Trade-offs in Organizational Architecture:

Information Systems, Incentives, and Process Design

Abraham Seidmann

William E. Simon Graduate School of Business Administration

University of Rochester, Rochester, NY 14627

Arun Sundararajan

Leonard N. Stern School of Business

New York University, New York, NY 10012

Revised, November 2002.

The authors wish to thank three anonymous referees and an anonymous associate editor, Professors Rajiv Banker, Anitesh Barua, Rajiv Dewan, Marshall Freimer, Harry Groenevelt, Eugene Kandel, John Long, Leslie Marx, Haim Mendelson, Roy Radner, Paul Schweitzer, participants in the Workshop on Information Systems and Economics, and seminar participants at Carnegie-Mellon University, Dartmouth College, London Business School, New York University, Purdue University, University of Florida, University of Minnesota, University of Rochester, University of Southern California, and the University of Texas at Dallas.

Abstract

We present a framework and a set of analytical models, which study functional and process-oriented organizational design. Drawing from theories of information technology and organizational design, queuing theory, and principal-agent theory, we model simultaneous trade-offs between different types of information systems, fixed and performance based incentives schemes, and variations in the design of work systems. The effects of crucial parameters such as knowledge intensity, information asymmetry, technology responsiveness, mass customization, and activity rates are also studied.

We demonstrate that process-oriented organization is desirable when processes are large, when customers demand significant customization, and when the nature of work is such that information asymmetry is naturally high. However, functional organization is optimal in a number of contexts, particularly when work is knowledge intensive, technology responsiveness is low, when there is high variation in activity rates, and when processes in firms have few steps. While typical changes in work design are complemented by changes in either information technology or employee incentive schemes, we show that the simultaneous introduction of information systems and performance-based pay is not necessarily optimal; rather than being complementary, these changes are substitutable drivers of improved organizational performance. Our framework and results also provide significant context-specific managerial insight into the value of technologies such as intranets, knowledge management systems, expert systems and workflow management systems. In addition, we provide a general formulation and solution technique that will enable the modeling of hitherto intractable problems involving a principal-agent problem in a tandem network of queues.

1. Introduction

Over the last decade, numerous companies have made significant investments in information technology and management services, in an attempt to transform themselves from functionally specialized organizations to process-oriented organizations. These efforts were initiated and driven by some common themes of business process reengineering[1] – a paradigm that called for radical change in work systems, performance control and information systems infrastructure. In this paper, we present a set of analytical models that study the capabilities and limitations of functional and process-oriented organizational architectures.

The ideas most closely associated with shift towards a process oriented organization include consolidating tasks to increase efficiency and accountability, eliminating handoffs, introducing performance-based pay, and, most importantly, enabling these changes with new information technology (Hammer and Champy, 1993). In a recent book, Bill Gates highlights similar themes, noting that in the new organization, the worker is no longer a cog in the machine, but an intelligent part of the process; that if you cut a job into too many pieces, and involve too many people, nobody can see the whole process; that too many handoffs create too many likely points of failure; and that digital technology makes it possible to develop much better processes (Gates, 2000).

Hammer and Champy (1993) and Davenport (1993) describe a few successful cases, upon which popular organizational and process design principles are based[2]. However, the results of transforming organizations have been mixed, at best. Having transformed themselves, some firms find that they are unable to handle their workload (Xerox Corporation and Boeing being significant examples). Other companies, such as Levi Strauss (King, 1998), have found that the new ideas of joint compensation and teamwork are markedly inferior to their prior system of functional specialization and piece-rate compensation.

Possibly, inadequate change management plays a role in the implementation failure of new organizational design. However, it is also likely that the design itself is optimal only in certain business contexts. The purpose of our paper is to study this rigorously. Specifically, we ask:

1. When is process-oriented organization preferred to functional organization?
2. What mix of work systems, information systems, and incentives is optimal in a particular business context?
3. What are the performance trade-offs between simultaneous changes *across* these organizational design factors?

Our approach is to focus on analyzing simultaneous changes in work systems, supporting information technology and performance incentives. Earlier work has considered a subset of these dimensions in isolation. For instance, the work of Buzacott (1996) and Seidmann and Sundararajan (1997) focuses on work system performance, analyzing different queuing configurations and concluding that variability and task asymmetry favor process-oriented organizations. Barua, Whinston, and Lee (1996) develop and axiomatic framework to prove that information technology and work system changes may be complementary; Brynjolfsson and van Alstyne (1997) have developed a framework and software package that enables managers to study such complementary interactions. More recently, Malone et al. (1999) have developed a comprehensive handbook and software package that enable managers to design their work processes based on a set of templates, incorporating the structure of a process, its tasks and their associated dependencies. What sets our research apart from these papers is a model that simultaneously captures the operational and queuing aspects of organizational work processes, the complexity of performance-based incentives for employee empowerment, and the central role of information systems that characterizes modern service organizations.

The rest of the paper is organized as follows. We outline our basic model in §2, and provide a general result that enables us to solve the more specific models of functional and process-oriented organizations, presented in §3 and §4. In §5 we contrast the analytical results from the prior sections, graphically illustrate the trade-offs between functional and process-oriented organization, and describe some of the complementarity and substitutability between information technology and the other organizational variables. Finally, we discuss the managerial insights of our research and conclude in §6.

2. Basic Model

Our basic model is of an organization with one principal and n identical agents. The principal attempts to maximize her net profit function by inducing optimal effort levels from the n agents. Each agent has the same utility function $u(x,\omega)$, where ω is the agent's effort level, and x is compensation paid to the agent. We assume that this utility function takes the form

$$u(x,\omega) = U(x - c(\omega)) \qquad (1)$$

where $c(\omega)$ is the cost (measured in units of x) of expending effort ω. The vector of effort levels $(\omega_1,\dots\omega_n)$ of the n agents is denoted $\mathbf{\Omega}$. This vector is not directly observable, but it influences a stochastic performance measure, which directly affects the profits of the principal. A closed-form expression for the density function of the distribution of the performance measure may not be known. However, the Laplace transform $F^*(s|\mathbf{\Omega})$ of this density function is known. The density function underlying the Laplace transform is assumed to has moments that exist and are finite.

The model is representative of service organizational units in which each agent is a worker who controls the processing rate of a queue of information-processing tasks, and the overall cycle

Notation	Explanation
$\Omega = (\omega_1, \omega_2, ..., \omega_n)$	Effort-level vector of the *n* workers. A higher value of ω_i represents a higher effort.
$[\Omega_{-i},\omega_0] = (\omega_1,..,\omega_{i-1},\omega_0,\omega_{i+1},...,\omega_n)$	Ω with the i^{th} component replaced by ω_0.
$c(\omega)$	Cost to each agent of exerting effort ω, measured in monetary units.
$u(x,\omega) = U(x - c(\omega))$	Identical utility function of each of the *n* agents in the organization; the argument *x* is monetary compensation.
U_0	Reservation utility of each workers.
$f(\cdot \mid \Omega)$	Density function of the performance measure if workers work at effort levels Ω.
$F^*(s \mid \Omega)$	Laplace transform of $f(\cdot \mid \Omega)$.
$m_i(\Omega) = (-1)^i \lim_{s\to 0} \frac{d^i}{dy^i} F^*(s \mid \Omega)$	i^{th} moment of $f(. \mid \Omega)$
$(\mathbf{a},\mathbf{b}) = (a_1, a_2,...,a_n),(b_1,b_2,...b_n)$	Contracts offered to the *n* agents.
$\pi(y)$	Gross profits (before accounting for compensation) to the principal if observed performance measure is *y*.

Table 1: Summary of notation for basic model

time of the process, which is the net sojourn time in the tandem queuing network (with a density function that is typically not analytically tractable), is potentially the only observable measure of performance.

We assume that the principal offers each individual agent a contract (a_i,b_i), which specifies a payment of $(a_i + b_i y)$ if the observed value of the performance measure is *y*. We are therefore assuming that the principal is restricted to offering linear contracts. This is a simplifying assumption, made for reasons of mathematical tractability. We discuss the justification and implications of this assumption in the concluding section of the paper.

The notation introduced thus far is summarized in Table 1. Given this setup, the principal's problem can be formulated as:

$$
\begin{aligned}
&\max_{a,b} \int_0^{\infty} \{\pi(y) - \sum_{i=1}^{n} (a_i + b_i y)\} f(y \mid \Omega^*) dy \\
&s.t.: \int_0^{\infty} U(a_i + b_i y - c(\omega_i^*)) f(y \mid \Omega^*) dy \geq U_0 \quad \forall i \\
&\omega_i^* \in \arg\max_{\omega \in E} \int_0^{\infty} U(a_i + b_i y - c(\omega)) f(y \mid [\Omega_{-i}^*, \omega]) dy \quad \forall i
\end{aligned}
\tag{2}
$$

The following proposition establishes that for a general class of utility functions and profit functions, the principal's problem has a closed-form reformulation, in terms of the Laplace transform and the moments of the unknown performance measure.

Proposition 1: If U(x) is expressible as a finite sum of negative exponentials, i.e. $U(x) = \alpha_0 + \sum_{i=1}^{m} \alpha_i e^{-\eta_i x}$, $\eta_i > 0$, and $\pi(y)$ *is polynomial in y, i.e.* $\pi(y) = \sum_{i=0}^{l} p_i y^i$, *then* (2) *has the following closed form formulation:*

$$
\begin{aligned}
&\max_{a,b} \sum_{j=1}^{n} [-a_j - b_j m_1(\Omega^*)] + \sum_{i=1}^{l} p_i m_i(\Omega^*) \\
&s.t. \\
&\sum_{i=1}^{m} \alpha_i [e^{-\eta_i \{a_j - c(\omega_j^*)\}} F^*(\eta_i b_j \mid \Omega^*)] \geq U_0 \quad \forall j \\
&\omega_j^* \in \arg\max_{\omega \in E} \sum_{i=1}^{m} \alpha_i [e^{-\eta_i \{a_i - c(\omega)\}} F^*(\eta_i b_j \mid [\Omega_{-j}^*, \omega])] \quad \forall j
\end{aligned}
\tag{3}
$$

The proof of this proposition is presented in Appendix A. This result, which allows the closed-form formulation of the optimal incentive contract when the performance measure is described only by a Laplace transform, has many potential applications in other service settings, as well as in the design of inter-organizational contracts for IT management (for instance, inter-network resource sharing for Internet routing). A commonly used method for approximating a concave function is Prony's method of interpolation by exponentials. One can show (omitted for brevity) that if any increasing, concave function is approximated using this method, then all the exponents are negative, and hence, Proposition 1 can be used to characterize the optimal contract approximately for any increasing, concave *U(.)*.

The next two sections apply this result to characterize a functional organization design, and to a process-oriented organization design.

3. Functional Organization

In a functional organization, the *n* agents are modeled as specialists who perform their tasks sequentially. Each specialist is assumed to have a utility function of the form:

$$u(x,\omega) = -e^{-\eta[x-c(w)]}. \qquad (4)$$

This specifies *U(.)* as the negative-exponential utility function, widely used in principal-agent problems involving moral hazard (for instance, Holmstrom and Milgrom, 1987). Customer jobs arrive at a Poisson rate λ. Each job consists of *n* tasks, and each specialist is assigned exactly one of these tasks (the one that the agent is *functionally specialized* in). When the specialist receives responsibility over a job, it joins the queue of pending work for that specialist. After completing his or her assigned task, the specialist passes job responsibility immediately[3] to the queue of the next agent.

