibrium called no effort equilibrium where no player ever provides effort.
(ii) Assume that n is a multiple of $P-1$. There exists a threshold (l) such that an l-effort equilibrium exists if and only if $c \leq \gamma(l)$, $\gamma(l) > 0$.
(iii) Assume that n is a multiple of $P-1$. If an l-effort equilibrium exists, where $1 \leq l \leq P-2$, then there exists an $l+1$-effort equilibrium.
(iv) The no effort equilibrium and the l-effort equilibria are the only equilibria of the game

[14] A full analysis taking into account integer constraints is available from the authors upon request.

played by the members of a centralized tree.
(v) The full effort equilibrium Pareto dominates all other equilibria whenever it exists.

PROOF (i) Obvious.

(ii) A lowest level player has an incentive to provide effort $\frac{n}{P-1}$ times given the strategies of the other players if and only if $\frac{n}{P-1}c \leq \left(k\left(n-(l-1)\frac{n}{P-1}\right)-k\left(n-l\frac{n}{P-1}\right)\right)$. It is obvious that it is a best response for the lowest level players not to provide effort if the top player does not provide effort when reading their partial results. The top player has an incentive to provide effort l times reading the partial results of his subordinates who provided effort if and only if $lc \leq k(n)-k\left(n-l\frac{n}{P-1}\right)$. Define $\gamma(l)=\min\left\{\frac{P-1}{n}\left(k\left(n-(l-1)\frac{n}{P-1}\right)-k\left(n-l\frac{n}{P-1}\right)\right),\ \frac{1}{l}\left(k(n)-k\left(n-l\frac{n}{P-1}\right)\right)\right\}$. It is obvious that it is a best response not to provide effort when reading the partial results of those subordinates who do not provide effort.

(iii) An l-effort equilibrium exists. Therefore we know that

(a) $c \leq \frac{P-1}{n}\left(k\left(n-(l-1)\frac{n}{P-1}\right)-k\left(n-l\frac{n}{P-1}\right)\right)$ and

(b) $c \leq \frac{1}{l}\left(k(n)-k\left(n-l\frac{n}{P-1}\right)\right)$.

An $l+1$-effort equilibrium exists if
$c \leq \gamma(l+1)=\min\left\{\begin{array}{c}\frac{P-1}{n}\left(k\left(n-l\frac{n}{P-1}\right)-k\left(n-(l+1)\frac{n}{P-1}\right)\right), \\ \frac{1}{l+1}\left(k(n)-k\left(n-(l+1)\frac{n}{P-1}\right)\right)\end{array}\right\}$ which is implied by (a) and (b) because $k(m)$ is weakly concave.

(iv) follows from Proposition 1 and part (ii) of the Proposition. Proposition 1 guarantees that there are no equilibria in which some managers process some data, but not the entire set of data assigned to them. Part (ii) eliminates putative equilibria in which the top manager provides effort when reading the file of a subordinate but the subordinate does not provide effort.

(v) Consider any agent i. First note that the full effort equilibrium payoff for agent i is larger than or equal to the payoff when agent i never provides effort while agents $-i$ always do. Next note that this payoff is higher than player i's payoff in any equilibrium where less players provide effort. Q.E.D.

Result (iii) is a direct consequence of the complementarity of effort. Due to the concavity of the $k(m)$ function an l-effort equilibrium exists for a wider range of cost parameters if l, the number of low level players who provide effort, is larger. This is due to the fact that that the marginal gain of effort provision is increasing in l for each of the lowest level players as well as for the top player. Note that if an l-effort equilibrium exists, then there exist another $\binom{P-1}{l}-1$ equilibria in which the same number of lowest level players provide effort but the identity of the effort players changes. If n is a multiple of $P-1$, the $(P-1)$-effort equilibrium called *full effort equilibrium* exists if and only if

$$c \leq \gamma(P-1) := \min\left\{c\left(\tfrac{n}{P-1}\right), \tfrac{1}{P-1}k(n)\right\}. \tag{8}$$

If the full effort equilibrium does not exist, then there exists no other equilibrium except for the no effort equilibrium. An l-effort equilibrium coexists with at least $\sum_{i=l}^{P-1}\binom{P-1}{i}$ other equilibria: the $\binom{P-1}{l}-1$ other l-effort equilibria, the $(l+i)_{i\in\{1,\ldots P-1-l\}}$-effort equilibria and the no effort equilibrium. According to part (v), we expect the full effort equilibrium to be the most obvious way to play the game whenever this equilibrium exists. Otherwise we expect players to play the no effort equilibrium.

