Pergamon

European Management Journal Vol. 16, No. 3, pp. 306–317, 1998

0263-2373/98 $19.00 + 0.00
PII: S0263-2373(98)00007-3

Operations Management and Reengineering

CHRISTOPH LOCH, *INSEAD, Fontainebleau, France*

Business Process Reengineering has been the most influential management movement of the 1990s, and like the quality movement of the 1980s, it has put management attention squarely on processes and operations. At first glance, however, it is hard to see any relationship between the precepts of Business Process Reengineering and traditional Operations Management teaching. The purpose of this article is two-fold. First, it offers some opinion as to what that relationship is by conceptualizing process design in terms of a design evaluated iteration: reengineering as represented by the most influential book on the topic, *Reengineering the Corporation* by Hammer and Champy (1993), concentrates almost exclusively on the design step, whereas traditional Operations Management teaching is heavily oriented toward evaluation, taking a design as given. Second, this article identifies and discusses the process design principle of integrated work enunciated by Hammer and Champy. It is demonstrated how the quantitative benefit of integrated work can be estimated using Operations Management tools, and that the benefit depends on flow control, an important additional process design lever that Hammer and Champy do not address. The article concludes with the observation that as corporations turn away from cost cutting to growth generation, Business Process Reengineering is disappearing from the headlines, but process redesign and improvements can be turned toward performance improvement and revenue generation just as effectively as they have been used for efficiency gains in the past. © 1998 Elsevier Science Ltd. All rights reserved

Introduction

Business Process Reengineering (BPR), although disappearing from the headlines of the business press, is undoubtedly the most influential development in management thinking in the 1990s. Although BPR seems to have run its course as companies turn from an efficiency focus to a search for new growth, reengineering concepts have been driving organizational change in many leading North American and European companies.

The popular conception of BPR was crystallized by Michael Hammer and James Champy in their 1993 best-seller *Reengineering the Corporation,* the most influential reengineering book, on which this article focuses, and hereafter referred to as H&C.[1] Two distinct and separable elements can be identified in H&C. On the one hand, they enunciate principles of business process design. In particular, they advocate a reintegration of industrial work, reversing the trend toward specialization and division of labor that has been with us since the early Industrial Revolution. On the other hand, Hammer and Champy advocate dramatic change, as opposed to an incremental or evolutionary approach, in implementing new process designs and associated organizational structures. Indeed, many managers' primary association with the term 'reengineering' is the bold approach to change management advocated by Hammer and Champy.[2]

Leaving aside the important subject of change management, this article focuses attention on principles of business process design, a central topic in the discipline of Operations Management (OM). A large body of knowledge associated with process design has been developed by practitioners and scholars over the last century. At this time when BPR is subsiding, it is natural to ask which of its precepts are likely to endure, and how they relate to the pre-existing body of knowledge that dominates OM teaching. This is the purpose of the present paper.

From the perspective of an OM professional, the reengineering movement has made an important contribution simply by putting in the foreground of top management concern the operations side of business through which work is routinely accomplished, without the wasted effort and firefighting that characterize inefficient operations. By focusing attention on processes as the means of achieving effective operations, reengineering leaders have reinforced a central theme of the 1980s quality movement.[3] To be effective, organizations must put creative energy into the design, documentation and maintenance of processes that satisfy customer needs on a routine basis. Workers must understand the overall function of core business processes, and performance must be measured in terms of value delivered to customers.

It seems, however, that BPR has illuminated some parts of the OM landscape while leaving other parts in shadow. In their treatment of process design, all of the influential books and articles cited earlier use vague language and lack cause-and-effect reasoning. In addition, their recommendations implicitly depend on a critical element of process design, namely intelligent flow control, but they never acknowledge it as a separate category of design levers.

The next section summarizes what Harrison (1997) calls the 'processing network paradigm,' a general process model that underlies much of OM teaching and from which one obtains precise language, causal reasoning, and analytical procedures to support process design. Section 3 summarizes the main precepts of BPR and explains, using the language of the processing network paradigm, how BPR and OM complement each other, while differing in both emphasis and texture. The last section of the paper demonstrates, in the example of integrated work, the benefit of flow management protocols and the usefulness of OM tools to estimate their value.

A Conceptual Framework of Business Process Design

To evaluate principles of process design, one needs to understand the causal relationship between design choice and bottom-line performance. For this kind of cause-and-effect reasoning, one first needs a vocabulary to describe business processes, including generic names for the elements that make up a process. Furthermore, any design problem worthy of discussion involves tradeoffs, whether it is a mouse trap or a business process that is being designed. Designers, therefore, invariably need to balance factors or objectives which cannot be achieved simultaneously.

To address these needs, operations management practitioners and scholars have developed over many decades a collection of generic process models and associated analytical techniques.[4] In this paper they will be referred to collectively as 'processing network models,' and their common structure will be called 'the processing network paradigm,' following Harrison (1997).

A process design in this paradigm consists of five basic elements – jobs, tasks, precedence constraints, resources, and flow management protocols. A sixth element will be proposed below. First, in this terminology the units to be processed are called *jobs.* Jobs are the entities that 'flow' through a process, and a single process may need to handle a variety of job types. As a second basic concept in the processing network paradigm, jobs are made up of *tasks.* In particular contexts, jobs might be called, for example, *transactions*, *manufacturing orders*, *projects*, or *customers*, and tasks are frequently referred to as *activities* or *operations.* Tasks are connected by *precedence relations*, that is to say, some tasks must be completed before others can be started.

Jobs, tasks and their precedence relationships can be represented in a conventional flow chart. Figure 1 shows the example of a flow chart of the product development process at ConnectCo, a manufacturer of electrical connectors. This example is described in detail in Adler *et al*. (1995, 1996). Jobs are development projects, tasks are depicted by boxes and precedence relationships by arrows in Figure 1. Precedences may be uncertain due to unforeseeable events: in the example, prototyping must be repeated in 75 per cent, and scale-up plus prototyping in 60 per cent of all cases because the design poses manufacturing problems which can be planned statistically, but not for an individual job.

The third basic element of a processing network model is a set of system performance criteria, which describe what the desired output of the process is (in Figure 1, it is new connector products), and how one can tell whether the process does well or not. Typical performance measures comprise a combination of: volume or frequency (e.g. 12 new products per year), throughput time (e.g. time-to-market of under 12 months on average), service level (with 90 per cent evidence, one can promise a new product in under 18 months), or costs (the whole process costs less than $10 M per year).

Together, the first three elements describe *what* the process should accomplish, namely its output as well as the work that needs to be performed in order to produce the output. In order to specify a complete process design, one must also determine *how* the work is to be performed. This is specified by resources, a flow management protocol and system status information.

