View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by DSpace@MIT

Centralizing Data Management with Considerations of Uncertainty and Information-based Flexibility

Chander Velu
Stuart Madnick
Marshall Van Alstyne

Working Paper CISL# 2013-02

January 2013

Composite Information Systems Laboratory (CISL)
Sloan School of Management, Room E62-422
Massachusetts Institute of Technology
Cambridge, MA 02142

Centralizing Data Management with Considerations of Uncertainty and Information-based Flexibility

Chander K. Velu
Cambridge Judge Business School
University of Cambridge
Trumpington Street
Cambridge CB2 1AG
United Kingdom
e-mail: c.velu@jbs.cam.ac.uk
Tel: +44 (0) 1223 765 879
Fax: +44 (0)1223 339 701

Stuart E. Madnick
Sloan School of Management
& School of Engineering
Massachusetts Institute of Technology
50 Memorial Drive
Cambridge, MA 02142
United States
e-mail: smadnick@mit.edu
Tel: 1-617-253-6671
Fax: 1-617-253-3321

Marshall W. Van Alstyne
Boston University School of Management
Boston University
595 Commonwealth Avenue
Boston, MA 02215
United States
e-mail: mva@bu.edu
Tel: 1-617-358-3571
Fax:1-617-353-5003

April 2013

Centralizing Data Management with Considerations of Uncertainty and Information-based Flexibility

Abstract

This paper applies the theory of real options to analyze how the value of information-based flexibility should affect the decision to centralize or decentralize data management under low and high uncertainty. This study makes two main contributions. First, we show that in the presence of low uncertainty, centralization of data-management decisions creates more total surplus for the firm as the similarity of business units increases. In contrast, in the presence of high uncertainty, centralization creates more total surplus as the dissimilarity of business units increases. The pivoting distinction trades the benefit of reduction of uncertainty from dissimilar businesses for centralization (with cost saving) against the benefit of flexibility from decentralization. Second, the framework helps senior management evaluate the trade-offs in data centralization that drive different business models of the firm. We illustrate the application of these propositions formally using an analytical model, and informally using case vignettes and simulation.

Keywords: economics of IS, real options, uncertainty, flexibility and information systems decentralization.

Introduction

Data centralization efforts within firms often follow a familiar pattern [1][8]: conversations with senior executives reveal numerous incidents whereby they start a major data-centralization project only to terminate it before completion. The frequency with which this happens can be partly explained by the failure of such efforts to fully take into account how data management enables a firm to react flexibly to uncertainty in the environment. This paper addresses this gap by applying the theory of real options to account for the value of flexibility in the context of uncertainty in order to decide when to centralize data management.

The degree of centralization within an organization can affect performance and innovation [18]. Centralization means that decision rights affecting interdependent groups are concentrated in a single authority, whereas decentralization distributes those rights, usually as authority granted to the groups nearest the data collection [25, 35].

Consider a canonical example based on interviews conducted at a financial services firm, Thomson Reuters.[1] One of Thomson Reuters' key businesses is that of infrastructure provider and information artery to the financial services industry. During the 1980s and 90s, the business was run geographically, with the business head of each country having the authority to collect data and serve the client needs of their own country. The data was held country by country for the plethora of products. In fact, there were over 2,500 products by the late 1990s, which resulted in a high cost of data maintenance. Thomson Reuters has multiple business units, each with hundreds of information systems and databases to support these operations. The decentralized systems and databases cater for the specialized requirements of that business unit or to adapt rapidly to meet local changing requirements. However, at the turn of the millennium, rapid changes in customer needs, financial innovation, and Internet access, created significant uncertainty and a new set of challenges for Thomson Reuters in gathering, disseminating and responding to new information. On the one hand, there was a need to centralize data management to improve coordination and reduce costs. On the other hand, decentralization enables better local tailoring. Therefore, management faces the constant challenge of whether the center or the business units should have the authority to collect data and use it to make decisions. In particular, how should management think about whether to centralize or decentralize data management in the context of environmental uncertainty? This choice affects the business model of the firm and its innovative capability. We will discuss resolution of the Thomson Reuters' challenge and its relationship to our model in subsequent sections.

We define data management as the authority to collect data, as well as the decision rights to use the data. In this paper, centralizing data management is the allocation of authority to collect data, as well as the rights to make decisions using the data to the central unit. Decentralization occurs when such decision rights reside with the business units. The allocation of such decision rights to the central unit or the business units influences data collection. Therefore, we assume that the rights to decide what to collect and to use the data

[1] The Thomson Corporation and Reuters Group plc combined in 2008 to form Thomson Reuters. Here, we use Thomson Reuters to refer to the firm both before and after the combination.

reside with the same unit. Hence, we focus on the decision rights to collect and use data and their effect on the total surplus of the firm.

Centralized versus decentralized decision-making has been examined extensively in the information systems and management literature [see 15 and 18]. At the heart of the discussion is the trade-off between the incentives at the center versus those of the business unit and the benefits from coordination versus costs. On the one hand, the center might obtain information from the business units to enhance coordination. On the other hand, the business units might have better information due to proximity to the local conditions and, hence, are better able to serve customers. Moreover, the center and business units might have different costs of information management. The decision to centralize or decentralize depends on the optimum trade-off between these benefits and costs [13]. Studies have shown that the information systems' design problem is intertwined with the organizational design issue of centralization versus decentralization as it relocates information among decision-makers [see 15, 17, 26]. Studies have examined centralization and decentralization with co-location of information and decision rights [4, 38], ownership of information assets [15], and the role of standardization on database design [42]. However, as illustrated by the challenges faced by Thomson Reuters, the degree of uncertainty in the environment also affects the decision to centralize or decentralize data management. The extant literature on information management has not examined how the degree of uncertainty affects the decision to centralize or decentralize data management, which is the question we address in this paper.

We classify the degree of uncertainty as low or high. We identify uncertainty here with variability of the environment. As in the case of information entropy [28][39], a more uncertain environment is less predictable. High (low) uncertainty represents the case of high (low) variance in possible outcomes.

Data management contributes to the total surplus of the firm by altering a firm's ability to react to an opportunity or limit a loss when change occurs. This is the value of information-based flexibility. Managers in industries experiencing significant environmental change (due to changes in competitive environments, consumer preferences or other shocks) often do not know what data they will need in order to react to change [31]. Collecting data in one period can help the firm leverage data collected in later periods. Local data can also benefit the business unit currently holding the data or other interdependent units that might need that data in the future. In this sense, the need to build in future requirements creates flexibility or option value for the firm [38, 11].

We build a simple model to develop our framework that incorporates information-based flexibility to help decide when it is optimal to centralize data management. The firm has to decide whether the center or the business units have the authority to collect data and the decision rights to use the data. For the purpose of our analysis, we assume that there are two business units and a central unit. The level of similarity between the business units determines the degree of cross-benefits between them whereby a higher degree of similarity implies higher joint pay-offs. The firm must collect some business-unit data in the first period in order to have the option to collect new data in the second period. Environmental uncertainty is resolved in the second period, after which the firm can collect further data in order to improve its response.

We model two decision-making structures. The centralized structure enables the center to coordinate between the two business units, taking into account the cross-benefit effects between business units as it decides the level of data collection. In contrast, in the decentralized structure, each individual business unit makes decisions to maximize its own profit without taking into account the cross-benefits to other units. The center, with its global view, accounts for the cross-benefit effects between business units and, with its scale, has lower costs. Yet, business units can react more flexibly to uncertainty than the center because local control permits closer tailoring to the local environment. Such closer tailoring of the data to suit the local environment entails higher associated costs. We then show how information-based flexibility arises as a result of uncertainty and irreversibility. Irreversibility in this context is the inability to undo a decision once it has been made. Therefore, the firm must consider uncertainty, irreversibility, and the cross-benefit effects between the business units when deciding whether to centralize data management. In addition, our model of data management seeks to analyze the benefits of coordination and diversity of information sources, as different structures produce different reductions in uncertainty.

A key result is that, in the presence of low uncertainty, centralization becomes optimal as the similarity of business units increases – benefits from coordination dominate. On the other hand, in the presence of high uncertainty, centralization becomes optimal as the dissimilarity of business units increases – benefits from reduced uncertainty dominate. Dissimilar business units enable new information to be obtained by combining diverse information from separate perspectives. Such new information reduces uncertainty by reducing the variance in possible states of the environment. Centralization becomes attractive when reducing uncertainty is more valuable than the flexibility of allowing opportunistic local business decisions. Ironically, business units that are more similar provide more correlated information, which limits the uncertainty-reducing benefit from diverse sources of information. On a cost basis alone, standard theory might predict that centralizing similar business units is optimal but this ignores the uncertainty-reducing benefit of centralizing information from dissimilar business perspectives. Decentralization, in contrast, becomes attractive when reducing uncertainty is less valuable than the flexibility of allowing local business decisions. If the environment is predictable and there are no cross-benefits of decisions, then decentralizing autonomous decisions can add more to the total surplus of the firm. These cases, defined by high/low uncertainty reduction and high/low flexibility, interact with business-unit similarity to push optimal decision structures more or less toward centralization. We illustrate the application of these propositions formally using an analytical model and informally using case vignettes and simulation.

We contribute to the information management literature by showing how the total surplus of the firm from information-based flexibility influences the decision to centralize data management. In particular, we show how such a decision depends on the degree of uncertainty (external to the firm) in the environment and the degree of similarity (internal to the firm) between business units. In addition, the framework helps senior management evaluate the criteria and trade-offs favoring data centralization across multiple decision contexts.

This paper is organized as follows. First, we review the relevant literature on information management and option value. Then, we develop the model and discuss the main

drivers of the total surplus of the firm from information-based flexibility. We then explore the effects of low and high uncertainty; and we draw implications for economies of scope in information management across business units with different degrees of similarities in the presence of low uncertainty. Subsequently, we extend the case of similarity of business units in the presence of high uncertainty. Finally, we discuss the managerial implications and conclude.

Relevant Literature

Our paper is related to, and borrows from, a variety of literature.

Information Management

Ownership of data can influence the effectiveness and success of information management [37]. Information is an intangible asset that benefits its owner [see 7, 23]. In the context of information management, the right to control is the ability to access, create and standardize data, as well as to determine access privileges for others[2] [42]. A firm's allocation of ownership rights on data between the center and the business units influences organizational design such as centralization and decentralization, which affects the total value of the firm.

Studies have examined centralization/decentralization in the context of information management. Anand and Mendelson [4] study the effect of alternative coordination mechanisms, such as centralized, decentralized and distributed structures, on the performance of firms that face uncertain demand. The optimal structure depends upon the relative importance of both central and local knowledge. As local knowledge is specific and not amenable to statistical aggregation, the results show that decentralization outperforms centralization in spite of the latter's superior ability to coordinate. Therefore, the study shows the benefit of co-locating decision rights with specific knowledge. In contrast, Nault [35] examines the impact of information technology on the profitability of different organizational design, such as hierarchy, decentralized and mixed models, when there is an information asymmetry between the central authority and the decentralized nodes. Moreover, there are global and local investments that are complementary and need to be coordinated. The authors show that co-location of investment decision rights and information might not be optimal when the cost of non-coordination is high.

Brynjolfsson [15] discusses when centralization and decentralization is optimal under an incomplete contracts framework based on ownership of information assets. The basic premise is that complementary assets should be owned by the same agent when complete contracts are infeasible. Therefore, when there is a need for centralized coordination, ownership of the information asset should be with the centralized authority in order to improve incentives for all. Van Alstyne et al. [42] extend the incomplete contracts reasoning to database management to show when centralization and decentralization are optimal. However, the extant literature on information management has not examined the degree of uncertainty and how this influences the decision to centralize. In order to do so, we draw upon options theory.