4.4 Inefficient decentralization

We have so far seen that in a centralized tree incentives for effort provision may be strengthened by increasing the number of lowest level managers, because incentives are related to the number of objects processed. However, there is a countervailing effect due to the increasing number of operations that have to be performed at the upper level.

Proposition 3 *Let $k(m)$ fulfill Property $P1$. Consider a centralized tree. There is a number $\tilde{P} < n+1$ such that it does not pay to increase the number of information processing agents beyond $\tilde{P}$. This would (i) increase the number of operations (ii) without reducing delay (iii) nor strengthening incentives.*

Proof (i) Obvious. (ii) The delay is minimized at $\hat{P} = \lfloor\sqrt{n}+1\rfloor$, in which case agents on both tiers (up to integer constraints) process the same number of objects. But in this situation the top-manager has better incentives than his subordinates, because his objects contain information of more data items. Thus it must be that $\tilde{P} > \hat{P}$.
(iii) The best incentives are provided if $\gamma(P-1)$ is maximized. By increasing the number of lowest level processors, incentives of lowest level players and the top player are driven in opposite directions. The maximum value for c for which the top manager is willing to provide effort is strictly decreasing in P, and the maximum value for which a lowest level manger is willing to provide effort is (weakly) increasing in P. Monotonicity is obvious. Q.E.D.

For the lowest level managers, the maximum value for c is non-increasing in P if $k(m)$ is linear. If $k(m)$ is strictly concave, it is non-increasing for some P due to integer restrictions with respect to the distribution of tasks. $\tilde{P}(n)$ is the smallest value of P which yields a full effort equilibrium in the centralized tree for the largest parameter range for c. To find

$\tilde{P}(n)$, add managers to the organization as long as $\gamma(P-1) = c\left(\left\lceil \frac{n}{P-1} \right\rceil\right)$. Stop at P' when $\gamma(P'-1) = \frac{k(n)}{P'-1}$. If $\frac{k(n)}{P'-1} > c\left(\left\lceil \frac{n}{P'-2} \right\rceil\right)$, $\tilde{P}(n) = P'$. Else $\tilde{P}(n)$ is the smallest value for P for which $c\left(\left\lceil \frac{n}{P-1} \right\rceil\right) = c\left(\left\lceil \frac{n}{P'-2} \right\rceil\right)$.

If $k(m)$ is strictly concave, then there exists a bound on the size n of the information processing problem which can be properly processed within a centralized tree. This is most easily seen if we abstract from integer problems and assume $\tilde{P}(n)$ is such that $c\left(\frac{n}{\tilde{P}-1}\right) = \frac{k(n)}{\tilde{P}-1}$. Assume we double the set of objects to be processed. To maintain the parameter range for which the lowest level managers are willing to provide effort, the number of lowest level managers has to be doubled. Then, the top manager has an incentive to provide effort if and only if $c \leq \frac{k(2n)}{2\tilde{P}-2} = \frac{1}{2}\frac{k(2n)}{\tilde{P}-1} < \frac{1}{2}\frac{2k(n)}{\tilde{P}-1}$. Hence, $\gamma(\tilde{P}(2n)-1) < \gamma(\tilde{P}(n)-1)$. Thus, for any cost of effort c, there exists an upper bound on n beyond which it is not possible to process the data set with a two-level hierarchy. This observation suggests that hierarchies with more layers may be useful in some cases.

5 Reduced tree equilibria

We now turn to the Nash Equilibria of the reduced tree. Recall the structure of the reduced tree from subsection 2.3. Again, we restrict attention to information processing problems in which integer problems do not arise, i.e. we assume $n = bP, b \in \mathbb{N}\backslash\{0,1\}$. We impose this assumption in order to avoid case distinctions in the proofs which do not add any further insights to the analysis.[15]

In contrast to the game induced by a centralized tree, the game induced by the reduced tree does not always (i.e. for all cost parameters c) have a no effort equilibrium. The reason is that each manager – including the manager who provides the final result – picks up some objects himself. Therefore, for sufficiently low costs c, the no effort equilibrium does not exist. However, in the reduced tree, there exists a class of coordination failure equilibria similar to the l-effort equilibria in the centralized tree described in Proposition 2 (ii). In these equilibria, subordinates do not work because their superiors do not and superiors do not work because their subordinates do not.