Processing *resources* are the units that execute the tasks, which make up the jobs. At ConnectCo the key resources are groups of development engineers and technicians. Resources are characterized by capabilities, such as the breadth of job types they can process, or whether they need to be supervised by other resources. For example, engineers have a wider range of capabilities and can work more independently than technicians, but they are also more expensive.

The fifth element of a process specification is a *flow management protocol*, the simplest example of which is a *route,* corresponding to a fixed order in which

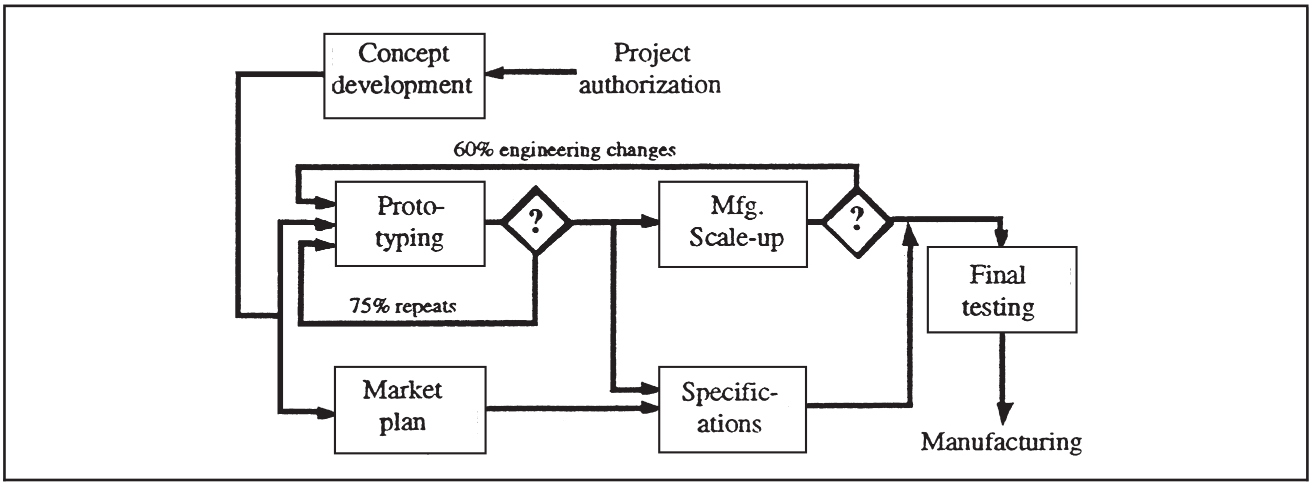


Figure 1 The Process Flow Chart

resources execute a job's constituent tasks. In general, however, management may have a great deal of latitude in assigning a job's tasks to different resources and the order in which tasks associated with different jobs are processed at a given resource. This may depend on job type and urgency, resource availability, or system status in general. Sophisticated flow control may require *system status information,* such as which jobs are where in the system, or which resources are busy or idle.

An important point to bear in mind is that tasks, resources and flow protocols are not independent. Assigning several tasks to one resource, or operating in a high-variability processing system, may require higher resource capabilities (such as cross-training, or universal processing machines). Dynamic flow protocols with tasks assigned to resources depending on the job type may even introduce new tasks, for coordination and monitoring, or for tracking system status information.

With the above five elements of the processing network model, the work which is to be accomplished and how it is to be done can be determined. Traditionally, OM regarded this specification as complete, implicitly viewing the processing resources as machines (see, for example, the characterization by Buzacott, 1996, p. 768) or as piece workers whose performance can be 'designed.' Recently, however, *incentive issues* have been incorporated into OM from organizational theory and economics, reflecting the realization that work cannot be monitored perfectly and that employee performance is heavily influenced by incentives and accountability. An example is the global coordination of supply chains by making upstream parties responsible for downstream ('echelon') inventories (e.g. Lee and Billington, 1992).

Incentives may even need to sacrifice some short-term performance of a process in order to ensure long-term improvement and employee motivation. For example, at Microsoft the customer help desk is an important entry level job, which thus needs to attract employees with high future development potential. Ensuring the motivation and career development of such employees requires a higher investment than may be necessary for short-term process efficiency. In the light of these considerations, we propose a *system of assessing individual contributions* as the final element of a process design.

These six elements provide a complete process *design specification.* A second step in process design is *evaluating* the proposed specification. Analogous to design prototype iterations in engineering (see Wheelwright and Clark, 1992, chap. 10), the cycle must be repeated until the evaluation indicates satisfactory performance. The process design cycle is summarized in Figure 2.

The first question one may be prompted to ask in evaluating a process specification is whether a given set of resources provides enough capacity to satisfy a given set of external demand rates on a sustained basis. An affirmative answer requires that total capacity consumption for each processing resource be less than or equal to the total capacity available. To estimate resource consumption rates, one needs to know the tasks involved in each type of job and the resources to which these tasks are assigned, but the ordering of the tasks is immaterial.

Resource consumption rates should include all non-productive time that is usage-driven, such as resource hours consumed in setups and in rework. Similarly, resource availability rates must be properly deflated to account for non-productive time that is not usage-driven, such as operator breaks and preventive maintenance. Assessing whether capacity is sufficient to satisfy a specified demand vector is often called *rough-cut capacity analysis.*

To understand where capacity is most scarce, one calculates the total capacity consumed as a percentage of the capacity available for each processing resource,