[2] It is useful to note that this right and its associated privileges accrue to the owners regardless of where such information is located. For example, data could be physically located in one place but there could be many owners, each with a residual right of control over specific parts of that data. Alternatively, data could be owned and controlled by one unit but physically located in a distributed manner at various business units.

Options Theory

An option is the right, but not the obligation, to take an action in the future [19]. The field of real options applies options reasoning to investment and commitment decisions, identifying the factors that influence the time at which investors choose whether or not to invest. The combination of an unfavorable outcome and the commitment to an irreversible course of action can be worse than being flexible by keeping an option open and acting only after further information is revealed. Therefore, the value of the option derives from two principal sources: uncertainty and irreversibility [12]. We argue that firms face many strategic decisions that display both uncertainty and irreversibility. Uncertainty stems from future conditions in the external environment; irreversibility is the inability to undo or reverse a decision once it has been made. In uncertain environments, firms have a short window of opportunity to react to unforeseen events. In these circumstances, versatility and the ability to adapt to the changing environment are critical sources of value. In a strategic context, flexibility is a function of being informed about opportunities as uncertainty unfolds. A firm becomes informed by collecting relevant data, and must decide whether to purchase the option value by investing in data and creating information-based flexibility value.

A number of studies have discussed the use of options reasoning in information systems decision-making [21]. Dos Santos [20] uses real options models for valuing new IS projects. The study shows how the pre-investment value could increase due to learning and, hence, future investments are treated as optional. Kumar [27] examines how options values of investments in new information technologies vary according to project risks. Taudes, Feurstein and Mild [40] show how strategic growth options are valuable in the case of software platforms when managers can intervene across the project's trajectory to create opportunities for follow-on investments. Benaroch and Kauffman [9] extend the use of options models to the development of point-of-sale (POS) debit services, where the issue is not *whether* to invest but *when* to invest. In a companion paper, Benaroch and Kauffman [10] show how options models can be used for strategic IS investment decisions by incorporating a project's idiosyncratic risks into the calculations. In addition, Benaroch et al. [11] use a case study of a data-mart consolidation in a global airline firm to discuss how to blend the technical aspects of valuing options with a strategy-focused approach. In this study, we build on previous research by applying the principles of option theory to study the impact of uncertainty on the value of information-based flexibility and the firm's decision to centralize data management.

Model Formulation

In this section we set up the basic model and discuss the key assumptions.

The Model

Local business-unit information, I, is defined as a state of information on which probability assignments will be made [33]. The value of information is its option value, implying $I > 0$, and that more information provides more possibilities for consideration. For example, a supermarket could collect information, I, about the types of product a customer purchases each week. The supermarket could decide to increase the amount of information collected about the customer by adding the precise time of day that the customer buys the product. The collection of such additional information increases costs but could help the supermarket

increase sales through more precise forecasting of the customer-buying pattern. We suppose that the firm consists of a central unit, *CU*, and several business units, *BU*. For ease of exposition, we initially examine the case of one central unit and one business unit, as this provides the main intuition. We add a second business unit when we examine the impact of economies of scope in information management across business units.

Consider the case where there are two periods, *t=0,* and *t=1*. In period *t=0*, the firm makes a decision about how much information to collect. The return from this information collection is realized in period *t=1*. There are two possible states of nature in terms of external environment, $\theta_N (N = 1,2)$. The environmental outcome could be due to changes in tastes, as from a change in fashion, or to change in competition, as from entry or exit. The state of the external environment is revealed at *t=1*. At this point, the firm can react by collecting further business-unit information in order either to leverage an opportunity better, or to limit its losses.[3] Local business-unit information can be collected via the central unit at cost C_0^{CU} at time *t=0*. If this is done, upon learning the true state of the environment in period *t=1,* the center can collect new information at cost C_1^{CU} to leverage the particular news. Alternatively, the business unit can collect the business-unit information in periods *t=0* and *t=1* at cost C_0^{BU} and C_1^{BU} respectively.

If state θ_1 prevails (the probability of this is *p*), then the revenue (excluding the cost of information collection) is R_+. If θ_2 prevails (the probability of this is by definition (1-*p*)), then the loss is R_-. Let θ_1 be the favorable outcome where revenue is positive, $R_+ > 0$ and θ_2 is the unfavorable outcome where losses are incurred and, hence, $R_- < 0$. However, the firm will only continue to be in business if the expected total surplus of doing so is positive, $E(R) = pR_+ + (1-p)R_- > 0$. When the firm collects business-unit information in period *t=0*, this decision allows it to collect further business-unit information in period *t=1* when the uncertainty is resolved to enable it either to leverage an opportunity better or to limit its losses.

Let $\alpha \in (0,1)$ be the extent to which the revenue or loss, *R*, can be altered strategically using the information, I. The favorable outcome, R_+, can be enhanced by a factor, $(1+\alpha)$. The unfavorable outcome, R_-, can be limited by a factor, $(1-\alpha)$. The revenue in favorable times, R_+, and losses in unfavorable times, R_-, can be viewed as the base revenue that can be enhanced or limited by factors $(1+\alpha)$ and $(1-\alpha)$ respectively. In this sense, $(1+\alpha)$ and $(1-\alpha)$ represent the additional benefit from information gained by the firm over and above the base revenue of the firm. For example, a firm should rationally plan for business of size *E(R)*, but if R_+ is realized the firm could hire additional staff, while if R_- is realized it could cancel outstanding orders. Therefore, the factors $(1+\alpha)$ and $(1-\alpha)$ can be thought of as the additional benefit accruing to the firm as a result of the additional information, helping it to amplify revenues or mitigate losses.

---

[3] Since there is uncertainty in the environment, it is not possible to fully contract on the type of data to be collected between the central unit and the business units ([4] and [35]).

The Assumptions

We model the optimal decision rights of centralization/decentralization under low and high uncertainty respectively. We first consider the case of low uncertainty before going on to examine the case of high uncertainty. Our assumptions are listed in Table 1.

Table 1. Assumptions

No.	Assumption
1	$TC^{CU} = (C_0^{CU} + C_1^{CU}) < TC^{BU} = (C_0^{BU} + C_1^{BU})$
2	$\alpha^{BU} > \alpha^{CU} \Rightarrow (1+\alpha^{BU}) > (1+\alpha^{CU}) \Rightarrow (1-\alpha^{BU}) < (1-\alpha^{CU})$
3	Total cost, $TC(I) = C_0(I) + C_1(I)$ where $TC'(I) > 0$ and $TC''(I) > 0$
4	Total revenue, $TR(I) = R_0(I) + R_1(I)$ where $TR'(I) > 0$ and $TR''(I) < 0$
5	Variance of W is: $Var(W) = a^2 Var(X) + b^2 Var(Y) + 2ab\,Cov(X,Y)$ and $Var(W^*) = Var(W) - V(X,Y,I^N)$, where $Cov(X,Y) = E(XY) - E(X)E(Y)$ and a and b are constants.

The interpretation of these assumptions is as follows:

(i) Assumption (1) states that the total cost of collecting the business-unit information at the central unit, $TC^{CU} = (C_0^{CU} + C_1^{CU})$, is cheaper than collecting it at the business unit, $TC^{BU} = (C_0^{BU} + C_1^{BU})$, over both periods. This can be interpreted as a scale economy for the central unit without the benefit of local tailoring.[4] For example, the benefit from scale economies for the central unit is due to savings from removing duplicated efforts of collecting the same information when each business unit collects its own information.[5] Hence, the savings from centralizing the information collection is the difference in costs, $G = (C_0^{BU} + C_1^{BU}) - (C_0^{CU} + C_1^{CU})$.

(ii) Assumption (2) implies that the business unit can react more opportunistically than the central unit. Since local information is better tailored, we assume that the extent to which the revenue can be altered strategically using the information, I, will be greater if the business unit collects the information, than if the information is centralized at the central unit. This is because of environmental proximity: the business unit, being closer to the customer and competitors, is better able to judge the type of information required in period *t=0*, and can react better to environmental changes in period *t=1*.

(iii) Assumption (3) states that the total cost of information collection and management is the sum of the costs for the two periods.[6] The marginal total cost is

4 Environmental proximity might provide the business unit with some cost advantage over the central unit. However, we consider the case where the scale benefits for the central unit are large enough that the total cost for the central unit is lower than that for the business units.

5 In addition to the factors outlined in (i), reasons in support of this assumption include the presence of more technology with lower utilization rates, more duplication, more people and more package purchases in a decentralized model [1].

6 We consider the data collection and maintenance costs. However, in an ongoing organization, firms usually already have in place existing information systems and databases. Hence, the firm needs to consider the additional set-up costs as a result of changes in the external environment. For example, this additional cost could be due to the cost of hardware for the database, database design and systems installation. Such one-off costs need to be added to the data-collection and

increasing with respect to information, twice differentiable and convex in information, I [see 34]. This implies that the cost increases as more information is collected, and does so more than proportionately to the increase in information.

(iv) Assumption (4) states that the total revenue is the sum of the revenues for the two periods. The marginal total revenue is increasing with respect to information, twice differentiable and concave in information, I.[7] This implies that the revenue increases as more information is collected, and does so less than proportionately to the increase in information.

(v) Assumption (5) states that the dispersion of the possible states of the environment is encapsulated by the random variable, W. The random variable, W, is defined as $W = aX + bY$, where a and b are constants (X and Y are information signals drawn from business unit A and business unit B respectively). The variance of W is denoted by $Var(W)$ and $Var(W^*)$ before and after incorporating the variance-reducing effect of the new information respectively. We explain this assumption in more detail later in the section on centralizing under high and low uncertainty.

We summarize the key notation used in the paper in Table 2.

Table 2. Modelling Notation: Definition and Comment

Notation	Definition	Comment
R_+, R_-	R_+ is revenue and R_- is loss.	Captures the base revenue and loss for the business units excluding the cost of information-collection.
TC	Total cost of collecting information.	Total cost consists of the sum of the costs for each of the two periods $(C_0 + C_1)$.
$(1+\alpha)$	Factor by which base revenue can be enhanced.	Additional benefit from information over and above the base revenue.
$(1-\alpha)$	Factor by which base loss can be reduced.	Additional benefit from information to limit the base loss.
I	Level of business-unit information.	Captures the level of investment in business-unit information.
θ_1, θ_2	Possible states of the world.	Captures the uncertainty in the environment.
p and $(1-p)$	Probability that the favorable or unfavorable states of the world will occur.	Based on the principle of information entropy, low uncertainty is when p is close to either 0 or 1, and high uncertainty is when p is close to 0.5 respectively.
$E(V)$	Expected total surplus.	Total expected total surplus is the surplus of the business units based on the information they have collected under the different levels of uncertainty.
ϕ	Correlation coefficient between business-unit returns.	Captures the level of similarities between the business units.

