Coordination and Decision-Making in a Market with Uncertainty

Ashish Arora
Vidyanand Choudhary
Karthik Kannan
Ramayya Krishnan
Rema Padman

Carnegie Mellon University,

Pittsburgh, PA – 15213

Abstract

Research comparing centralized and decentralized coordination schemes is often contradictory. For example, Oskar Lange (1969) among others points out that the centralized coordination scheme can outperform decentralized schemes while Arrow and Debreu (1954) show otherwise. This paper focuses on this issue in the context of agent-based marketplaces. Specifically, we use a computational approach to study the trade off between imperfect decision-making by the central coordinating agent (the central authority) and the imperfect coordination among decentralized seller agents. Using social welfare as a metric, we study how the correlation in the quality of the seller agents, the marginal costs of product building, decision-making costs and the fraction of consumer utility transferred as compensation to the winning seller agent affect the terms of this trade-off. We find that the decentralized scheme with its parsimonious use of information and simplistic decision rule does very well in comparison to the centralized scheme, which internalizes the externalities. This is surprising since one may expect the centralized scheme to always perform better than the decentralized scheme. This paper analyzes the results and provides intuition for this apparent anomaly.

Acknowledgements: This work was funded in part by NSF CISE/IIS/KDI 9873005. Authors would like to thank Amy Greenwald of Brown University for her valuable comments

1. Introduction

We are interested in examining the performance consequences of different modes of coordination in a reverse marketplace for consumer-specific products that have ex ante quality uncertainty. Consumer-specific products are products built to a particular consumer's specification and cannot be resold to another consumer. Other features of this marketplace are: non-negligible marginal costs of production, decision-making costs and substantial ex ante heterogeneity in the match between the product creation task and the capability of the seller agent (hereafter referred to as a *node*) to perform the task. These characteristics are observed in web-based reverse marketplaces for services. For example, in eLance (http://www.elance.com), buyers post information about their web design projects in the manner of a RFQ (request for quotation) (see Figure 1). Developing a bid in response to the RFQ is costly (Snir and Hitt 2000) and the value of the bid to the consumer is uncertain ex ante. Understanding the performance consequences of different modes of coordinating the nodes that can generate the bids is important to the design of marketplaces for such services.

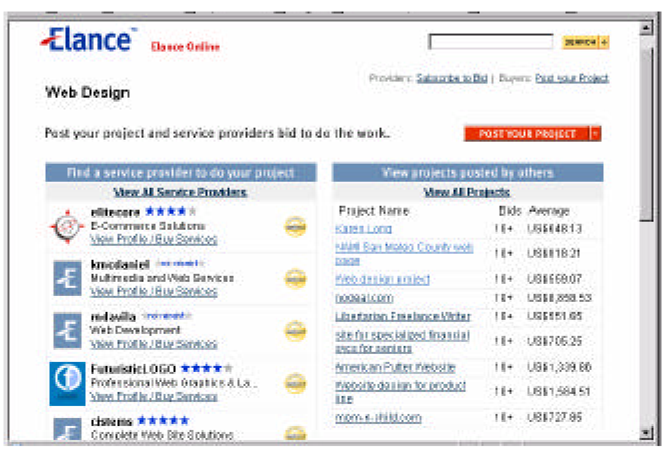


Figure 1 Screenshot of eLance web site

In any coordination scheme, despite the redundancy of additional designs, ex ante uncertainty about the quality of the designs will lead to demand for multiple designs. In other words, once the designs have been built and evaluated, only the best is used and the rest are discarded. Even so, since one does not know which design will turn out to be the best, there is value in being able to select from more than one design. On the other hand, increasing the number of nodes also increases costs. How many and which 'design' nodes should be tasked is the key question. This question is a particular instance of the broader economic question of how the decisions of nodes should be coordinated to achieve the greatest social welfare.

In this context, coordination has two dimensions – information and decision making authority. The two polar forms of coordination are a) Universal or global information and centralized decision-making, and b) Localized information and localized decision-making. Intermediate forms such as universal information and decentralized decision-making are also possible.

In this paper, we address the question of how well a "decentralized scheme" – using only local information (e.g., nodes using local information about their own capabilities) and deciding without explicit coordination with others – performs relative to centralized coordination schemes, and more importantly, how the difference in performance is affected by factors such as changes in marginal costs of production, decision-making costs and correlation in the quality of the nodes.

The rest of the paper is organized as follows. We motivate our research question, review the relevant literature, and discuss the rationale underlying our methodological approach in Section 2. In Section 3, we formally define social welfare, our evaluation metric, and present the alternative coordination schemes we study, with an emphasis on the decision problems that need to be solved to implement them. We present the results of our computational study in Section 4, and conclude in Section 5 with a discussion of the implications of our work for implementing reverse marketplaces.

2. Literature Review and Theoretical Foundations

Research comparing the performance of centralized coordination schemes and decentralized coordination schemes (also referred to as a market scheme) is contradictory. Although it is now widely accepted that market schemes are the best way to achieve coordination, in the 1930s and 1940s, the efficacy of a centralized coordination scheme vis a vis that of a decentralized scheme was analyzed in debates on "The Plan versus The Market". Economists such as Oskar Lange (1969) and others pointed out that a centralized computer could match demand and supply more efficiently than the adjustment processes that characterize actual markets. In response, von Hayek and others pointed out decentralized systems greatly economize on the amount of time required by decision makers since they coordinate using market prices (Arrow and Debreu 1954) and can therefore do without the large computing requirements of centralized systems.

There are many exceptions to this general principle. It is widely recognized that externalities (positive or negative) may to lead to market failure (Tirole 1990). For instance, where many nodes "race" for a single prize, decentralized schemes may lead to excessive entry because nodes disregard the negative externality on other nodes of their decision to enter (Dasgupta and Stiglitz 1980).

Organization theory literature has addressed this issue as a comparison between hierarchies and markets. Malone (1987) compares the impact of different costs – production, coordination and vulnerability costs – across coordination schemes. Using simple models, Malone (1987) shows that the vulnerability costs are higher for the centralized schemes but that coordination costs are higher in the decentralized scheme. Malone (1987) also analyzes the impact of these costs when the number of nodes increases. Tan and Harker (1999) extend Malone's work and find that from an 'expected-cost' perspective, a decentralized scheme using an auction mechanism outperforms a centralized scheme. However, these papers do not model interactions among the nodes.

Nault (1999) assumes that the information asymmetry between "non-competing" nodes and the centralized authority lowers performance in a centralized scheme while the lack of a global view impacts the decentralized scheme. The optimization problem is to decide on global and local investments to maximize the total organizational profits. The schemes compared are 1) Centralized system where both global and local investments decisions are made by the centralized authority. 2) Decentralized system where decisions about local and the global investment decisions are made by the individual nodes, and 3) Mixed mode where decisions about global investments are made by the centralized authority but the local investment decisions are made by the decentralized nodes. Nault (1999) provides sufficient conditions under which hierarchy (centralized) is better than markets (decentralized) and similarly conditions when hierarchy (centralized) is better than the mixed mode.