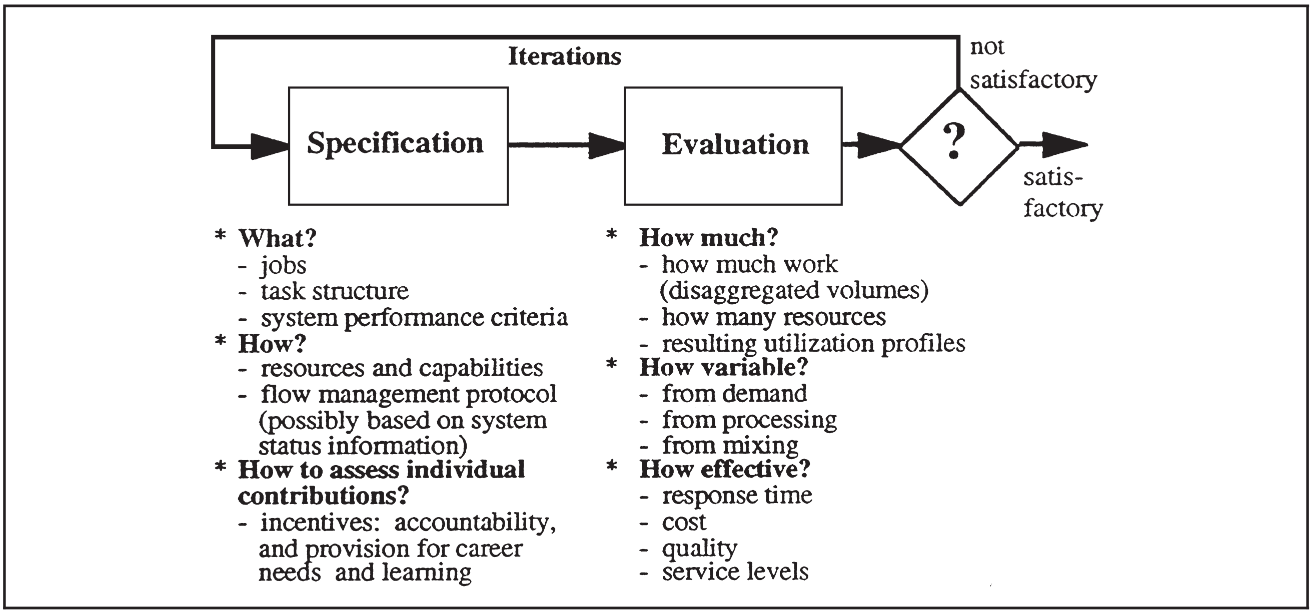


Figure 2 The Iterative Cycle of Process Design

based on a particular demand scenario. This ratio is called 'utilization', and by the above description it includes time consumed in non-productive activities like rework and setups. By breaking total utilization into its constituent parts one obtains a *utilization profile* of the resources. Resources for which the utilization rate is highest are often called process *bottlenecks*.

The second question to ask in evaluating a process specification involves the effects of *variability* on performance. A crucial aspect of the processing network paradigm is the conception of work as a flow of jobs over time. Jobs are not processed in isolation, but arrive in an overlapping way, with the pattern of arrivals depending on many factors, some controllable and some not. It is an unfortunate fact of life that demand for products and services varies hourly, daily, and seasonally. Moreover, in most cases the tasks required for the completion of a job are not completely foreseeable, as is evident in the ConnectCo example above. Jobs may have to compete for resources during periods when demand is high or processing is unexpectedly slow. Variability degrades system performance by leading to queuing delays and inventories, or to lost throughput, both of which are a focus of attention in OM teaching.

Queuing delays extend the total time that a job spends within a processing network, referred to hereafter as its overall *throughput time*. In cases where jobs represent requests or orders from customers, long throughput times obviously represent poor service, and in make-to-stock manufacturing they hinder rapid response to changing market conditions, leading to stockouts and high inventory costs. It should be emphasized that throughput time is itself highly variable in most processing networks, and its average is seldom an adequate measure of responsiveness. In other words, the throughput time or response time of a business process must be considered as a *distribution,* a range of values with associated probabilities, and the mean of that distribution is seldom a satisfactory summary of process performance. For example, in make-to-order manufacturing, a manager may wish to ensure that a particular promised lead time is achieved with 95 per cent reliability, in which case, concern centers on the 95th percentile of the throughput time distribution, not on its mean.

Once utilization profiles and variability are understood, the performance of a process can be evaluated. Work flow performance modeling has been a dominant concern of OM, and a large set of sophisticated tools for performance evaluation, such as capacity modeling (e.g. linear programming) and flow modeling (e.g. queuing models, simulation), is available. However, a number of qualitative process specification principles have also emerged, centering on *reducing utilization and reducing variability,* which we summarize in Table 1. These are now well accepted, not only in manufacturing (Hopp *et al.*, 1990), but also in services (Fitzsimmons and Fitzsimmons, 1994) and product development (Adler *et al.*, 1995, 1996). We concentrate on process design for response time performance, which is also the emphasis of most examples in H&C.

The New Conventional Wisdom of Business Process Design

In this section, we briefly summarize H&C's view of process design in their own words. We then show that their process design principles are consistent with the specification design step of the processing network paradigm.

Table 1 Process Design Principles for Response Time Improvement

Specification Step	Design Principle
What?	
Jobs (taking volumes as given)	**Reduce variability**: Demand: reduce peaks, negotiate with customers to make demand more regular, processing: product/service standardization, variant reduction
Task structure	**Reduce utilization**: remove steps through design-to-manufacture (DFM), or design-for-logistics, eliminate activities through automation (e.g. IT support) **Reduce variability**: standard operating procedures (SOPs), in product development separation of research and development or reliance on proven technologies
How?	
Resources and capabilities	**Reduce utilization**: add resources, involve workers (e.g., reduce absenteeism and turnover), make capacity more flexible (e.g., overtime) **Reduce variability**: statistical process control (SPC), quality circles and worker training **Reduce *both*:** minimize scrap, set-ups and breakdowns (e.g. preventive maintenance), material shortages (e.g. supplier certification), provide timely information and standard procedures through IT; **Pool** (make resources interchangeable) in order to equalize resource utilizations (however, this may reduce efficiency or increase system variability)
Flow management protocols	**Smooth** production flows, in just-in-time processing (JIT) **Schedule** around bottlenecks to increase effective bottleneck capacity **Sequence** work with the shortest processing time first (SPT rule), or use priorities to concentrate congestion on 'background work', or sequence and route jobs based on system status reporting systems (toward resources that are idle or not congested)
Incentives	Make upstream party (e.g. production) **responsible** for effects of decisions on downstream (e.g. inventories) to increase global system performance

H&C formulate four general themes of reengineering in Chapter 2: process orientation, the creative use of Information Technology (IT), ambition, and rule-breaking. We focus on the first two themes, which are concerned with process design, leaving aside the latter two, which are associated with change management. Chapter 3 formulates nine principles of reengineered (or 're-thought') processes. Chapter 4 describes how the authors believe their process design principles will influence the 'nature of work,' namely worker capabilities, performance measurement and values. Chapter 5 argues that IT enables the implementation of the new process and work characteristics, by making information available in a decentralized way and by automating routine tasks. Re-thought processes and the new world of work are referred to as 'process orientation' and summarized in Figure 3.