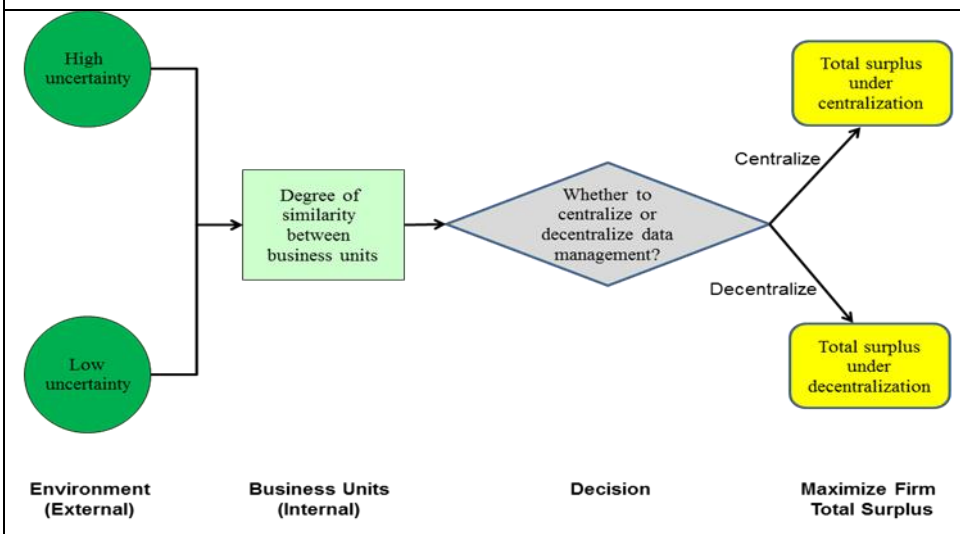
Figure 1 provides a conceptual framework for the paper. We examine how low and high uncertainty (external to the firm) respectively, and the degree of similarity between business

maintenance costs above. In such a case, our main analysis and propositions will still hold but some of the conditions will be stricter for centralization to be optimal.

[7] This assumption is similar for a concave revenue function with reference to investment in input such as talent [see 5].

units (internal to the firm), influence the decision to centralize or decentralize data management. Our approach compares the relative total surplus of the firm under centralization and decentralization in order to decide on the optimal structure.

Figure 1. Conceptual Framework



Centralization and Flexibility

In this section we discuss the tension between efficiency and flexibility related to data centralization. The key insight from this section is that uncertainty and irreversibility give rise to option-based flexibility value and, hence, ignoring them could result in sub-optimal decisions. We analyze the decision about whether to centralize in order to see how the main sources of optionality, namely uncertainty and irreversibility, affect this decision.

We define flexibility in terms of decision-making flexibility and its relation to how information is collected and managed in the organization [32]. In this sense, the flexibility of decision-making is greater the larger the choice-set under consideration or the more alternatives are available for a decision that, in turn, gives rise to more total surplus. To explore the tension more formally, let us examine the decision to centralize in the case where there is uncertainty.

The Decision to Centralize

The expected total surplus of centralization and decentralization can be denoted by $E(V^{CU})$ and $E(V^{BU})$ respectively.[8] We consider the case of both centralization and decentralization to show how the option value arises. In particular, we compare the total surplus of the firm under both centralization and decentralization in order to decide which structure creates more total surplus. Let us consider the case of centralization:

---

[8] Following Ponssard [36, p. 740] the expected value of information should be denoted as $E(V^{CU} / I)$, where $E(V^{CU} / I)$ is the incremental gain obtained by making a decision based on information, *I*. However, for simplicity we write $E(V^{CU} / I)$ as $E(V^{CU})$.

$$E(V^{CU}) = [pR_{+}(1+\alpha^{CU}) + (1-p)R_{-}(1-\alpha^{CU})] - TC^{CU} \quad (1)$$

$$E(V^{BU}) = [pR_{+}(1+\alpha^{BU}) + (1-p)R_{-}(1-\alpha^{BU})] - TC^{BU} \quad (2)$$

Taking the difference in expected total surpluses between the center and the business unit and substituting for *G* gives:

$$E(V^{CU}) - E(V^{BU}) = G + pR_{+}(\alpha^{CU} - \alpha^{BU}) + (1-p)R_{-}(\alpha^{BU} - \alpha^{CU}) \quad (3),$$

where $G = TC^{BU} - TC^{CU}$.

The decision to centralize implies that the expected total surplus from centralization must be larger than the expected total surplus from decentralization, (i.e. (3) must be positive):[9]

$$E(V^{CU}) - E(V^{BU}) = G + pR_{+}(\alpha^{CU} - \alpha^{BU}) + (1-p)R_{-}(\alpha^{BU} - \alpha^{CU}) > 0$$

Hence, in order for centralization to increase total surplus of the firm:

$$G > (\alpha^{BU} - \alpha^{CU})[pR_{+} - (1-p)R_{-}] \quad (4)$$

Inequality (4) implies that the savings from centralization, *G*, must be larger than the difference in the total surplus from information-based flexibility between the center and the business unit, in order for centralization to be beneficial. Hence, $(\alpha^{BU} - \alpha^{CU})[pR_{+} - (1-p)R_{-}]$ (which is a positive value) translates into flexibility value, or the option value from decentralization. The value of such flexibility must be smaller than *G* for centralization to be worthwhile, given that it involves foreclosing the option. We show below that there could be an excess tendency to centralize if the firm were to ignore irreversibility and uncertainty.

The Effect of Irreversibility and Uncertainty

To see why this is the case, first consider the roles that irreversibility and uncertainty play in the decision to centralize. Irreversibility arises because the investment costs are usually sunk, in the sense that the costs of collecting and storing data cannot be retrieved when the environment changes [38].

Irreversibility implies that the firm would not be able to collect information in period *t=1* if it did not collect the relevant information in period *t=0.*[10] Therefore, ignoring irreversibility implies that the firm myopically thinks that it will be able to leverage opportunities or limit losses just as effectively as when the relevant information at issue was collected by the central unit or the business unit (i.e. $\alpha^{BU} = \alpha^{CU}$, hence $(1+\alpha^{CU}) = (1+\alpha^{BU})$ and $(1-\alpha^{CU}) = (1-\alpha^{BU})$). Clearly, if it is cheaper to collect the information at the level of central unit than that of business unit, and the ability to react strategically is the same, then it makes sense for the firm to centralize data management at the central unit.

---

[9] We could consider the weaker condition that (3) must be non-negative as there could be other non-specified options that arise around centralization efforts. However, for simplicity we consider the stricter case that (3) must be positive.

[10] For example, a grocer might be unable to gather consumer purchasing data having not previously invested in product barcode data.

Now consider the effect of ignoring uncertainty: the profit would be known for certain regardless of the type of information or where it is collected. Since there is no uncertainty, there is no opportunity to enhance the revenue or mitigate losses. As the returns are certain, the firm would be better off centralizing, as this is the cheaper structure for the firm. The effects of ignoring irreversibility and uncertainty are shown formally in Appendix 1. In summary, we show that the firm would have an excess tendency to centralize if it were to ignore irreversibility and uncertainty respectively.

Empirical Illustrations

Marchand et al. [30] distinguish between moderate competition and hypercompetitive environments in formulating a framework for data centralization. In moderately competitive environments, the demand pattern is stable and, hence, less emphasis is placed on uncertainty. In contrast, in hypercompetitive environments, the demand pattern changes continuously and, hence, the firm needs to place more emphasis on uncertainty in its data-centralization decision. In moderately competitive markets, the focus on supply-chain management (e.g. production planning or inventory management systems) can bring substantial operational benefits. On the other hand, in a hypercompetitive environment, where the demand pattern is continually changing, the focus must be on managing customer interaction, account management, and order processing. In the latter environments, the emphasis shifts from supply-chain management, with detailed management reporting and controls, to demand-chain management by simplifying detailed management reporting, and empowering managers to act in response to the changing environment. As the competitive environment changes, the need for flexibility to serve customer requirements becomes more important than the need for centralized data management. In highly competitive environments, the data management is decentralized to the business units facing the environment in order to leverage the benefit from the total surplus from the information-based flexibility. On the other hand, in less competitive environments the focus is on centralizing data management in order to reduce costs.

The case of a leading elevator company illustrates how firms manage the tension between centralizing and decentralizing data management while maintaining business flexibility [29]. The firm's business focused on providing local services to customers in 22 countries. This involved 22 local operations, which were managed centrally. Competitors were entering their markets with a more decentralized operating model, which was affecting the firm's performance. The newly appointed CEO decided to divide the 22 country operations into 3 regions and to operate with a regional, rather than a country, focus. The business-information systems were also organized regionally, enabling the regional business units to react in a more agile manner to opportunities that arose, compared to the previous, more centralized, approach. In effect, decentralizing data management, by moving away from being driven by the center to the regions, enabled a fuller realization of the benefit from the value of information-based flexibility at the regional level. This allowed the elevator company to "leverage its reputation, deep business knowledge and human capabilities to target real sources of competitive differentiation locally in elevator markets and at the same time leverage regional and global expertise where it creates technological, product and operational advantages" [29]. The company achieved a better balance between business-unit flexibility and centralization of data management, which improved its operating performance.

Centralization under High and Low Uncertainty

This section extends the above framework to analyze the decision to centralize data when there are cross-benefit effects across the different business units. In the next section we discuss a case where there are cross-benefit effects from similar business units when there is low uncertainty. We then go on to discuss the case when there is high uncertainty.

Similarity of Business Units in the Presence of Low Uncertainty

The key insight of this section is that under low uncertainty centralization becomes more attractive as the similarity of the business units increases. This is because coordination benefits are leveraged more fully while the benefit from diversity of information sources in reducing uncertainty is minimal.

In this section we are concerned with low uncertainty. Based on the principle of information entropy [28][39], low uncertainty is when p is close to either 0 or 1, where the firm becomes certain about the true state of the environment (equivalently, p and $(1-p)$ are further away from 0.5). It follows that high uncertainty is when the firm is less certain about the true state of the environment and, hence, p and $(1-p)$ are closer to 0.5. Consider two business units, A and B, denoted by BU_A and BU_B. Part of business unit B's profits is correlated with business unit A's profits, and vice versa. The profit directly attributed to BU_A is R_+^A and R_-^A, respectively, for the two states of the environment (θ_1 and θ_2 with probabilities p and $(1-p)$ respectively). Similarly, the profits directly attributed to BU_B are R_+^B and R_-^B, respectively, for the two states of the environment. However, there are cross-benefit effects across the business units. Let X and Y be two random variables that are information signals drawn from business unit A and business unit B respectively about their client needs and the state of the environment. Let X and Y be drawn from distributions $f(X)$ and $f(Y)$ respectively. The extent of the cross-benefit is captured by the correlation coefficient, $\phi \in (0,1)$, where $\phi = \frac{Cov(X,Y)}{Var(X)Var(Y)}$. The cross-benefit effect captures the level of similarity between the businesses.[11]

Business units that are similar, for example, due to having similar customers or product lines, have a higher correlation coefficient due to higher cross-benefit effects. The sharing of a common customer often results in one business unit collecting information that is beneficial to another. For example, the payment history of a bank's credit card customer might be helpful in assessing the same customer's credit rating for a mortgage application, and vice versa. Business units that are similar have cross-benefit effect, ϕ closer to 1. When business units are similar, with ϕ closer to 1, the distributions $f(X)$ and $f(Y)$ are more homogeneous. This implies that the firm gets limited new information from the information signals X and Y due to the redundancy of the dependent sources [16]. In contrast, business units that are dissimilar have cross-benefit effects ϕ closer to zero, and, hence, the distributions $f(X)$ and $f(Y)$ are more heterogeneous. This implies that the firm gets most

[11] We define the level of similarity with reference to the commonality of information requirements across the business units due to, for example, common customers or serving a common market segment. In this sense, business units that are more similar display a higher overlap in their information requirements in leveraging an opportunity or limiting a loss.

new information from the information signals X and Y due to the high independence of the sources [17]. We discuss further the case when business units are dissimilar in the section on high uncertainty below.