Before a specialist begins to perform a task, he or she must first spend time understanding the job's specifications and customization needs, inferring any special requirements of the relevant customer (perhaps based on past jobs) and understanding the nature of what has already been performed (by reading summary reports and other information from preceding tasks). After the nature of the job has been understood, the specialist performs the task. Finally, she may prepare summary information to be transferred to and read by the succeeding workers.

Our model divides the time taken for these activities into two categories. One of these is the *task processing time* of the agent, which we assume to be exponentially distributed. When the worker works at effort level ω^i, the parameter of this exponential distribution is μ_i. The workers choose from two levels of effort, ω^L and ω^H (which correspond to processing rates of μ_L and μ_H). The other component of the time a job spends with a specialist is the *handoff delay* σ. This consists of the time-consuming, but fairly standard, pre- and post-processing activities. Rather than by the effort level of the worker or the actual work to be performed, the value of σ is influenced primarily by the nature of the custom specifications, or the *degree of customization* demanded on the job, which in turn determines the magnitude of *information that 'flows with the job'.* Higher levels of customization correspond to lengthier and more complex customer specifications, and consequently, more pre-processing time. A widely documented drawback of functional organization (Davenport, 1993, Hammer and Champy, 1993) is that it causes each sequential worker to repeatedly have to incur the same handoff delays as the others. If all tasks in the process were performed by the same worker, then there would be no need for these inter-specialist information transfers.

We assume that σ is deterministic, and independent of the effort level of the agent. This is a simplifying assumption to allow σ to highlight the factor that we believe is its most important determinant – the magnitude of the information transfers between tasks. Making σ variable or a function of effort level is likely to strengthen our results further; this discussed briefly in section 6.

The cycle time of the job at each specialist is modeled as the sojourn time of an *M/G/1* queue. The queues have an arrival rate equal to the rate of arrivals into the system (λ) and a processing

Notation	Explanation
$U(x-c(\omega)) = -e^{-\eta[x-c(w)]}$	Utility function for each of the *n* specialists.
λ	Arrival rate of jobs to the organization
μ_i	Exponential processing rate of a specialist at effort level $\omega^i \in \{\omega^L, \omega^H\}$.
σ	Handoff delay due to 'information that flows with the job', which shifts the processing time distribution by a deterministic amount.
c_i	Cost of effort per task for a specialist at effort level $\omega^i \in \{\omega^L, \omega^H\}$.
$g(y;\mu_i) = g_i(y)$	Server processing-time density at effort level $\omega^I \in \{\omega^L,\omega^H\}$; $g_i(y) = \mu_i e^{-\mu_i(y-\sigma)}$.
$T_s(\mu_i,\lambda)$	Expected sojourn time in an *M/G/1* queue with arrival rate λ and processing time distribution $g(y;\mu_i)$.
q	Indicator variable; $q=0$ corresponds to partial asymmetry, and $q=1$ corresponds to complete asymmetry.

Table 2: Summary of notation introduced in Section 3.

time that is distributed as a *shifted exponential*. If the effort level of a worker is ω^i (yielding a processing rate μ_i) and the handoff delay is σ, then the processing time of each queue is has the density function

$$g(y;\mu_i) = \mu_i e^{-\mu_i(y-\sigma)}. \tag{5}$$

For subsequent notational simplicity, we denote $g(y;\mu_i)$ as $g_i(y)$, and define the values of the cost of low and high effort as $c_j = c(\omega^j)$, $\omega^j \in \{\omega^L, \omega^H\}$. Denote the expected cycle time of this queue with a handoff delay of σ, an effort level of μ_I, and an arrival rate of λ as $T_s(\sigma,\mu_i,\lambda)$.We have assumed for tractability that the arrival process to each of the queues in tandem is Poisson. In an exact model, the inter-departure time of our *M/G/1* queue would be a mixture of shifted exponentials, with variance lower than that of an exponential (we derive this in the Appendix, and discuss some implications of this approximation). Our model may therefore overestimate the cycle time of subsequent stages in the process, relative to an exact queuing model. However, the inter-departure times will not be i.i.d., which precludes exact analysis of this system, though some bounds are available for the overall sojourn time (Seshadri and Pinedo, 1998).

We analyze two forms of information asymmetry. With the first, *complete asymmetry*, the principal can only observe the cycle time of the *entire job*, and each agent can observe only the processing time of his or her own task. With the other, *partial asymmetry*, both the principal, and each agent can observe that individual agent's cycle time[4]. In either case, the value of σ is known to both the principal, and to all the agents, before contracting[5]. Figure 2 contrasts these two information asymmetry situations. Only each agent can see what is inside her box. All the agents and the principal can see what is outside the dotted box. An indicator variable q is used to represent information asymmetry. $q = 0$ represents partial asymmetry, and $q = 1$ represents complete asymmetry. The notation introduced in this section is summarized in Table 2.

Now, the Laplace transform of the processing time of each server i is:

$$G^*[s;\omega^i] = \int_{\sigma}^{\infty} e^{-sx}\mu_i e^{-\mu_i(x-\sigma)}dx = \frac{\mu_i e^{-s\sigma}}{\mu_i + s}, \tag{6}$$

and the activity rate at each server is

$$\rho_i = \lambda\left(\frac{1}{\mu_i} + \sigma\right) = \frac{\lambda}{\mu_i} + \lambda\sigma. \tag{7}$$

We use the Pollaczek-Khinchin transform equation (Kleinrock, 1976), to derive the Laplace transform for the distribution of system time at each server:

$$G_S^*[s;\omega^i] = \frac{s(\mu_i - \lambda - \lambda\mu_i\sigma)}{(s+\mu_i)(s-\lambda)e^{s\sigma} + \lambda\mu_i}. \tag{8}$$

The expected service time at each server is therefore:

$$T_s(\mu_i,\lambda,\sigma) = \lim_{s\to 0} -\frac{\partial G_S^*[s;\omega^i]}{\partial s} = \frac{(1-\sigma\lambda+\sigma\mu_i - \frac{\sigma^2\lambda\mu_i}{2})}{(\mu_i - \lambda - \lambda\mu_i\sigma)}. \tag{9}$$

The principal may offer the agents either activity-based compensation (a flat fee for each job completed), or a linear contract as described in Section 2. In the former case (fixed-fee compensation), the workers all work at effort level μ_L, and are paid a constant wage *a,* which covers their reservation utility:

$$-e^{-\eta(a-c_L)} = U_0 \Rightarrow a = c_L + \frac{1}{\eta} log(\frac{-1}{U_0}). \tag{10}$$

In the latter case (performance-based compensation), one can use Proposition 1 to derive the optimal contract for each agent. Since there are two effort levels, and the agents seek pure-strategy Nash equilibria, the only rationale for a performance-based incentive scheme is if it induces workers to work at ω^H. The agents have identical preferences and identical job sizes; hence, by symmetry, the optimal performance-based contracts offered to workers will be identical, inducing identical effort levels. Using Proposition 1, the contract *(a,b)* chosen must satisfy:

$$-e^{-\eta[a-c_H]} F^*[\eta b|(\omega^H,\omega^H,...,\omega^H)] \geq -e^{-\eta[a-c_L]} F^*[\eta b|(\omega^H,\omega^H,...,\omega^L)], \tag{11}$$

and

$$-e^{-\eta[a-c_H]} F^*[\eta b|(\omega^H,\omega^H,...,\omega^H)] = U_0. \tag{12}$$

Now, under partial asymmetry,

$$F^*[\eta b|(\omega^H,\omega^H,\omega^H,..,\omega^H)] = G_S^*[\eta b;\omega^H]. \tag{13}$$

Therefore, solving (12) for *a* yields:

$$a = c_H + \frac{1}{\eta} log(\frac{-1}{U_0}) + \frac{1}{\eta} log(G_S^*[\eta b;\omega^H]) \tag{14}$$

Also, under partial asymmetry, (11) reduces to:

$$-e^{-\eta[a-c_H]}\, G_S^*[\eta b;\omega^H] = -e^{-\eta[a-c_L]}\, G_S^*[\eta b;\omega^L], \tag{15}$$

which yields the following expression for *b:*

$$\frac{(\mu_L-\lambda-\lambda\mu_L\sigma)e^{c_L\eta}}{e^{\eta b\sigma}(\eta b+\mu_L)(\eta b-\lambda)+\lambda\mu_L}=\frac{(\mu_H-\lambda-\lambda\mu_H\sigma)e^{c_H\eta}}{e^{\eta b\sigma}(\eta b+\mu_H)(\eta b-\lambda)+\lambda\mu_H}. \tag{16}$$

On the other hand, under complete asymmetry,

$$F^*[\eta b,(\omega_1,\omega_2,..,\omega_n)]=G_S^*[\eta b;\omega_1].G_S^*[\eta b;\omega_2]....G_S^*[\eta b;\omega_n], \tag{17}$$

because the system described is a product-form queuing network with n servers in tandem. Equation (11) now reduces to:

$$-e^{-\eta[a-c_H]}(G_S^*[\eta b;\omega^H])^n=-e^{-\eta[a-c_L]}(G_S^*[\eta b;\omega^H])^{n-1}G_S^*[\eta b;\omega^L], \tag{18}$$

which solves to the same expression for b as in equation (16). Furthermore, equation (12) yields:

$$a=c_H+\frac{1}{\eta}log(\frac{-1}{U_0})+\frac{1}{\eta}log(G_S^*[\eta b;\omega^H])^n \tag{19}$$

Therefore, depending on whether there is partial (q=0) or complete (q=1) asymmetry of information, the optimal contract is specified by:

$$a=c_H+\frac{1}{\eta}log\frac{-1}{U_0}+\frac{n^q}{\eta}log[\frac{\eta b(\mu_H-\lambda-\lambda\mu_H\sigma)}{e^{\eta b\sigma}(\eta b+\mu_H)(\eta b-\lambda)+\lambda\mu_H}], \tag{20}$$

and

$$\frac{(\mu_L-\lambda-\lambda\mu_L\sigma)e^{c_L\eta}}{e^{\eta b\sigma}(\eta b+\mu_L)(\eta b-\lambda)+\lambda\mu_L}=\frac{(\mu_H-\lambda-\lambda\mu_H\sigma)e^{c_H\eta}}{e^{\eta b\sigma}(\eta b+\mu_H)(\eta b-\lambda)+\lambda\mu_H}. \tag{21}$$

Note that the factor b has the same value, irrespective of the information asymmetry. This does not mean that variable compensation is the same in both situations. Under complete asymmetry, the performance measure is the cycle time of the entire job; hence, the magnitude of the realized variable compensation (b multiplied by the observed cycle time of the entire job) is far higher than in the case of partial asymmetry (b multiplied by the observed cycle time of a single task).

This component increases approximately linearly with n, which is intuitively appealing, because as n increases the problem of free riding becomes increasingly more difficult[6].

The principal chooses the contract (incentive or fixed-pay) that minimizes total costs. The costs of the principal are the sum of compensation costs and cycle-time delay costs, for which we assume a rate c_D. It is straightforward to show that the principal's costs reduce to:

$$min[\{ nc_D T_s(\mu_L,\lambda,\sigma) + n[c_L + \tfrac{1}{\eta}log(\tfrac{-1}{U_0})]\}, \\ \{nc_D T_s(\mu_H,\lambda,\sigma) + n^{(q+1)} bT_s(\mu_H,\lambda,\sigma) + na]\}], \tag{22}$$

where a and b are as specified in equations (20) and (21). The first expression in equation (22) represents costs when all agents work at the lower effort level, and the second one represents the total costs to the principal when all agents work at the higher effort level. Recall that $q=0$ corresponds to partial asymmetry, and $q=1$ to complete asymmetry, and that the variable compensation per worker is proportional to expected task time for $q=0$, and to the sum of n expected task times for $q=1$.