Chapter 4 contains the essence of H&C's process design principles (for the detailed descriptions, we refer the reader to pp. 51–64). The principle of process steps performed in their natural order refers to following the 'natural work precedence rather than the artificial one introduced by linearity.' Reduced checks and controls are achieved by replacing detailed controls by aggregate (and thus fewer) ones as long as it is economical, 'allowing limited abuse.' Minimizing reconciliation aims to reduce the external contact points of a process, reducing the risk of inconsistent data being received. Task combination assigns several tasks to one worker (resource), thus eliminating handoffs.[5] 'Workers making decisions' integrates formerly managerial activities into everyone's work by delegating a certain degree of decision-making. Multiple process versions separate jobs into 'standard' and 'hard' cases through 'triage' and then process these separately. Having work done 'where it makes the most sense' shifts activities, even across organizational boundaries, to customers and suppliers whenever efficiencies can be gained. Hybrid centralization/decentralization allows local decision-making (e.g. sales contracts signed on the spot), while at the same time maintaining global information updates and control. Finally, the case manager provides a single co-ordinating interface to the customer, behaving 'as if he/she were responsible for the whole process, even though that is not the case.'

The 'new world of work' refers to 'implications (of fundamental changes in business processes) for many other parts and aspects of an organization' (p. 65). These implications are in the organization (flatter and process- rather than functional-oriented), the workers' capabilities, performance measurement, and the organization's culture and values. It is, however, misleading to view these changes only as *results* of process reengineering; they are at the same time essential *enablers* of reengineered processes. Workers cannot perform integrated work or make decisions without first acquiring the capabilities to take on the additional functions. Similarly, without appropriate

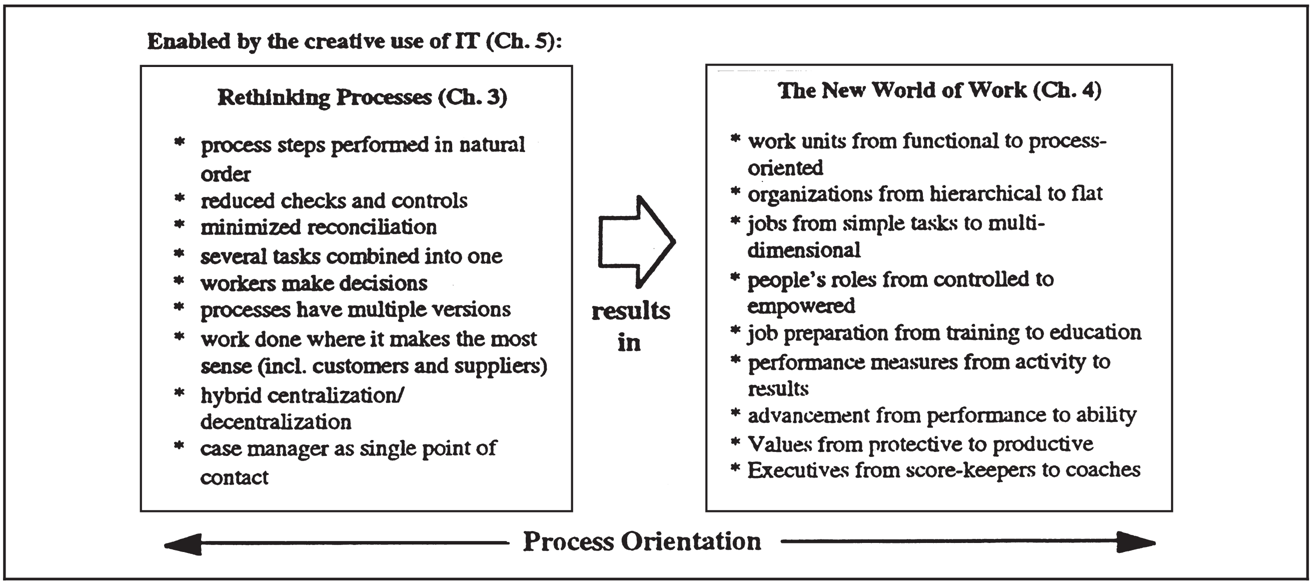


Figure 3 H&C's Principles of Process Design

customer-oriented performance measures that relate an individual's activities to the overall process purpose, workers may revert to pursuing narrow and possibly dysfunctional goals. Finally, in a decentralized process with delegated decision power, it is impossible to fully monitor everyone's individual contribution. If culture and values do not support responsible behavior, decentralization and delegation may lead to prohibitive abuse. Indeed, H&C introduce the business diamond (p. 80), where process design, jobs, performance measurement, and values are depicted as forming an iterative cycle.[6]

How, then, do H&C's process design principles relate to OM and the processing network paradigm? In Figure 4, we group H&C's principles (excluding the IT enablers) under the process specification questions from Figure 2. The first three principles relate to defining a task structure, such as performing steps in a natural (e.g. parallel) order. We also find principles about upgrading resource capabilities, assigning tasks to resources (e.g. combining tasks and delegating), system status information (hybrid centralization via IT), and performance measurement.

The case manager also relates to the processing network paradigm, but in a more complicated way by combining several elements from Figure 2. The task structure is altered by adding explicit coordination tasks (for the benefit of the inquiring customer). A dedicated resource is created for these tasks, with the authority to impose situation-specific flow protocols. For example, the case manager may determine that a certain individual customer order is so important that it needs to be given preemptive processing priority over all other jobs. This process is complicated, and possibly more costly, but very flexible in reacting to specific customer needs.

It becomes clear that the principles in the first six categories in Figure 4 harmonize with the process specification step in Figure 2, but they are more 'general' than the OM design principles formulated in Table 1. H&C's principles mainly relate to the left-hand step of the process design iteration in Figure 2. Thus, BPR is heavily oriented toward process *specification*, but has little to say about process design evaluation. On the other hand, the BPR principles of values and culture, as well as the shape of the organizational structure, are outside the processing network paradigm. They are clearly important since organizational structure may influence the ease of co-ordination across units, and values may impact how well workers are willing to perform without being monitored. Traditionally, these issues have not been examined within the field of OM.

Based on this comparison between H&C's process design principles and the processing network paradigm, we come to the conclusion that BPR has made a very important contribution to the specification aspects of process design. First, BPR has elevated process management and operations to the attention of top management. Scholars and practitioners in strategy and organizational behavior have adopted the concept of a process in their thinking.[7] Second, BPR has emphasized that in many cases major improvements can be achieved by creatively rethinking the basic structure of processes which have not been examined for a long time. OM has tended to take basic process structure as given and to concentrate on sophisticated but formal improvement tools. Third, the BPR literature has collectively illuminated the fact that IT is a key enabler for effective processes and has the potential of shifting tradeoffs in the decentralization of decisions and in work automation and simplification,[8] and fourth, BPR has attempted to present the manager with an 'integrated approach' by tackling together the issues of process design ('where to go') and change management ('how to get

Defining task structure
* process steps performed in natural order
* reduced checks and controls
* minimized reconciliation

Determining resource capabilities
* jobs from simple tasks to multi-dimensional
* people's roles from controlled to empowered
* job preparation from training to education

Assigning tasks to resources
* several tasks combined into one
* workers make decisions
* work done where it makes the most sense (incl. customers and suppliers)

Devising a method of system status reporting
* hybrid centralization / decentralization

Case manager as single point of contact

Assessing individual contribution
* performance measures from activity to results
* advancement from performance to ability

Fostering values and culture
* Values from protective to productive
* Executives from score keepers to coaches

Process-centering the organization
* work units from functional to process-oriented
* organizations from hierarchical to flat

Covered–in Processing Network Paradigm

Figure 4 H&C's Principles in the Processing Network

there'). These are conceptually separate, but represent two parts of the same problem for the manager facing severe performance problems.