As a result of the cross-benefit effect, the profit of BU_B is given by $R_+^B + \phi R_+^A$ and $R_-^B + \phi R_-^A$, where ϕ is a correlation coefficient.[12] Similarly, the profit of BU_A has cross-benefit effects from the profits of BU_B, such that $R_+^A + \phi R_+^B$ and $R_-^A + \phi R_-^B$. Hence, the business-unit information, I for BU_A benefits BU_B, and vice versa. Table 3 summarizes the returns.

Table 3. Similarity of Business Units and Returns

States of Nature $\theta_N (N = 1,2)$	**Business Unit A,** BU_A	**Business Unit B,** BU_B
θ_1 with probability p	$R_+^A + \phi R_+^B$	$R_+^B + \phi R_+^A$
θ_2 with probability $(1-p)$	$R_-^A + \phi R_-^B$	$R_-^B + \phi R_-^A$

In addition, the similarity of the business units also affects the cost advantage of centralization relative to decentralization of data management. This is because centralization removes duplicated efforts of collecting the same information, as noted in assumption (1). The higher the business unit similarity, the higher will be the cost advantage of centralization. We capture this through the relationship between total costs of the central unit and the business units respectively, $TC^{CU} = (1-\lambda\phi)[TC^A + TC^B]$, where λ is scale parameter, $0 < \lambda < 1$ that captures the extent of cost advantage from centralization due to the degree of similarity of the business units.

Impact of Low Uncertainty on the Decision to Centralize

For simplicity, let us consider the case where $\alpha^A = \alpha^B = \alpha^{BU}$. This implies that the ability of both business units to leverage an opportunity and limit a loss with respect to the level of information is more or less the same. In order to decide whether to centralize or decentralize, we compare the expected total surplus of the firm with centralized and decentralized data management respectively. First, we examine the role of cross-benefit effects on coordination across business units. Second, we examine the role of diversity of information sources on new information generation and its influence on the degree of uncertainty. Without discounting, the expected total surplus of centralized and decentralized business units' data collection can be denoted by $E(V^{CU})$ and $[E(V^A) + E(V^B)]$ respectively, where:

$$E(V^{CU}) = (1+\phi)[p(1+\alpha^{CU})(R_+^A + R_+^B) + (1-\alpha^{CU})(1-p)(R_-^A + R_-^B)] - TC^{CU} \quad (5)$$

and:

[12] There could be some cases of negative correlation when business units act as a hedge, whereby increase in revenue for one business implies a decrease in revenues for another business. We do not consider these in our analysis.

$$[E(V^A)+E(V^B)]=[p(1+\alpha^{CU})(R_{+}+R_{+}^{B})+(1-p)(1-\alpha^{CU})(R_{-}^{A}+R_{-}^{B})]+\phi[K^A+K^B] \\ -[TC^A+TC^B] \quad (6)$$

where:

$$E(V^A)=p(1+\alpha^{BU})R_{+}^{A}+(1-p)(1-\alpha^{BU})R_{-}^{A}+\phi[p(1+\alpha^{BU})R_{+}^{B}+(1-p)(1-\alpha^{BU})R_{-}^{B}] \\ -TC^A \quad (6a)$$

$$E(V^B)=p(1+\alpha^{BU})R_{+}^{B}+(1-p)(1-\alpha^{BU})R_{-}^{B}+\phi[p(1+\alpha^{BU})R_{+}^{A}+(1-p)(1-(1-\alpha^{BU})R_{-}^{A}] \\ -TC^B \quad (6b)$$

and:

$\mathrm{K}^B=[p(1+\alpha^{BU})R_{+}^{B}+(1-p)(1-\alpha^{BU})R_{-}^{B}]$ and:

$\mathrm{K}^A=[p(1+\alpha^{BU})R_{+}^{A}+(1-p)(1-\alpha^{BU})R_{-}^{A}]$.

The individual business units A and B maximize the expected total surplus $E(V^A)$ and $E(V^B)$ by taking surpluses K^B and K^A as given respectively, as they do not have influence over the decision about how much to invest in the information of the other business unit. This implies that because individual business units are optimizing their own profit, they do not factor in the cross-benefit effects of their own decisions on other business units.[13] On the other hand, the central unit is able to factor in the cross-benefit effects and, hence, maximizes the total (joint) surplus subject to such cross-benefit effects. Let the optimized investment level for the central units and business units A and B be $(I^{CU_A^*},I^{CU_B^*})$ and $(I^{BU_A^*},I^{BU_B^*})$ respectively.

In order to decide whether to centralize or decentralize, we compare the total surplus of the firm under both conditions. Taking the difference in expected total surpluses between CU and BU gives (double star denotes optimized value for CU and single star for BU respectively):

$$E(V^{CU})-[E(V^A)+E(V^B)]=(1+\phi)[p(1+\alpha^{CU})(R_{+}^{A**}+R_{+}^{B**})+(1-\alpha^{CU})(1-p)(R_{-}^{A**}+R_{-}^{B**})] \\ -[[p(1+\alpha^{BU})(R_{+}^{A*}+R_{+}^{B*})+(1-p)(1-\alpha^{BU})(R_{-}^{A*}+R_{-}^{B*})]+\phi[K^{A*}+K^{B*}]]+[(TC^{A*}+TC^{B*})-TC^{CU**}] \quad (7)$$

The decision to centralize implies that the expected total surplus from centralizing must be larger than the expected total surplus from decentralizing (i.e. (7) must be positive):

$$(1+\phi)[p(1+\alpha^{CU})(R_{+}^{A**}+R_{+}^{B**})+(1-\alpha^{\mathbf{B}U})(1-p)(R_{-}^{A**}+R_{-}^{B**})] \\ -[[p(1+\alpha^{BU})(R_{+}^{A*}+R_{+}^{B*})+(1-p)(1-\alpha^{BU})(R_{-}^{A*}+R_{-}^{B*})]+\phi[K^{A*}+K^{B*}]]+[(TC^{A*}+TC^{B*})-TC^{CU**}]>0.$$

This requires:

[13] An alternative specification might include a parameter that captures the extent to which a business unit is biased towards its own division in relation to the other division [2]. However, for simplicity we assume that such a bias does not exist and the business division optimizes its own surplus.

$$[(TC^{A^*} + TC^{B^*}) - TC^{CU^{**}}] > -(1+\phi)[p(1+\alpha^{CU})(R_+^{A^{**}} + R_+^{B^{**}}) + (1-\alpha^{CU})(1-p)(R_-^{A^{**}} + R_-^{B^{**}})]$$
$$+[[p(1+\alpha^{BU})(R_+^{A^*} + R_+^{B^*}) + (1-p)(1-\alpha^{BU})(R_-^{A^*} + R_-^{B^*})] + \phi[K^{A^*} + K^{B^*}]]$$

In Appendix 2 we show that when businesses face a low uncertainty environment, higher similarity between the business units calls for more centralization. Moreover, higher similarity between business units implies higher correlated information signals X and Y respectively, due to the distributions $f(X)$ and $f(Y)$ being more homogeneous. Hence, this implies that the firm gets limited new information from information signals X and Y. In addition, since the degree of uncertainty is already low, any new information does not help in reducing the degree of uncertainty further. Moreover, since correlated signals do not reduce information variance of the random variable, W (denoted by $Var(W) = a^2Var(X) + b^2Var(Y) + 2abCov(X,Y)$, as noted in assumption (5) of Table 1), and the degree of uncertainty is already low, there is limited value from decentralization to benefit further from improved flexibility. Therefore, we posit the following proposition:

PROPOSITION 1 (Low Uncertainty and Business-Unit Similarity Proposition)

In the presence of low uncertainty, centralization contributes to a higher increase in total surplus for the firm compared to decentralization as the similarity of the business units increases (i.e. ϕ increases) and:

(i) does so at the same rate as the advantage of the business units relative to the central unit in their ability to react to information increases (*i.e. the larger is* $(\sigma = \frac{\alpha^{BU}}{\alpha^{CU}})$);

(ii) does so at a higher rate as the cost of the central unit relative to the business unit decreases (*i.e. the larger is λ, the cost advantage of centralization over decentralization).*

Proposition 1 implies that centralization enables similar businesses to benefit from each other's decision to collect information. However, if business units are allowed to act independently, there is a risk of increased costs from duplication of the information collected by each business unit, as well as the opportunity cost of not being able to coordinate actions to maximize the total surplus of the firm. Therefore, the Uncertainty and Business Unit Similarity Proposition (P1) implies that the benefits from centralization are larger when businesses are similar, as the center is better able to internalize the cross-benefit effects between the business units. The more similar the business units' information requirements, the higher the cross-benefit will be between them, resulting in a larger total surplus for the firm from centralization. Moreover, the proportionate increase in total surplus from centralization over decentralization as the business-unit similarity increases remains the same as the advantage of the business from local tailoring becomes stronger. In addition, Proposition 1 states that the cross-benefit and cost advantage realized through centralization of decision-making become even stronger as the center's cost advantage from less duplication becomes stronger. A direct corollary of Proposition 1 is that the lower the similarity between business units, the lower the cross-benefit between businesses and, hence, the lower the

benefits from centralization.[14] Therefore, the corollary implies that as the business units become more dissimilar, decentralizing data management becomes optimal.

Empirical Illustrations

The case of Wal-Mart provides an illustrative example of a business with similar stores across the country, allowing for data to be managed centrally. When Wal-Mart analyzed its centralized shopper history data after Hurricane Charlie hit Florida in August 2004, it found that, somewhat surprisingly, people tended to buy strawberry pop tarts and beer in Wal-Mart stores that were in the eye of the hurricane just before the storm hit [41]. Acting on this information, Wal-Mart filled its trucks with toaster pastries and six-packs and stocked its other Florida stores with similar buyer profiles that were in the pathway of Hurricane Frances a few weeks later in September 2004. Most of the stock was sold before the second hurricane. Wal-Mart was able to benefit because of two factors. First, by centralizing its data management, Wal-Mart was able to share information across stores because of the similarity of buying patterns across some of its Florida stores. Second, it was able to eliminate duplication of information on its supplier side by capturing information about customer purchases centrally. Such customer-purchase information reduces costs by consolidating the delivery of the product categories in demand to stores ahead of the hurricane and reducing other product categories.

On the other hand, when business units become more dissimilar, they prefer to have more independent control of business-unit information because choices made by the corporate center diverge from local business-unit needs. For example, within a pharmaceutical firm, a division that competes on cost, based on delivering commodity products such as intravenous fluid bags or other supplies, takes an approach to data management that is different from the approach taken by another division that focuses on research and development with a high-margin business [5]. The commodity-based product division will be concerned with information related to managing the manufacturing and divisional operating costs, while the R&D development division needs to track and manage information about new ideas and discoveries that help its scientists' research efforts. Any attempt to centralize data management of these dissimilar divisions is likely to compromise one or both business units' objectives. As discussed earlier, this is because the firm will not benefit as much from centralizing data management due to dissimilar businesses, while reducing the ability of the business units to leverage their respective total surplus from information-based flexibility.

In summary, under low uncertainty firms need to be cognizant of the level of similarity across business units when deciding whether to centralize data across the firm. Next we look at how this decision is affected in the case of high uncertainty.