Under complete asymmetry, equations (11) and (12) are intended to specify the exact conditions under which the contract (a,b) offered to each agent makes $(\omega^H,\omega^H,...,\omega^H)$ a symmetric *Nash* equilibrium of the game in which the n agents choose effort levels. However, as it turns out, these conditions are sufficient to ensure that if (a,b) is offered to each agent, then ω^H is in fact the *dominant strategy* for each agent. This observation is proved in Appendix A, and ensures the implementability of the incentive contract (a,b) in the case of complete asymmetry.

Some further observations based on equation (22):

(1) At a particular level of agent effort, costs due to delay are independent of the information asymmetry, and are linear in the number of workers.

(2) If it is optimal to use a performance-based compensation scheme, then

(a) Under partial asymmetry, agency costs are *linear* in the number of workers.

(b) Under complete asymmetry, agency costs are *quadratic* in the number of workers.

If the expression for *a* from equation (20) is substituted into equation (22), one gets an expression for agency costs of the form

$$n^{(q+1)}\left(\frac{1}{\eta}log[\frac{\eta b(\mu_H-\lambda-\lambda\mu_H\sigma)}{e^{\eta b\sigma}(\eta b+\mu_H)(\eta b-\lambda)+\lambda\mu_H}]+bT_s(\mu_H,\lambda,\sigma)\right)+n^q\left(c_H+\frac{1}{\eta}log\frac{-1}{U_0}\right) \quad (23)$$

Clearly, *b* is always negative, since it is a penalty imposed on the worker per unit of delay. However, can be verified (a short proof is provided in Appendix A) that the term multiplying n^{q+1} is positive (otherwise, the agents would not get their minimum reservation utility from the contract). This is the driver of observation (2) above. The implications are clear – the impact of agency costs becomes increasingly more important in a functional organization with complete information asymmetry. This convexity of agency costs in the number of workers partially explains why long sequential systems are inefficient, even when specialization gains are significant. In addition, there is significant value from an information system that moves a firm from complete to partial asymmetry. Workflow automation software is one example of such a system. Apart from reduced reconciliation costs and better co-ordination, the accompanying reduction in agency costs is another potentially significant driver of value from these systems.

Further insights based on numerical optimization are presented in §5.

4. Process-oriented organization

Under a process-oriented organizational architecture, the *n* agents are modeled as case managers who work in parallel. When a job arrives in the organization, it is sent to one of the *n* case managers, who performs all the *n* tasks of the job. This model depicts a common work system in process-oriented corporations (a well-known example being that of IBM Credit from Hammer and Champy, 1993). While this design eliminates inter-agent handoff delays, it presents the new performance issues. Due to a loss of specialization and barriers to accessing specialized information systems, at any given effort level, the processing rate per task of the case managers will not be as high as it was in the case of functional specialists. Specifically, in our model, the average rate of completing any specific task reduces by a factor α, relative to the functional specialist. In addition, their skills are not broad enough for them to handle all the jobs that come to them; they need the intervention of an expert for a fraction β of jobs, which represent the exceptions (Hammer and Champy, 1993).

Typically, issues are alleviated by using two kinds of *information systems*: *speed-enhancing systems* (productivity tools and decentralized information access systems like regular client-server systems or intranets[7]) and *skill-expanding systems* (expert systems and knowledge management systems). The level of technology in each of these types of systems is denoted θ_1 and θ_2 respectively. The effects of these information systems on the task-processing rates are illustrated in Figure 3.

We model these two effects – loss in productivity and skill reduction – separately. α, which represents the drop in processing rate per task when tasks are consolidated, depends on α_0 (the reduction in productivity when there is no new IS support, representative of the *knowledge intensity* of the tasks) and *t*, the rate of returns from information technology. Its explicit functional form is $\alpha = 1 - \alpha_0 e^{-t\theta_1^2}$.

β, which represents the exception rate of the process, i.e. the percentage of arrivals diverted to an expert due to inadequate worker knowledge about all tasks, depends on β_0 (the exception rate when there is no IS support) and θ_2. Its functional form is $\beta(\theta_2) = \beta_0 e^{-t\theta_2^2}$. In both these cases,

t is an absolute measures of the *technology responsiveness* of the organization or, in other words, the returns per dollar due to increased technology levels.

Since the workers are processing n tasks per job, their processing rate is also factored down by

Notation	Explanation
α	Reduction in the agent's processing rate due to loss of specialization.
β	Percentage of jobs rerouted to an expert.
α_0	Reduction in the agent's processing rate in the absence of supporting IS, and a measure of knowledge intensity.
β_0	Exception rate in the absence of supporting IS.
t	A measure of the rate of returns to investment in information technology.
c_E	Additional cost per unit time of the jobs rerouted to the expert.
μ_E	Processing rate of the expert.
θ_1,θ_2	Investment in information systems for enhancing speed and expanding skills.
$U_n = -(-U_0)^n$	Reservation utility corresponding to n tasks

Table 3: Summary of notation introduced in Section 4.

$1/n$. Hence, at an effort level ω^i, the processing rate per job of the generalist will be $\alpha\mu_i/n$. In addition, the workers need to examine job requirements, specific information pertaining to a particular customer as well. However, they need to do this *exactly once*, and this results in a single shift of σ. The processing time distribution at each server is therefore a shifted exponential distribution with parameters $\alpha\mu_i/n$ and σ.

There is only one information asymmetry situation: partial asymmetry. The principal can observe the cycle time of each agent and therefore contracts with each of them individually. The cost per job to the agent is the sum of the costs of the n individual tasks[8] and is therefore nc_i at an effort level ω^i. To facilitate comparison between the systems, the expert is assumed external to the agency conflict[9]. A clean way to model this scenario is to assume that the principal does this work; it is done at a processing rate μ_E. Since the information relating to the specific job instance must be transferred to the expert every time a job is sent, the expert is delayed by σ per job, and has a resulting processing time that has a shifted exponential distribution with parameters σ and μ_E. The personal cost (due to spending time that could have been spent on other value-adding work) is measured at a rate c_E, over and above the delay cost borne by the firm. The additional notation introduced in Section 4 is summarized in Table 3.

Now, since the processing rate at each case manager is $\alpha\mu_i/n$, and the arrival rate to each is λ/n, using the Pollaczek-Khinchin transform equation as in Section 3 yields the Laplace transform of the total sojourn time:

$$G_P^*[s;\omega^i] = \frac{s(n\alpha\mu_i - n\lambda - \lambda\alpha\mu_i\sigma)}{(ns+\alpha\mu_i)(ns-\lambda)e^{s\sigma} + \lambda\mu_i}. \tag{24}$$

Correspondingly, the incentive compatibility and rationality constraints that the optimal contract *(a,b)* must satisfy are:

$$-e^{-\eta[a-c_H]}\, G_P^*[\eta b;\omega^H] \geq -e^{-\eta[a-c_L]}\, G_P^*[\eta b;\omega^L], \tag{25}$$

and

$$-e^{-\eta[a-c_H]}\, G_P^*[\eta b;\omega^H] = U_n \tag{26}$$

Equations (25) and (26) are similar in form to equations (11) and (12), under partial asymmetry, in section 3. Proceeding similarly, one characterizes the optimal performance-based contract:

$$a = nc_H + \frac{1}{\eta}\left(\log\frac{-1}{U_n} + \log\frac{\eta b(n\alpha\mu_H - n\lambda - \lambda\alpha\mu_H\sigma)}{e^{\eta b\sigma}(n\eta b + \alpha\mu_H)(n\eta b - \lambda) + \lambda\mu_H} \right), \tag{27}$$

$$\frac{(n\alpha\mu_L - n\lambda - \lambda\alpha\mu_L\sigma)e^{nc_L\eta}}{e^{\eta b\sigma}(n\eta b + \alpha\mu_L)(n\eta b - \lambda) + \lambda\mu_L} = \frac{(n\alpha\mu_H - n\lambda - \lambda\alpha\mu_H\sigma)e^{c_H\eta}}{e^{\eta b\sigma}(n\eta b + \alpha\mu_H)(n\eta b - \lambda) + \lambda\mu_H}. \tag{28}$$

Furthermore, adding up the expected delay and compensation costs, and accounting for the additional exception-related costs, yields the total cost for the principal as:

$$min\left\{\begin{aligned}&\left(c_D[T_s(\frac{\alpha\mu_L}{n},\frac{\lambda}{n},\sigma)+\beta T_s(\mu_E,\beta\lambda,\sigma)]+c_E\beta T_s(\mu_E,\beta\lambda,\sigma)+nc_L+\frac{1}{\eta}log\frac{-1}{U_n}\right),\\&\left(c_D[T_s(\frac{\alpha\mu_H}{n},\frac{\lambda}{n},\sigma)+\beta T_s(\mu_E,\beta\lambda,\sigma)]+c_E\beta T_s(\mu_E,\beta\lambda,\sigma)+a+bT_s(\frac{\alpha\mu_H}{n},\frac{\lambda}{n},\sigma)\right)\end{aligned}\right\}. \quad (29)$$

From equations (27) and (29), it is clear that in the process-oriented organizations, agency costs are linear in the number of workers – this suggests that their variation will be qualitatively similar to those under partial asymmetry in the functional organization. The nature of delay costs is not immediately clear, however, since it seems likely that there will be a trade-off between the reduced handoffs, and the increase in delay due to α and β. However, it is clear that a decrease in α or an increase in β is likely to result in a convex increase in costs. These issues are explored further in section 5.

5. Analysis and trade-offs

We explore the analytical results from sections 3 and 4 further via a sensitivity analysis of these results. Since there are limited insights one can get from standard analytical comparative statics (owing to the complexity of the contract and queuing delay expressions), we have conducted numerical optimization using these closed-form expressions as a starting point.

We start with an analysis of cycle time in isolation, to highlight the trade-offs between the handoff delays and loss of specialization across functional and process-oriented organizations, alluded to at the end of section 4. Figures 3 and 4 show some results from this analysis. For a fixed set of values of *n,* λ and μ, figures 3(a) and 3(b) plot the total expected cycle time per job for functional and process-oriented work systems respectively, as the magnitude of handoff delay σ and loss in specialization α vary. In general, we vary one of the two parameters α or β, and tie the second to the first (so in this case, we assume that β=1-α); we also assume that $c_E = 0$, and normalize μ_E to μ/n. The difference in expected cycle time across the two designs (that is, the expected cycle time for a functional work system subtracted from the expected cycle time for a process-oriented system) is plotted in figure 3(c). When this difference is positive, the functional work system is superior (i.e., has a lower cycle time), and when it is negative, the process-oriented work system is superior – the (σ,α) pairs where it is zero plot the curve of indifference between the performance of the two work systems.

Figure 3(d) projects these points onto the σ-α plane. The figure shows that, as one would expect, there is a trade-off between reduced handoff delays and specialization, leading to a range of values where each work system is superior. Moving forward, most of our charts are similar to Figure 3(d), which compactly captures most of the relevant information (the relative benefits of each design choice) from the more detailed 3-D charts. In general, the magnitude of superiority of the chosen design (in this case, for instance, functional or process-oriented work system) *increases* monotonically as one moves away from the curve of indifference.

