Supplier Selection Problem in an e-Marketplace

Context

Ashish Arora
Vidyanand Choudhary
Karthik Kannan
Ramyya Krishnan
Rema Padman

Carnegie Mellon University
Abstract

Research comparing centralized and decentralized coordination schemes is often contradictory. 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

Different supplier-selection methods are employed in e-Marketplaces. For example, in Freemarkets.com, the market-maker exploits its knowledge about the suppliers to screen and select a subset of them to participate in its auction (we refer to this scheme as the Centralized Coordination Scheme). In contrast, the market-maker of eLance.com acts only as an intermediary facilitating the marketplace. Suppliers in this marketplace independently decide about their participation (we referred to this scheme as the Decentralized Coordination Scheme). Note that these two schemes differ along the two dimensions - information and decision making authority. The Centralized Scheme uses global information and centralized decision-making, whereas the Decentralized Scheme uses only localized information and localized decision-making. These differences have differing impact on social welfare.

Our interest lies in comparing the two schemes from the perspective of a social welfare maximizing market-maker. In fact, a market-maker whose reverse-auction marketplace is owned by a consortium of buyers and suppliers would maximize social welfare. Examples of such marketplaces are: ForestExpress.com, an e-marketplace for the forest products industry, and Transora, an e-marketplace for consumer product goods owned by Coca-Cola Company, Proctor & Gamble and Grocery Store Association of America. In this reverse-marketplace, we assume that only web-site designs are transacted. Although we use this specific example context, our results are more generally applicable to marketplaces owned by a consortium of buyers and suppliers with the following characteristics: a) Bids having ex ante quality uncertainty. b) Buyer-specific bids: Buyer-specific products are products built to a particular buyer’s specification and cannot be resold to another buyer - they differ in the ability to satisfy user needs. c) Developing bids is costly.
d) Suppliers preparing bids have substantial heterogeneity. e) Participation decision-making costs are non-negligible. These characteristics are applicable to the web-site design reverse-marketplace (Snir & Hitt (2002)).

Notice that despite the redundancy of additional designs, ex ante uncertainty about the quality of the designs will lead to demand for multiple web site designs in this marketplace. 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 suppliers also increases costs. The Centralized Scheme balances these opposing demands while optimizing social welfare. Viewed differently, the Centralized Scheme minimizes the cost of under-participation or over-participation which otherwise might occur in the Decentralized Scheme. But, the Decentralized Scheme is simpler in practice since it requires no explicit coordination. This is important especially since the performance of the Centralized Scheme deteriorates with increase in decision-making costs. This result is intuitive and Hayek (1945) has argued about it. But, these arguments are valid only when the qualities of the suppliers are independent. Our contribution is to examine the impact on the performance of the coordination schemes due to the correlation in qualities of products.

To provide some intuition, imagine a scenario where the buyer specifies the willingness to pay as a function of the quality of the product and the suppliers compete on quality. In such a case, the participation of suppliers whose qualities are highly correlated is wasteful. Redundancy because of excess participation can be eliminated in the Centralized Scheme. If the savings realized due to avoiding the redundancy is higher than the decision-making cost, then the performance of the Centralized Scheme is improved. In fact, we demonstrate in this paper that it is not just the value
of correlation that is important; it is also the richness of the correlation matrix (correlation among
different suppliers) that plays a significant role in deciding the performance of one scheme over
the other. Interesting insights about the impact of correlation on social welfare is gained from this
analysis. Further, this comparison is applicable to real-world context since typically suppliers may
have their product qualities correlated - coal suppliers from the same geographic region may have
similar qualities highly correlated. Similarly, firms using similar tools for web-site designs, say
Macromedia’s Dreamweaver, may have their qualities correlated when compared to firms using
some other tool, say Microsoft’s Visual Studio.

The rest of this paper is organized as follows. We motivate our research question and review
the relevant literature in Section 2. We discuss the market mechanism and the rationale underlying
our methodological approach in Section 3. In Section 4, we present the alternative coordination
schemes that are studied, with an emphasis on the decision problems that need to be solved. We
present the results of our computational study in Section 5, and conclude in Section 6 with a
discussion of the implications of our work for implementing reverse marketplaces.