Table 1 Related Work

	Tan and Harker (1999)	Nault (1999)	Our Work
Product heterogeneity	Heterogeneous	Homogeneous	Heterogeneous
Competition among nodes	No	No	Yes
Demand Function	Unit	Generic Representation	Unit
Performance Metric	Total Expected Cost	Organization Profits	Social Welfare[1]

Table 1 positions this paper with respect to prior work in this area. Our work is distinctive in its consideration of a) competition between nodes, b) modeling of the ex ante uncertainty that nodes have about their qualities and therefore, their profits, and c) an explicit modeling of decision-making costs. We also compare an imperfect-information based decentralized scheme to two variants of a centralized scheme with complete information – one where the decision making is optimal (but possibly with high decision-making costs) and the other with heuristic, imperfect decision-making (but with possibly lower decision-making costs). The rationale for our focus on these coordination schemes and our use of a computational approach to investigate them is presented later in this section. We begin with an analytical formulation and establish that it would be intractable to implement.

2.1 Market Mechanism

We consider a reverse marketplace with three types of participants – the buyer, the broker, and nodes. The buyer needs a web site design[1] and would like to accept a design that offers him the highest "quality". Nodes have different assets that can be used to build site-designs in response to a buyer's request. However, the "quality" of the design is unknown until the designs have been evaluated. The broker assesses the "quality" of the designs developed by the nodes based on the specification in the RFQ provided by the buyer. For tractability, we assume that the "quality" can be mapped to a value in the range [0,1].

[1] The model is applicable to a wide variety of settings and products. We use the example of web site design for clarity of exposition.

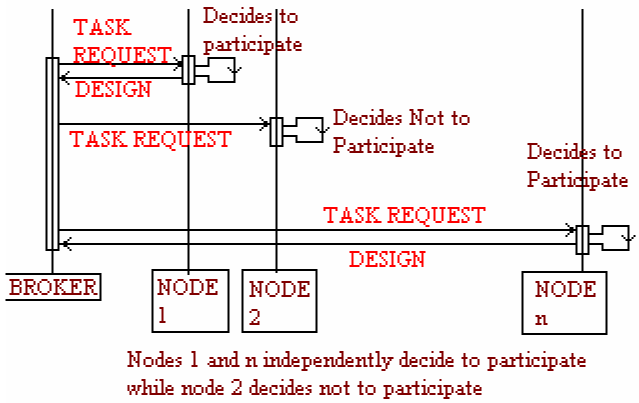


Figure 2 Decentralized Scheme

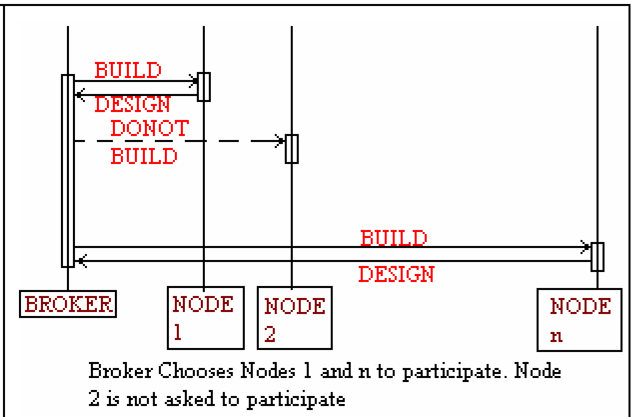


Figure 3 Centralized Scheme

A *market-session* is initiated when the buyer submits the desired characteristics of the "product" (e.g., the web site design) as part of the RFQ to the broker. Based on the coordination scheme adopted, nodes are either chosen (centralized scheme) or independently decide (decentralized scheme) to participate. Participation entails production costs since a product (design) needs to be developed in response to the RFQ provided by the buyer. Figure 2 and Figure 3 illustrate the locus of decision-making in each of the two schemes. The broker is the decision maker in the centralized scheme while the node is the decision maker in the decentralized scheme. Under each scheme, nodes that have either been selected or which choose to participate, build site-designs in response to the RFQ. After receiving the designs built by the nodes, the broker evaluates the designs, and returns the highest quality design to the buyer and compensates the winning node under either scheme. The effect on social welfare (our performance metric) of differences in the information available to the decision maker, and the manner in which decision problems related to participation are solved under each scheme are the foci of this paper.

2.2 Theoretical Framework

Consider the case where a consumer is willing to pay V for the best available design and where the quality of the product supplied by any node has distribution function $F(q)$, where q ranges from $[Q_{\min}, Q_{\max}]$. In our analysis, $Q_{\min} = 0$, $Q_{\max} = 1$. Ignoring decision-making cost for now, let C be the marginal cost of producing a design. Since the distributions are ex-ante identical, each node that produces the product has an equal chance of winning. If k_d nodes participate in the decentralized scheme, the expected revenue for each node is V / k_d. The number of nodes producing the designs (ignoring the integer nature of k_d) is given by the zero profit condition, $V / k_d = C$.

The entry of a node benefits the consumer because it increases the expected quality of the best product. If k nodes participate and X denotes the quality of the best product i.e., $X(k) = \max\{q_1, q_2, \cdots q_k\}$, and if $G_k()$ represents the distribution function of $X(k)$, then the distribution of the max order statistic of $G_k(x) = F^k(x)$ (David 1981). Let the consumer's utility function be $X(k) - V$. The expected social welfare generated is therefore $E(X(k)) - kC$, where E is the expectation operator. This can be written as

$$\int_{Q_{\min}}^{Q_{\max}} \mathrm{x}\, d\, \mathrm{G}_k(\mathrm{x}) = 1 - \int_{Q_{\min}}^{Q_{\max}} \mathrm{G}_k(\mathrm{x}) d\, \mathrm{x} = 1 - \int_{Q_{\min}}^{Q_{\max}} \mathrm{F}^k(\mathrm{x}) d\, \mathrm{x}$$

A centralized decision maker can maximize expected social welfare by choosing k_c. The gain in quality from adding a node must be greater than the cost i.e., the optimal k_c must satisfy

$$\int_{Q_{\min}}^{Q_{\max}} F^{k_c}(x)\,dx - \int_{Q_{\min}}^{Q_{\max}} F^{k_c+1}(x)\,dx \geq C$$

$$\int_{Q_{\min}}^{Q_{\max}} F^{k_c+1}(x)\,dx - \int_{Q_{\min}}^{Q_{\max}} F^{k_c+2}(x)\,dx \leq C$$

Suppose V is equal to $E(X(k))$, so that the entire social surplus accrues to the wining node under either scheme. In the decentralized scheme, social welfare is, $E(X(k_d)) - k_d C = E(X(k_d)) - V = 0$; this is because competition among the nodes dissipates all rents. The number of participating nodes is restricted in centralized coordination, $k_c \leq k_d$ and therefore, social welfare is higher under centralized coordination.

2.3 Need for a Computational Approach

This example shows that decentralized outcomes may be inefficient and the extent of inefficiency depends on how nodes are compensated. When the nodes are differentiated and aware of their task capabilities, more complex game theoretic formulations are possible. Such a formulation would involve each node estimating not only its expected quality but also the likely competition from other nodes. Each node would have to form beliefs about the design capabilities of other nodes as well as estimates of their beliefs about its own task capabilities and so on. Further, these beliefs may be functions of the nature of the task (e.g. B2B site design or a furniture store site design). In general, such a game-theoretic model would rapidly become intractable particularly if nodes are heterogeneous and their qualities correlated.