The cross-sensitivity of changes in the parameters σ and α and changes in the number of tasks n in the job, and the total workload λ/μ., is explored in Figure 4. Each of these charts is similar to that in Figure 3(d). As in Figure 3, the objective here is to compare the relative cycle times of the two work systems. As expected, process-oriented organization is favored at higher values of σ as well as higher values of α. A couple of other interesting trends emerge from this analysis. As the number of tasks increases, the range of α values over which *process-oriented* work systems are superior begins to widen, though at a diminishing rate, and as n becomes large, the range begins to shrink, as shown in Figures 4(a) and 4(b). This is because as the number of tasks increases, the performance effects of the exceptions (the jobs diverted to the expert) begin to become increasingly significant. This effect is substantially higher when the overall workload is higher, as shown in figure 4(b). While the set of parameter values over which each type of system is preferable becomes far more sensitive to the number of tasks, increasing workload makes the cycle-time performance of a *process-oriented* work system superior overall. This is illustrated in Figures 4(b) and 4(d).

Next, Figure 5 isolates the nature and magnitude of agency costs across the two organization designs. For a sample set of parameter values, with a relatively high handoff delay, figures 5(a) and 5(b) chart the variation in the terms of the incentive contract *(a,b)*, as the number of tasks varies. The optimal contract is unaffected under functional organization (with partial information asymmetry). As expected, the fixed component of the contract a increases under process-oriented organization. However, as depicted in Figure 5(a), as n increases, the total payout from this component of incentive pay is progressively lower than the corresponding amount under functional organization. Recall that each of the n agents gets paid a in the latter case.

A different trend is observed for the variable rate b. As the number of tasks increases, the value of b under process-oriented organization increases (becomes less negative), to compensate for the fact that the expected value of the performance measure (total cycle time) is increasing. However, this increase is at a diminishing rate, suggesting that the total penalty that the agent incurs under process-oriented organization (which goes back to the principal) is likely to be higher. This is not unambiguously clear, since the total expected cycle time under functional organization is also higher. Figure 5(c) shows that the total incentive-related costs per job are lower under functional organization for a lower number of tasks, but rise more rapidly, therefore making process-oriented organization preferable for higher values if n.

A similar analysis for varying workload is illustrated in Figure 5(d), for a moderate number of tasks. Only the incentive-related costs per job are shown – they increase with workload under both organization designs, but more rapidly under functional organization. As workload increases, process-oriented organization is clearly increasingly superior along this cost-related measure.

What emerges from our analysis of queuing costs and agency costs in isolation is that while responses to changes in handoffs, loss from specialization and so on are as one would expect, that increases in workload, as well as increases in the number of tasks per job, favor process-oriented organization along both the queuing and agency cost dimensions. To examine *simultaneous* changes in the impact of information technology and the other organizational parameters, we continue to plot two-dimensional projections of the parameter space. As these parameters vary, the regions in which different combinations of information systems, compensation schemes, and work design are optimal are examined.

Without loss of generality, and for notational simplicity, we normalize the value of μ_L to 1 and μ_H to h. We also restrict the value of σ to lie in [0,1] and interpret this value as the handoff delay measured as a fraction of the expected processing time, $1/\mu_L$. While σ may actually be greater than 1, that interval is uninteresting, as it corresponds to situations with unduly large handoff delays, resulting in a clear case for process-oriented organization. λ is restricted to the interval $\left(0, \frac{1}{(1+\sigma)}\right)$ to ensure that queuing delays are finite under both the absence and the presence of incentive compensation. We then normalize c_L to 1 (this is without loss in generality, since it simply represents a change in currency units) and assume that the cost of effort is linear in the effort; i.e., $c_H = h$, where h can be interpreted as a measure of the difficulty of the task. We also normalize U_0 to -1, which[10] implies that $log\frac{-1}{U_0}=log\frac{-1}{U_n}=0$. Finally, we normalize η to 1. A change in the value of η changes the absolute value of the optimal contract, but qualitatively, the results remain the same.

We can now set a level of information technology (θ_1,θ_2) for the process-oriented organization, and solve for the optimal value of *b* for the functional and process-oriented organizations, given a certain set of parameter values. Following this, the values of *a* and the cost functions for each work system are derivable. We compare these figures within each work system to derive the optimal level of θ_1 and θ_2, and then compare these optimal cost figures across organization designs. One of the four choices — (functional/process oriented, fixed/incentive compensation) is chosen. Varying the parameters two at a time enables us to plot *transition charts*, that divide the 2X2 parameter space into regions where respective designs are optimal (this is done by seeing where they become cost-equivalent, as was done in Figures 3(d) and 4, for just cycle time. These successive points of equivalence define a curve of transtion from one organizational design to another).

A few of these transition charts are shown in Figures 6 and 7; they focus on the five parameters that we found to be the most significant determinants of organization design choice. Figures 6(a) and 6(b) illustrate the effects of level of customization (σ), the work intensity (λ), and the technology responsiveness (*t*) on the optimal choice of organization. As shown, an increase in either the level of *customization* or the *work intensity* causes *process-oriented* organization to become optimal. This effect is more pronounced when the technology responsiveness *t* increases, as shown by a much larger area of optimality for the process-oriented organization in Figure 6(b). Figures 6(c) and 6(d) plot similar results while varying the knowledge intensity (α and

$\beta)$[11], the work intensity (λ), and the returns to I.T. (t). Note that an increase in knowledge intensity (a decrease in α) results in a *functional* organization being more desirable. Again, this effect is compounded by the increase in returns to I.T.

A very interesting result here is the relationship between *information technology* and *performance-based incentives.* As illustrated by Figure 6(a) and 6(c), when returns from I.T. are low, the optimal process-oriented organization often features incentive pay. The corresponding values of θ_1 and θ_2 (derived as described earlier) are low, corresponding to a low level of I.T. investment. These values of θ_1 and θ_2 increase significantly when the returns from I.T. are high – there is a higher investment level in technology, which is not surprising by itself – but this is accompanied by the elimination of incentive pay in process-oriented organizations, as shown in Figures 6(b) and 6(d). Information technology and incentives therefore appear to be **substitutes** in a process-based organization.

Figure 7 strengthens the conclusions that higher customization and activity rates favor process-oriented organization, while higher knowledge intensity favors a functional organization (note that this trend is illustrated in all four charts). In addition, at the same level of work intensity and customization/knowledge intensity, an increase in the number of tasks results in the process-oriented organization being optimal in a larger number of cases. The figures illustrate the case of partial asymmetry – complete asymmetry would strengthen this result further, as observed in Section 3. This is consistent with the isolated results we had in Figures 4 and 5.

The analysis yielded another significant result, which is also evident from a closer examination of Figure 6 and 7. Both *incentives* and *information technology complement* the move to a process-based organization. The former can be seen clearly in figures 6(a), 6(c), and 7(a)-(d), where a majority of the area where process-based organization is favorable also features incentives as a part of the optimum. The latter follows from our result that when incentives don't appear as a feature of the optimal organizational design, (such as in 6(b) and 6(d)), information technology levels are high. As mentioned earlier, these two are almost mutually exclusive, suggesting a strong substitutability between technology investments and performance-based incentives.

It is important to note that the desirability of fixed-fee compensation is accentuated in our model by the presence of just two candidate agent effort levels ω^H and ω^L. If there were multiple possible effort levels, fixed-fee compensation would be less frequently optimal. However, this does not qualitatively change the discussion above. For instance, at high levels of IT, rather than fixed-fee compensation, the model would be prescribing either fixed pay or lower powered incentives, and more leveraged incentives at lower IT levels. The effect of IT and incentives being substitutes (and of the other complementary relationships) would still be present.

6. Managerial Insights and Conclusions

Perhaps the most significant result of our analysis is that *typical changes in work organization, incentive compensation, and information systems do not always appear to be complementary.* In fact, in a very large number of cases, while the adoption of process-oriented organization and information technology appear to complement one another, as do the adoption of process-oriented organization and performance-based incentives, information systems and performance-based incentives appear to be substitutable drivers of process performance improvement. This contrasts with recent results (Bresnahan, Brynjolfsson and Hitt, 2002); and with most prescriptive essays on information technology-enabled design in service organizations (e.g. Hammer and Champy, 1993), which consistently recommend simultaneously adopting a process orientation, introducing output-based pay and using advanced information technology to enhance skills and productivity. As our results indicate, this approach is frequently not optimal, and choosing on two out of the three is typically the best solution. While this does not constitute a formal proof of the complementarity or substitutability of different organizational design initiatives, it is a new testable hypothesis that our modeling results strongly support.

Although this result seems very counter-intuitive, we can explain why it arises. There is a fixed cost of sorts to switching to performance-based incentives; it is not a continuous variable cost (you either have them, or you don't). If the returns from information technology are low, it is optimal to bear this cost, as its benefits are more cost-effective when compared to introducing advanced information technology. As the returns to technology increase, this option becomes relatively less favorable, until a breakpoint is reached, where it becomes sub-optimal to bear this cost. Now, given that one is not introducing incentives, the desirability of technology-based performance enhancers increases significantly; hence the optimal process design is characterized by high levels of information technology.

An aspect of information technology that we have not captured explicitly in our models is its use in management control. Often, IT is used to aid in the more precise measurement of employee performance (for instance, in call centers). By making more targeted performance-based incentives feasible, this kind of measurement technology may increase the desirability of incentive compensation. Our model is of spend-enhancing and skill expanding systems – the systems that commonly accompanied process-oriented work systems, to directly aid in the performance of work, rather than in its measurement.

Another salient result is our demonstration of the *optimality* of *functional organization* in a number of situations. The reengineering revolution of the early 1990's has resulted in the almost absolute rejection of the value of functional specialization. Yet it can be ideal in a number of situations: specifically, when work is knowledge-intensive, when jobs have few tasks, and when work intensity is naturally low. Following the widespread move to process-based organization, many companies are currently wrestling with the problem of managing their knowledge and are consequently implementing expensive knowledge management information systems. A functional organization provides a natural environment for managing functional knowledge – this is an insight that could help many of these companies in their efforts in this direction.

A summary of further managerial insights from our models is presented below:

- A process-oriented organization is desirable when the typical job in the firm has a large number of tasks, when the firm's customers demand increased customization, and when there is significant information asymmetry between workers and management about the extent of each worker's contribution to observed job performance.
- A functional organization is desirable when jobs are knowledge-intensive, when there are low returns from information technology, when seasonal variation requires low activity rates, and when the typical process in the firm has few sub-parts.
- Performance based incentives are optimal in a process-oriented organization when jobs have a large number of tasks, and the constituent tasks require lower specialization. They tend to be sub-optimal when returns from technology are higher. Also, they tend to complement process-oriented work redesign.
- In functional organizations, technology such as workflow systems, which track where a job is at any given time has benefits beyond job control and co-ordination – it also significantly increases the cost-effectiveness of performance-based incentive compensation.

We are working on relaxing our model's assumption of linear compensation. A number of earlier papers in the principal-agent literature have discussed how compensation contracts are often linear in practice. As a consequence, some authors (for instance, Ross, 1973, Holmstrom and Milgrom, 1987, McAfee and McMillan, 1987) have established conditions under which linear contracts are optimal in the standard principal-agent problem, while a number of others have restricted their solution sets to linear contracts (for instance, Stiglitz, 1974, Mathewson and Winter, 1985, Gallini and Lutz, 1992). Our model falls in the latter category. Relaxing the assumption of risk-aversion among agents is likely to make the analysis of general nonlinear contracts analytically feasible, and we continue to explore this problem.