2 Literature Review

Research comparing the performance of centralized coordination schemes and decentralized coor-
dination 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 Lange (1969) and others pointed
out that a centralized computer could match demand and supply more efficiently than the adjust-
ment processes that characterize actual markets. In response, Hayek (1945) and others pointed out that decentralized systems greatly economize on the amount of time required by decision makers since they coordinate using market prices (Arrow & 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 lead to market failure (Tirole (1990)). For instance, where many suppliers “race” for a single prize, decentralized schemes may lead to excessive entry because suppliers disregard the negative externality on other suppliers of their decision to enter Dasgupta & Stiglitz (1980).

Apart from economists, a few organization theorists have 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 as the number of suppliers increases. Tan & 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.

A similar problem is addressed by Nault (1998). Nault (1998) compares the following schemes on organizational profits: 1) Centralized system where both global and local investment-decisions are made by the central authority. 2) Decentralized system where decisions about local and the global investment-decisions are made by the individual suppliers, and 3) Mixed mode, where decisions about global investments are made by the central authority but the local investment decisions are made by the decentralized suppliers. Using analytical model, Nault (1998) determines suf-
Table 1: Related Work

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Product Heterogeneity</td>
<td>Heterogeneous</td>
<td>Homogeneous</td>
<td>Homogeneous</td>
<td>Heterogeneous</td>
</tr>
<tr>
<td>Competition among supplier</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Demand Function</td>
<td>Unit</td>
<td>Generic Function</td>
<td>Uncertainty about Demand</td>
<td>Unit</td>
</tr>
<tr>
<td>Performance Metric</td>
<td>Total Expected Cost</td>
<td>Organization Profit</td>
<td>Organization Profit</td>
<td>Social Welfare $^1$</td>
</tr>
</tbody>
</table>

ficient conditions under which hierarchy (centralized) is better than markets (decentralized) and similarly conditions when hierarchy (centralized) is better than the mixed mode.

Summary of how this paper is different from Anand & Mendelson (1997), Nault (1998), Tan & Harker (1999) is tabulated and shown in Table 1. Notice that the primary factor differentiating our work from others is that we assume competition among suppliers. This also means that we need a model different from those in the papers mentioned above. For our analysis, we use a model similar to that of Snir & Hitt (2002). Their paper analytically models the behavior of suppliers (firms) in eLance.com - the decentralized scheme. Based on this analytical model, they hypothesize the bidding and participation behavior of the suppliers, which they validate empirically. Our paper extends their framework along two dimensions. One, in addition to analyzing the decentralized scheme, we also study the centralized scheme. Two, we compare the schemes by relaxing the assumptions used by Snir & Hitt (2002): the qualities of the suppliers are independent and suppliers are homogeneous. When these assumptions are relaxed, the problem becomes analytically intractable. We argue about this intractability result in the next section to make a case for the computational methodology used in this paper.
Before we make a case for computational methodology, we describe the market-mechanism used in this framework. In our reverse marketplace, there are three types of participants - the buyer, the broker, and suppliers. The buyer needs a “product” (web site) and would like to accept a bid (web site design) that offers him the highest “quality”. The buyer initiates a market session by submitting the desired characteristics of the product as a part of the RFQ to the broker. The broker acts as an intermediary for a marketplace involving multiple suppliers. Each supplier has a different asset that it can use to build site-designs in response to a buyer’s request. Building the bid entails participation costs. Further, the “quality” of the bid is unknown until the end of the market session. For a specific market session, the suppliers are either chosen (centralized scheme) or independently decide (decentralized scheme) to participate based on the coordination scheme adopted. Figure 1 and Figure 2 illustrate the locus of decision-making in each of the two schemes. The broker is the decision maker in the centralized scheme while each supplier is the decision maker in the decentralized scheme. Under each scheme, suppliers that have either been selected or which choose to participate, build site-designs in response to the RFQ.
After receiving the bids from the suppliers, the broker evaluates the bids for the buyer, returns the highest quality to the buyer and compensates the winning supplier under both schemes. For tractability, we assume that the “quality” can be mapped to a value in the range \([0, 1]\). This mechanism is used to study the effect on social welfare due to differences in the manner in which decision problems related to participation are solved under each scheme. We begin by defining our metric for comparison - social welfare.