Similarity of Business Units in the Presence of High Uncertainty

We now analyze the impact of high uncertainty on the decision to centralize data management. The key insight of this section is that centralization can become more attractive

[14] We defined similarity of the business in terms of common customer or market segment being served. However, although the business units could be serving different customer segments, they might still be jointly affected by co-integrated exogenous factors such as the state of the economy. We do not model this aspect of the similarity between the business units.

as the dissimilarity of the business units increases. This is because dissimilar businesses enable the firm to reduce uncertainty, which in turn enhances the value of cross-benefit effects and hence, increases the surplus for the centre more than for the business units.

High uncertainty means the firm is less certain about the true state of the environment [28][39]. Hence, high uncertainty implies that there is higher variance in the possible outcomes. However, business units that are dissimilar enable the firm to get new information about the state of the environment by combining information from diverse sources. We next formalize how the reduction in uncertainty influences the decision to centralize data management.

Impact of the Similarity of Business Units

One approach to reducing high uncertainty is to compare information across business units. Similarity of business units implies commonality in terms of information requirements in order to benefit from opportunities arising from existing businesses. The low uncertainty environment arises from relatively known events, such as the probability of whether or not the customer is likely to purchase the product. For example, this could be the case with the mortgage and credit divisions of a bank that might share the same potential customer. However, in a highly uncertain environment, the uncertainty arises from macro changes, such as new technological breakthroughs. Such a highly uncertain environment often calls for diversity in the types of information required in order to reduce the high uncertainty and leverage opportunities. For example, the uncertainty could be about changes in customer preferences as a result of a new technology, such as mobile technology, on retail banking. Such uncertainty calls for diverse information about mobile technology and payment systems to reduce the high level of uncertainty.

Therefore, when there is high uncertainty, the possibilities to reduce the variance of outcomes are increased, by comparing information between business units that are dissimilar relative to business units that are similar. Various sources of research support this assertion. First, studies show that organizations combine diverse data to reduce ambiguity or high uncertainty in equivocal contexts [see 18]. Second, Clemen and Winkler [16] show that when information sources are dependent, the resulting information from these sources will be redundant. Therefore, the reduction in variance or gain in precision afforded by dependent information sources will be less than that from independent sources. Third, the benefit of combining dissimilar information is supported by research that shows that creative solutions to a problem are more likely to come from diverse sources [22, 24]. As discussed earlier, when business unit A and business unit B are dissimilar, the distribution from which information signals X and Y are drawn, $f(X)$ and $f(Y)$ respectively, is more heterogeneous. Such heterogeneous distribution implies diversity and, therefore, the combination of signals X and Y provides new information. Hence, we assume that in this context, when business units are more dissimilar, $\phi \to 0 \Rightarrow X \cup Y$ creates more new information, I^N, which helps reduce the variance of the possible states of the environment.

Impact of High Uncertainty on the Decision to Centralize

To incorporate the situation of high uncertainty into our model, we consider the case where the firm is less certain about the true state of the environment and, hence, p and $(1-p)$ are

closer to 0.5. Such a dispersion of the possible states of the environment is encapsulated by the random variable, W, with corresponding distribution, $f(W)$, as per assumption (5). Therefore, the expected total surplus of W is $E(W) = aE(X) + bE(Y)$. The variance of W, denoted by $Var(W) = a^2Var(X) + b^2Var(Y) + 2abCov(X,Y)$, is as noted in assumption (5) of Table 1. When the business units are more dissimilar, namely $\phi \to 0$, then $Cov(X,Y) \to 0$, which helps reduce the variance of the possible states of the environment. An improvement in information signals reduces the variance of W to $Var(W^*)$, where $Var(W^*) = Var(W) - V(X,Y,I^N)$, as noted in assumption (5) of Table 1. We capture the effect of the new information in reducing uncertainty via a reduction in variance through the term $V(X,Y,I^N)$ [3]. Therefore, the term $V(X,Y,I^N)$ can be interpreted as the gain in precision from the new information, I^N [17]. Intuitively, this says that the new information reduces uncertainty in the random variable, W, due to the combined information provided by information signals X and Y jointly. Such a gain in precision due to the new information in turn reduces uncertainty by moving p and $(1-p)$ away from 0.5 to either 0 or 1. Hence, the gain in precision from the new information makes the outcome more certain for the firm.

In order to decide whether to centralize or decentralize, we compare the expected total surplus of the firm with centralized and decentralized models respectively. As $\phi \to 0$, the business units become most dissimilar and hence, the information signals X and Y become more independent. Therefore, the high degree of uncertainty is reduced more as business units become more dissimilar compared to when they are similar. In order to decide how such a reduction in uncertainty affects the decision to centralize or decentralize data management, we revisit the decision criteria as provided by inequality (7):

$$[(TC^A + TC^B) - TC^{CU}] > -(1+\phi)[p(1+\alpha^{CU})(R_+^A + R_+^B) + (1-\alpha^{CU})(1-p)(R_-^A + R_-^B)]$$
$$+[[p(1+\alpha^{BU})(R_+^A + R_+^B) + (1-p)(1-\alpha^{BU})(R_-^A + R_-^B)] + \phi[K^A + K^B]]$$

Recollect that when the left-hand side is larger than the right-hand side, the firm needs to centralize. We examine the role of diversity of information sources on new information generation and its influence on the degree of uncertainty. In order to do so, we examine the strength of the cross-benefit effect on coordination across business units relative to the benefit of local tailoring.

A reduction in uncertainty causes an interaction between the coordination benefit accruing to the central unit and the local tailoring benefit accruing to the business units. As discussed earlier, the reduction in high uncertainty is higher when the business units are more dissimilar compared to when they are similar. When dissimilarity is high, the probability of an event occurring changes from $p \approx 0.5$ to a higher value if the favorable outcome is more likely, or to a lower value if an unfavorable outcome is more likely. The case when the unfavorable outcome becomes more likely implies that the firm is less likely to continue in the business because the returns are less attractive and might fall below the minimum threshold. Therefore, we examine the decision to centralize or decentralize data management when the reduction in uncertainty makes the favorable outcome more likely.

We are interested primarily in how the decision about whether to centralize or decentralize data management varies as we change the parameter values. Therefore, we need to compare how the relative surplus between centralization and decentralization changes when the business units are dissimilar compared to when they are more similar. We conduct numerical analysis in Appendix 3 in order to do this. In particular we examine the decision to centralize or decentralize due to the similarity of the business units, and examine how such an effect changes as we change the relative advantage of the business units to the central unit in their ability to react to information and the relative cost advantage of the central unit to the business units.

The numerical analysis shows that as the business units become more dissimilar, the change in probability towards the more favorable outcome implies that more weight is placed on the favorable outcome compared to the unfavorable outcome. However, the increased weight placed on the favorable outcome means the coordination benefit becomes stronger, which favors centralization. On the other hand, the positive weight placed on the favorable outcome means the benefit of local tailoring is also correspondingly higher, which favors decentralization. The numerical analysis shows that when business-unit dissimilarity increases, the larger reduction in uncertainty influences the cross-benefit effect from coordination more than the benefit from local tailoring. This means that as the business units become more dissimilar, centralization becomes more favorable compared to decentralization. The analysis also shows that such an effect becomes stronger as the advantage of the business unit relative to the central unit in their abilities to react to information increases. In contrast, the incentive to centralize remains the same when the cost of the central unit relative to the business unit decreases. The corollary of this result is that as the dissimilarity of the business units increases, centralization becomes less attractive and, hence, increasingly favors decentralization. Therefore, we posit the following proposition:

PROPOSITION 2 (High Uncertainty and Business-Unit Dissimilarity Proposition)

In the presence of high uncertainty, centralization contributes to a higher increase in total surplus for the firm compared to decentralization as the dissimilarity of the business units increases (i.e. ϕ *decreases) and:*

(i) does so at a higher rate as the advantage of the business units relative to the central unit in their ability to react to information increases (i.e. the larger is $\lambda = \frac{\alpha^{BU}}{\alpha^{CU}}$ *);*

(ii) does so at the same rate as the cost of the central unit relative to the business unit decreases (*i.e. the larger is* λ *, the cost advantage of centralization over decentralization).*

Therefore, the High Uncertainty and Business-Unit Dissimilarity Proposition (P2) implies that the benefits from centralization could be larger when businesses are more dissimilar. This is because dissimilar businesses enable the generation of new information in order to reduce high uncertainty. Such a reduction in high uncertainty increases the benefit of coordination more than the corresponding benefit from local tailoring, which in turn makes centralization optimal. Such a reduction in uncertainty makes the cross-benefit and cost advantage realized through the centralization of decision-making even stronger as the

advantage of the business unit from local tailoring becomes stronger. In contrast, such a reduction in uncertainty makes the cross-benefit and cost advantage realized through the centralization of decision-making change proportionally as the center's cost advantage from less duplication becomes stronger. A direct corollary of Proposition 2 is that the higher the similarity between the business units, the lower the reduction in uncertainty and, hence, the lower the benefits from centralization. Therefore, the corollary implies that as the business units become more similar, decentralizing data management becomes more optimal.

Empirical Illustrations

An illustrative example of this is the pharmacovigilance activities of a pharmaceutical company [37]. Pharmacovigilance activities include the systematic detection, assessment, understanding and prevention of adverse drug reactions. A major problem that pharmaceutical companies face in the pharmacovigilance area is the possibility that a drug might be taken by a patient in a manner for which it was not intended or clinically tested. This might arise due to the biological differences between the patient and the participants in the clinical trial, the combination of medicines taken with other therapies, or the medicine being prescribed for an indication that is different from the one that the medicine was approved for. These factors increase the odds of adverse effects that might not be detected at the clinical stage but could manifest after the launch of the medicine. For example, the US Food and Drug Administration shows an average growth rate of 12 per cent annually in adverse events reported between 1995 and 2004.

Pharmaceutical firms face an ambiguous situation regarding the possibility of adverse events. High uncertainty is due to the complex combination of factors outlined above between product lines across various regulatory jurisdictions. As a result, the need to accommodate regulatory requirements or cultural imperatives calls for centralized data management. Moreover, much data is often collected and analyzed by discrete business units that have different therapy or compound-specific working groups using different collection approaches with broad data needs. Therefore, centralizing data management across the different business units enables data-sharing for more efficient patient safety, consistent regulatory compliance, and controlled risk management, while enabling scale by systematically following a set of procedures. For example, clinical trial data might not be leveraged for use for patient-safety purposes because they were not collected with future purposes in mind. In such cases, centralizing data management allows for a more coordinated use across departments to identify safety trends, resulting in reduced uncertainty and, consequently, lower adverse consequences. Therefore, in the case of high uncertainty, centralized data management allows for the benefits of managing overall risk, which enables data-sharing to narrow the possible outcomes and, hence, benefit from the value of flexibility simultaneously. In summary, high uncertainty situations might call for more centralized data management in the case of dissimilar business units in order to control broad risks and enable a reduction in the level of uncertainty through information-sharing.

Discussion and Managerial Implications

Our research develops a framework to help firms to decide more accurately when it is optimal to centralize data by considering the value of information-based flexibility in the presence of uncertainty. We show that firms need to consider not only costs but also the value

of information-based flexibility in evaluating the benefits of centralizing data management. We first consider the case of low uncertainty and then we go on to discuss the case of high uncertainty. In doing so, we show that, in the presence of low uncertainty, centralization creates more total surplus as the similarity of the business units increases. In contrast, in the presence of high uncertainty, centralization creates more total surplus as the dissimilarity between business units increases when new information reduces the high uncertainty and in turn enhances the surplus from coordination relative to the surplus of the superior ability of the business units to react to uncertainty.