Relaxing the assumption of effort-independent handoff delays is unlikely to change our results significantly. If the handoff delays are dependent on effort, then the benefits of inducing higher effort levels are likely to be relatively higher in a functional organization (because the delay is incurred *n* times rather than just once) than in a process-oriented organization. Therefore, one is more likely to see incentives in a functional (rather than process-oriented) organization – which would highlight the substitutable nature of IT and incentives in complementing process redesign. Making them variable is likely to increase the fraction of cases in which process oriented organization is favored. Our current work aims to add to and generalize the set of organizational settings analyzed using the model in Section 3. This paper provides a new step towards a comprehensive theory of IT-enabled organizations. We hope its results will be used by practitioners, and built on by future researchers.

A. Appendix

Proof of Proposition 1: The objective function in (2) is

$$\int_0^\infty \{\pi(y) - \sum_{i=1}^{n}(a_i + b_i y)\} f(y \mid \Omega^*) dy \tag{30}$$

By definition, $m_i(\Omega^*) = \int_0^\infty y^i f(y|\Omega^*)dy$, and $\int_0^\infty f(y|\Omega)dy = 1$. Therefore, the above expression reduces to the objective function of (3).

The first set of constraints of (2), referred to as constraint set 1, are:

$$\int_0^\infty U(a_j + b_j y - c(\omega_j^*))f(y|\Omega^*)dy \geq U_0 \quad \forall j, \tag{31}$$

which is equivalent to:

$$\int_0^\infty \{\alpha_0 + \sum_{i=1}^m \alpha_i e^{-\eta_i[a_j + b_j y - c(\omega_j^*)]}\} f(y|\Omega^*)dy \geq U_0 \quad \forall j, \tag{32}$$

or:

$$[\alpha_0 + \sum_{i=1}^m \int_0^\infty \alpha_i e^{-\eta_i[a_j - c(\omega_j^*)]} f(y|\Omega^*)dy + \sum_{i=1}^m \int_0^\infty \alpha_i e^{-b_j \eta_i y} f(y|\Omega^*)dy] \geq U_0 \quad \forall j. \tag{33}$$

From the definition of the Laplace transform,

$$\int_0^\infty \alpha_i e^{b_j \beta_i y} f(y|\Omega^*)dy = \alpha_i F^*(-b_j \beta_i | \Omega^*). \tag{34}$$

Equation (34) implies that constraint set 1 has reduced to:

$$[\alpha_0 + \sum_{i=1}^m \alpha_i e^{\beta_i[a_j - c(\omega_j^*)]} + \sum_{i=1}^m \alpha_i F^*(-b_j \beta_i | \Omega^*)] \geq U_0 \quad \forall j, \tag{35}$$

which is the first set of constraints in (3).

Finally, the second set of constraints of (2), referred to as constraint set 2, are:

$$\omega_j^* \in \underset{\omega \in E}{arg\,max} \int_0^\infty U(a_j + b_j y - c(\omega))f(y|[\Omega_{-j}^*, \omega])dy \quad \forall j, \tag{36}$$

or, for all *j*:

$$\omega_j^* \in \underset{\omega\in E}{arg\,max}[\alpha_0 + \sum_{i=1}^{m}\int_0^{\infty}\alpha_i e^{-\eta_i[a_j - c(\omega)]} f(y\,|\,[\Omega_{-j}^*,\omega])dy + \sum_{i=1}^{m}\int_0^{\infty}\alpha_i e^{-b_j\eta_i y} f(y\,|\,[\Omega_{-j}^*,\omega])dy], \quad (37)$$

which easily reduces to:

$$\omega_j^* \in \underset{\omega\in E}{arg\,max}\sum_{i=1}^{m}\alpha_i[e^{-\eta_i\{a_i - c(\omega)\}}F^*(\eta_i b_j\,|\,[\Omega_{-j}^*,\omega])] \quad \forall j. \quad (38)$$

Therefore, the objective functions and all constraints of (2) and (3) are identical. This completes the proof.

Departure process of M/G/1: The queue in question has a Poisson arrival rate of λ, and a shifted exponential processing time distribution with parameters μ and σ. For simplicity, we drop the subscript on μ. From equation (6), we know that the Laplace transform of the processing time distribution is

$$G^*(s) = \frac{\mu e^{-s\sigma}}{\mu + s}. \quad (39)$$

Adapting Gross and Harris (1985), pp 283-284, the Laplace transform of the distribution of inter-departure times is:

$$H^*(s) = \rho G^*(s) + (1-\rho)G^*(s)\frac{\lambda}{\lambda + s} \quad (40)$$

which reduces to:

$$H^*(s) = \frac{\lambda e^{-s\sigma}}{\lambda + s} + \frac{\lambda\mu\sigma s e^{-s\sigma}}{(\lambda + s)(\mu + s)}. \quad (41)$$

Inverting the transform yields the density function of the inter-departure time, which is a mixture of shifted exponentials:

$$h(t) = \lambda e^{-\lambda(t-\sigma)} + \frac{\mu\lambda\sigma}{\mu - \lambda}\left(\mu e^{-\mu(t-\sigma)} - \lambda e^{-\lambda(t-\sigma)}\right) \quad (42)$$

The restriction that activity rate $\rho<1$ ensures that $\mu-\lambda > \mu\lambda\sigma$ (see equation (7)), which means that the fraction in the second term of the RHS of (42) is always less than 1. An intuitive interpretation of this distribution is as follows: when $\sigma = 0$, the departure process is Poisson; as σ increases, the inter-departure times become a shifted exponential with parameter λ, along with a fraction of the density replaced by a shifted exponential with parameter μ.

Comparing this departure process to a Poisson arrival process with rate λ: the mean inter-departure time is $1/\lambda$, the same as that of the inter λ. However, the variance of the inter-departure time is $\frac{2}{\lambda^2}-\frac{\sigma(2+\mu\sigma)}{\mu}$, which is less than $\frac{2}{\lambda^2}$, that of the exponential distribution.

Dominant strategies for inter-agent game: $(\omega^H,\omega^H, \ldots,\omega^H)$ is a symmetric Nash equilibrium for the inter-agent game if equation (11) is satisfied:

$$-e^{-\eta[a-c_H]}F^*[\eta b|(\omega^H,\omega^H,\ldots,\omega^H)] \geq -e^{-\eta[a-c_L]}F^*[\eta b|(\omega^H,\omega^H,\ldots,\omega^L)]\,. \tag{43}$$

This reduces to the indifference equation (18):

$$-e^{-\eta[a-c_H]}(G_S^*[\eta b;\omega^H])^n = -e^{-\eta[a-c_L]}(G_S^*[\eta b;\omega^H])^{n-1}\,G_S^*[\eta b;\omega^L]\,, \tag{44}$$

which is equivalent to:

$$-e^{-\eta[a-c_H]}(G_S^*[\eta b;\omega^H]) = -e^{-\eta[a-c_L]}\,G_S^*[\eta b;\omega^L] \tag{45}$$

Now, for ω^H to be a dominant strategy, it should be preferable to ω^L irrespective of the actions of the other *n-1* agents. That is:

$$-e^{-\eta[a-c_H]}F^*[\eta b|(\omega_1,\omega_2,\ldots,\omega_{n-1},\omega^H)] \geq -e^{-\eta[a-c_L]}F^*[\eta b|(\omega_1,\omega_2,\ldots,\omega_{n-1},\omega^L)]\,, \tag{46}$$

for any $(\omega_1,\omega_2, \ldots,\omega_{n-1})$, and choosing the n^{th} agent as the candidate agent, without loss of generality. However, equation (46) is equivalent to:

$$-e^{-\eta[a-c_H]}.G_S^*[\eta b;\omega^H].\prod_{i=1}^{n-1} G_S^*[\eta b;\omega_i] \geq -e^{-\eta[a-c_L]}.G_S^*[\eta b;\omega^L].\prod_{i=1}^{n-1} G_S^*[\eta b;\omega_i], \qquad (47)$$

or

$$-e^{-\eta[a-c_H]}(G_S^*[\eta b;\omega^H]) \geq -e^{-\eta[a-c_L]}\,G_S^*[\eta b;\omega^L] \qquad (48)$$

which is implied by the Nash condition in equation (45). Furthermore, equation (45) implies that

$$e^{\eta c_H}(G_S^*[\eta b;\omega^H]) = e^{\eta c_L}\,G_S^*[\eta b;\omega^L], \qquad (49)$$

which implies that

$$(G_S^*[\eta b;\omega^H]) < G_S^*[\eta b;\omega^L] \qquad (50)$$

for any values of *b* which is part of an optimal contract. Now, the participation constraint in equation (12):

$$-e^{-\eta[a-c_H]}F^*[\eta b|(\omega^H,\omega^H,\ldots,\omega^H)] = U_0 \qquad (51)$$

is satisfied by *(a,b)*, that is:

$$-e^{-\eta[a-c_H]}(G_S^*[\eta b;\omega^H])^n = U_0 \qquad (52)$$

Equations (50) and (52) ensure that the candidate agent will also participate under *(a,b)*, even if the other agents deviate. This ensures that the optimal contract is implementable.

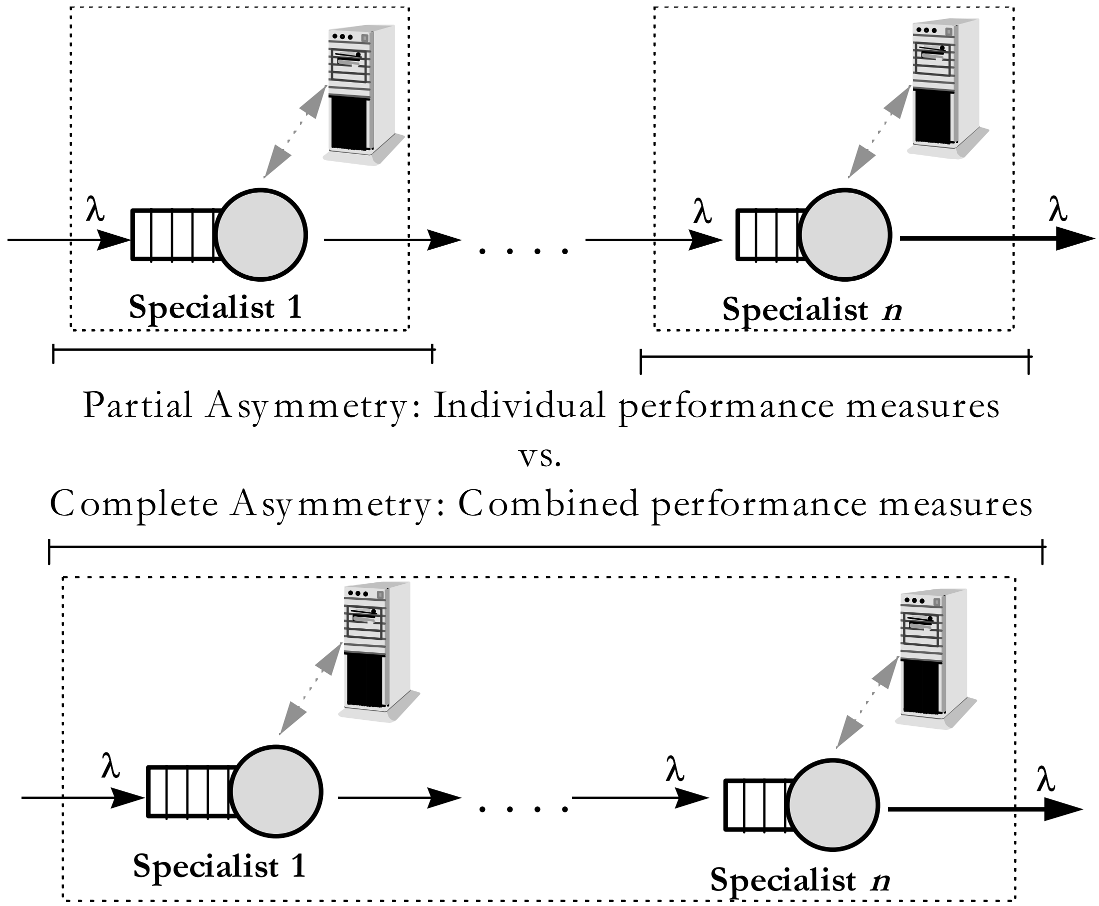


Figure 1: Work design and information asymmetry in functional/process-oriented organizations