### 3.1 Social Welfare

Let \(s_i\) denote the subset of all suppliers \(T\) i.e., \(s_i \in \{s \mid s \subseteq T\}\). Also, let \(n = \#(T)\) represent the number of suppliers in the market. Let \(Q_{s_i}\) denote the best quality when suppliers in subset \(s_i\) are choose to participate (in either scheme), and let \(k = \#(s_i)\) represent the cardinality of that set. Let the buyer’s willingness to pay equal the best quality. If we assume that \(p\) (exogenous) share of the buyer utility is paid to the supplier, then the buyer surplus generated is given by

\[
B = (1 - p)Q_{s_i} \tag{1}
\]

Let \(q_y\) represent the quality produced by any supplier \(y\). If supplier \(y\) is the winner generating the best quality in the marketplace \(q_y = Q_{s_i}\), it receives a payment from the buyer while other suppliers generate zero revenue for that market session. In general,

\[
\Pi_y = \begin{cases} 
q_y p & \text{if } q_y = Q_{s_i} \\
0 & \text{otherwise}
\end{cases}
\]

If \(C\), an exogenous parameter, represents the production cost for each supplier, and is assumed
to be a constant then

$$\Pi_y = \begin{cases} 
q_y p - C & \text{if } q_y = Q_{s_i} \\
-C & \text{otherwise}
\end{cases}$$

If $D$ is the total decision-making cost, the social welfare generated when suppliers in subset $s_i$ participate is

$$SW = Q_{s_i} - kC - D$$  \hspace{1cm} (2)$$

When $p = 1$, buyer surplus generated is zero and in such a case, the total social welfare generated in the decentralized scheme corresponds to the industry profit.

### 3.1.1 Coordination Schemes and Associated Decision Problems

Using the social welfare metric defined in the earlier subsection, we compare the two schemes. Decision problems solved under each scheme is described below. In the Centralized Scheme, the broker is the decision-making authority that tasks the best combination of suppliers to optimize social welfare. If the broker knows the distribution, $f(Q_{s_i})$, of the best quality, $Q_{s_i}$, when suppliers in subset $s_i$ participate, then the decision problem is

$$\max_{s_i} \left\{ \int_{Q_{s_i}}^{Q_{s_i}^{\max}} f(Q_{s_i}) Q_{s_i} dQ_{s_i} - #(s_i) C - D \right\}$$  \hspace{1cm} (3)$$

where the integration term represents the expected best quality, $Q_{s_i}$, when suppliers in subset $s_i$ participate. Note that this is a stochastic combinatorial optimization problem. After solving the decision problem, the broker invites a subset of suppliers to participate. Although, the suppliers may sometimes reject the invitation, our assumption precludes that i.e., supplier participates so long as it is invited.$^2$

---

$^2$We base this assumption on our ability to analytically demonstrate that it is sub-optimal for suppliers to reject invitations from the market-maker so long as firms are homogeneous.
In decentralized coordination scheme, suppliers independently decide if they should participate in each RFQ. Since each supplier is profit maximizing, it decides to participate in an RFQ, only if its expected profits from its participation in that RFQ is non-negative. In our case, the expected profits for any seller $x$ is given by

$$
\Pi_{DC} = p E(q_x|q_x = Q_{s_i}) - C - d
$$

where $E(q_x|q_x = Q_{s_i})$ is the expected quality conditional on winning, $d$ is the decision-making cost and $C$ is the production cost. If we assume that quality distribution, $f_x(q_x)$, and the probability of winning $\phi_x(q_x)$, are independent, then

$$
\Pi_{DC} = \int_{Q_{\min}}^{Q_{\max}} p f_x(q_x)\phi_x(q_x)dq_x - C - d
$$

3.2 Need for a Computational Approach

equation 5 and equation 3 can be analytically computed if we assume that the suppliers are homogeneous and their qualities are independent. Without these, the problem becomes analytically intractable. We consider each assumption separately in the context of the Centralized Scheme’s decision problem and use relevant literature in Statistics to make our argument. Recall that in the Centralized Scheme we are interested in deriving the first order statistic of the quality distributions which distributed both non-identically and non-independently. Only a very few results are available to compute order statistics for non-independent random variables “because the situation is considerably complex” (Siegel (1993), pp.77). In fact, all these results are applicable only to the case when the random variable was distributed normally (Rawlings (1976); Hill (1976); Owen & Steck (1962)). When distributions are non-identical, no tractable analytical formulation is possible (Cao & West (1997)).
To augment our case further, we cite David (1981) which states that no tractable analytical formulation is possible to solve a combinatorial problem similar to the the broker’s decision problem even under zero decision-making cost:

\[
\max_{s_i} \{ \hat{Q}_{s_i} - \#(s_i) C \} \tag{6}
\]

where \( \hat{Q}_{s_i} \) denotes the expected best quality when suppliers in subset \( s_i \) are tasked. When decision-making costs, \( D \), to arrive at the optimal set are considered, the optimization problem for the centralized decision maker is,

\[
\max_{s_i} \{ \hat{Q}_{s_i} - \#(s_i) C - D \} \tag{7}
\]