Framework for Summarizing Recommendations

The two main considerations for firms to take into account in deciding between centralization and decentralization are uncertainty (external to the firm) and the extent to which businesses are similar (internal to the firm). First, the level of uncertainty determines the value of flexibility from an unknown future state. Second, the level of similarity determines to what extent businesses can share useful information to increase joint flexibility value.

Figure 2. Summary Recommendations

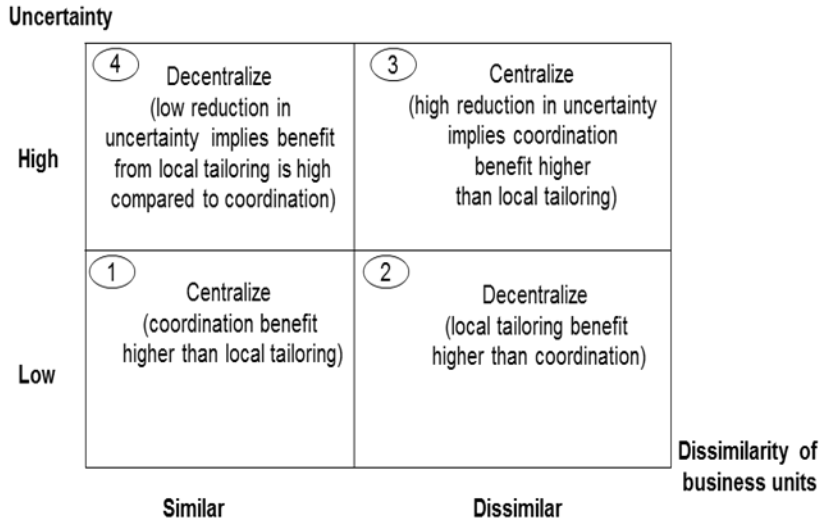


Figure 2 helps to provide a cohesive summary of recommendations. The vertical axis measures the level of uncertainty and the horizontal axis measures the degree to which the businesses are dissimilar. First, we consider the case of low uncertainty, which is displayed in the bottom two rows of Figure 2. The two alternatives are as follows:

(i) Case (1) with low uncertainty and similar businesses implies that centralization creates more total surplus than decentralization, as prescribed by Proposition 1. The benefits of decentralization in order to leverage information-based flexibility due to improved local tailoring would be minimal compared to the loss of coordination benefit because of the similarity of the business units and the low level of uncertainty.

(ii) Case (2) with low uncertainty and dissimilar businesses implies that decentralization creates more total surplus than centralization, as prescribed by the corollary implied by

Proposition 1. The benefits from centralization due to better coordination are low because the business units are dissimilar and there is a low level of uncertainty.

Second, we consider the case of high uncertainty. High uncertainty calls for the sharing of data to create new information and, hence, reduce the level of uncertainty. The top row of Figure 2 discusses the case of high uncertainty. The two alternatives are as follows:

(i) Case (3) with high uncertainty and dissimilar businesses implies that centralization could create more total surplus than decentralization, as prescribed by Proposition 2. Centralization becomes more attractive when the reduction in high uncertainty due to dissimilar businesses (because of the new information) for the central unit increases the benefit from coordination more than the benefit from local tailoring for the business units. In such cases, the benefit from decentralization would be minimal because the reduction in uncertainty is less responsive to the total surplus of business unit's ability to benefit from local tailoring than the benefit from central coordination.

(ii) Case (4) with high uncertainty and similar businesses implies that decentralization creates more total surplus than centralization, as prescribed by the corollary implied by Proposition 2. This is because high uncertainty benefits the business units more than the central unit due to a relatively higher degree of information-based flexibility value due to improved local tailoring. The benefit from centralization as a result of new information reducing uncertainty is low because of the similar nature of the business units.

Application of Framework to the Thomson Reuters Case

The challenges faced by Thomson Reuters, as discussed in the Introduction, provide an illustrative example of this framework. In the 1980s and the 90s, Thomson Reuters had a very decentralized model in order to cater for each country's specific needs. Thomson Reuters' management thought that they were operating in quadrant (2) with dissimilar businesses, low benefits of central coordination, and an uncertain country-specific customer demand. However, with the decentralized data-management approach, due to the plethora of products, Thomson Reuters found it incredibly difficult to serve its large global clients when the rate of innovation in the financial markets increased in the late 1990s. This led to a period of centralized data management between early 2000 and 2005, with the introduction of a 'Fast Forward' programme. The business transformation programme centralized data management by transforming the product architecture from being organized around geography to customer groupings. In addition, the programme consolidated the client base in order to focus on a smaller set of key clients. In effect, Thomson Reuters' management realized that they were in quadrant (1), where coordination across customer grouping was key to managing the velocity of change in the financial markets. The 'Fast Forward' programme streamlined the product portfolio to enable the center to coordinate better between business units.

However, the rapid innovation in financial instruments, coupled with rapid changes in technology in the late 1990s and the new millennium (2000–9), reduced the competitive advantage of Thomson Reuters considerably. In particular, the pervasiveness of the Internet challenged the value of the private network, the Integrated Data Network built by Thomson Reuters to relay information. Moreover, the adoption of Internet technology provided the impetus for a rapid shift in industry borders and created a highly uncertain environment for Thomson Reuters. These developments meant that Thomson Reuters found it hard to keep

pace with such rapid change in order to identify information, disseminate it and respond to opportunities or threats. Continued market evolution thus caused Thomson Reuters to face significant uncertainty in the development of its markets and the ability to respond with appropriate products and services. Therefore, Thomson Reuters' management realized it was in quadrants (3)–(4) and needed to strike a balance between centralized and decentralized data management in order to reduce high uncertainty depending on the degree of similarity of its business units.

Reuters Group's strength lay in the broker–dealer bank (the 'sell-side') market, while the Thomson Corporation's strength lay in a different market, namely the investor market (the 'buy-side'), such as asset-management firms. The merger of the two groups to form Thomson Reuters in 2008 provided an opportunity to combine the different business units to reduce the high degree of uncertainty. In order to grow through innovation in such a dynamic market place, Thomson Reuters recognizes that it needs to move away from product-centric thinking to a model based on platform innovation where there is collaboration across its business units. This is because Thomson Reuters' product-centric business model was perceived to be slow in identifying information, disseminating it and responding to an opportunity.

The collaborative model calls for a more coordinated approach to information-gathering, dissemination and response. This coordinated approach implies greater collaboration with a common platform architecture and frequent data-sharing across its businesses in order to reduce high uncertainty and increase innovative activity. In order to facilitate such collaboration, Thomson Reuters is centralizing its platform-based data management to enable interactions among multiple business units, customers and suppliers. Therefore, Thomson Reuters recognizes it is operating in quadrant (3), where there are dissimilar business units combined with high uncertainty. Hence, Thomson Reuters is beginning to embrace a more collaborative innovation business model by building a common platform architecture across the sell-side and buy-side business units with centralized data management. On the other hand, where the business units are relatively similar, the benefits of centralized data management are reduced in a high-uncertainty environment and, hence, Thomson Reuters is embracing a more decentralized data-management model, as prescribed in quadrant (4).

Conclusion

The results derived in this paper provide a framework for how managers can decide on when it is optimal to centralize decisions about data management in the presence of low and high uncertainty. In particular, we show how there is a positive relationship between business-unit similarity and centralization under low uncertainty and how this relationship could be reversed in the case of high uncertainty.

We recognize certain limitations of this study, which warrant further research. First, we do not consider the impact of standards on the decision to centralize or decentralize data management. A central assumption in much of the literature is that there is an association between centralization and standardization [1] [6] [25]. Standardization is the adoption of voluntary formal rules or procedures that pertain to a set of stakeholders [14]. Standardization would enable better communication between business units or with the center [7]. An extension of our model could incorporate how data standards could influence the decision to centralize or decentralize data management. Second, we assume that the

correlation coefficient is positive between the business units. However, in the case where business units act as a hedge, whereby an increase in revenue for one business implies a decrease in revenues for another business, the correlation coefficient could be negative. An extension of the paper could analyze how the results would change in this case.

Acknowledging these limitations, we argue that our model provides a useful framework for modelling the important decisions about centralizing data management. We show how the decision to centralize data management in the context of uncertainty in the environment has implications for the design of business models of firms and, hence, their innovative capabilities.

Acknowledgments: The authors are grateful to Xitong Li, Sriya Iyer, the Editor-in-Chief and two anonymous reviewers, who provided constructive comments on earlier versions of this article.

References

1. Allen, B., and Boynton, A. Information architecture: In search of efficient flexibility. *MIS Quarterly*, 15, 4 (1991), 435–445.
2. Alonso, R.; Dessein, W.; and Matouschek, N. When does coordination require centralization? *American Economic Review*, 98, 1 (2008), 145–179.
3. Amari, S., and Nagaoka, H. *Methods of Information Geometry: Translations of Mathematical Monographs*. New York: Oxford University Press, 2000.
4. Anand, K.S., and Mendelson, H. Information and organization for horizontal multimarket coordination. *Management Science*, 43, 12 (1997), 1609–1627.
5. Applegate, L. M.; Robert, D. A.; and Warren McFarlan, F. *Corporate Information Strategy and Management: The Challenge of Doing Business in a Network Economy*. 6th ed., New York: McGraw-Hill, 2003.
6. Atkinson, S.; Stanley, L.; and Tschirhart, J. Revenue sharing as an incentive in an agency problem: An example from the national football league. *Rand Journal of Economics*, 19, 1 (1988), 27–43.
7. Avadikyan, A.; Llerena, P.; Matt, M.; Rozan, A.; and Wolff, S. Organisational rules, codification and knowledge creation in inter-organisation cooperative agreements. *Research Policy*, 30 (2001), 1443–1458.
8. Bader, J.; Hayward, C.; Razo, J.; Madnick, S.; and Siegel, M. An analysis of data standardization across a capital markets/financial services firm. *MIT Sloan School of Management Working Paper*, 4108, Cambridge, MA, 1999.
9. Benaroch, M., and Kauffman, R.J. A case for using real options pricing analysis to evaluate information technology project investments. *Information Systems Research*, 10, 1 (1999), 70–86.
10. Benaroch, M., and Kauffman, R.J. Justifying electronic banking network expansion using real options analysis. *MIS Quarterly*, 24, 2 (2000), 197–225.
11. Benaroch, M.; Jeffery, M.; Kauffman, R.J.; and Shah, S. Options based risk management: A field study of sequential IT investment decisions. *Journal of Management Information Systems*. 24, 2 (2007), 103–140.
12. Bernanke, B. S. Irreversibility, uncertainty, and cyclical investment. *The Quarterly Journal of Economics*, 98, 1 (1983), 85–106.
13. Bolton, P., and Farrell, J. Decentralization, duplication and delay. *Journal of Political Economy*, 98, 4 (1990), 803–826.
14. Brunsson, N., and Jacobsson, B. *A World of Standards*. Oxford: Oxford University Press, 2000.
15. Brynjolfsson, E. Information assets, technology and organization. *Management Science*, 40, 12 (1994), 1645–1662.
16. Clemen, R.T. and Winkler, R.L. Limits for the precision of information from dependent sources, *Operations Research*, 33, 2 (1985), 427–442.
17. Clemons, E.; Reddi, S.P.; and Row, M.C. The impact of information technology on the organization of economic activity: The "Move to the Middle" hypothesis. *Journal of Management Information Systems*, 10, 2 (1993), 9–35.
18. Daft, R.L. and Lengel, R.H. Organizational information requirements, media richness and structural design. *Management Science*, 32, 5 (1986), 554–571.