Call manager i's immediate superior $s(i)$ and let $s^n(i)$ be i's n^{th} superior. Call the top manager T and let $s(T) = T$. Let $S_i := \{i \cup k : k = s^n(i), n \leq T\}$ be the ordered set of mangers who are working on the calculation path starting with agent i.

[15] The full analysis is available from the authors upon request.

Manager i provides effort only if all managers $k \in S_i$ provide effort when reading the object containing i's information. Providing effort if one of his superiors does not read the information does not change the result but involves costs for i. For the same reason, a manager never provides effort when reading the information provided by a direct subordinate if the subordinate never provided effort.

If the cost function has Property P1, the full effort equilibrium exists whenever any equilibrium exists in which some effort is provided. The necessary and sufficient condition for the existence of a full effort equilibrium can be characterized as follows when P2 holds in addition.

Proposition 4 *Let $k(m)$ fulfill Property $P1$ and Property $P2$.*
(i) Consider a reduced tree with P managers, where $\frac{n}{P} = b \in \mathbb{N}$. A full effort equilibrium exists if and only if

$$c \leq c(b). \tag{9}$$

(ii) The full effort equilibrium Pareto dominates all other equilibria whenever it exists.
(iii) There is a lower bound on c below which no equilibrium exists where no player provides effort.

Proof (i) Necessity of the above condition for the existence of a full effort equilibrium is obvious. To show sufficiency, we proceed in two steps. First, consider the case where the number of managers is a power of 2, $P = 2^a$, $a \in \mathbb{N}^+$.
Consider the top player. Assume all others provide effort. His best response will include to provide effort when reading the object provided by the last subordinate[16] independently of what he did before if

$$c \leq k(n) - k\left(\tfrac{n}{2}\right) = k\left(2^a b\right) - k\left(2^{a-1} b\right). \tag{10}$$

The rest of the top player's strategic situation is equivalent to the one of his last subordinate. Therefore, it will be rational for him to provide effort if it is rational for his subordinate to provide effort, and if the above condition holds.
Now, consider the top player's last subordinate. Assume again that all other players provide effort. The same reasoning as above applies: This player will provide effort when reading the information provided by his last subordinate independently of what he did before if

$$c \leq k\left(\tfrac{n}{2}\right) - k\left(\tfrac{n}{4}\right) = k\left(2^{a-1} b\right) - k\left(2^{a-2} b\right). \tag{11}$$

[16] Recall from the description of the reduced tree in subsection 2.3 that this report contains half the data, given the subordinate provided effort.

The remaining strategic situation is again equivalent to the one of his last subordinate. We conclude: There exists a full effort equilibrium if

$$c \leq \min \left\{ c(b), \min_{j} \left\{ k(2^j b) - k(2^{j-1} b) \right\}_{j=1 \ldots a} \right\}. \tag{12}$$

When Property P2 holds the thresholds $k(2^j b) - k(2^{j-1} b)$ are nondecreasing in j, which implies that the sufficient condition for effort provision reduces to

$$c \leq \min \left\{ c(b), k(2b) - k(b) \right\}. \tag{13}$$

Moreover, under P2,

$$c(b) \leq k(2b) - k(b), \tag{14}$$

which reduces the sufficient conditions to the necessary and sufficient condition

$$c \leq c(b). \tag{15}$$

Next, consider the case $P \neq 2^a$. After the raw data processing, each manager's objects are ordered in a way such that each report contains the information of at least twice as many raw data items as the object previously read (given that all other players provide effort). Hence, the above conditions are stricter than what is needed such that every agent in this hierarchy provides effort as mutually best responses.