Observe that the decision-making cost adds another layer of complexity to an already intractable combinatorial problem. Given these constraints in analytically comparing the schemes, we make the context richer by relaxing both these assumptions and by computationally comparing the schemes.

4 Computational Methodology

After having argued for the computational model, we revisit the decision problem under each coordination scheme and describe the strategies used by the decision makers to solve them. We begin with the Centralized Scheme

4.1 Centralized Coordination Scheme

Recall that the broker is the decision-making authority and the decision problem is

\[
\max_{s_i} \{ \int_{Q_{\min}}^{Q_{\max}} f(Q_{s_i})Q_{s_i} dQ_{s_i} - \#(s_i) C - D \} \tag{8}
\]
To solve the problem, we propose two alternative strategies. 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.

4.1.1 Exhaustive Search Strategy

A straightforward solution is to exhaustively search the power set of $n$ suppliers - $2^n - 1$ combinations - and choose the optimal set of suppliers 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

$$\max_{s_i} \{ \int_{Q_{min}}^{Q_{max}} f(Q_{s_i}) Q_{s_i} dQ_{s_i} - \#(s_i) \ C \}$$

For solving this, the broker must know how to compute the expected task capability $\hat{Q}_{s_i}$ for any supplier-combination $s_i$. In our case, this knowledge is assumed to be available to the broker in the form of a regression equation 9 whose independent variables correspond to the task-specific characteristics in the RFQ. Instead of regression, any form of learning mechanism can be used.

$$Q_{s_i} = \alpha_{0,s_i} + \text{task characteristics} \ \alpha_{1,s_i}$$

Independent variables in these equations correspond to task characteristics specified by the buyer. Regression coefficients, $\alpha$, represent the knowledge possessed by the broker about the quality of suppliers 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 supplier responds to the RFQ with a design (its product) which is evaluated and assigned a quality. A sample calibration-run is shown in Figure 3 for $n = 3$. Columns 2, 3 and 4 represent the ex post
qualities of participating suppliers. 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 suppliers in the market as a function of the task characteristics specified in the RFQ. Columns 4, 5, 6, and 7 in Figure 3 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 3 for each session.

In response to a new RFQ, the broker uses the regression model (equation 9 with coefficient estimates from the calibration run) to compute the expected quality for all supplier combinations. Based on these estimates, the broker chooses the supplier combination that generates the highest expected social welfare and tasks only the specific suppliers in the combination. The algorithm is presented in Figure 4. 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.
Input: n suppliers and regression coefficients to estimate qualities for each element of the power set of suppliers
Output: Element of the power set (SCombo) that offers maximum expected social welfare
Procedure-Begin:

\[
\begin{align*}
\text{maxSW} & = 0 ; \text{Maximum expected social welfare} \\
\text{SCombo} & = NULL ; \text{Selected supplier combination} \\
\text{For } s_i \text{ in power set of } n \\
\text{Compute } \hat{Q}_{s_i} \text{ using regression coefficients} \\
\text{SW} & = \hat{Q}_{s_i} - \#(s_i) C \\
\text{maxSW} & = \max \text{maxSW, SW} \\
\text{if } \text{maxSW} == \text{SW} \\
\text{SCombo} & = s_i \\
\end{align*}
\]

End-if
End-For
Procedure-End

Figure 4: Algorithm for Exhaustive-search centralized scheme.

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 supplier combination and comparing supplier-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,

\[
D = \frac{m \#_{\text{search}}}{2^n - 1}
\]

where \(m\) is an exogenous variable that represents the total decision-making cost for the exhaustive search given \(n\). \(#_{\text{search}}\) represents the number of comparisons made and when \(#_{\text{search}} = 2^n - 1\), \(D = m\). Relative to other schemes, the decision-making costs of the exhaustive search scheme increase exponentially with \(n\), the number of suppliers.