On the other hand, a centralized scheme faces a number of challenges – keeping track of the various nodes and their capabilities, and developing algorithms for computing the optimal k_c. The central decision maker's problem is not simply to calculate the number nodes to task, but also which specific ones to task. Further complicating the problem is the possible correlation among the outcomes across nodes. This lack of independence in outcomes and the diverse nature of the nodes imply that analytical solutions are not possible. Indeed, with heterogeneous nodes, the central- decision-maker's problem is a combinatorial problem of deciding which subset of the set of potential nodes to task. If s_i denotes the subset of all nodes T i.e., $s_i \in \{s \mid s \subset T\}$ and $X(s_i)$ denotes the best quality when nodes in subset s_i are tasked, the decision problem of the central decision-maker is

$$\max_{s_i} E(X(s_i)) - \#(s_i).C$$

where $\#(s_i) = k$ stands for the cardinality of the set *s*. In order to solve the problem, one would need to know the distribution of the first order statistic of every subset of the set of nodes. In general, there is no tractable analytical formulation for this (David 1981).

This sets up a simple trade-off that we cannot study analytically but one that we can analyze with a computational test-bed. A centralized coordination scheme takes into account the negative externality implicit in markets for custom tasks. To do so, however, requires not only extensive information on the distribution of qualities for individual vendors, but also solving a combinatorial problem that will become rapidly intractable as the number[2] of nodes increases. Thus, the quality of the information available to the central decision maker and the quality of the decision rules the latter follows should determine whether a centralized or a decentralized coordination scheme yields higher social welfare. Although this applies to the decentralized coordination scheme as well, the implications are less restrictive. Individual nodes need only form some estimate of the likelihood of winning in order to decide whether to participate or not.

Even so, the individual decision rules can vary in sophistication, particularly if the node has to potentially choose among several competing projects.

3. Decision-Making under each Coordination Scheme

Since our analysis is from the perspective of a market designer whose interest lies in comparing social-welfare under each scheme, we begin this section by defining social welfare. Following that, we describe the decision problems that arise under each of the coordination schemes we analyze.

3.1 Social Welfare

Let k out of n nodes in the marketplace be tasked or choose to participate and let $Q = X(k)$ represent the best quality generated. If consumer's willingness to pay, $V = Q$, the consumer surplus, where p the (exogenous) share of the consumer utility paid to the node, is

$$Consumer\ Surplus = (1-p)Q \quad \cdots\cdots (1)$$

Let q_y represent the quality produced by any node y. If node x is the winner then $q_x = Q$, the best quality generated in the marketplace. Node x receives payment from the consumer while other nodes generate zero revenue for that market session. In general, the revenue for any node y is given by:

$$Revenue\ \ for\ Node\ y = \begin{cases} q_y\, p & if\ Q = q_y \\ 0 & otherwise \end{cases} \quad \cdots\cdots (2)$$

If C, an exogenous parameter, represents the production cost for each node, then

$$Profit\ \ for\ \ Node\ y = \begin{cases} q_y p - C & if\ \ Q = q_y \\ -C & otherwise \end{cases} \cdots\cdots(3)$$

If D stands for the decision-making cost incurred, then social welfare generated is

$$Social Welfare = Q - kC - D \quad \cdots\cdots(4)$$

In this set-up, if $p = 1$, consumer surplus generated is zero. In the decentralized scheme, since the decision-making costs are borne by the nodes, the total social welfare generated when $p = 1$ is the industry profit. Decision-making costs are further discussed in Section 3.2.1.1.

3.2 Coordination Schemes and Decision Problems

Having defined the metric for our analysis, we proceed to describe the decision problems that arise under each coordination scheme and the strategies used to solve them by the decision makers.

3.2.1 Decision Problem faced by Broker under the Centralized Scheme

The broker is the decision-making authority that tasks the best combination of nodes based on the social welfare metric. If the broker knows the distribution, $f(Q_{s_i})$, of the best quality, Q_{s_i}, when the nodes in subset s_i are tasked, then the decision problem is

$$\max_{s_i} \left\{ \int_{Q_{\min}}^{Q_{\max}} Q_{s_i} \ \ f(Q_{s_i})\ dq - \#(s_i) * C - D \right\} \cdots\cdots(5)$$

where the first term represents the expected best quality, $\hat{Q}_{s_i}$, from tasking nodes in subset s_i.

This is a stochastic combinatorial optimization problem. We next present two alternative strategies that could be used to solve the problem. The first is an exhaustive search strategy that is ex-ante optimal ignoring decision making costs. The second is a heuristic strategy that is sub-optimal but which incurs lower decision-making costs.

3.2.1.1 Exhaustive Search Strategy

A straightforward solution is to exhaustively search the power set of n nodes – $2^n - 1$ combinations – and choose the optimal set of nodes to be tasked. Note that this decision is, by assumption, independent of the decision-making cost of finding the optimal combination. In subsequent analysis, we analyze how social welfare varies as the decision-making cost is taken into account. Simply put, we ignore the problem of "deciding to decide". Since we ignore decision-making costs, the decision problem can be written as $\max_{s_i}\{\hat{Q}_{s_i} - \#(s_i) * C\}$. In our set-up, the task capability, $\hat{Q}_{s_i}$ for any node-combination s_i, is estimated using regression (equation 6) as a function of task-specific characteristics in the RFQ.

$$Q_{s_i} = a_{0s_i} + (task\ characteristics)\ a_{1s_i} \cdots\cdots\cdots (6)$$

Broker Calculates this information

Session #	A	B	C	AB	BC	CA	ABC	A	B	C
1	60	70	50	70	70	60	70	0	Wir	0
2	70	80	90	80	90	90	90	0	0	Win
3	80	60	70	80	70	80	80	Win	0	0
4	90	50	95	90	95	95	95	0	0	Win

Broker uses this information

Figure 4 Calibration Run data and information provided to the broker.

Independent variables in these equations correspond to task characteristics specified by the consumer. Regression coefficients, a, represent the knowledge possessed by the broker about the quality of nodes and their combinations. To estimate the regression, we collect data using what we refer to as a "calibration run". In the calibration run, RFQs are generated randomly and each node responds to the RFQ with a design (its product) which is evaluated and assigned a quality. A sample calibration-run is shown in Figure 4 for $n = 3$. Columns 2, 3 and 4 represent the ex post qualities of participating nodes. This raw data can be used by the broker to create a table with the highest quality generated by each element of the power set of the set of nodes in the market as a function of the task characteristics specified in the RFQ. Columns 4, 5, 6, and 7 in Figure 4 are generated by the broker and used for estimating the regression coefficients. To reduce clutter, task characteristics of the RFQ are not shown in the table of Figure 4 for each session.

Input*: n nodes and regression coefficients to estimate qualities for each element of the power set of nodes*

Output*: Element of the power set ($SelectedCombination$) that offers maximum expected social welfare*

Procedure-Begin:

$MaxSW = 0$ *; Maximum expected social welfare*

$SelectedCombination = NULL$ *; Selected node combination*

For s_i *in power set of n*

Compute $\hat{Q}_{s_i}$ *using regression coefficients*

$SW = \hat{Q}_{s_i} - \#(s_i) * C$

$MaxSW = \max\{MaxSW, SW\}$

if $(MaxSW = SW)$

$SelectedCombination = s_i$

End-if

End-For

Procedure-End

Figure 5 Algorithm for Exhaustive-search centralized scheme

The regression model (equation 6 with coefficient estimates from the calibration run) enables the broker to estimate – in response to a new RFQ – the expected quality from all node combinations. Based on these estimates, the broker chooses the node combination that generates the highest social welfare and tasks only the specific nodes in the combination. The algorithm is presented in Figure 5. This exhaustive search scheme is the ex-ante social welfare maximizing scheme when the decision-making costs are ignored.