To see this, consider the top player. Assume that all other players provide effort. His last subordinate's report contains more than half of the information to be processed (to be precise, he reports the aggregate of $2^{\lceil \log_2 P \rceil - 1} b$ raw data items). Thus, the top manager will provide effort when reading the report no matter what he did before, if $k(Pb) - k\left(2^{\lceil \log_2 P \rceil - 1} b\right) > c$. Since this condition is implied by $c < c(b)$ because of Property P2, he knows that his best response includes to provide effort on the last object. The same reasoning applies to the next-to-last object. It contains at least half of the information of the remaining set. Thus, the sufficient conditions are implied by the necessary condition in this case as well.

(ii) The proof of Proposition 2 (v) applies to the reduced tree as well.

(iii) Suppose that no player provides effort. If $c < \frac{P}{n}\left(k(n) - k\left(n - \frac{n}{P}\right)\right)$, then there exists a profitable deviation from the supposed equilibrium strategy for the top player. He will choose to provide effort when he reads the objects which are directly assigned to him. Q.E.D.

Thus, in order to induce all players in the reduced tree to provide effort, one only has to take care of the incentives of lowest level players. An immediate consequence is stated in Corollary 1.

Corollary 1 *If $k(m)$ has Properties P1 and P2, the parameter range for which full effort equilibria exist can be increased by increasing the number of managers processing information in the reduced tree, until $\frac{n}{P} = 2$.*

Hence, by decentralizing information processing tasks in the way proposed by Radner (1993), proper incentives can be provided by involving enough agents. Note that for information processing problems which satisfy Properties P1 and P2, if it is not possible to provide proper incentives to the members of a reduced tree, then there exists no hierarchy design with which this is possible. This is due to the fact that in the reduced tree it suffices to provide incentives to the lowest level managers, which is a necessary condition for effort provision in any network. Since the number of lowest level managers is maximal in the reduced tree, it is not possible to provide better incentives.

Another consequence of our analysis concerns the efficient size of an organization. Radner (1993) points out that some hierarchy sizes, i.e. the number of managers P, are inefficient for the aggregation of n data items, since the same delay may be achieved with a smaller number of managers. For example, if 10 objects are to be processed, the reduced tree yields a delay of 5 units of time for any $P \in \{3, 4, 5\}$. Hence, a hierarchy employing more than 3 managers is inefficient for this task in Radner's framework. When taking into account incentives for information provision, employing 5 managers may be efficient. This is the case if $c \in (c(3), c(2)]$, that is if the reduced tree with 3 managers does not have a full effort equilibrium, but the reduced tree with 5 managers does.

In our extended framework, in an organization in which every player provides full effort, the number of managers P is inefficiently high, only if there is a smaller number of managers achieving the same delay and yielding a full effort equilibrium. That is the case if (i) $\lfloor \frac{n}{P} \rfloor + \lceil \log_2 (P + n \bmod P) \rceil = \lfloor \frac{n}{P-1} \rfloor + \lceil \log_2 (P - 1 + n \bmod (P-1)) \rceil$ and (ii) $c\left(\lceil \frac{n}{P-1} \rceil\right) \geq c$.

6 Efficient organization design

Proposition 4 enables us to characterize the set of efficient organizations. It is important to note that if the cost function $k(m)$ has Property P1, every organization which does not have a full effort equilibrium does not have any equilibrium in which effort is provided. Therefore, it can be replaced by the trivial organization in which the aggregate state of the world x is guessed without delay. Hence, we may restrict the analysis of efficient organizations to those which have a full effort equilibrium.[17]

Our main result is that any efficient outcome can be achieved with a reduced tree. Any programmed network $\tilde{N}$ which has a full effort equilibrium for an information processing task (n, c) can be replaced by a reduced tree $R(\tilde{N})$ with the same number of processors that also has a full effort equilibrium for this task. Therefore $R(\tilde{N})$ generates the same classical surplus and a (weakly)[18] faster decision. We prove this result for the subclass of information processing problems for which the cost function $k(m)$ has Properties P1 and P2. Again, we restrict attention to the case that n is a multiple of P, $n = bP, b \in \mathbb{N}^{+} \backslash \{0, 1\}$.

Let there be a programmed network $\tilde{N}$ of P managers which has a full effort equilibrium for processing $n = bP$ objects. In this network, there must be at least one agent handling at least b raw data items. Since the cost function $k(m)$ has Property P1, the maximum benefit for these tasks is $k(b)$. Hence, it must be true that $c \leq c(b)$. This in turn is the sufficient condition for the existence of a full effort equilibrium in a reduced tree $R(\tilde{N})$ with the same number of managers, if the information processing problem has Property P2.