13
4.1.2 Heuristic decision making strategy

Similar to the previous scheme, the broker is again the decision-making authority that decides which subset $s_i$ of the $n$ suppliers to task to maximize social welfare. Although the exhaustive search scheme provides a mechanism for choosing the best set of suppliers 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 suboptimal-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 suppliers, the broker has to decide whether to add another supplier 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 supplier $x \notin s_i$ to add to this set such that the maximum expected increment in quality, the positive contribution to social welfare, is at least as much as the cost incurred by the participation of the additional supplier, the loss in social welfare i.e.,

$$\max_{s_i}\{\hat{Q}_{s_i+x} - \hat{Q}_{s_i}\}$$

$$\text{s.t. } \hat{Q}_{s_i+x} - \hat{Q}_{s_i} \geq C$$

This selection process is repeated until all suppliers are selected or until the stopping condition is met.

Note that this algorithm implicitly takes into account the correlation among the qualities of the suppliers. It begins with the broker selecting the supplier with the highest expected quality.
Input: $n$ suppliers and regression coefficients to estimate qualities for any combination of suppliers
Output: Subset of suppliers ($S_{Combo}$) that offers maximum expected social welfare

Procedure-Begin:

$s' = NULL$ ; temporary variable representing an element of the power set

do

$S_{Combo} = s'$

$MaxQual = 0$

for $x = 1$ to $n$

Compute $\hat{Q}_{S_{Combo} + x}$ using regression coefficients

If $\hat{Q}_{S_{Combo} + x} > MaxQual$

$MaxQual = \hat{Q}_{S_{Combo} + x}$

$s' = S_{Combo} + x$

end-if

while $Q_{s'} - \hat{Q}_s \geq C$

Procedure-End

Figure 5: Algorithm for the heuristic search scheme variant of the centralized scheme.

In our framework, the expected quality is estimated using the regression models as discussed in the previous section. The regression models are also used to estimate the value of adding a new supplier $x$ to the current best combination of suppliers i.e., $\hat{Q}_{S_{Combo} + x} - \hat{Q}_{S_{Combo}}$. The savings in the decision-making cost comes from not having to compute the estimates for the entire power set, but only for those combinations that involve subset $s_i$ with each of the other suppliers $x \notin s_i$.

The algorithm for this scheme is shown in Figure 5. 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 suppliers. This corresponds to searching $n = 4$ points. Let A be the agent with the highest expected quality; the broker selects supplier A. To pick the next supplier, it estimates and compares supplier combinations AB, AC and AD. This search involves $n - 1 = 3$ combinations. Suppose the combination AB generates the maximum incremental quality over supplier A and this is higher than the production cost incurred by B, the broker adds supplier
B to the set of selected suppliers. 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 \( k = \#(s_i) \) suppliers, the total number of searches made is \( n + (n - 1) + \ldots + (n - k) \) which simplifies to \( \#_{search} = (k + 1)(n - k/2) \). When all suppliers are selected, \( k = n \), then, the number of searches made is \( \#_{search} = n(n + 1)/2 \).

Therefore the decision-making cost for this scheme is

\[
D_{Heuristics}^{s_i} = \begin{cases} 
\frac{m(k+1)(n-k/2)}{2^{n-1}} & \text{if } k < n \\
\frac{m n (n+1)}{2(2^{n-1})} & \text{if } k = n 
\end{cases}
\]

Note that the algorithm presented here is greedy, suboptimal and it searches only combinations that involve the highest-quality supplier.

### 4.2 Decentralized Coordination Scheme

Recall that each supplier independently decides if it should participate in each RFQ and the expected profit in equation 5 is rewritten here

\[
\Pi_{DC} = \int_{Q_{\min}}^{Q_{\max}} p_{x}(q_{x})q_{x}\phi_{x}(q_{x})dq_{x} - C - d
\]  

(11)

In this expression, knowledge of the probability of winning requires not only knowledge of the distribution of the quality of other suppliers but also the likelihood that they will participate.