On the other hand, H&C have little to say about process performance evaluation, the second step of the design iteration in Figure 2. Their treatment of process design lacks evaluation tools, and the term 'tradeoff' is not mentioned. Thus, their design principles amount to assertions that lack cause-and-effect reasoning. None of H&C's principles are universally applicable, and yet no methods that would allow one to evaluate a particular design strategy in a particular situation are explained. The OM principles in Table 1, in contrast, point to the tradeoffs involved: if utilization can be reduced without worsening variability, or *vice versa*, response time improves. In addition, H&C fundamentally miss the importance of intelligent flow control as a vital process design lever, and they fail to acknowledge it as a distinct category, although most of their principles implicitly require it in order for them to be of benefit.

BPR, in its essence, represents a very useful enrichment of the professional discipline of OM, and in order to achieve full value for the practitioner, it needs the evaluation and modeling tools that OM provides. In addition, OM has developed a qualitative theory of how capacity, variability, and dynamic flow control together determine response time performance. Such a qualitative theory can help to build intuition about design tradeoffs in BPR efforts, as is demonstrated with the example of integrated work in the next section.

Benefits of Integrated Work

In Chapter 2, H&C describe a radical redesign of the core business process at IBM Credit Corporation, whose 'jobs' to be processed were requests from customers for credit to finance their purchases. The credit authorization process originally involved steps performed sequentially by dedicated specialists, with queuing at each stage (see System 1 in Figure 5). For example, one specialist would check the potential borrower's creditworthiness, another would prepare a properly specialized loan covenant, another would price the loan, and so on.

In the redesigned process, individual workers were given comprehensive responsibility for structuring loans, supported by a sophisticated computer system that gave access to all the relevant databases and further provided technical guidance of various

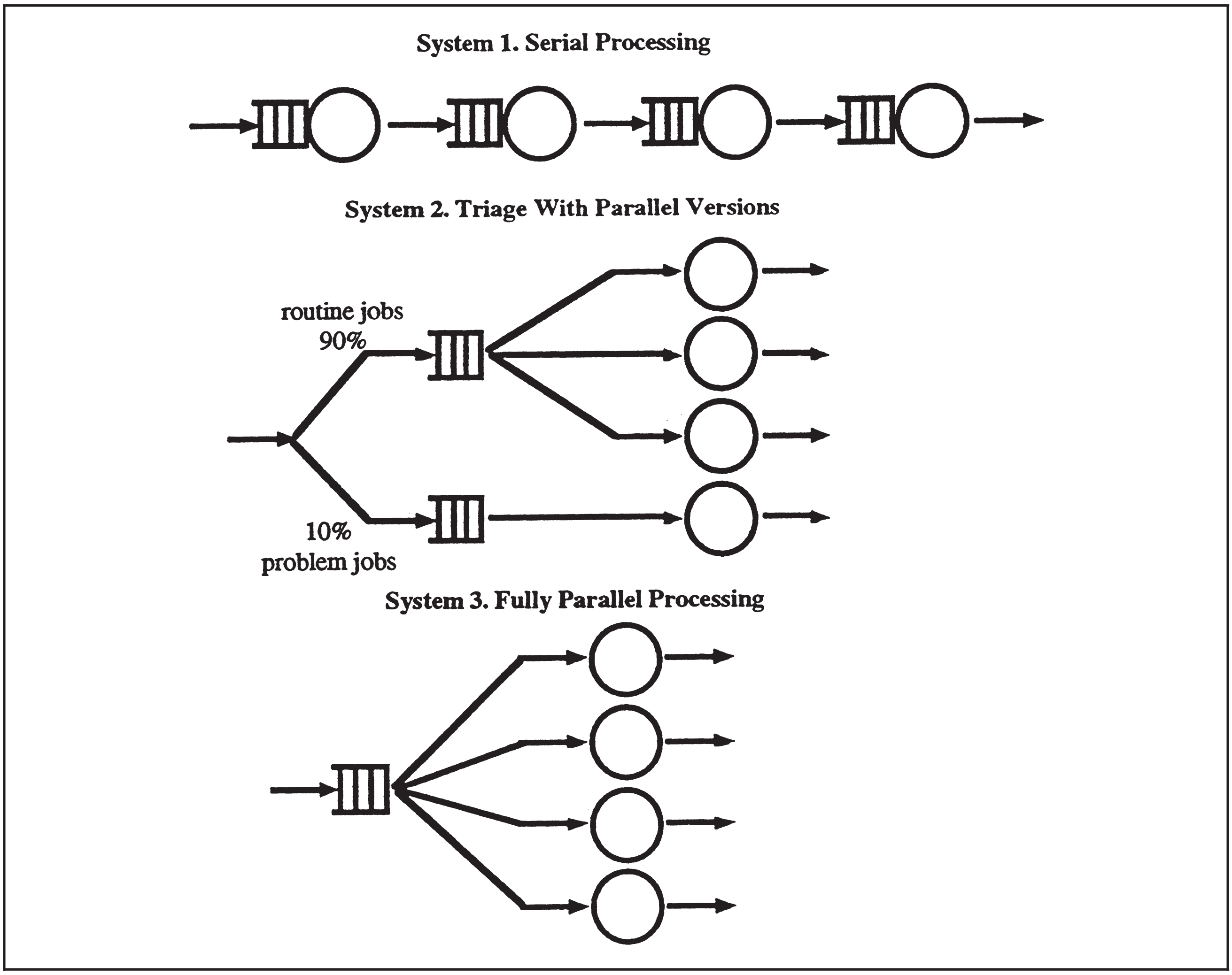


Figure 5 Three Processing Networks With Pooling

kinds.[9] The resulting process architecture had a single queue of waiting requests and a number of essentially interchangeable employees who worked in parallel. However, the process was not fully parallel (as System 3 in Figure 5 is), because provisions had to be made for redirecting difficult cases to a small pool of workers having expertise in specific functions (System 2 in Figure 5). That is, most workers became loan-structuring generalists in the new process, and, as they could not hope to handle all aspects of the most difficult customer requests, a small pool of 'specialists' was available to provide help as needed. However, the generalists were able to handle most requests without such help, and H&C report dramatic performance improvements: average throughput time was cut from seven days to four hours, and labor productivity (presumably this means the number of requests processed per day or week per employee) increased by a factor of 100.