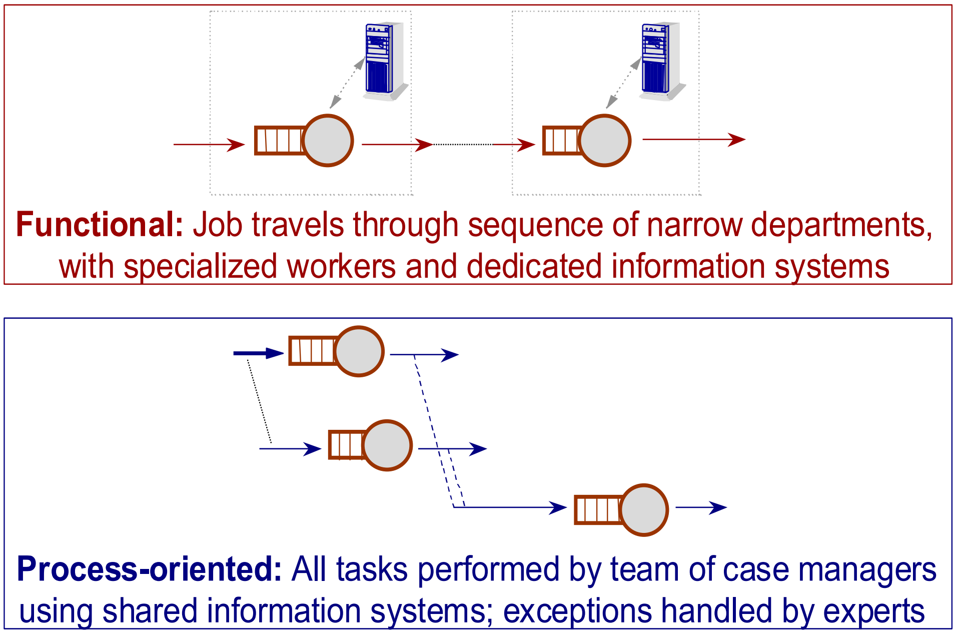


Figure 2(a): Functional vs. process oriented organization

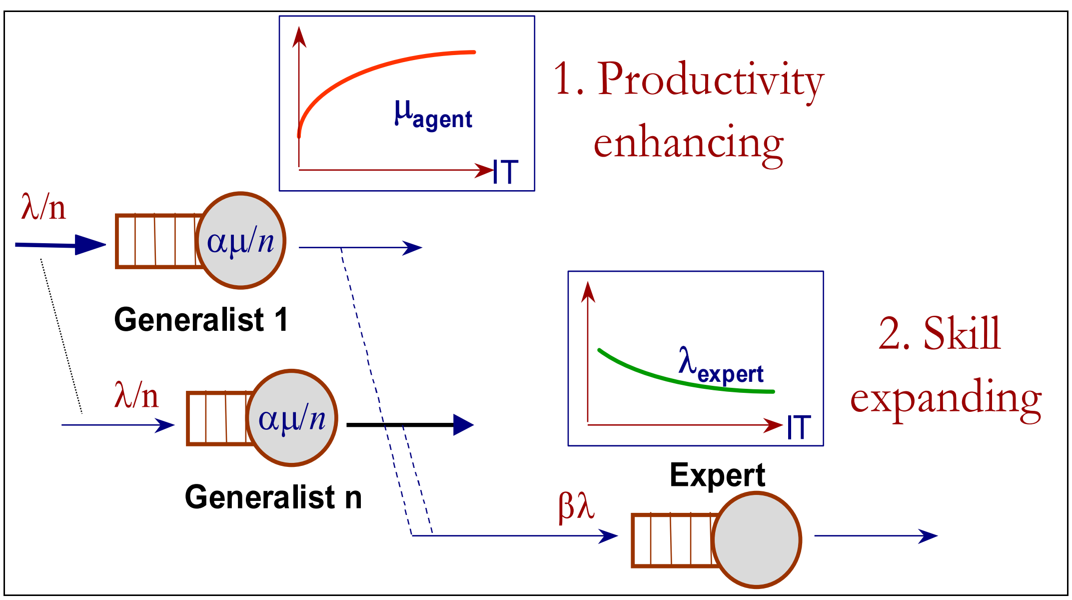
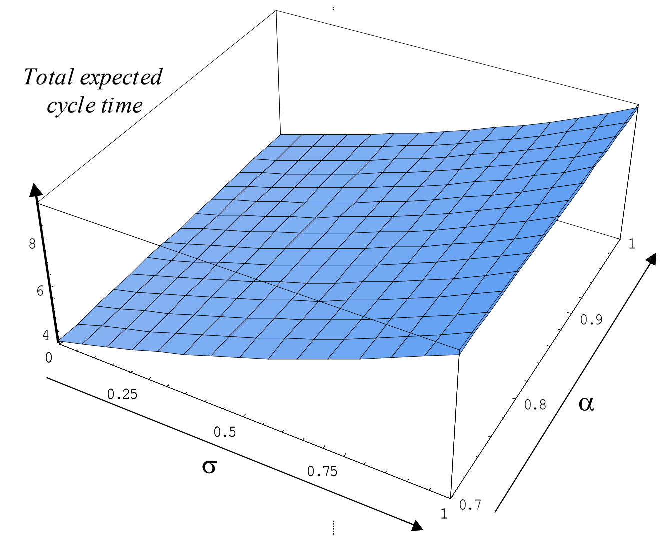
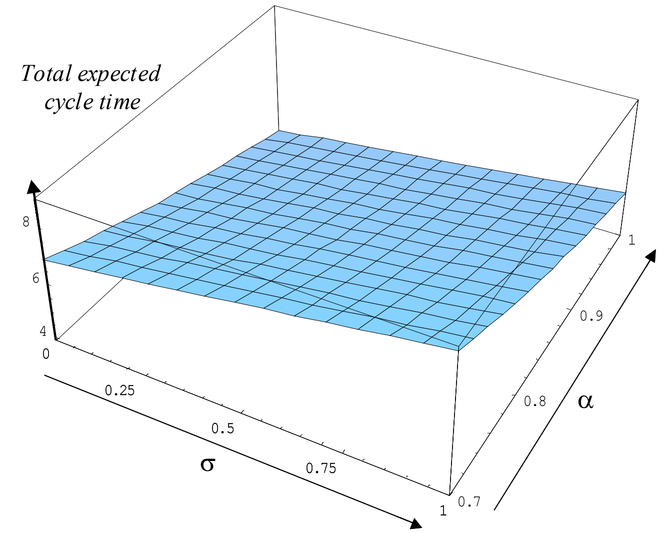


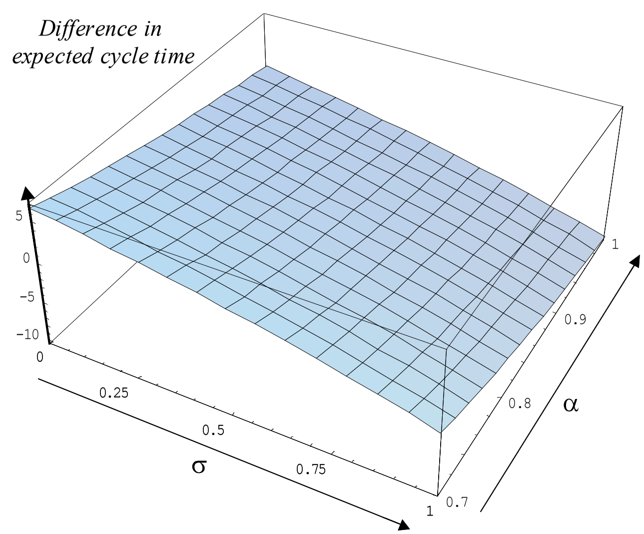
Figure 2(b): The impact of information technology on the performance of case managers



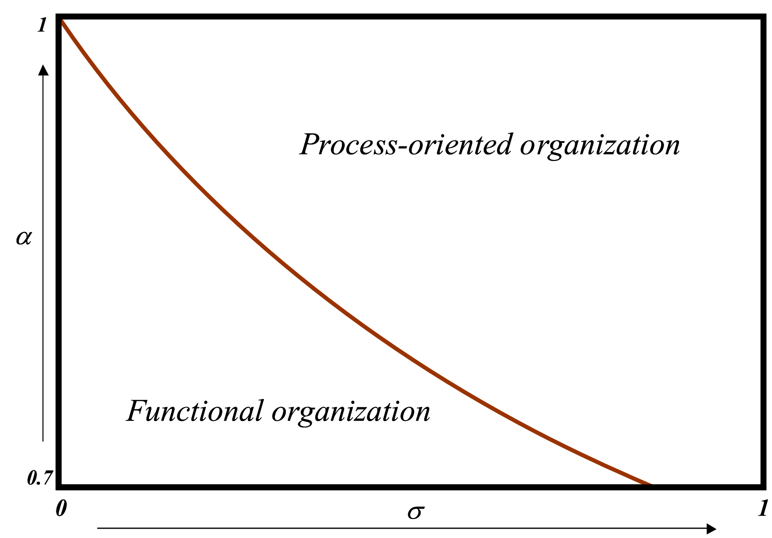
3(a) Cycle time for functional work system:
$nT_s(\mu,\lambda,\sigma)$



3(b) Cycle time for process-oriented work system:
$T_s(\frac{\alpha\mu}{n},\frac{\lambda}{n})+\beta T_s(\frac{\mu}{n},\beta\lambda)$



3(c) Difference in cycle times
(process-oriented) – (functional)



3(d) Range of (σ,α) values over which each
work system has the lower cycle time

Figure 3: Cycle time trade-offs between handoff delays σ and loss in specialization α