#### 4.2.1 Decision Strategy

In general, suppliers may not know the capabilities of other competing suppliers. Each supplier may be limited to knowing about its own capabilities and its likelihood of winning for each RFQ.
If using the local information, each supplier evaluates its expected quality as $\hat{q}$ and its likelihood of winning as $\hat{\pi}$ then the expected profit for the supplier is $p\hat{\pi}\hat{q} - C - d$ where $d$ represents the decision-making cost incurred by the individual supplier. In the decentralized scheme, we use the same idea explained in the centralized scheme. We conduct “calibration runs” and using data gathered from those runs, estimate regression models which mimic the knowledge gained by the individual suppliers. The regression models relate individual supplier capabilities such as quality and likelihood of winning (equations 12 and 13) as a function of task characteristics - the independent variables - specified in the RFQ.

$$q = \gamma_0 + \text{task characteristics } \gamma_1$$  \hspace{1cm} (12)

$$\pi = \alpha_0 + \text{task characteristics } \alpha_1$$  \hspace{1cm} (13)

Regression coefficients, $\alpha$’s and $\gamma$’s, represent the knowledge possessed by the suppliers. Although both the decentralized and centralized schemes use the same “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 6. As before, columns 2-4 represent ex-post qualities of individual suppliers. The key difference is that the broker receives information about all the suppliers. In contrast, each supplier knows only
the qualities it produced for all market sessions, based on which it can estimate the coefficients of equation 12. It also knows whether it was the winning agent for each market session using which it can estimate the coefficients of equation 13. In this manner, the decentralized scheme implements the concept of decision making with local information.

Each supplier uses its set of coefficients to estimate its quality, \( \hat{q} \), and its probability of winning, \( \hat{\pi} \), 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., \( p\hat{q}\hat{\pi} - C > 0 \). Note that estimates of \( \hat{\pi} \) takes into account the correlation among the qualities of the suppliers. Also, note that the decision-making cost is a sunk cost that does not affect the supplier’s decision to participate. However, decision-making costs appear in the expressions for social welfare and supplier profits involve. Based on our definition, the decision-making cost for an individual supplier \( d = m/(2^n - 1) \) since only one comparison is made; and the total decision-making cost \( D = n m/(2^n - 1) \) since all \( n \) suppliers incur this cost.

5 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 suppliers. 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 supplier 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, \( \rho \), 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 suppliers 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 suppliers (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, $\rho$, the production cost, $C$, the decision-making cost, $m$, and the percentage of buyer utility paid as remuneration to the winning supplier, $p$, are the exogenous variables in this setup. We assume $C = 5$ unless otherwise explicitly mentioned and present the results of our analysis.

### 5.1 Effect of Percentage of Buyer Utility Paid as Prize

As a first step, we analyze the effect of changing the percentage of buyer utility paid as prize to the winning supplier, $p$, on social welfare. We study this in a set-up where the correlation is set to a modest value of $\rho = 0.43$. Altering $p$ does not have a direct effect on social welfare generated in the marketplace (refer to equation 3). It affects only the individual profits of the suppliers. Since the broker’s decision, in the heuristic-search scheme, is based on the collective social welfare,

<table>
<thead>
<tr>
<th>Percentage of the buyer utility</th>
<th>Average Quality (%)</th>
<th>Average Social Welfare</th>
<th>Efficiency (Social Welfare/Ex-post optimal social welfare)</th>
</tr>
</thead>
<tbody>
<tr>
<td>100%</td>
<td>73.83 (0.27)</td>
<td>55.83 (0.27)</td>
<td>80.60</td>
</tr>
<tr>
<td>75%</td>
<td>72.85 (0.27)</td>
<td>58.91 (0.27)</td>
<td>85.05</td>
</tr>
<tr>
<td>50%</td>
<td>72.85 (0.27)</td>
<td>58.91 (0.27)</td>
<td>85.05</td>
</tr>
<tr>
<td>25%</td>
<td>41.02 (0.91)</td>
<td>38.41 (0.85)</td>
<td>54.9</td>
</tr>
</tbody>
</table>

Table 2: Social Welfare in the decentralized scheme for different values of $p$. 
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 $p = 0.75$.