In their discussion of IBM credit and other examples, H&C suggest that a process configuration like that depicted in Systems 2 or 3 in Figure 5, where individual workers or work teams have comprehensive responsibility, is always preferable to a sequential process configuration as shown in System 1. But there are obviously tradeoffs to be evaluated. For example, integrated work requires broader skills and greater judgment from employees, and presumably such workers can command higher wages. Integrated work also requires more extensive training, which makes employee turnover more damaging. Moreover, one cannot simply dismiss Adam Smith's observation that workers specializing in a single task become highly proficient at that task. Volvo and other companies have repeatedly tried broadening job responsibilities in their auto assembly plants, and these experiments have been consistent failures, cf. Adler and Cole (1993). Finally, making specific reference to the IBM credit example, one can well imagine that it cost a small fortune to develop and maintain the sophisticated, easy-to-use computer system required to support loan-structuring generalists, and such expenditures must be balanced against whatever benefits are derived from the broadening of worker responsibilities.

Figure 6 compares the three systems from Figure 5 with respect to the generalist work characteristics required. The 'traditional' serial system has workers

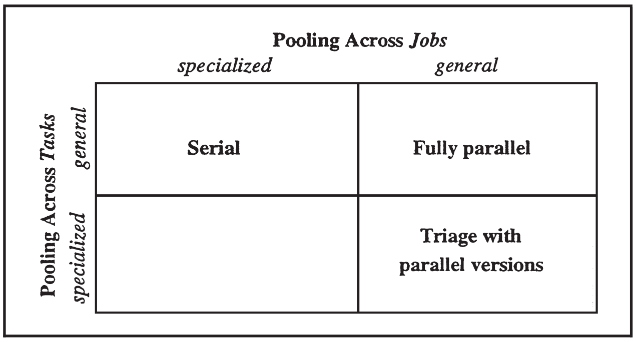


Figure 6 Types of Pooling in H&C

specialized in their tasks (every worker performs only one step), although they are able to handle all types of jobs processed in the system. Triage requires workers to be cross-trained for all tasks, but reduces their job generality by sending complicated cases to an expert. The fully parallel system requires workers to be full generalists with respect to jobs and tasks. Is more generalization always better?

In order to explore this question more rigorously with OM-style evaluation methodology, we analyze the systems in Figure 6 with a simulated numerical example. Suppose that jobs arrive at a rate of $\lambda = 0.75$ per hr. Each arrival has four tasks, or operations, to be done, the durations of which are independent of one another and are hyperexponentially distributed: with 97.4 per cent probability, a task is 'trouble-free' with an exponentially distributed duration of mean one hour, and with 2.6 per cent probability, a task has a 'problem,' implying a mean duration of 8.69 hr. These data imply that overall, 90 per cent of all jobs are routine, or trouble-free in all four operations, and 10 per cent of all jobs are non-routine, with at least one problem task. As a result, the overall average task duration is 1.2 hr (and thus the overall average processing time per job, 4.8 hr), and the utilization of all servers in System 1 is 90 per cent. The reader may note that although only 10 per cent of all arriving jobs are non-routine, they account for 25 per cent of the total workload. This is a fairly typical workload distribution in processing systems.

In System 2, workers are pooled across operations, but all non-routine tasks are identified at the beginning and routed to a specialist.[10] That is, each routine worker performs four operations in series with an exponentially distributed task time of mean 1, while the specialist performs four operations in series with a hyperexponentially distributed task time of mean 3.[11] In System 3, all workers are fully cross-trained, capable of handling all jobs. They perform four operations with hyperexponentially distributed task times in series.

Table 2 summarizes the average throughput times and standard deviations for the three systems (and of variants with priorities, as discussed below), based on simulations (25 independent runs of 30,000 customers each). The results indeed confirm H&C's claims: the serial system performs very badly, with a throughput time that is over twenty times as long as the pure processing time, while triage and fully parallel processing reduce the throughput time to four and three times pure processing, respectively. So, what is the value of doing OM-style process evaluation in addition to H&C's design principles? We see three important benefits of OM methodology, which we discuss in turn:

- a precise estimation of what the benefits of pooling are and what their size is
- recognition of additional design principles (see Table 1). A particular principle not at all acknowledged by H&C is flow control
- a comprehensive qualitative understanding of tradeoffs involved in work integration.

First, OM methodology allows not only ranking the systems in terms of their performance, but also estimating the extent of the benefit, as demonstrated in Table 2.[12] In addition, OM methodology allows assessing the whole throughput time distribution, which is important in comparing service levels: Figure 7 shows that System 3 not only has a lower mean throughput time, compared to System 1, but the whole distribution is shifted to the left. For example, the 95 per cent percentile is at 70 hr for the fully parallel system, but at 210 hr for the serial system, a huge difference for the purpose of offering service levels to customers.

Second, Systems l(a), 2 and 3(a) in Table 2 assume simple first-in-first-out sequencing of jobs at every station. H&C do not mention or acknowledge flow control as a process improvement lever, but the OM discipline has proposed many ways of flow control to improve throughput time performance of processing networks, ranging from simple rules such as the shortest processing time sequencing, over static priorities, to sophisticated system-status-dependent routing of jobs. As an example, we assume in Systems l(b) and 3(b) that lower priority be given to problem-tasks at every step. Table 2 shows that, in the example, the response time improvements from such priorities are significant and amount to over 50%. Priority schemes have the advantage that they can often be implemented very quickly, without the training or IT systems investment required for redesigned processes with triage or fully parallel processing. If, as in the example above, priority sequencing achieves a large part of the response time benefit offered by the redesign, it may even be preferable to avoid the cost and risk of a major reengineering effort and simply continue with improved flow control.