In reality, decision-making costs can be considerable. There are two types of decision-making costs involved. First is the one time cost of gathering and analyzing the "calibration-run" data. The other cost is incurred for each market session that is initiated in response to a customer RFQ. This is the cost for estimating expected best quality for each node combination and comparing node-combinations to select the optimal set. In our paper, we restrict our attention to the second type of decision-making cost.

Since the broker in the exhaustive-search scheme exhaustively searches i.e., estimates and compares the regression estimates for all $2^n - 1$ combinations, the decision-making cost is the highest among the schemes we analyze. In our study, we characterize the decision-making cost for other schemes based on the decision-making cost for this scheme,

$$D = (\#_{search}\ m)/(2^n - 1) \cdots\cdots (7)$$

where m is an exogenous variable representing the total cost for the exhaustive search for a given n and $\#_{search}$ represents the number of comparisons made. For this scheme, $\#_{search} = 2^n - 1$, so D = m. Relative to other schemes, the decision-making costs of the exhaustive search scheme increase exponentially with n, the number of nodes.

3.2.1.2 Heuristic decision making strategy

As with the previous scheme, the broker is again the decision-making authority that decides which subset s_i of the n nodes to task to maximize social welfare. Although the exhaustive search scheme provides a mechanism for choosing the best set of nodes to be tasked, it may not maximize social welfare due to the decision-making cost involved. In this section we provide a heuristic solution that maximizes social welfare by exploiting the trade-off between a sub-optimal-search and its associated lower decision-making cost. Clearly, other heuristic procedures with their associated decision making costs can be devised. The value of the approach is in its ability to evaluate alternative proposals using social welfare as a metric.

The decision problem faced by the broker can be modeled as a stopping problem with a finite horizon (Ferguson 1989). Having selected a set of nodes, the broker has to decide whether to add another node or to stop the selection process.

The solution for the broker's decision-problem is remarkably simple. Suppose the broker has already selected a subset s_i. It selects a node x (that is not in the set s_i) to add to this set that produces the maximum expected increment in quality among the nodes – the positive contribution to social welfare – which is at least as much as the cost incurred by the participation of the additional node – the loss in social welfare – i.e., $\max_x(\hat{Q}_{s_{i+x}} - \hat{Q}_{s_i})$ such that $(\hat{Q}_{s_i+x} - \hat{Q}_{s_i}) \geq C$. This selection process is repeated until all nodes are selected or until the stopping condition is met. Analytic solution for this problem is possible if nodes are distributed independently and identically. Since in our case, the nodes are heterogeneous and their qualities correlated, we directly implement the numerical solution. In this scheme, the broker will begin by first selecting the node with the highest expected quality using the regression models estimated as discussed in the previous section. The regression models are also used to estimate the value of adding a new node x to the current best combination of nodes i.e., $(\hat{Q}_{s_i+x} - \hat{Q}_{s_i})$. This is done in a greedy manner with the regression models used to compute expected maximum quality only for those combinations that involve subset s_i with the other nodes $x \notin s_i$.

***Input**: n nodes and regression coefficients to estimate qualities for any combination of nodes*
***Output**: Subset of nodes $(SelectedCombination)$ that offers maximum expected social welfare*
Procedure-Begin:
$s' = \{NULL\}$ *; temporary variable representing an element of the power set*
do
    $SelectedCombination = s'$
    MaxQual=0
    for x=1 to n
        Compute $\hat{Q}_{selectedCombination+x}$
        if $\hat{Q}_{selectedCombination+x}$ *>MaxQual*
            MaxQual=$\hat{Q}_{selectedCombination+x}$
            $s' = SelectedCombination + x$
        end-if
while $(\hat{Q}_{s'} - \hat{Q}_s) \geq C$
Procedure-End

Figure 6 Algorithm for the heuristic search scheme variant of the centralized scheme

The algorithm for this scheme is shown in Figure 6. To explain this heuristic search with a simple example, imagine $n = 4$ and the four agents are A, B, C and D. The broker has to evaluate and compare the regression estimates for all nodes. This corresponds to searching $n = 4$ points. Let A be the agent with the highest expected quality; the broker selects node A. To pick the next node, it estimates and compares node combinations AB, AC and AD. This search involves $n-1=3$ combinations. Suppose the combination AB generates the maximum incremental quality over node A and this is higher than the production cost incurred by B, the broker adds node B to the set of selected nodes. With this combination, AB, the broker performs a similar search comparing combinations ABC and ABD – searches $n-2=2$ combinations. If neither combination generates an incremental quality over combination AB that is higher than the participation cost, then, the broker stops the search. Thus, for selecting $k=\#(s_i)$ nodes, the total number of search points is $(n)+(n-1)+...+(n-k)$ which simplifies to $\#_{search}=(k+1)\,(n-k/2)$. When all nodes are selected, $k=n$, then, the number of search point becomes $\#_{search}=n\,(n+1)/2$. Therefore the decision-making cost for this scheme is

$$D^{s_i}_{heuristic}=\begin{cases}\dfrac{m\,(k+1)\,(n-k/2)}{2^n-1} & if\ k<n\\[2ex] \dfrac{m\,n\,(n+1)}{2\,(2^n-1)} & if\ k=n\end{cases}$$

Note that the algorithm presented here is greedy, sub-optimal and it searches only combinations that involve the highest-quality node.

3.2.2 Decision problem faced by each node under Decentralized Coordination Scheme

In this coordination scheme, nodes independently decide if they should participate in each RFQ. Since each node is profit maximizing, it decides to participate in an RFQ only if its expected profit from its participation in the RFQ is non-negative. Expected profit perceived by any node x, if it knows its quality distribution as $f_x(q)$, its probability of winning as $f_x(q)$, its production cost C and its cost of decision-making d, is given by

$$\int_0^1 p\,f_x(q)\,q\;\,f_x(q)\,dq-C-d\,.$$

Knowledge of the probability of winning requires not only knowledge of the distribution of the quality of other nodes but also the likelihood that they will participate.