19. Dixit, A.K., and Pindyck R.S. *Investment Under Uncertainty*. Princeton: Princeton University Press, 1994.
20. Dos Santos, B.L. Justifying investments in new information technologies, *Journal of Management Information Systems*, 7, 4 (1991), 71–90.
21. Fichman, R.G.; Keil, M.; and Tiwana, A. Beyond valuation: "Options Thinking" in IT project management. *California Management Review*, 47, 2 (2005), 74–96.
22. Gebert, D.; Boerner, S.; and Kearney, E. Fostering team innovation; Why is it important to combine opposing action strategies? *Organization Science*, 21, 3 (2010), 593–608.
23. Han, K.; Kauffman, R.J.; and Nault, B.R. Information exploitation and interorganizational systems ownership. *Journal of Management Information Systems*, 21, 2 (2004), 109–135.
24. Hennessey, B. and Teresa, A., The conditions of creativity in *The Nature of Creativity: Contemporary Psychological Perspectives*, Sternberg, R. (ed.), New York: Cambridge University Press, 1988, pp. 11–38.
25. Jain, H.; Ramamurthy, K.; Ryu, H.S.; and Yasai-Ardekani, M. Success of data resource management in distributed environments: An empirical investigation. *MIS Quarterly*, 22, 1 (1998), 1–29.
26. Kauffman, R. and Mohtadi, H. Proprietary and open systems adoption in e-procurement: A risk-augmented transaction cost perspective. *Journal of Management Information Systems*, 21, 1 (2004), 137–166.
27. Kumar, R.M. A note on project risk and option values of investments in information technologies. *Journal of Management Information Systems*, 13, 1 (1996), 187–193.
28. Manning, C. and Schutze, H. *Foundations of Statistical Natural Language Processing*, Cambridge, MA: MIT Press, May 1999.
29. Marchand, D.A. Seeking global advantage with information management and information technology capabilities. In *The Blackwell Handbook of Global Management – A Guide to Managing Complexity*, H.W. Lane; Maznevski, M.L.; Mendenhall, M.E.; and McNett, J. (eds), Oxford: Blackwell Publishing, 2004, pp. 300–321.
30. Marchand, D.A.; Kettinger, D.; and Rollins, D. *Information Orientation: The Link to Business Performance*. Oxford: Oxford University Press, 2001.
31. Melville, N.; Kramer, K.; and Gurbaxani, V. Information technology performance: An integrative model of IT business value, *MIS Quarterly*, 28, 2 (2004), 283–322.
32. Merkhofer, M.W. The value of information given decision flexibility. *Management Science*, 23, 7 (1977), 716–727.
33. Miller, A. The value of sequential information. *Management Science*. 22, 1 (1975), 1–11.
34. Moscarani, G., and Smith, L. The optimal level of experimentation. *Econometrica*, 69, 6 (2001), 1629–1644.
35. Nault, B. Information technology and organizational design: Locating decisions and information. *Management Science*, 44, 10 (1998), 1321–1325.
36. Ponssard, J.P. On the concept of the value of information in competitive situations. *Management Science*, 22, 7 (1976), 739–747.
37. PricewaterhouseCoopers. *Unlocking the Powers of Pharmacovigilance: An Adaptive Approach to an Evolving Drug Safety Environment.* PwC Views, 1 January 2007.
38. Sambamurthy, V.; Bharadwaj, A.; and Grover, V. Shaping agility through digital options: Reconceptualizing the role of information technology in contemporary firms. *MIS Quarterly*, 27, 2 (2003), 237–263.

39. Shannon, C.E., A mathematical theory of communication. *Bell System Technical Journal*, 27, (1948), 379–423.
40. Taudes, A.; Feurstein, M.; and Mild, A. Options analysis of software platform decisions: A case study. *MIS Quarterly*, 24, 2 (2000), 227–243.
41. *The New York Times.* What they know about you, 14 November 2004.
42. Van Alstyne M.; Brynjolfsson, E.; and Madnick, S. Why not one big database? Principles for data ownership. *Decision Support Systems,* 15, (1995), 267–284.

Appendix 1. Irreversibility and Uncertainty

Irreversibility implies that in period *t=1* the business-unit information can be collected just as efficiently, so that there is no difference in the impact on total surplus. If decisions are reversible, then $\alpha^{BU} = \alpha^{CU}$ and, hence, $(1+\alpha^{BU}) = (1+\alpha^{CU})$ and $(1-\alpha^{BU}) = (1-\alpha^{CU})$. This implies that (3) becomes: $E(V^{CU}) - E(V^{BU}) = G$ (1A)

Since *G*>0, (3) has a positive value, the optimal decision is to centralize data management.

If the firm ignored uncertainty then the optimal decision would be such that the stochastic net profit problem becomes deterministic. If the expected total surplus (excluding costs) were believed to be known with certainty to be ψ, then $pR_1(1+\alpha^i) + (1-p)R_2(1-\alpha^i) = \psi^i > 0$, where $i \in (CU, BU)$. Effectively this means there is no opportunity to enhance the surplus or reduce losses as there is no uncertainty. This implies that (3) becomes: $E(V^{CU}) - E(V^{BU}) = G + (\psi^{CU} - \psi^{BU})$ (1B)

Since G is a positive amount and $\psi^{CU} = \psi^{BU}$, (1B) has a positive value; therefore, the optimal decision is to centralize data management.

Appendix 2. Low Uncertainty

For both BU_A and BU_B, the expected total surplus, as before, is:[15]

$$E(V^A) = p(1+\alpha^{BU})R_+^A + (1-p)(1-\alpha^{BU})R_-^A + \phi[p(1+\alpha^{BU})R_+^B + (1-p)(1-\alpha^{BU})R_-^B] - TC^A \quad (2A)$$

$$E(V^B) = p(1+\alpha^{BU})R_+^B + (1-p)(1-\alpha^{BU})R_-^B + \phi[p(1+\alpha^{BU})R_+^A + (1-p)(1-\alpha^{BU})R_-^A] - TC^B \quad (2B)$$

$\mathrm{K}^B = [p(1+\alpha^{BU})R_+^B + (1-p)(1-\alpha^{BU})R_-^B]$ and $\mathrm{K}^A = [p(1+\alpha^{BU})R_+^A + (1-p)(1-\alpha^{BU})R_-^A]$

Under decentralization, since each business unit treats K^B and K^A as fixed, the first-order conditions for BU_A and BU_B are respectively:

$$E'(V^A) = p(1+\alpha^{BU})R_+'^A + (1-p)(1-\alpha^{BU})R_-'^A - TC'^A = 0 \quad (2C)$$

$$E'(V^B) = p(1+\alpha^{BU})R_+'^B + (1-p)(1-\alpha^{BU})R_-'^B - TC'^B = 0 \quad (2D)$$

This implies:

$$p(1+\alpha^{BU})R_+'^A + (1-p)(1-\alpha^{BU})R_-'^A = TC'^A \quad (2E)$$

$$p(1+\alpha^{BU})R_+'^B + (1-p)(1-\alpha^{BU})R_-'^B = TC'^B \quad (2F)$$

Under centralization, the expected total surplus from (5) is:

$$E(V^{CU}) = (1+\phi)[p(1+\alpha^{CU})(R_+^A + R_+^B) + (1-\alpha^{CU})(1-p)(R_-^A + R_-^B)] - TC^{CU} \quad (2G)$$

The first-order condition for the central unit is:

$$E'(V^{CU}) = (1+\phi)[p(1+\alpha^{CU})(R_+'^A + R_+'^B) + (1-\alpha^{CU})(1-p)(R_-'^A + R_-'^B)] - TC'^{CU} = 0 \quad (2H)$$

This implies:

---

[15] Revenues and loss (R_+, R_-), as well as costs (TC), are functions of I. However, for simplicity we list them without reference to I.

$$(1+\phi)[p(1+\alpha^{CU})(R'^{A}_{+}+R'^{B}_{+})+(1-\alpha^{CU})(1-p)(R'^{A}_{-}+R'^{B}_{-})]=TC'^{CU} \quad (2I)$$

One of the major differences between the central unit making the decision and the business units making the decision can be understood from examining the first condition in (2I) compared to (2E) and (2F). When the central unit makes the decision, it factors in the cross-benefit accruing to each business unit as a result of the information-collection decision made by the other unit, which is captured by the correlation coefficient in the first-order condition in (2I). However, when the business units make the decision, although it affects overall surplus, as noted in equations (6a) and (6b), each business unit is optimizing its own surplus without considering the cross-benefit effects (because it is unable to influence the decision of the other business unit), as can be observed by the absence of the correlation term in first-order conditions (2E) and (2F).

We had established that for centralization to be optimal, the net total surplus, as defined by (7), must be positive,[16] $E(V^{CU})-[E(V^{A})+E(V^{B})]>0$, which implies:

$$\begin{aligned}&(1+\phi)[p(1+\alpha^{CU})(R^{A}_{+}+R^{B}_{+})+(1-\alpha^{CU})(1-p)(R^{A}_{-}+R^{B}_{-})]-TC^{CU}>\\&[[p(1+\alpha^{BU})(R^{A}_{+}+R^{B}_{+})+(1-p)(1-\alpha^{BU})(R^{A}_{-}+R^{B}_{-})]+\phi[K^{A}+K^{B}]]-[(TC^{A}+TC^{B})]\end{aligned} \quad (2J)$$

At the optimal point I^{*}, $\frac{\partial E(R)}{\partial I}=\frac{\partial TC(I)}{\partial I}$ and $\frac{\partial^{2}E(V)}{\partial I^{2}}<0$. Let the optimal information for CU and business units A and B respectively be:

$(I^{CU^{*}_{A}},I^{CU^{*}_{B}})$ and $(I^{BU^{*}_{A}},I^{BU^{*}_{B}})$, which are the solutions for (2I), (2E) and (2F) respectively. In order to explore the comparative static of how the total surplus under centralization and decentralization changes as the values of the different parameters vary, we need to specify functional forms for the revenue and cost. Based on the assumptions in Table 1 and the associated discussion, we assume the following convex cost and concave revenue functional forms:

Business unit cost, $TC^{A}=TC^{B}=\frac{1}{2}I^{2}$ (2K)

Therefore, $TC'^{A}=TC'^{B}=I$

Revenue, $R^{A}_{+}=R^{B}_{+}=10I^{\frac{1}{2}}$ (2L)

Therefore, $R'^{A}_{+}=R'^{B}_{+}=5I^{-\frac{1}{2}}$

and the loss is defined as a negative proportion of revenues.

Loss, $R^{A}_{-}=R^{B}_{-}=-\Omega 10I^{\frac{1}{2}}$, where $0<\Omega<1$ (2M)

Therefore, $R'^{A}_{-}=R'^{B}_{-}=-5\Omega I^{-\frac{1}{2}}$

As discussed earlier, the total cost for the central unit is related to the business units' costs,

$TC^{CU}=(1-\lambda\phi)[TC^{A}+TC^{B}]$, where $0<\lambda<1$ (2N)

---

16 Alternatively we can compare the ratio of the total surplus for centralization relative to the total surplus for decentralization, which needs to be larger than one for centralization to be optimal. We perform such an analysis in the numerical section below.

and λ captures the rate of cost reduction for the central unit due to savings from duplications.