Third, the discipline of OM has developed a large body of knowledge about the tradeoffs of pooling and work integration, which can help to evaluate the above-mentioned tradeoffs neglected by H&C. The

Table 2 Throughput Time Comparison of the Pooled Systems

Processing network	Average throughput time (hr)	Throughput time standard deviation (hr)
1(a) Serial	110.6	22.90
1(b) Serial with priorities	49.2	4.10
2 Triage with parallel versions	17.8	3.20
3(a) Fully parallel processing	13.1	2.20
3(b) Fully parallel processing with priorities	5.2	0.04

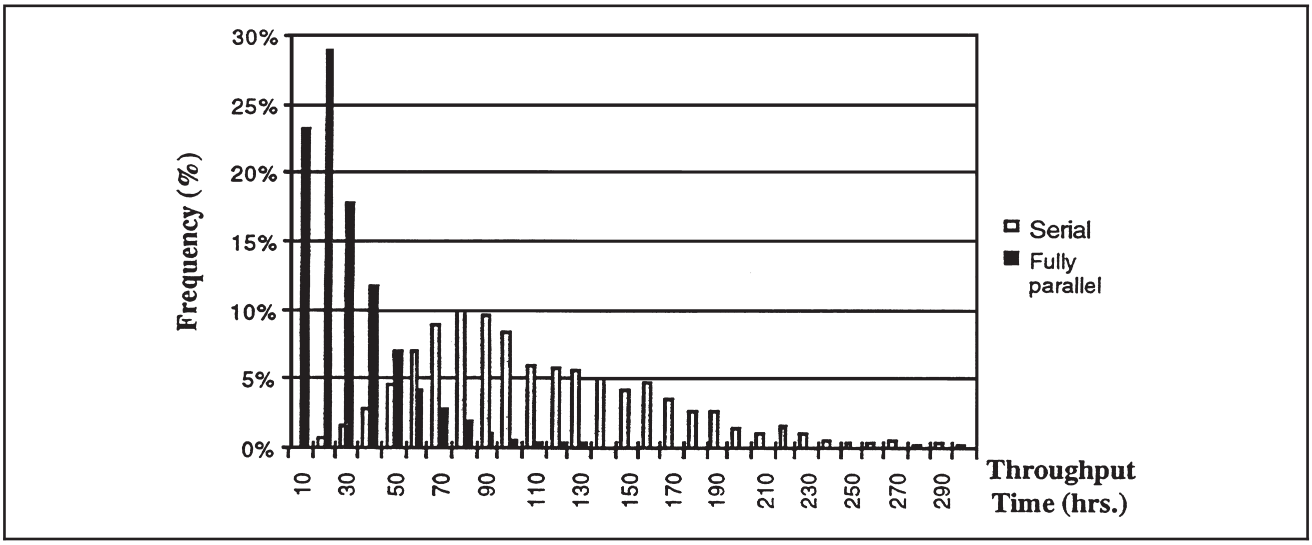


Figure 7 Throughput Time Distributions

fundamental benefit from pooling consists of making capacity use more efficient by preventing the possibility of one resource sitting idle while another faces a backlog of work, thus equalizing the utilizations of the pooled resources and eliminating bottlenecks (this has been shown for quite general arrival and service processes, e.g. Smith and Whitt, 1981).

On the other hand, the pooling benefit may not be as high as hoped because even in a non-pooled system, customers often 'jockey' from long queues to shorter queues, which basically results in response times identical to a pooled system. In the IBM example, this may not be possible, but it is worth contemplating whether one could implement some managed 'jockeying of jobs' across workers even in the case of claim processing (Rothkopf and Rech, 1987). In addition, pooling may have obvious downside effects, because the pooled resources require broader capabilities (e.g. cross-trained workers or universal machines), which may make them more expensive and/or less efficient in processing. Similarly, the switching back and forth between different job types may introduce additional setups. The efficiency loss may be quite subtle, such as through lower accountability or weaker identification of a server with a specific set of jobs or customers. Moreover, in a customer service context, pooling may deprive management of motivating desired customer behavior by offering a fast server for special jobs ('fast checkout lane for fewer than 10 items in a supermarket'). The net effect of pooling must, therefore, be decided on a case-by-case basis (Rothkopf and Rech, 1987).

But even if no cost or efficiency issues arise, there may be a more subtle downside. In a simple system such as our example above, with tasks following in series, pooling is always advantageous (in the absence of efficiency issues, see Mandelbaum and Reiman, 1996), but the benefit decreases markedly with the variability in the system: if processing is highly regular, Systems 2 and 3 above will yield only marginal advantages over System 1. In addition, in more general flow configurations (such as parallel tasks or loops, for example), pooling may actually *worsen* response time because the mixing of different job types increases the effective variability (Smith and Whitt, 1981; Mandelbaum and Reiman, 1996, p. 12). Again, the net benefit must be assessed on a case-by-case basis.

Finally, pooling servers with very different speeds may hinder, rather than help, response time performance, because it may actually be faster for a job to wait for the faster server and then get processed quickly rather than go to the immediately available slow server (Rubinovitch, 1985a). However, this problem can be overcome by intelligent flow control, which allows the customer to choose between direct service and waiting, based on system status (Rubinovitch, 1985b).

In summary, the example of integrated work in this section has shown that the design principles advocated by H&C may very well be valuable, but three questions must be resolved before such principles should be implemented: first, the extent of the benefit does not necessarily justify the investment and must therefore be estimated based on knowledge of cause-and-effect relationships. In particular, there may be situations where the tradeoffs in work integration are such that there is no benefit. Finally, flow control, such as simple priorities, may be used to get a similar benefit more cheaply, or to enhance the benefits gained from integration.

Summary and Conclusion

This paper has argued that BPR has made an important contribution to OM by elevating processes to the center of management attention, and by focusing on simple process design principles that are easily communicated and implemented in practical situations. However, BPR lacks systematic cause-and-effect reasoning in its design recommendations, and it misses a key design lever, namely flow management protocols, which is implicitly required in order for many BPR principles to work. The OM discipline has developed a set of design principles and process evaluation tools that can be used to elevate the BPR method to higher professional rigor.

Unfortunately, BPR has in most of its practical use become a downsizing and cost cutting tool. As companies are beginning to shift their focus from raising efficiency to generating new growth, BPR has now run its course and is disappearing from the list of corporate priorities. But process performance improvements can be targeted toward response time or service level improvements just as well as for efficiency gains. Therefore, there is reason to believe that the lessons from BPR will stay with the OM discipline, and our challenge is to demonstrate to the management community that process design is a powerful weapon to be turned toward service performance and growth generation.