5.2 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 suppliers plays an important role. Figure 7 and Figure 8 show the variation of social welfare for each setting of correlation, $\rho$, 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 suppliers 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 suppliers 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 supplier. In the limit, when the suppliers are perfectly correlated, having ten suppliers is no better than having one supplier since they all produce identical designs.
and only one supplier should be tasked. This logic also implies that efficiency (Social welfare/Ex-post optimal social welfare; discussed further below) increases with highly correlated suppliers since tasking fewer suppliers entails lower marginal costs of producing alternative designs. By contrast, with independent suppliers (highly diverse), each supplier is likely to produce a different design providing a range of quality levels to the buyer 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 7.
Table 3: Sensitivity to Cost. In this table, \( \hat{Q} \) represents the average Quality, \( \hat{SW} \) represents social welfare, \( \eta \) represents Efficiency while standard deviations are represented in parentheses.

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

On the other hand, in the decentralized scheme decisions are made by suppliers lacking information about other suppliers (and specifically, about the correlation across suppliers) and ignoring the impact of their actions on the payoffs of other suppliers, 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 7 to Figure 8, 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 suppliers that should participate decreases). Further insights 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 supplier that produces the highest quality ex post.
Based on Figure 9 and Figure 10, 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 $\rho$, the efficiency of the decentralized scheme decreases with $\rho$. Further, the gap between the two types of schemes becomes more pronounced, when $C$, the cost of building a design, increases. Indeed, the centralized schemes are relatively more efficient than the decentralized schemes even at a modest value of $\rho = 0.43$ for higher values of $C$ (see Table 3. At $C = 15$, suppliers in the decentralized scheme perceive very low expected profits and a few market-sessions elapse with no participation from any of the suppliers.

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

### 5.3 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 $\rho$. 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 suppliers into two equal sized groups. Suppliers were correlated with others in their group by a coefficient $\rho$. Suppliers in each group were negatively correlated with suppliers in the other group by the same coefficient. The
Table 4: Social Welfare when correlations are both positive and negative. In this table, \( \hat{Q} \) represents the average Quality, \( \hat{SW} \) represents social welfare, \( \eta \) represents Efficiency while standard deviations are represented in parentheses.

<table>
<thead>
<tr>
<th>( \rho )</th>
<th>Exhaustive Search Scheme</th>
<th>Heuristic Search Scheme</th>
<th>Decentralized Scheme</th>
</tr>
</thead>
<tbody>
<tr>
<td>( Q ) (%)</td>
<td>( SW ) ($)</td>
<td>( \eta ) (%)</td>
<td>( Q ) (%)</td>
</tr>
<tr>
<td>0.25</td>
<td>72.56 (0.28)</td>
<td>64.56 (0.29)</td>
<td>89.88</td>
</tr>
<tr>
<td>0.5</td>
<td>74.1 (0.26)</td>
<td>65.1 (0.25)</td>
<td>90.57</td>
</tr>
<tr>
<td>1</td>
<td>75.24 (0.19)</td>
<td>66.24 (0.19)</td>
<td>92.98</td>
</tr>
</tbody>
</table>

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 suppliers A, B, C and D. Let suppliers A and B be in group-1 and suppliers C and D in group 2. Suppliers within each group are correlated by a factor and are negatively correlated with members of the other group by a factor \(( -\rho )\). However, even with this set-up, the performance of the heuristic search scheme was identical to the ex ante optimal (see Table 4).

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 suppliers that were positively and negatively correlated as discussed above, we introduced suppliers that were uncorrelated with either set. As an illustration, consider five suppliers A, B, C, D and E, such that A and C form one group, and suppliers B and E form another group. As before, suppliers in a group are correlated with each other by a factor of \( \rho \) but correlated by a factor of \(( -\rho )\) with suppliers in the other group. Supplier D is uncorrelated with suppliers of either group. This
TABLE 5: Social Welfare when correlations are both positive and negative. In this table, $\hat{Q}$ represents the average Quality, $\hat{SW}$ represents social welfare, $\eta$ represents Efficiency while standard deviations are represented in parentheses.

<table>
<thead>
<tr>
<th>$\rho$</th>
<th>Exhaustive Search Scheme</th>
<th>Heuristic Search Scheme</th>
<th>Decentralized Scheme</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$Q$ (%)</td>
<td>$SW$ ($)</td>
<td>$\eta$ (%)</td>
</tr>
<tr>
<td>0.25</td>
<td>71.52 (0.31)</td>
<td>64.5 (0.29)</td>
<td>90.46</td>
</tr>
<tr>
<td>0</td>
<td>71.54 (0.31)</td>
<td>64.53 (0.29)</td>
<td>90.56</td>
</tr>
<tr>
<td>1</td>
<td>73.22 (0.27)</td>
<td>65.22 (0.25)</td>
<td>91.89</td>
</tr>
</tbody>
</table>

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 $\rho$ (see Table 5). At low values of $\rho$, 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.