For business unit A, replace (2K) to (2M) into (2E) for the optimal information-collection:[17]

$$I^{BU_A^*} = [5p(1+\alpha^A) - (5\Omega)(1-p)(1-\alpha^A)]^{\frac{2}{3}} \quad \text{(2P)}$$

Similarly for business unit B from (2F),

$$I^{BU_B^*} = [5p(1+\alpha^B) - (5\Omega)(1-p)(1-\alpha^B)]^{\frac{2}{3}} \quad \text{(2Q)}$$

For the central unit, replace (2K) to (2N) into (2I) for the optimal information-collection:

$$I^{CU^*} = [\frac{(1+\phi)[5p(1+\alpha^{CU}) - 5\Omega(1-p)(1-\alpha^{CU})]}{(1-\lambda\phi)}]^{\frac{2}{3}} \quad \text{(2R)}$$

We provide numerical solutions to compare how the total surplus under centralization and decentralization vary in order to examine how (i) the similarity of the business units, ϕ; (ii) the advantage of the business unit relative to the central unit, $\sigma = \frac{\alpha^{BU}}{\alpha^{CU}}$; and (iii) the cost advantage of the central unit relative to the business units, λ, influences the decision to centralize or decentralize data management. Since the firm only operates in the business when the expected value of doing so is positive, we consider the low uncertainty case when the probability of the favorable outcome, p, is close to 1.[18] The surplus from decentralization can be computed by replacing the values from (2P) and (2Q) into surplus of business unit A, (2A) and surplus of business unit B, (2B) respectively. The surplus for the firm from decentralization comes from the sum of the surpluses of business units A and B respectively. The surplus from centralization can be computed by replacing the value from (2R) into the surplus of the central unit, (2G). In order to examine how the similarity of the business units affects the total surplus when the central unit makes the decision, compared to when the business units make the decision, we compare the benefit from centralization when business-unit similarity is low,[19] ϕ_L, to when it is high, ϕ_H. We do so by comparing the ratio of the total surplus from centralization and decentralization (i) when business-unit similarity is low, ϕ_L, with (ii) when business-unit similarity is high, ϕ_H. Table 4 reports this ratio for different parameter values. The region of interest for Table 4 does pass the Hessian test for optimization.

---

17 The central unit's optimal level of information collection ($I^{CU_A^*}, I^{CU_B^*}$) is obtained through the solution for I^{CU^*}. The numerical analysis shows that (i) the total collection of information by the central unit could be higher or lower than the business units, depending on the parameter values; and (ii) the total surplus for the central unit could be higher or lower than the business units, depending on the parameter values.

18 The low uncertainty case where the probability of the unfavorable outcome is high, p is close to 0, implies that the expected total surplus is negative and, hence, the firm is unlikely to continue investing in the business. Although we analyze the case where the probability of the favorable outcome is high, it is straightforward to examine the case where low uncertainty implies the unfavorable outcome is high. The basic result of Proposition 1 would not change.

19 Since the degree of uncertainty is low, any new information from dissimilar businesses (low ϕ) reducing the degree of uncertainty is negligible.

Table 4. Ratio of Expected Total Surplus from Centralization and Decentralization
($p = 0.9, \Omega = 0.8, \alpha^{CU} = 0.5$)

No.	Parameter values when business-unit similarity is low	$\frac{EV^{C\phi_L}}{[EV^{A\phi_L} + EV^{B\phi_L}]}$ when business-unit similarity is low	Parameter values when business-unit similarity is high	$\frac{EV^{C\phi_H}}{[EV^{A\phi_H} + EV^{B\phi_H}]}$ when business-unit similarity is high
1	$\phi_L = \mathbf{0.4;}$ $\frac{\alpha^{BU}}{\alpha^{CU}} = 1.1;$ $\lambda = 0.3$	0.97	$\phi_H = \mathbf{0.6;}$ $\frac{\alpha^{BU}}{\alpha^{CU}} = 1.1;$ $\lambda = 0.3$	1.06
2	$\phi_L = 0.4;$ $\frac{\alpha^{BU}}{\alpha^{CU}} = \mathbf{1.2;}$ $\lambda = 0.3$	0.96	$\phi_H = 0.6;$ $\frac{\alpha^{BU}}{\alpha^{CU}} = \mathbf{1.2;}$ $\lambda = 0.3$	1.05
3	$\phi_L = 0.3;$ $\frac{\alpha^{BU}}{\alpha^{CU}} = 1.1;$ $\lambda = \mathbf{0.4}$	0.98	$\phi_H = 0.6;$ $\frac{\alpha^{BU}}{\alpha^{CU}} = 1.1;$ $\lambda = \mathbf{0.4}$	1.09

Table 4 shows how the ratio of the expected total surplus from centralization over the expected value from decentralization varies when business-unit similarity is low compared to when it is high. When the ratio is larger than one, centralization creates more value than decentralization, and vice versa. Row 1 of Table 4 shows that as one increases business-unit similarity (i.e. $\phi_L = 0.4$ to $\phi_H = 0.6$), the ratio increases from below 1 to above 1 by 9.3 per cent (0.97 to 1.06). We redo such an analysis in row 2 of Table 4, whereby we increase the advantage of the business units relative to the central unit in their ability to react to information decreases (i.e. the higher is $\lambda = \frac{\alpha^{BU}}{\alpha^{CU}}$). The percentage increase in the ratio from low business-unit similarity to high similarity is the same compared to row 1 of Table 4 at 9.3 per cent (0.96 to 1.05). We redo such analysis in row 3 of Table 4, whereby the cost of the central unit relative to the business unit decreases (i.e. the larger is λ, the cost advantage of centralization over decentralization). The percentage increase in the ratio from low business-unit similarity to high similarity is higher compared to row 1 at 11.2 per cent (0.98 to 1.09). The results of Table 4 are robust to other combinations of parameter values. These results give us Proposition 1. Proposition 1 also implies the reverse relationship, whereby a decrease in business-unit similarity means that decentralization becomes optimal. Such a reverse relationship gives us the corollary that Proposition 1 implies.

Appendix 3. High Uncertainty

We next examine the impact of high uncertainty on the decision to centralize data management. In order to do so, we need to specify the functional form for how business-unit similarity affects uncertainty in the environment, whereby the reduction in uncertainty is

higher when business units are dissimilar compared to when they are similar. In order to capture this relationship, we assume the following functional form:

$$p = \frac{1}{\exp(\phi^2)} \text{ if } 0 \le \phi \le 0.5 \text{ and } p = \ln(\phi^{0.05}) \text{ if } 0.5 < \phi \le 1 \qquad (3A)$$

The above functional form has the property that when business-unit similarity is low (i.e. $0 \le \phi \le 0.5$), the reduction in uncertainty is high compared to when business-unit similarity is high (i.e. $0.5 < \phi \le 1$). Let $p = 0.5$ when uncertainty is high. Then, using (3A), we get a relatively large increase in probability to $p = 0.85$ (i.e. $\Delta p = 0.35$) when business-unit similarity is low ($\phi_L = 0.4$). In contrast, we get a relatively small increase in probability to $p = 0.53$ (i.e. $\Delta p = 0.03$) and when business similarity is high ($\phi_H = 0.6$). We examine the decision to centralize or decentralize when the reduction in uncertainty makes the favorable outcome more likely.[20] Since uncertainty changes with the degree of business-unit similarity (e.g. ϕ_H and ϕ_L), we need to compare the percentage change in the ratio of the total surplus from centralization and decentralization as we vary the parameter values. Table 5 provides the details of the percentage change in such a ratio for different parameter values. The region of interest for Table 5 does pass the Hessian test for optimization.

Table 5. Percentage Change in Ratio of Expected Total Surplus from Centralization and Decentralization ($p = 0.5, \Omega = 0.8, \alpha^{CU} = 0.5$)

No.	Parameter values when business-unit similarity is low $\Delta p = 0.35$, to $p = 0.85$	Percentage change in $\frac{EV^{C\phi_L}}{[EV^{A\phi_L} + EV^{B\phi_L}]}$ due to reduction in uncertainty when business-unit similarity is low	Parameter values when business-unit similarity is high $\Delta p = 0.03$ to $p = 0.53$	Percentage change in $\frac{EV^{C\phi_H}}{[EV^{A\phi_H} + EV^{B\phi_H}]}$ due to reduction in uncertainty when business-unit similarity is high
1	$\phi_L = \mathbf{0.4;}$ $\frac{\alpha^{BU}}{\alpha^{CU}} = 1.1;$ $\lambda = 0.3$	5.41%	$\phi_H = \mathbf{0.6;}$ $\frac{\alpha^{BU}}{\alpha^{CU}} = 1.1;$ $\lambda = 0.3$	0.89%
2	$\phi_L = 0.4;$ $\frac{\alpha^{BU}}{\alpha^{CU}} = \mathbf{1.2;}$ $\lambda = 0.3$	10.48%	$\phi_H = 0.6;$ $\frac{\alpha^{BU}}{\alpha^{CU}} = \mathbf{1.2;}$ $\lambda = 0.3$	1.67%
3	$\phi_L = 0.3;$ $\frac{\alpha^{BU}}{\alpha^{CU}} = 1.1;$ $\lambda = \mathbf{0.4}$	5.41%	$\phi_H = 0.6;$ $\frac{\alpha^{BU}}{\alpha^{CU}} = 1.1;$ $\lambda = \mathbf{0.4}$	0.89%

[20] As discussed earlier, the case with an unfavorable outcome arguably implies that the firm is less likely to continue investing because the returns are less attractive and might fall below the minimum threshold.

Table 5 shows the percentage change in the ratio of the expected total surplus from centralization over the expected total surplus from decentralization as uncertainty reduces when business-unit similarity is low compared to when it is high. A high percentage implies a larger increase in the ratio, implying that centralization creates more total surplus compared to decentralization when uncertainty reduces. Row 1 of Table 5 shows that as one increases business-unit similarity (i.e. $\phi_L = 0.4$ to $\phi_H = 0.6$), the percentage increase in ratio is low (0.89%) when business-unit similarity is high, compared to when business-unit similarity is low (5.41%). Moreover, when business-unit similarity is low the ratio changes from less than one to more than one as the uncertainty reduces by a larger amount, which implies that the decision to centralize becomes stronger. We redo such an analysis in row 2 of Table 5, whereby we increase the advantage of the business units relative to the central unit in their ability to react to information decreases (i.e. the higher is $\lambda = \frac{\alpha^{BU}}{\alpha^{CU}}$). The percentage increase in the ratio from low business-unit similarity to high similarity is higher compared to row 1 of Table 5 (1.67% and 10.48% for high and low business-unit similarity respectively). We redo such an analysis in row 3, whereby the cost of the central unit relative to the business unit decreases (i.e. the larger is λ, the cost advantage of centralization over decentralization). The percentage increase in the ratio from low business-unit similarity to high similarity is the same compared to row 1 of Table 5 (0.89% and 5.41% for high and low business-unit similarity respectively). In general, the larger reduction in uncertainty when business-unit similarity is low compared to when it is high enables the total surplus for centralization to increase more than decentralization, which favors centralization. The results of Table 5 are robust to other combinations of parameter values. These results give us Proposition 2. Proposition 2 also implies the reverse relationship, whereby an increase in business-unit similarity means that decentralization becomes optimal. Such a reverse relationship gives us the corollary that Proposition 2 implies.