This yields Proposition 5.

Proposition 5 *Let $k(m)$ have Properties P1 and P2 and let $P = \frac{n}{b}, b \in \mathbb{N}^{+}$.*
(i) If the reduced tree with P managers does not have a full effort equilibrium for the processing task, then there exists no programmed network with P managers which does.
(ii) Every reduced tree produces at least the same speed and classical surplus as any other organization with P managers.

Hence, any organization producing classical surplus $\bar{v}$ coexists with a reduced tree that also produces $\bar{v}$ with a weakly lower delay. However, as argued above, these organizations are not efficient if the same delay can be achieved with a smaller number of managers without destroying incentives.

In particular an organization with P managers is efficient for an information processing problem (n, c) which satisfies Properties P1 and P2 only if there exists no $P' < P$ such that (i) $\left\lfloor \frac{n}{P'} \right\rfloor + \lceil \log_2 (P' + n \bmod P') \rceil = \left\lfloor \frac{n}{P} \right\rfloor + \lceil \log_2 (P + n \bmod P) \rceil$ and (ii) $c\left(\left\lceil \frac{n}{P'} \right\rceil\right) \geq c$. We call such an organization *P-efficient.* We have that:

[17] However, a time constraint may force an organization, for which a full effort equilibrium exists, to limit the amount of information which is processed properly to $n' < n$. We briefly discuss this issue at the end of this section.

[18] Remember from Section 2.3 that the reduced tree is in general not the unique delay minimizing organization.

Corollary 2 *Let $k(m)$ have Properties P1 and P2.*
(i) Every P-efficient reduced tree with a full effort equilibrium is efficient.
(ii) Every organization which is not P-efficient is dominated by a P-efficient reduced tree.

In order to determine the part of the efficiency frontier for which the whole set of information is processed properly, we may restrict attention to P-efficient reduced trees with a full effort equilibrium. A P-efficient reduced tree may coexist with other organizations producing the same classical surplus with the same delay. However, there is a class of problems for which the efficient organization is unique.

Corollary 3 *Consider an information aggregation problem (n, c) which has Properties P1 and P2, $n = b2^a, b, a \in \mathbb{N}^+ \backslash \{0, 1\}$, and $c \leq c(b2^{a-j})$, where $j \in \{1...a\}$. An organization with 2^j managers is efficient if and only if it is a reduced tree.*

Proof From Section 2.3 we know that there exists no organization with 2^j managers except for the reduced tree which achieves the minimum delay. It remains to be shown that the reduced tree with 2^j managers is P-efficient. It is enough to show that $\left\lfloor \frac{b2^a}{2^j-1} \right\rfloor + \lceil \log_2 (2^j - 1 + b2^a \bmod (2^j - 1)) \rceil > b2^{a-j} + j$. Note that $\left\lfloor \frac{b2^a}{2^j-1} \right\rfloor \geq b2^{a-j}$ and that $\lceil \log_2 (2^j - 1 + b2^a \bmod (2^j - 1)) \rceil = j$ only if $b2^a \bmod (2^j - 1) \leq 1$ and $> j$ else. If $b2^a \bmod (2^j - 1) \leq 1$, then $\left\lfloor \frac{b2^a}{2^j-1} \right\rfloor > b2^{a-j}$. Q.E.D.

Note that the minimum delay within which a set of n items can be processed is $1 + \lceil \log_2 n \rceil$.[19] If a decision has to be made within a shorter period of time, the organization is forced to only partially (if at all) process the set of relevant information. If the cost function $k(m)$ is strictly concave, incentives are diluted by restricting the information processing to a subset of the relevant information.

Suppose the decision must be made till time unit $T < 1 + \lceil \log_2 n \rceil$. Given our assumption that society is large enough such that the welfare gain of better decisions always outweigh the processing cost, it is clear that in the limited time span as much information as possible should be processed. In T units of time, 2^{T-1} information items can be processed properly in a reduced tree (with 2^{T-2} managers) if the cost of effort provision is sufficiently small such that a full effort equilibrium for this task exists. This defines the efficiency frontier for delays smaller than $1 + \lceil \log_2 n \rceil$.