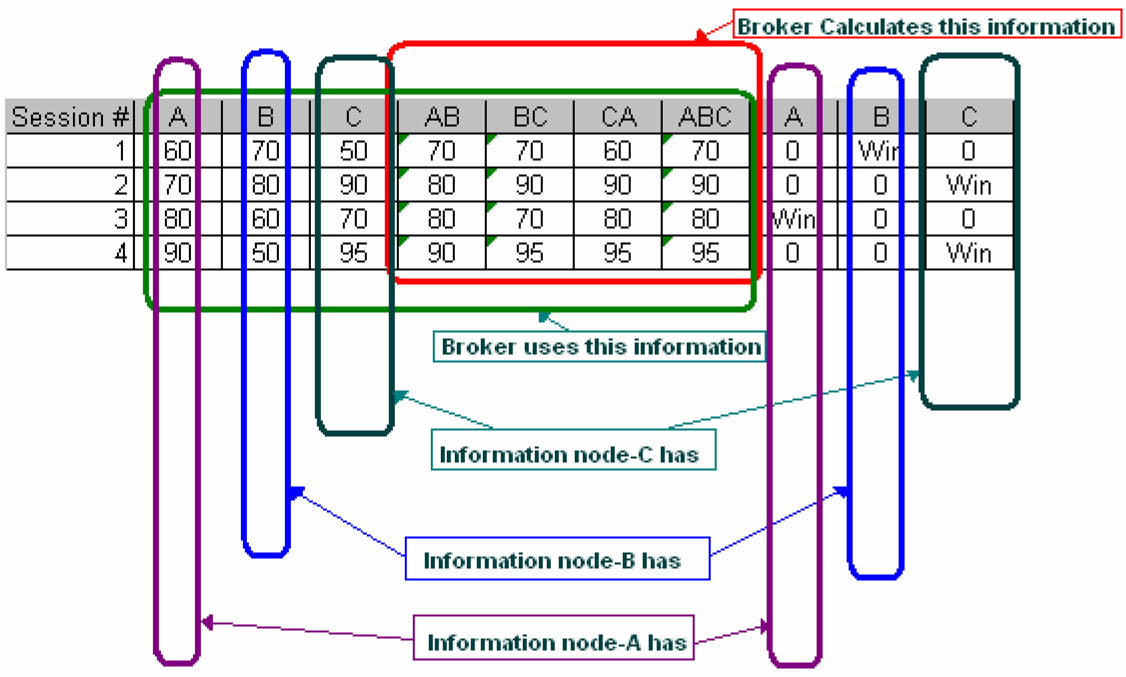
3.2.2.1 Decision strategy

In general, nodes may not know the capabilities of other competing nodes. Each node may be limited to knowing about its own capabilities and its likelihood of winning for each RFQ. If using the local information, each node evaluates its expected quality as $\hat{q}$ and its likelihood of winning as $\hat{p}$, then the expected profit for the node is $p\,\hat{p}\,\hat{q} - C - d$ where d represents the decision-making cost incurred by the individual node. In the decentralized scheme, we use the same conceptual move explained in Section 3.2.1.1 with the centralized schemes. We conduct a "calibration run" and using data gathered from that run, estimate regression models. The models relate individual node capabilities such as quality and likelihood of winning (equations 9 and 10) as a function of task characteristics – the independent variables – specified in the RFQ.

$$q = a_0 + (task\ characteristics)\, a_1 \cdots\cdots\cdots (9)$$

$$p = g_0 + (task\ characteristics)\, g_1 \cdots\cdots\cdots (10)$$

Regression coefficients, a's and g's, represent the knowledge possessed by the nodes. Although both the decentralized and centralized schemes use the "calibration-run", the data set used to estimate the regression models in the decentralized scheme is different from the data set available to the broker in the centralized schemes. The difference is highlighted in Figure 7. As before, columns 2-4 represent ex-post qualities of individual nodes. The key difference is that the broker receives information about all the nodes. In contrast, each node knows only the qualities it produced for all market sessions, based on which it can estimate the coefficients of equation 9. It also knows whether it was the winning agent for each market session using which it can estimate the coefficients of equation 10. In this manner, the decentralized scheme implements the concept of decision making with local information.



Session #	A	B	C	AB	BC	CA	ABC	A	B	C
1	60	70	50	70	70	60	70	0	Win	0
2	70	80	90	80	90	90	90	0	0	Win
3	80	60	70	80	70	80	80	Win	0	0
4	90	50	95	90	95	95	95	0	0	Win

Figure 7 Calibration-run and information revealed to individual nodes

Each node uses its set of coefficients to estimate its quality, $\hat{q}$, and its probability of winning, $\hat{p}$ before deciding about its participation in an RFQ. It participates in the RFQ only if its expected revenue is higher than the production cost i.e., $\hat{q}\hat{p} - C > 0$. Note that we assume that the decision-making cost does not affect the decision to participate. But the expressions for social welfare and node profits involve decision-making costs. Based on our definition, the decision-making cost for an individual node $d = m/(2^n - 1)$ since only one comparison is made; and the total decision-making cost $D = (m\ n)/(2^n - 1)$ since all n nodes incur this cost.

4. Results and Discussion

Given these choices of coordination schemes, what design choices should a market-designer make in order to maximize social welfare? To answer this question, we construct a market with eight nodes. We are limited by the cost of collecting and analyzing the calibration data for all $2^8 - 1$ combinations. Ex post quality values for each node is obtained by sampling a distribution. Our set-up provides a mechanism for manipulating these distributions such that they are correlated with one another by a desired factor. This mechanism is detailed in the appendix.

Assuming a specific correlation, r, we generate the distributions Y_i for $i = 1,2,...8$. Then, the set-up is executed for 'calibration run' or the *first phase*, by sampling the quality distributions Y_i to generate ex post qualities for 1500 market-sessions. The objective is to create a database that can be used to endow the nodes with knowledge about their own capabilities (in the form of regression models as explained earlier) and the broker with the knowledge about the capabilities of all the nodes (also in the form of regression models). In the *second phase*, the distributions used in phase 1 are sampled for another 1500 market-sessions. The correlation, r, the production cost, C, the decision-making cost, m, and the percentage of consumer utility paid as remuneration to the winning node, p, are the exogenous variables in this setup. We assume $C = 5$ unless otherwise explicitly mentioned and present the results of our analysis.

4.1 Effect of Percentage of Consumer Utility Paid as Prize

Table 2 Social Welfare in the decentralized scheme for different values of p.

Percentage of the consumer utility	Decentralized Scheme		
	Average Quality (%)	Average Social Welfare ($)	Efficiency (%) (Social welfare / Ex-post optimal social welfare)
100%	73.83 (0.27)	55.83 (0.27)	80.60
75%	72.85 (0.27)	58.91 (0.27)	85.05
50%	72.85 (0.27)	58.91 (0.27)	85.05
25%	41.02 (0.91)	38.41 (0.85)	54.9

As a first step, we analyze the effect of changing the percentage of consumer utility paid as prize to the wining node, p, on social welfare. We study this in a set-up where the correlation is set to a modest value of $r = 0.43$. Altering p does not have a direct effect on social welfare generated in the marketplace (refer to equation 4). It affects only the individual profits of the nodes. Since the broker's decision, in the heuristic-search scheme, is based on the collective social welfare generated, varying p does not influence the outcome of the heuristic-search scheme. However, in the decentralized scheme, decreasing p has conflicting indirect effects. The first effect is the reduction in the negative externality due to excess participation. The second effect, on account of lowered participation, is the reduction in the quality of the design. Table 2 shows that the decentralized scheme performs the best at an intermediate value of p. In subsequent analysis, we set $p = 0.75$.

4.2 Comparing the different Coordination Schemes

We first compare the schemes when the decision making cost, $m = 0$. In this case, the exhaustive search scheme is ex ante optimal. Using this framework, we compare the different schemes and find that correlation among nodes plays an important role. Figure 8 and Figure 9 show the variation of social welfare for each setting of correlation, r, at costs, $C = 1$ and $C = 5$, respectively. 95% confidence intervals are also shown in the figure for each setting. Before we analyze the effect of correlation on social welfare, observe that the heuristic search and the exhaustive search schemes perform identically. This is because when nodes are ex ante identical the performance of the heuristic matches the ex ante optimal. Further discussion on this issue is deferred to the next subsection.

The correlation in the performance of the nodes can be interpreted as a measure of the diversity among them. A high level of correlation would imply that all designs (products) are likely to be of similar quality. An increase in correlation (a reduction in diversity) reduces the ex ante benefits of tasking more than one node. In the limit, when the nodes are perfectly correlated, having ten nodes is no better than having one node since they all produce identical designs and only one node should be tasked. This logic also implies that efficiency (Social welfare/Ex-post optimal social welfare; discussed further below) increases with highly correlated nodes since tasking fewer nodes entails lower marginal costs of producing alternative designs. By contrast, with independent nodes (highly diverse), each node is likely to produce a different design providing a range of quality levels to the consumer and the broker can identify the best design. Thus we see that average social welfare tends to fall as correlation increases (diversity decreases). This is true for all coordination schemes, as shown in Figure 8.