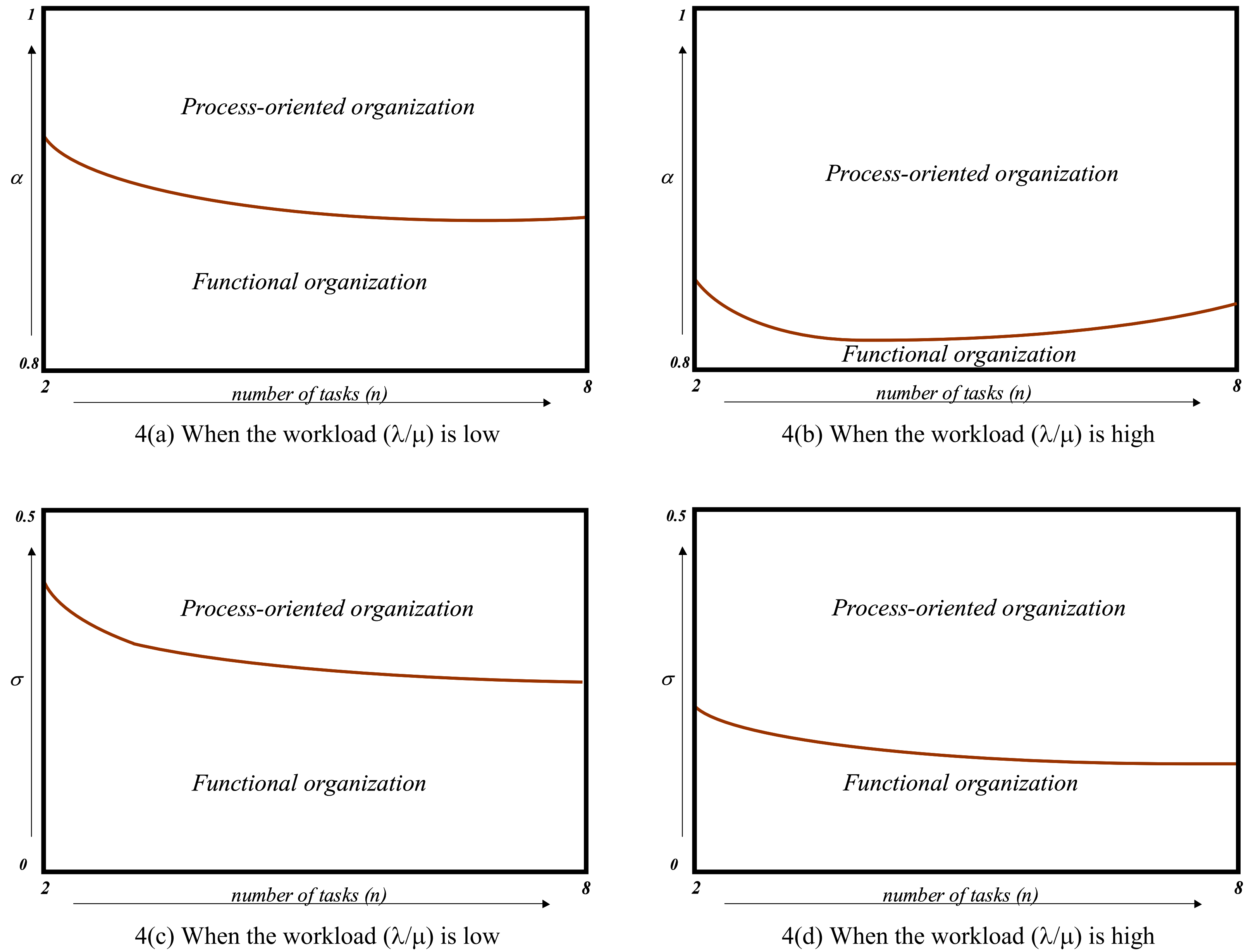
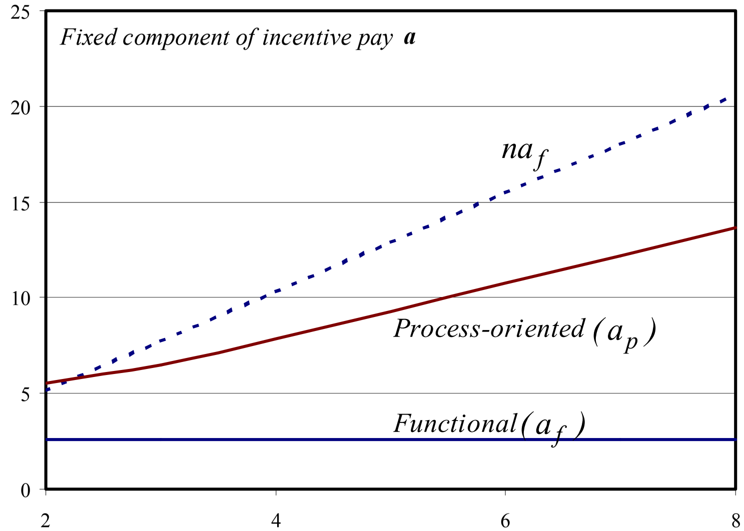
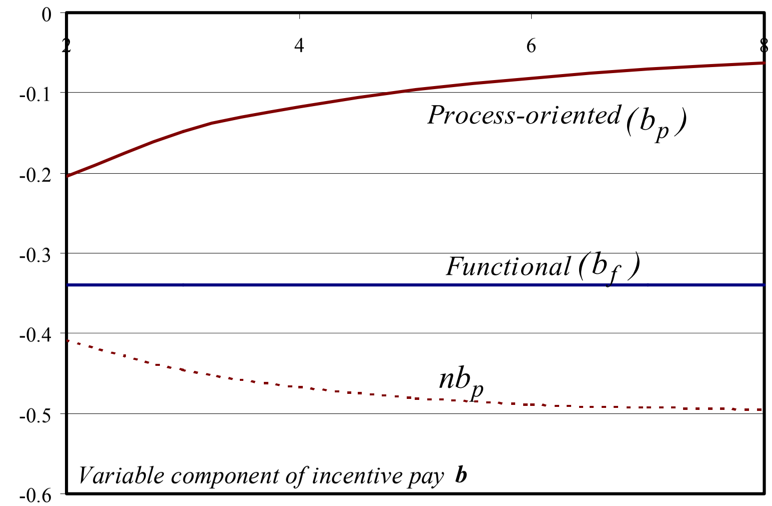


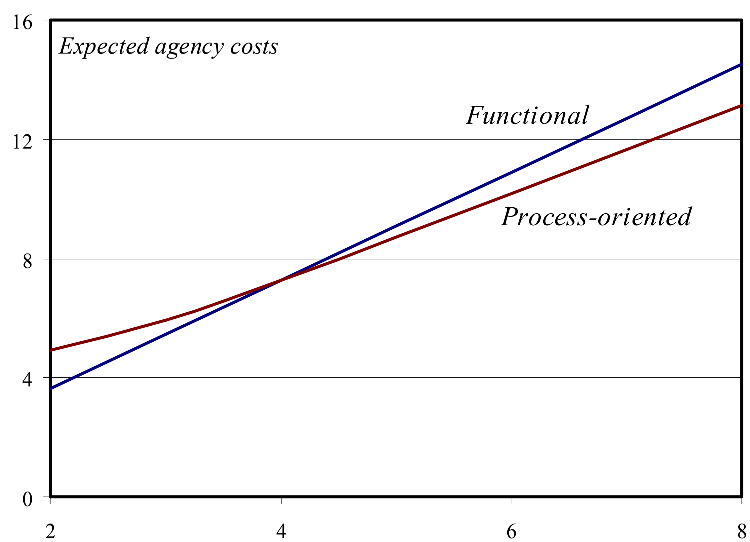
Figure 4: Trade-offs between functional and process-oriented work systems as α, σ and *n* vary



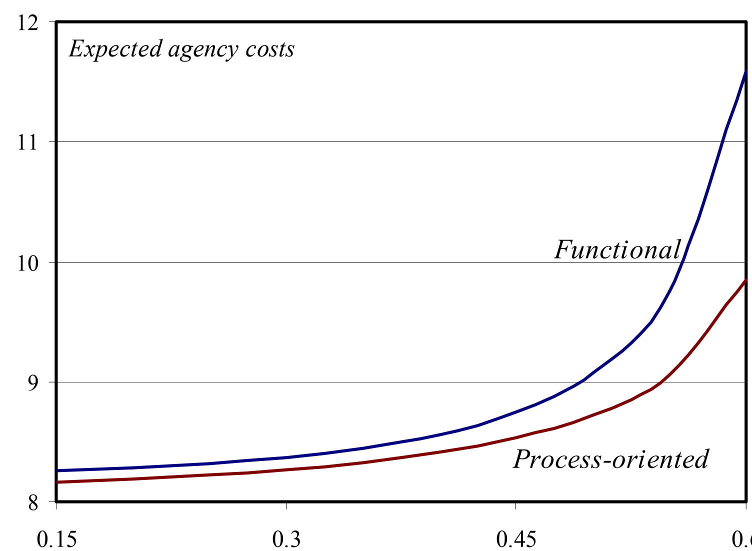
5(a) Change in fixed incentive pay *a* as *n* increases