Notes

1. Other well-known publications on BPR include Kaplan and Murdock (1991), Harrington (1991) and CSC Index (1994). The most important BPR work from the IT perspective is Davenport and Short (1990), later extended in Davenport's 1993 book *Process Innovation*. None has been as influential as H&C and its precursor, Hammer (1990).
2. Change management is important and difficult, which prompted a subsequent book mainly devoted to this topic (Hammer and Stanton 1995).
3. While being consistent with the quality movement, BPR puts greater emphasis on the discontinuous and cross-departmental character of improvements. For a discussion, see Cole (1994).
4. See, e.g. Blackburn and Van Wassenhove (1995) and Kanigel (1997), who demonstrates how Frederick Taylor himself was already concerned with processes, not only with individual tasks.
5. Note that H&C use the term 'task' differently from us, namely in the sense of the combination of a task, as in Section 2, and a worker (resource). In the language of the processing network, this principle reads: 'Assign several tasks to the same resource.' Throughout the paper, we use the term 'task' in this sense.
6. This theme is further developed in Hammer (1996), chap. 3 and 13.
7. Hammer (1996) writes (p. 194 f.): 'It is a fundamental principle of the process-centered organization that execution is key. What a company *does* is central to deciding what it is, and where and how it should compete.' Michael Porter has recently argued that processes are *not* everything in strategy because they focus on efficiency rather than competitive advantage (Porter, 1996). There can, however, be no doubt that the first few to harness powerful process improvements will reap a competitive advantage.
8. Davenport and Short (1990) and Davenport (1993) have made a significant contribution in this area.
9. The discussion of integrated work is not new, it has a long history in OM under the name of work enrichment (see, e.g. Wild, 1971, chap. 7). However, this discussion centered on work satisfaction and mentioned productivity only on the side. H&C introduced the 'quantum leap' performance potential of work integration.
10. Note that this is only possible if the problems can indeed be identified at the beginning, when the job arrives.
11. This is a different conceptualization of triage than the one chosen by Buzacott (1996): he interprets the specialist in our System 2 as a 'manager' making decisions, while our System 3 in his view corresponds to full delegation where workers 'make decisions' (p. 773). We believe that this interpretation does not correspond in essence to H&C's triage principle.
12. Note in particular that the first of H&C's claims is supported, which is a significant throughput time reduction. We are, however, at a loss to explain the productivity improvement by a factor of 100 without the use of heavy automation, which H&C do not mention in their example.

References

Adler, P.S. and Cole, R.E. (1993) Designed for learning: a tale of two auto plants. *Sloan Management Review* **Spring**, 85–94.

Adler, P.S., Mandelbaum, A., Nguyen, V. and Schwerer, E. (1995) From project to process management: an empirically based framework for analyzing product development time. *Management Science* **41**, 458–484.

Adler, P.S., Mandelbaum, A., Nguyen, V. and Schwerer, E. (1996) Getting the most out of your development process. *Harvard Business Review* **Mar/Apr.**, 134–152.

Blackburn, J.D. and Van Wassenhove, L.N. (1995) From industrial engineering to just-in-time to time-based competition to business process reengineering: coming full circle. INSEAD working paper.

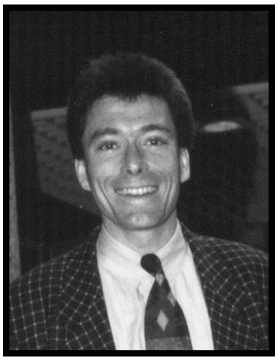
Buzacott, J.A. (1996) Commonalities in reengineered business processes: models and issues. *Management Science* **42**, 768–782.

Cole, R.E. (1994) Reengineering the corporation: a review essay. *Quality Management Journal* **July**, 77–85.

CSC Index (1994) *State of the Reengineering Report*. Cambridge.

Davenport, T.H. and Short, J.E. (1990) The new industrial engineering: information technology and business process redesign. *Sloan Management Review* **Summer**, 11–27.

Davenport, T.H. (1993) *Process Innovation: Rengineering Work Through Information Technology*. Harvard Business School Press, Boston.
Davenport, T.H. and Stoddard, D. (1994) Reengineering: business change of mythic proportions?. *MS Quarterly* **18**, 121–127.
Fitzsimmons, J.A. and Fitzsimmons, M.J. (1994) *Service Management for Competitive Advantage*. McGraw Hill, New York.
Goldratt, E. and Cox, J. (1989) *The Goal*. Gower Publishing, Aldershot, UK.
Hammer, M. (1990) Reengineering work: don't automate, obliterate. *Harvard Business Review* **Jul/Aug.**, 104–112.
Hammer, M. (1996) *Beyond Reengineering*. Harper Business, New York.
Hammer, M. and Champy, J. (1993) *Reengineering the Corporation: A Manifesto for Business Revolution*. Harper Business, New York.
Hammer, M. and Stanton, S.A. (1995) *The Reengineering Revolution*. Harper Business, New York.
Harrington, H.J. (1991) *Business Process Improvement*. McGraw Hill, New York.
Harrison, J.M. (1997) The processing network paradigm in operations management. Working paper, Graduate School of Business, Stanford University (in preparation).
Hopp, W.J., Spearman, M.L. and Woodruff, D.L. (1990) Practical strategies for lead time reduction. *Manufacturing Review* **3**, 78–84.
Juran, J.M. (1989) *Juran on Leadership for Quality*. The Free Press, New York.
Kaplan, R.B. and Murdock, L. (1991) Core process redesign. *The McKinsey Quarterly* **Summer**, 27–43.
Kanigel, R. (1997) *The One Best Way – Frederick Winslow Taylor and the Enigma of Efficiency*. Viking, New York.
Lee, H.L. and Billington, C. (1992) Managing supply chain inventory: pitfalls and opportunities. *Sloan Management Review* **Spring**, 65–73.
Mandelbaum, A. and Reiman, M.I. (1996) On pooling in queueing networks. Working paper Technion/Bell Labs, forthcoming in *Management Science*.
Porter, M.E. (1996) What is strategy?. *Harvard Business Review* **Nov/Dec.**, 61–78.
Rothkopf, M.H. and Rech, P. (1987) Perspectives on queues: combining queues is not always beneficial. *Operations Research* **35**, 906–909.
Rubinovitch, M. (1985a) The slow server problem. *Journal of Applied Probability* **22**, 205–213.
Rubinovitch, M. (1985b) The slow server problem: a queue with stalling. *Journal of Applied Probability* **22**, 879–892.
Smith, D.R. and Whitt, W. (1981) Resource sharing efficiency in traffic systems. *Bell System Technical Journal* **60**, 39–55.
Wheelwright, S.C. and Clark, K.B. (1992) *Revolutionizing Product Development*. The Free Press, New York.
Whitt, W. (1992) Understanding the efficiency of multi-server service systems. *Management Science* **38**, 708–723.
Wild, R. (1971) *The Techniques of Production Management*. Holt, Rinehart and Winston, London.



CHRISTOPH LOCH, *INSEAD, Boulevard de Constance, 77305, Fontainebleau, Cedex, France.*

Dr. Christoph Loch is Associate Professor of Operations Management in the Technology Management Group at INSEAD. His research and teaching are concerned with the management of new product and process development. He is also involved in a multi-year study of management quality in manufacturing plants, a co-operation between INSEAD's Operations Group, and the magazines L'Usine Nouvelle *(France) and* Wirtschaftswoche *(Germany).*