However, the centralized schemes, which take into account the externalities across nodes, respond better to the increased correlation. As production cost increases, the number of nodes selected by the broker in both the exhaustive-search scheme and the heuristic-search scheme decreases and finally becomes one.

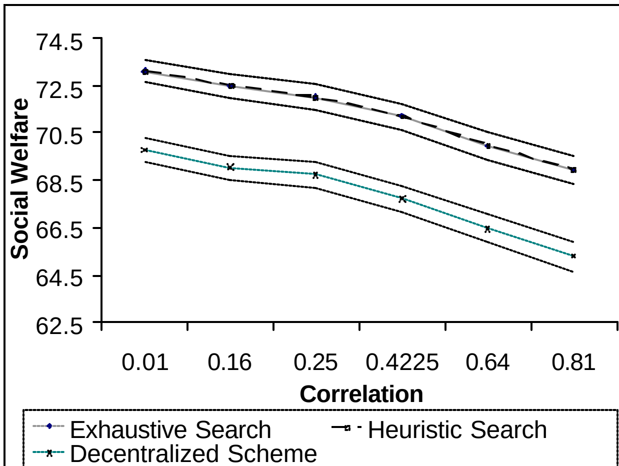


Figure 8 Social Welfare varying r at $C=1$

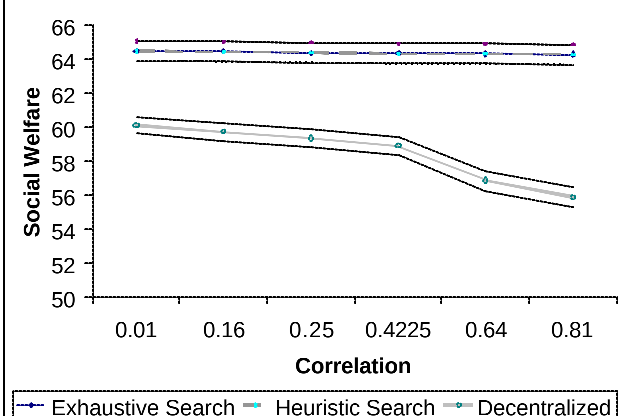


Figure 9 Social Welfare varying r at $C=5$

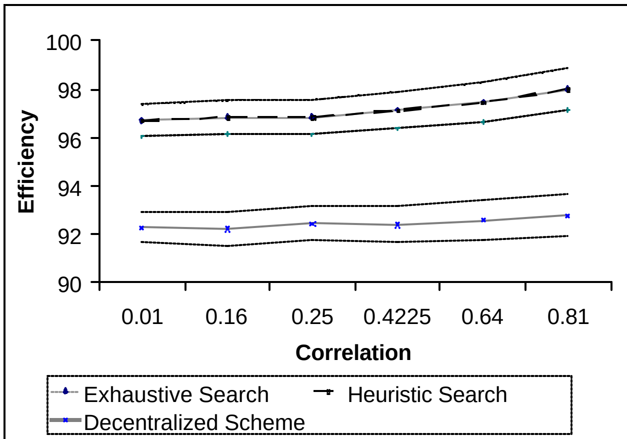


Figure 10 Efficiency varying r at $C = 1$

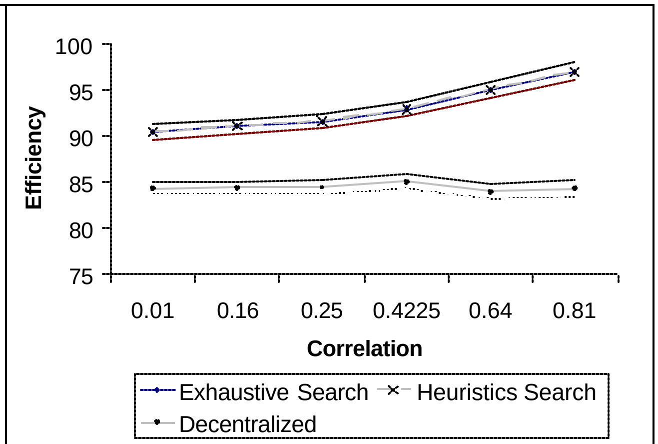


Figure 11 Efficiency varying r at $C = 5$

On the other hand, in the decentralized scheme decisions are made by nodes lacking information about other nodes (and specifically, about the correlation across nodes) and ignoring the impact of their actions on the payoffs of other nodes, which may result in excessive participation. This excess participation is especially important when marginal costs of production are high and when correlation is high. Comparing Figure 8 to Figure 9 one sees that the average gap between the centralized and the decentralized is larger with higher production cost and this gap increases with the correlation coefficient when production costs are high. (Recall that with higher correlation, the optimal number of nodes that should participate decreases.)

Further insight can be obtained by normalizing social welfare by the "ex post" welfare. Define *efficiency* of a scheme as the ratio of social-welfare generated under the setting relative to the ex post optimal, which is the social welfare produced by tasking the node that produces the highest quality ex post. Based on Figure 10 and Figure 11, we observe that none of the schemes achieve 100% efficiency and this is due to the ex ante uncertainty. Further, we also observe that whereas the efficiency of centralized coordination schemes (both exhaustive search and heuristic) increase with r, the efficiency of the decentralized scheme decreases with r.

Further, the gap between the two types of schemes becomes more pronounced, when C, the cost of building a design, increases. Indeed, as Table 3 shows, at high values of C, the centralized schemes are relatively more efficient than the decentralized schemes even at a modest value of $r = 0.43$. At $C = 15$, nodes in the decentralized scheme perceive very low expected profits and a few market-sessions elapse with no participation from any of the nodes.

Table 3 Sensitivity to Cost (standard deviations in parentheses)

Cost	Exhaustive Search Scheme			Heuristic Search Scheme			Decentralized Scheme		
	Average Quality (%)	Average Social Welfare ($)	Efficiency (%)	Average Quality (%)	Average Social Welfare ($)	Efficiency (%)	Average Quality (%)	Average Social Welfare ($)	Efficiency (%)
1	73.77 (0.28)	71.17 (0.28)	97.13	73.77 (0.28)	71.17 (0.28)	97.13	74.11 (0.27)	67.70 (0.26)	92.40
5	72.85 (0.27)	58.91 (0.27)	85.05	72.85 (0.27)	58.91 (0.27)	85.05	74.31 (0.29)	59.36 (0.26)	84.41
15	69.35 (0.30)	64.35 (0.30)	92.90	69.35 (0.30)	64.35 (0.30)	92.90	41.02 (0.91)	38.41 (0.85)	54.9

To summarize, social welfare decreases with increase in the correlation across nodes and with increases in the production cost. However, the relative efficiency of centralized schemes that takes into account the externalities across nodes increases with increases in r and with increase in the production cost.