5(b) Change in variable incentive pay rate *b* as *n* increases

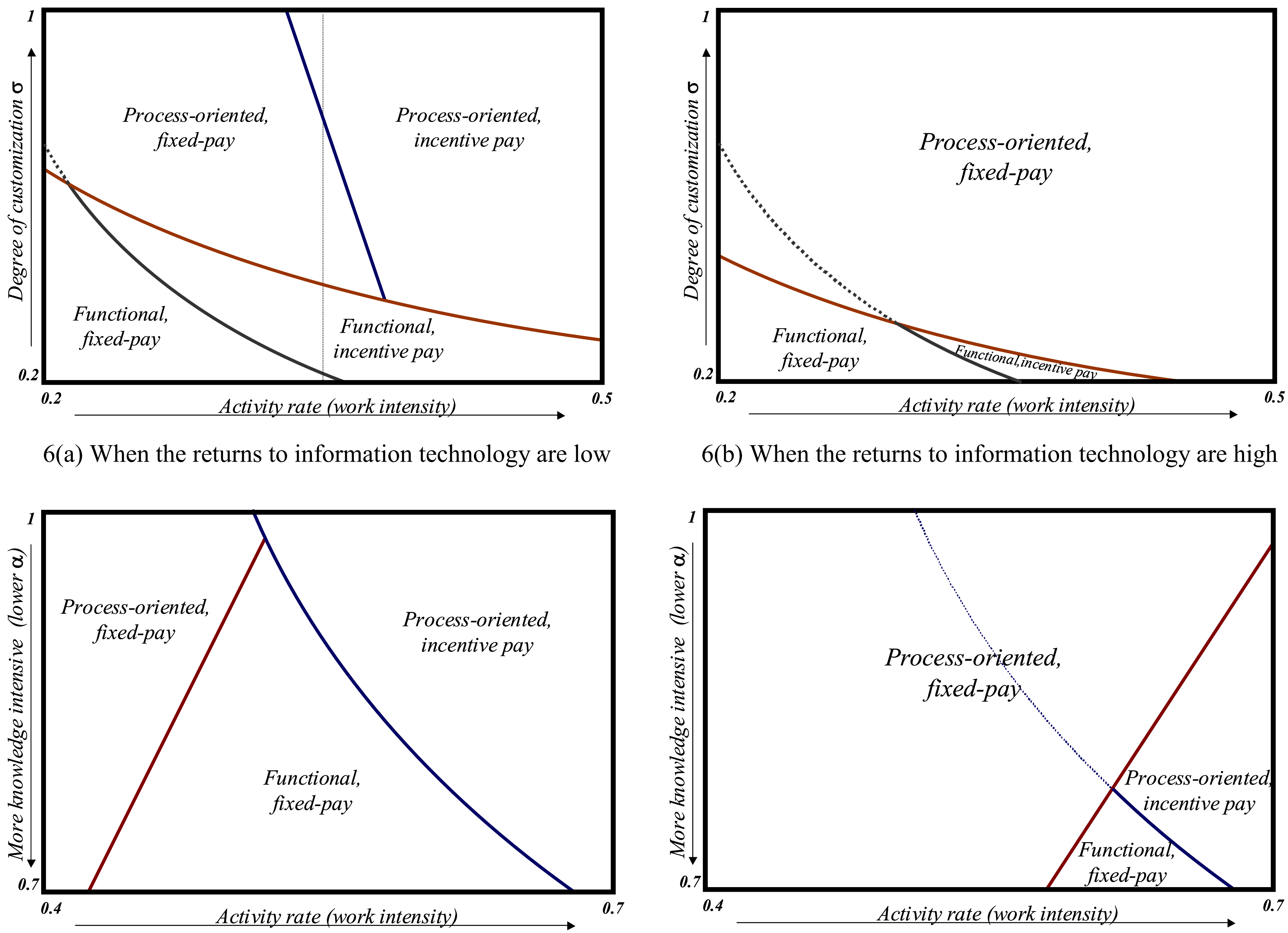


5(c) Total expected incentive pay as number of tasks *n* varies



5(d) Total expected incentive pay as workload λ varies

Figure 5: Optimal contracts and agency costs



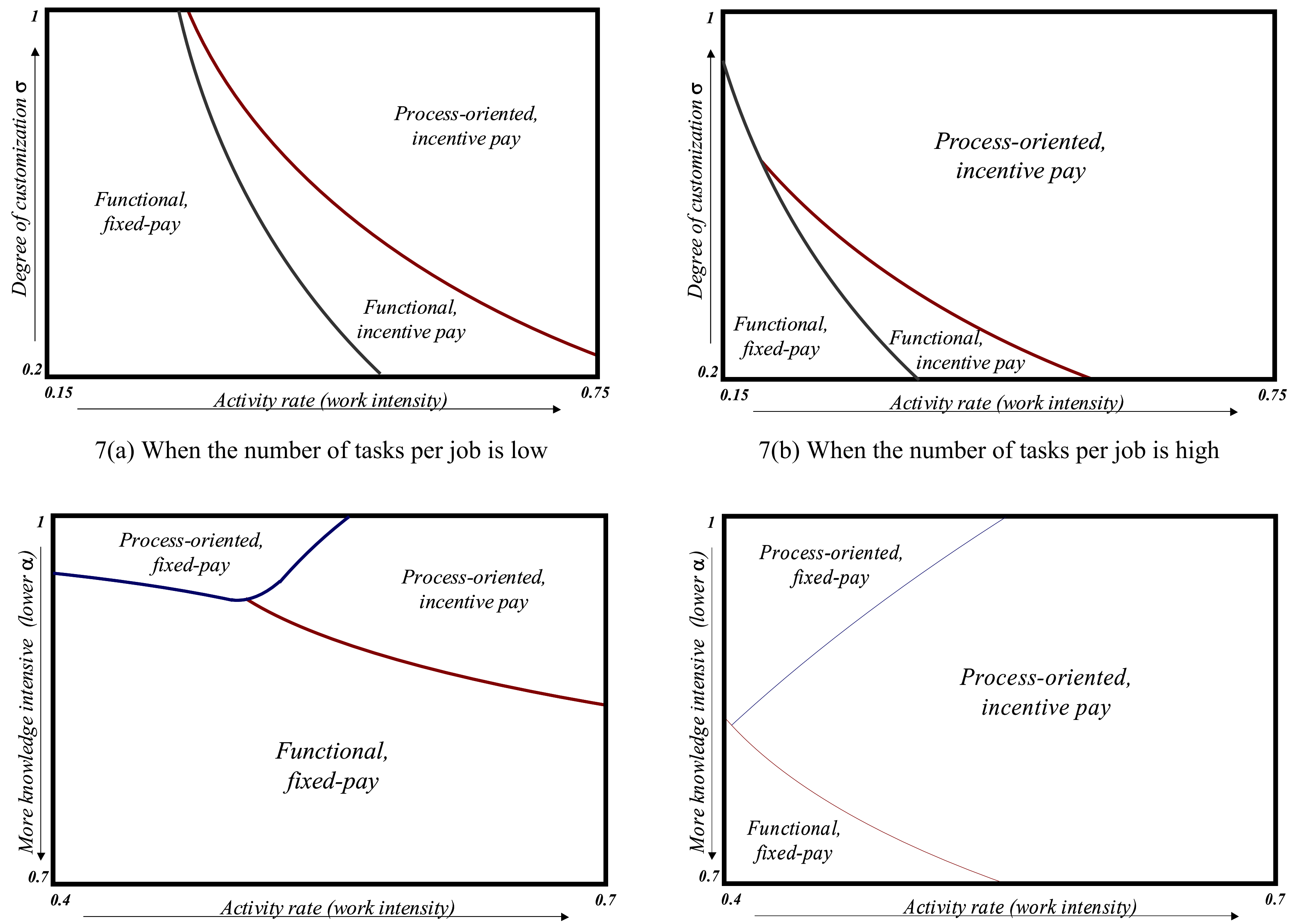
6(a) When the returns to information technology are low

6(b) When the returns to information technology are high

6(c) When the returns to information technology are low

6(d) When the returns to information technology are high

Figure 6: Optimal organization – work intensity, knowledge intensity, customization, and returns to I.T.



7(a) When the number of tasks per job is low

7(b) When the number of tasks per job is high

7(c) When the number of tasks per job is low

7(d) When the number of tasks per job is high

Figure 7: Optimal organization – work intensity, knowledge intensity, customization, and number of tasks per job

References

1. Ballou, R.H. Reengineering at American Express: The Travel Services Group's Work in Process. *Interfaces* (May-June 1995), 22-29.

2. Barua, A., Whinston, A. B., and Sophie Lee, C. H. The Calculus of Reengineering. *Information Systems Research* 7 (1996), .

3. Brickley, J., Smith, C., and Zimmerman, J. *Managerial Economics and Organizational Architecture.* Richard D. Irwin, Inc., 1997.

4. Brynjolfsson, E., Renshaw, A and Van Alstyne, M. The Matrix of Change. *Sloan Management Review* (Winter 1997).

5. Buzacott, J. Commonalties in Reengineered Processes: Models and Issues. *Management Science* (May 1996).

6. Byrne, J. A. The Horizontal Corporation. *Business Week* (Dec. 20th, 1993), 76-82.

7. Clark, T. Proctor and Gamble: Improving Consumer Value Through Process Design. HBS Case 195-126 (1995).

8. Clemons, E.K., Thatcher, M.E. and Row, M.C. Identifying Sources of Reengineering Failures: A Study of the Behavioral Factors Contributing to Reengineering Risk. *Journal of Management Information Systems* (Fall 1995), 9-36

9. Davenport, T.H., and Short, J.E. The New Industrial Engineering: Information Technology and Business Process Redesign. *Sloan Management Review* (Summer 1990), 11-27

10. Davenport, T.H., *Process Innovation*. Harvard Business School Press, 1993

11. Gallini, N., and Lutz, N. Dual Distribution and Royalty Fees in Franchising. *Journal of Law, Economics and Organization* 8 (1992), 471-501.

12. Gates, W. *Business @ the Speed of Thought: Succeeding in the Digital Economy*. Warner Books, 2000.

13. Gross, D., and Harris, C. *Fundamentals of Queuing Theory*. John Wiley and Sons, 1985.

14. Gurbaxani, V. and Whang, S. The Impact of Information Systems on Organizations and Markets. *Communications of the ACM* (January 1991), 59-73

15. Hammer, M. Re-engineering Work: Don't Automate, Obliterate. *Harvard Business Review* (July-August 1990), 104-112

16. Hammer, M. and Champy, J. *Reengineering the Corporation: A Manifesto for Business Revolution.* Harper Business, 1993.

17. Holmstrom, B. and Milgrom, P. (1987) Aggregation and Linearity in the Provision of Inter-temporal Incentives. *Econometrica* 55(2), 303-28.

18. Humphrey, S. Bell Atlantic Reengineers Payment Processing. *Enterprise Reengineering*, Oct/Nov, 1995.

19. King, R. T. Jeans Therapy: Levi's Factory Workers Are Assigned to Teams, And Morale Takes a Hit -- Infighting Rises, Productivity Falls as Employees Miss The Piecework System. *Wall Street Journal,* May 20, 1998, A1.

20. Kleinrock, L. *Queuing Systems.* John Wiley & Sons, 1976.

21. Malone, T. and Smith, S. Modeling the Performance of Organizational Structures. *Operations Research* 36 (1988), 421-436.

22. Mathewson, G. and Winter, R. The Economics of Franchise Contracts. *Journal of Law and Economics* 36, 33-70.

23. McAfee, P., and McMillan, J. Competition for Agency Contracts. *RAND Journal of Economics* 18 (1987), 296-307.

24. Ross, S. The Economic Theory of Agency: The Principal's Problem. *American Economic Review* 63 (1973), 134-139.

25. Seshadri, S. and Pinedo, M. Bounds on the Mean Delay in Multiclass Queuing Networks Under Shortfall-Based Priority Rules. *Probability in the Engineering and Informational Sciences* 12 (1998), 329-350.

26. Seidmann, A. and Sundararajan, A., Competing in Information Intensive Services: Analyzing the Impact of Task Consolidation and Employee Empowerment.. *Journal of Management Information Systems* Vol.14, No.2 (1997), 33-56.

27. Smith, M. G. *Laplace Transform Theory.* D. Van Nostrand Company Ltd., 1966.

28. Stalk, G. The Time Paradigm. *Forbes ASAP* (November 30th, 1998), 213-214.

29. Stiglitz, J. Incentives and Risk-sharing in Sharecropping. *Review of Economic Studies* 41 (1974), 219-255.

30. Zell, D. Changing By Design: *Organizational Innovation at Hewlett Packard.* Cornell University Press (1997)

---

[1] See Davenport and Short (1990), Davenport (1993), Hammer (1990), Hammer and Champy (1993), or Brickley, Smith and Zimmerman (1997) for extensive discussion on some of the new organizational design philosophies.

---

[2] Some other cases are described in Ballou, 1995, Byrne, 1993, Clark, 1995, Humphrey, 1995, and Zell, 1997.

[3] This immediacy reflects the advent of workflow automation software and intranet technology in today's workplaces and allows us to focus on more subtle issues, such as the need for information transfers and the effects of specialization

[4] In this case, it is immaterial whether the agents can observe each others' cycle time, as the principal is free to contract with each agent individually

[5] Although the principal has full information about the value of σ, the best measure on which a contract can be written is *the entire cycle time of the agent's task,* as the presence of the handoff delay influences queuing time as well. A measure based solely on μ cannot therefore be separated out.

[6] The fact that b turns out to have identical values is due to a property of the Laplace transform of the density function of a sum of independent random variables.

[7] The barriers to information access are modeled as reducing the effective rate at which the tasks are processed; hence information access systems here have the effect of increasing processing rates.

[8] It is easy to be misled into thinking that this is equivalent to assuming a linear cost structure, but this is not the case. The convex nature of the cost of effort is based on the amount of work done in a fixed time period, and the personal cost increases at an increasing rate in the amount of work an agent does *in a working day*. In this case, the agent does the same volume of work in a given time period – instead of doing n instances of the same task, she does n potentially different tasks. Hence the cost of doing these n tasks remains the same. There may be a cost associated

with switching between tasks; this can be assumed to balance the reduction in cost due to lowering of monotony.

[9] Normally, the expert is a person who is employed for the purpose of doing other value-adding work. Rerouting exceptions to this expert costs the firm at a rate equal to the value of her time.

[10] Normalizing log[$-1/U_0$] to zero is equivalent to an identical linear shift in the expected cost under any process design, and leaves the results unaltered, since a cost of log[$-1/U_0$] is borne per employee irrespective of the nature of the organization design.

[11] While both α and β are inverse measures of knowledge intensity, the results obtained from varying β are highly sensitive to the processing rate of the expert μ_E (an increase in β causes both μ and λ to reduce by a factor of (1-β), which does not change agent activity rates). We can therefore conclude that a slow expert causes functional organization to be optimal and vice versa. The more interesting parameter is α, and hence the graphs plot variations in α for a fixed β.