### 5.4 Effect of Decision-Making Cost

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 suppliers increases, the relative performance of the decentralized scheme is likely to improve. If the number of suppliers were to double to 16, the marginal search cost, $m$, is 0.0007 which is less than 1/1000th of the participation cost. Thus, if the welfare gap remains relatively stable as the number of suppliers increases, decentralized scheme is likely to dominate the centralized schemes.

Table 6 also shows that the threshold value of $m$ varies with $\rho$. As Figure 12 and Figure 11 show, this is due to two factors. First, the difference in decision-making costs decreases with $\rho$ because at higher $\rho$, the number of searches decreases under the heuristic-search scheme (though not in the exhaustive search scheme). Second, the welfare gap increases with $\rho$. Thus the threshold value of $m$ for the decentralized scheme to outperform either centralized scheme increases with $\rho$. This threshold value is more responsive to $\rho$ in the heuristic scheme.
6 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 suppliers to optimize social welfare. In the exhaustive search scheme, the broker searches all combinations to select the best combination of suppliers to task. On the other hand, in the heuristic-search scheme, it solves a stopping problem with a finite horizon. However, since fewer suppliers are examined, decision-making cost is reduced. These centralized schemes are compared against the decentralized scheme where suppliers 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 suppliers and with increase in production costs.

Our results highlight two important issues. First, we demonstrate that as the number of suppliers 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 suppliers. 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 buyer and producer surplus), we can readily extend the analysis to other metrics such as buyer surplus, and suppliers’ 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 suppliers and each supplier 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.
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 $\epsilon_i$ (if $\epsilon_i$s are identically distributed and independent of each other and also are identical and independent of $X$), in the following manner

$$Y_{ik} = \lambda_1 X + \lambda_2 \epsilon_i + C_{ik} \quad (C_{ik} \text{ is a constant}),$$

correlation between any two distributions $Y_{ik}, Y_{jk}$ will be

$$\rho = \frac{\lambda_2}{\lambda_1 + \lambda_2}.$$

In our set-up, we assume $X, \epsilon_i \sim \beta(a, b)$ with parameters $a$ and $b$ set to 1 and 3 respectively. The choice of $\beta$ distribution is to bind the range for ex post quality in the range $(0, 1)$. $C_{ik}$, is a constant that is assumed to vary with supplier $i$ and task type $k$. For our simulation, we assume that there are 6 task types, so $k = 1, 2, \ldots 6$.

With this framework, the mean for $Y_{ik}$ is $E(Y_{ik}) = \lambda_1 X + \lambda_2 \epsilon_i + C_{ik}$ and the variance is

$$Var(Y_{ik}) = \lambda_1^2 Var(X) + \lambda_2^2 V ar(\epsilon_i).$$

Since $X$ and $\epsilon_i$ are identically distributed, these expressions can be written

$$E(Y_{ik}) = (\lambda_1 + \lambda_2) \epsilon_i + C_{ik},
\text{Var}(Y_{ik}) = (\lambda_1^2 + \lambda_2^2) V ar(\epsilon_i).$$

When changing $\rho$ we retain both mean and variance as constants. Mean value for $Y_{ik}$ is randomly set for each supplier, for each task. Variance is always set as $Var(Y_{ik}) = 0.25 V ar(\epsilon_i)$. Then, correlation between distributions is $\rho = \frac{\lambda_2}{0.25}$. To achieve the desired correlation, we manipulate $\lambda_1$. $\lambda_2$ is then calculated based on that as $\lambda_2 = \sqrt{0.25 - \lambda_1^2}$. Finally, $C_{ik}$ is adjusted to retain the mean value a constant. Using a similar set-up, one can also achieve a correlation of $-\rho$ between suppliers A and B. For this, first set $\lambda_1a, \lambda_2a$ to achieve the desired correlation. Then set $\lambda_1a = -\lambda_1b$ and $\lambda_2a = \lambda_2b$.

But for a third supplier, C, that is uncorrelated to suppliers A and B but has the same variance of A and B, we set $\lambda_1c = 0$ and let $\lambda_2c = \sqrt{\lambda_1^2 + \lambda_2^2}$. 

30
References