Figure 3 depicts the efficiency frontier for an information aggregation problem with properties P1 and P2. In this example, there are $n = 16$ objects to be aggregated. The maximal

[19]See Radner (1993).

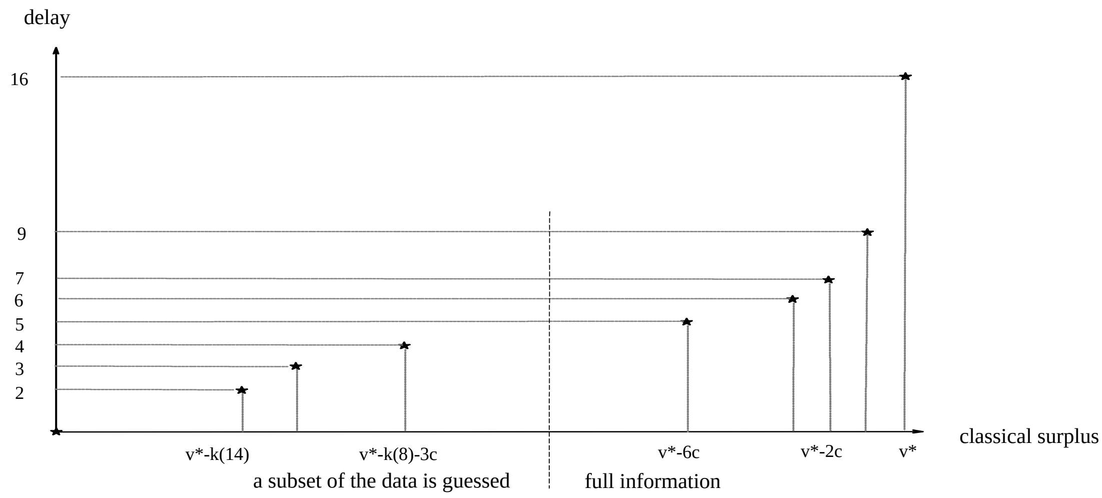

Figure 3: Efficiency frontier for an information processing problem with Properties P1 and P2, $n = 16$.

classical surplus v^* is obtained if a single manager works on the whole set of data items. Adding a manager decreases the delay of decision making to 9, but also reduces the classical surplus to $v^* - c$, since there is an additional object (the report) to be read. The delay cannot be reduced below 5 by increasing the number of managers beyond 7 if all data items shall be processed. In order to obtain a smaller delay, the organization is forced to guess part of the data and limit information processing activities to the data which can be handled within the given period of time. Note that c must be small enough to sustain effort provision as an equilibrium, because the marginal benefit of effort provision is smaller if part of the data is not processed properly.

7 Conclusion

Our analysis merges two recent strands of the literature on organization design: the game-theoretical analysis of institutions and the theory of the efficient design of programmed networks. We identify an elementary advantage of decentralized structures: these structures provide better incentives for self-interested managers. For a large class of managers' preferences and stochastic structures of the information aggregation problem, the division of tasks favors equilibria where all players provide effort. The reason is the complementarity of effort of different managers which arises endogenously in our model.

Our second main result is that the speed of a decision procedure and the quality of the decision need not be conflicting objectives when it comes to the evaluation of organization designs. We find that reduced trees à la Radner (1993) outperform other arrangements along both dimensions. Information is aggregated in parallel. Therefore, in the course of a manager's processing activity, the information content of the objects he is supposed to read is increasing. Those managers who have to work most, also have the most important jobs. Hence, Radner's efficiency result is robust with respect to more complex behavioral assumptions (see also Schulte and Grüner, 2004).

Our main results hold for all underlying stochastic structures and preferences that guarantee (i) the monotonicity of the cost functions $k(m)$ and $c(m)$ and (ii) the (not to strong) concavity of $k(m)$. An interesting extension of our work would be to consider environments in which the concavity property is not satisfied. Moreover, it would be interesting to study a framework in which the cost of guessing is not independent of information previously acquired, e.g. a framework with Bayesian updating. In such an information aggregation problem, the value of an information item is path-dependent, and it can be beneficial to approach the agents sequentially. It has been shown in Smorodinsky and Tennenholtz (2003) and in Gershkov and Szentes (2005) that limiting the agents' information (concerning their position in the sequence and the values of partial results) enhances their incentives for information acquisition. In both papers incentive problems only arise when learning the initial data whereas the aggregation operations are not subject to incentive constraints. Moreover, decision delay is not costly per se. It is certainly worth studying this class of information aggregation problems in our extended framework.