4.3 Effect of Richness of the Correlation Data

An interesting question is why does the heuristic search scheme perform as well as the ex ante optimal? We find that the heuristic decision-making matches ex ante optimal only because the ex post qualities, we assume, are correlated by the same factor r. With richer variance-covariance data for ex post qualities a heuristic search scheme may not do as well. To investigate this question, first we executed the following simulation experiment. We divided the set of nodes into two equal sized groups. Nodes were correlated with others in their group by a coefficient r. Nodes in each group were negatively correlated with nodes in the other group by the same coefficient. The object was to determine if the greedy, sub-optimal nature of the heuristic search would incorrectly identify subsets due to the richer variance-covariance structure and yield sub-optimal performance. For example, suppose there are four nodes A, B, C and D. Let nodes A and B be in group-1 and nodes C and D in group 2. Nodes within each group are correlated by a factor r and are negatively correlated with members of the other group by a factor $(-r)$. However, even with this set-up, the performance of the heuristic search scheme was identical to the ex ante optimal (see Table 4).

Table 4: Social Welfare when correlations are both positive and negative

$\|r\|$	Exhaustive Search Scheme			Heuristic Search Scheme			Decentralized Scheme		
	Average Quality (%)	Average Social Welfare ($)	Efficiency (%)	Average Quality (%)	Average Social Welfare ($)	Efficiency (%)	Average Quality (%)	Average Social Welfare ($)	Efficiency (%)
0.25	72.56 (0.28)	64.56 (0.29)	89.88	72.56 (0.28)	64.56 (0.29)	89.88	75.38 (0.25)	60.43 (0.26)	84.13
0.5	74.10 (0.26)	65.10 (0.25)	90.57	74.10 (0.26)	65.10 (0.25)	90.57	76.30 (0.24)	60.34 (0.21)	83.95
1	75.24 (0.19)	66.24 (0.19)	92.98	75.24 (0.19)	66.24 (0.19)	92.98	72.06 (0.28)	60.15 (0.25)	84.43

To investigate the question further, we executed a third set of experiments with a richer covariance structure; three different values of correlation were used. In addition to the two sets of nodes that were positively and negatively correlated as discussed above, we introduced nodes that were uncorrelated with either set. As an illustration, consider five nodes A, B, C, D and E, such that A and C form one group, and nodes B and E form another group. As before, nodes in a group are correlated with each other by a factor of r but correlated by a factor of $(-r)$ with nodes in the other group. Node D is uncorrelated with nodes of either group. This richer variance-covariance structure results in sub-optimal performance of the heuristic method. Consider the three nodes, A, B and D. Node D is uncorrelated with A and B. Nodes A and B are perfectly negatively correlated and identically distributed. Let us assume that the (quality of the) nodes are uniformly distributed with means E(D) = 0.55, and E(A)=E(B)=0.5. If a maximum of quality of two nodes (recall, that qualities are random variables) is to be chosen, then the heuristic scheme picks the combination (D, A) or (D, B), since it always begins with the highest expected quality node, in this case D. However, the optimal combination is (A, B), since Max {A, D} = 0.71 (approx) is lower than Max {A, B} = 0.75.

Table 5 Social Welfare when correlations are positive, negative, and zero.

$\mid r \mid$	Exhaustive Search Scheme			Heuristic Search Scheme			Decentralized Scheme		
	Average Quality (%)	Average Social Welfare ($)	Efficiency (%)	Average Quality (%)	Average Social Welfare ($)	Efficiency (%)	Average Quality (%)	Average Social Welfare ($)	Efficiency (%)
0.25	71.52 (0.31)	64.50 (0.29)	90.46	71.52 (0.31)	64.5 (0.29)	90.46	74.99 (0.26)	60.75 (0.26)	85.20
0.5	71.54 (0.31)	64.53 (0.29)	90.56	71.54 (0.31)	64.53 (0.29)	90.56	75.56 (0.26)	58.58 (0.25)	82.23
1	73.22 (0.27)	65.22 (0.25)	91.89	72.94 (0.29)	64.92 (0.27)	91.88	74.66 (0.28)	59.67 (0.24)	84.45

With this set-up, we observe divergence between the exhaustive search scheme and the heuristic search scheme but this occurs only at high values of r (see Table 5). At low values of r, the performances of the two schemes are virtually identical.

By comparing Table 3, Table 4 and Table 5, one can observe that the efficiency of the heuristic-search scheme can be impaired by increasing the "richness" of the correlation structure in the data. Also, one can observe that the relative performance of the decentralized scheme improves. This suggests that richer correlation structures will tend to reduce the performance gap between centralized and decentralized schemes.

4.4 Effect of Decision-Making Cost

Table 6 Decision Cost Thresholds for the Decentralized Scheme to perform better

r	$C = 1$		$C = 5$	
	Decentralized Scheme > Heuristic Scheme	Decentralized Scheme > Exhaustive Search Scheme	Decentralized Scheme > Heuristic Scheme	Decentralized Scheme > Exhaustive Search Scheme
0.01	47.12	3.42	134.29	4.49
0.16	48.79	3.60	145.15	4.86
0.25	48.69	3.36	155.65	5.21
0.4225	54.82	3.57	198.19	5.59
0.64	63.43	3.61	272.38	7.68
0.81	78.75	3.77	307.45	8.68

Till now, we assumed that the decision-making cost is zero or negligible. However, in reality this is not the case. When decision-making costs (m) are high, exhaustive search may not be optimal. Table 6 shows the threshold values of m, the decision-making cost, when the decentralized scheme performs better than the centralized schemes. To provide some perspective, note that the welfare gap (the difference between the centralized and decentralized schemes) is of the order of \$5. A value of m = 50 implies that the cost of one additional search is about \$0.2 (since $50/(2^8 - 1) \approx \$\,0.2$). Thus, Table 6 shows that when we neglect the one-time costs of information gathering and processing, decision-making costs include only search costs. In this case, the marginal search costs have to be more than 20-25% of production costs (C) in order for the decentralized scheme to outperform the heuristic centralized scheme. However, the exhaustive search centralized scheme is inferior even at marginal search costs below 2% of production costs. Note also that as the number of nodes increases, the relative performance of the decentralized scheme is likely to improve. If the number of nodes were to double to 16, m implies that the marginal search cost is \$0.0007, or less than 1/1000th of the participation cost. Thus, if the welfare gap remains relatively stable as the number of nodes increases, decentralized schemes are likely to dominate centralized schemes.

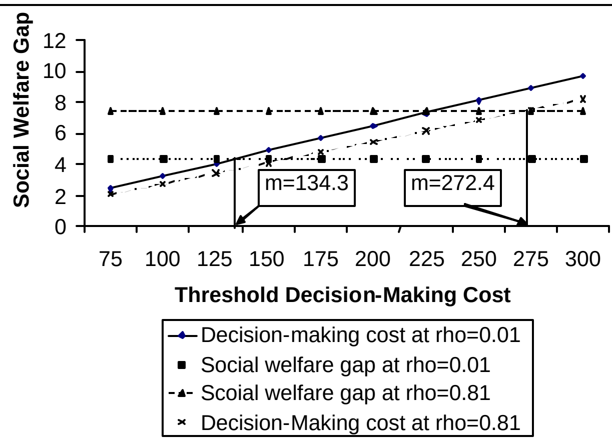


Figure 12 Threshold value when comparing against the heuristic search scheme.

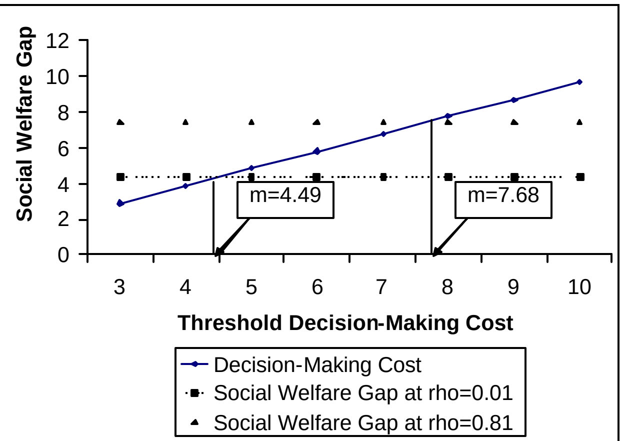


Figure 13 Threshold value when comparing against the exhaustive search scheme.