There are several further useful extensions of the present framework. One would be to include different conflicts of interest. In the present paper the only conflict of interest among agents arises from the individual disutility from providing effort. In many interesting applications there is also some disagreement about the best decision even if there is perfect information. A second approach would be to look at alternative verifiability structures. If some data is verifiable ex post then monetary incentives may be used and the set of efficient organizations might look different. One could also link the cost of information aggregation to the complexity of the task. Finally, it would be interesting to study operations that cannot be permuted without altering the results. All these extensions may prove useful for extending the range of applications of the present framework.

References

[1] Aghion, Philippe and Jean Tirole (1997): Formal and Real Authority in Organizations, *Journal of Political Economy* 105(1):1-29.

[2] Bolton, Patrick and Mathias Dewatripont (1994): The Firm as a Communicating Network, *Quarterly Journal of Economics* 109:809-839.

[3] Crémer, Jacques (1980): A Partial Theory of the Optimal Organization of a Bureaucracy, *Bell Journal of Economics* 11:683-93.

[4] DeGroot (1970): Optimal Statistical Decisions, McGraw-Hill Book Company, New York.

[5] Dewatripont, Mathias and Jean Tirole (2004): Modes of Communication, IDEI Working Paper 323.

[6] Garicano, Luis (2000): Hierarchies and the Organization of Knowledge in Production, *Journal of Political Economy* 108(5), 874-904.

[7] Gershkov, Alex and Balász Szentes (2005): Optimal Voting Schemes with Costly Information Acquisition, *mimeo.*

[8] Jehiel, Philippe (1999) Information Aggregation and Communication in Organizations, *Management Science*, 45:659-669.

[9] Meagher, Kieron, Hakan Orbay and Timothy Van Zandt (2001): Hierarchy Size and Environmental Uncertainty, CEPR Discussion Paper No. 2839.

[10] Meagher, Kieron and Timothy Van Zandt (1998): Managerial Costs for One-shot Decentralized Information Processing, *Review of Economic Design* 3: 329-345.

[11] Melumad, Nahum D., Dilip Mookherjee and Stefan Reichelstein (1995): Hierarchical Decentralization of Incentive Contracts, *RAND Journal of Economics*, 26:654-672.

[12] Mookherjee, Dilip and Stefan Reichelstein (1997): Budgeting and Hierarchical Control, *Journal of Accounting Research* 35(2):129-155.

[13] Orbay, Hakan (2002): Information Processing Hierarchies, *Journal of Economic Theory* 105: 370-407.

[14] Prat, Andrea (1997): Hierarchies of Processors with Endogenous Capacity, *Journal of Economic Theory* 77:214-222.

[15] Radner, Roy (1992): Hierarchy: The Economics of Managing, *Journal of Economic Literature* 30:1382-1415.

[16] Radner, Roy (1993): The Organization of Decentralized Information Processing, *Econometrica* 61(5): 1109-1146.

[17] Radner, Roy and Timothy Van Zandt (1992): Information Processing in Firms and Returns to Scale, *Annales d'economie et de statistique* 25/26:265-298.

[18] Rosen, Sherwin (1982): Autority, control, and the distribution of earnings, *Bell Journal of Economics* 13: 311-323.

[19] Schulte, Elisabeth and Hans Peter Grüner (2004): Speed and Quality of Collective Decision Making: Imperfect Information Processing, forthcoming in *Journal of Economic Theory.*

[20] Smorodinsky, Rann and Mose Tenneholtz (2003): "Overcoming Free Riding in Multi-Party Computations – The Anonymous Case", *mimeo*, Technion.

[21] Van Zandt, Timothy (1997): The Scheduling and Organization of Periodic Associative Computation: Essential networks, *Review of Economic Design* 3:15-27.

[22] Van Zandt, Timothy (1998): The Scheduling and Organization of Periodic Associative Computation: Efficient networks, *Review of Economic Design* 3:93-127.

[23] Winter, Eyal (2004): Incentives and Discrimination, *American Economic Review* 94:764-773.