Table 6 also shows that the threshold value of m varies with r. As Figure 12 and Figure 13 show, this is due to two factors. First, the difference in decision-making costs decreases with r because at higher r, the number of searches decreases under the heuristic-search scheme (though not in the exhaustive search scheme). Second, the welfare gap increases with r. Thus the threshold value of m for the decentralized scheme to outperform either centralized scheme increases with r. This threshold value is more responsive to r in the heuristic scheme.

5. Conclusion

The comparison between the centralized and the decentralized coordination schemes is very important but results from prior literature are contradictory. This paper is distinctive in analyzing the interactions between the coordination and information in a market with ex-ante *uncertainty* about product quality. This paper uses a computational approach to study this issue. We studied three different coordination schemes a) centralized exhaustive search scheme b) centralized heuristic search scheme and c) decentralized scheme.

In both the centralized schemes, the broker possesses ex ante information about all nodes to optimize social welfare. In the exhaustive search scheme, the broker searches all combinations to select the best combination of nodes to task. On the other hand, in the heuristic-search scheme, it solves a stopping problem with a finite horizon. However, since fewer nodes are examined, decision-making cost is reduced. These centralized schemes are compared against the decentralized scheme where nodes optimize on individual profits using ex ante local information about their individual capabilities.

The ex post optimal social welfare is never reached in any coordination scheme including the exhaustive search scheme because of ex-ante uncertainty. The performance of the exhaustive search scheme, which, ignoring decision-making costs, is ex ante optimal, degenerates rapidly with increase in unit decision-making costs, improves with correlation across nodes and with increase in production costs.

Our results highlight two important issues. First, we demonstrate that as the number of nodes in the market increases, the threshold marginal decision-making cost for the decentralized scheme to outperform the heuristic scheme falls considerably. Second, the relative performance of the decentralized scheme improves with complex correlation structure between nodes. These results can provide a market designer with valuable insights that can be extended beyond the computational test-bed to understand the impact of their strategies and policies in the marketplace. Although the analysis in this paper has been limited to social welfare (sum of consumer and producer surplus), we can extend the analysis to other metrics such as consumer surplus, and nodes' profit.

Finally, there are elements of our approach that need to be further refined. As discussed, regression models estimated using a "calibration run" are used to make decisions in "real market sessions" in response to a RFQ. A more realistic analysis would require the use of an adaptive learning technique by each agent. In this scenario, the broker learns about the performance of all its registered nodes and each node learns from its own performance. This would make the decision processes dynamic and provide opportunity to study adaptive marketplace architectures.

In conclusion, we believe that the computational approach is a useful means to understand a problem that is central to all markets including the emerging electronic markets. We propose to continue investigating this line of research to create computational test beds that can be used to quickly instantiate and analyze alternative e-market designs.

References

Arora, A., Choudhary, V., Kannan, K., Krishnan, R., and Padman, R. 2000. Analyzing market schemes for software agents. 10th Annual Workshop on Information Technology and Systems, Brisbane, Australia.

Arora, A., Cooper, G., Krishnan, R., Padman, R. 2000. IBIZA: E-market Infrastructure for Custom-built Information Products. Information Systems Frontier Vol 2, No 1.

Arrow, K., and Debreu, G. 1954. Existence of an equilibrium for a competitive economy. Econometrica. 22. 265-290.

David, H. A. 1981. Order Statistics. John Wiley, New York.

Dasgupta, P., and Stiglitz, J. June 1980. Industrial Structure and the Nature of Innovative Activity. Economic Journal. 90(358). 266-93.

Ferguson, T.S. 1989. Who solved the secretary problem? Statistical Science. 4. 282-296.

Lange, O. 1969. The Computer and The Market. Socialism, Capitalism and Economic Growth. Edited by C.H. Feinstein, Cambridge University Press, 158-161.

Malone, T. 1987. Modeling Coordination in organizations and markets. Management Science. 33(10), 1317-1332.

Nault, B. Oct 1998. Information Technology and Organization Design: Locating decisions and information. Management Science. 44(10). 1321-1335.

Snir and Hitt, 2000. he Emerging Knowledge Economy: Exchange in Internet Spot Markets for IT Services. Working Paper, The Wharton School, University of Pennsylvania.

Tan, J.C., and Harker, P. 1999. Designing Workflow Coordination: Centralized Versus Market-Based Schemes. Information Systems Frontier. 10(4), 328-342.

Tirole, J. 1990. Theory of Industrial Organization. MIT Press, Cambridge, Masschussets.

---

[1] Our set-up can be modified to analyze a set-up for comparing organization profit

[2] Even with 8 nodes, the number of node combinations to be searched is $2^8 - 1$.

Appendix A. Creating Distributions with Desired Correlation

The key idea used to create these distributions is, when we combine any "base" distribution X with other distributions e_i (if e_is are identically distributed and independent of each other and also are identical and independent of X), in the following manner $Y_{ik} = l_{1i} X + l_{2i} e_i + C_{ik}$ (C_{ik} is a constant), correlation between any two distributions Y_{ik}, Y_{jk} will be $r = l_1^2/(l_1^2 + l_2^2)$.

In our set-up, we assume $X, e_i \sim b(a,b)$ with parameters a,b set to 1 and 3 respectively. The choice of beta distribution is bind the range for ex post quality in the range (0, 1). C_{ik}, is a constant that is assumed to vary with node i and task type k; its value is randomly set. For our simulation, we assume that there are 6 task types, so that k = 1,..6.

With this framework, the mean for Y_{ik} is $E(Y_{ik}) = l_{1i} \hat{X} + l_{2i} \hat{e}_i + C_{ik}$ and the variance is $Var(Y_{ik}) = l_{1i}^2 Var(X) + l_{1i}^2 Var(e_i)$. Since we assume X and e_i to be identically distributed, these expressions can be simplified $E(Y_{ik}) = (l_{1i} + l_{2i})\hat{e}_i + C_{ik}$, $Var(Y_{ik}) = (l_{1i}^2 + l_{1i}^2) Var(e_i)$. When changing r we retain both mean and variance as a constant. Mean value for Y_{ik} is randomly set. Variance is set to be $Var(Y_{ik}) = (0.25)\ Var(e_i)$. Then, correlation between distributions is $r = l_{1i}^2/(0.25)$. To achieve the desired correlation, we manipulate l_{1i}. l_{2i} is then calculated based on that as $l_{2i} = \sqrt{(0.25)^2 - l_{1i}^2}$. Finally, C_{ik} is adjusted to retain the mean value a constant. Using a similar set-up, one can also achieve a correlation of $(-r)$ between nodes A and B. For this, first set l_{1a}, l_{2a} to achieve the desired correlation. Then set $l_{1b} = -l_{1a}$, $l_{2a} = l_{2b}$. But for a third node, C, that is uncorrelated to nodes A and B but has the same variance of A and B, we set $l_{1c} = 0$ and let $l_{2c} = \sqrt{l_{1d}^2 + l_{1d}^2}$.