

**The Improvement Paradox:  
Three Essays on Process Improvement Initiatives**

by

Nelson P. Repenning

B.A., Economics, The Colorado College, 1990

Submitted to the Alfred P. Sloan School of Management  
in partial fulfillment of the requirements for the Degree of

Doctor of Philosophy in Management

at the

Massachusetts Institute of Technology

June 1996

**ARCHIVES**

MASSACHUSETTS INSTITUTE  
OF TECHNOLOGY

**JUN 03 1996**

LIBRARIES

© 1996 Massachusetts Institute of Technology  
All Rights Reserved

Signature of  
Author.....

Department of Operations Management and System Dynamics  
May 1996

Certified by.....

John D. Sterman  
Professor of Management  
Thesis Supervisor

Accepted by.....

Birger Wernerfelt, Chair, Ph.D. Committee  
Sloan School of Management

# **The Improvement Paradox: Three Essays on Process Improvement Initiatives**

by

Nelson P. Repenning

Submitted to the Alfred P. Sloan School of Management  
in partial fulfillment for the Degree of  
Doctor of Philosophy in Management

## **Abstract**

### **Essay #1: Modeling the Failure of Productivity Improvement Programs**

This paper develops a simple model of a manufacturing firm in which a successful productivity improvement program is implemented. The model is used to show how a successful improvement program can fail to significantly improve a firm's financial performance. It is argued that the potential rates of improvement in the firm's capabilities can differ substantially based on the intrinsic complexity of those processes. The spread of improvement skills and commitment to the effort is modeled as a diffusion process among employees in a given area. The allocation of resources to support that commitment is represented as a dynamic adjustment process. The formulation, with the assumption of locally rational decision rules, results in differential rates of improvement in the capacity and demand generating areas of the firm. If excess capacity results, interactions with traditional accounting, pricing, and human resource policies can create unanticipated side effects that result in sub-standard performance or failure of the program. Policies for mitigating these problems are discussed and analyzed.

### **Essay #2: Agency Problems in Process Improvement Efforts**

In this paper I study the problem faced by a firm that tries to induce its workforce to reveal information leading to productivity improvements when those improvements may lead to lay-offs or 'downsizing'. The analysis begins with a discussion of the conditions under which productivity improvements are likely to lead to lay-offs. I then develop a model in which the firm attempts to extract productivity improving information from its workforce by providing monetary incentives for such revelations. The impact of different contractual and institutional assumptions on the firm's ability to implement such programs is investigated. There are two main results of the analysis. First, the employees' ability to collude or participate in binding side agreements – to write contracts with each other or to join a union – is a critical determinant of the firm's cost of implementing new programs. Second, the program's perceived impact on the firm's survival strongly influences the firm's cost and the ability of employees to profitably collude. These results allow me to explain the differing experiences of firms that use such programs, and to generate some insight into the effect that a firm's financial health has on its ability to implement programs like TQM.

### **Essay #3: A Tale of Two Improvement Efforts: Towards a Theory of Process Improvement and Redesign**

The purpose of this paper is to lay the foundation for a theory of process improvement and redesign that accounts for both the physical *and* the behavioral components of the

environment in which improvement is taking place. The main tools for theory development are intensive case study research, the development of stock/flow and feedback diagrams, and the analysis of existing literature. The results from two intensive case analyses of process improvement efforts with a major U.S. manufacturing company are reported. The main thrust of the argument is that, contrary to the popular conception, TQM and re-engineering are complementary activities, and a more general improvement and redesign methodology can be developed using precepts from each theory. TQM offers an organizational structure and decision making methodology, and re-engineering provides a tool for challenging the dominant mental models that guide the organization.

Thesis Supervisors: John D. Sterman (chair)  
Professor of Management Science

Stephen C. Graves  
Professor of Management Science

Julio J. Rotemberg  
Professor of Applied Economics

Dr. Roger B. Saillant  
Director of Technology and Process Development  
Automotive Components Division, The Ford Motor Company

## Acknowledgments

Although it may not have always appeared so, I enjoyed writing this thesis. The quality of the experience is due in large part to the people with whom I have had contact during the process.

I had the “dream team” of thesis committees. While the phrase “an economist, a system dynamicist, a management scientist, and a manager were all sitting in a room together...” sounds like the beginning of a bad joke, in my case it was the beginning of a enjoyable research effort. My thesis utilized a diverse set of tools, and it is a testament to the open minds of my committee that the research proceeded without a hitch.

Roger Saillant took time away from his very busy schedule to read drafts and give me detailed feedback. Besides, opening doors within the research site and providing wide access to his organization, Roger made a heroic effort to read every paper, work through every equation, and keep me focused on the concerns and problems faced by real managers. I hope this is just the beginning of a long and productive relationship.

Julio Rotemberg spent the year teaching me economics and convincing me that my model did not need one more equation. He too had to cope with my peculiar combination of disciplines, and his contributions added a critical dimension to the thesis that would not have been there otherwise.

Although much of my work strayed far from his normal domain of expertise, Steve Graves contributed many valuable comments and insights. He often suggested a simpler and more satisfying interpretation to the problems I faced, and I am very grateful for his contributions.

You couldn't ask for a better teacher, advisor, thesis chair, friend, or colleague than John Sterman. John took a chance in allowing me to come to MIT and then admitting me to the doctoral program, and for that I am eternally grateful. Since then John has supported every aspect of my work whether it be discussing new ideas, critiquing models, editing papers, or developing talks. He has contributed to every facet of this research. My thesis is a direct product of our ongoing collaboration and I truly hope that it will continue for many years to come.

Beyond my committee, many many people provided invaluable help along way. The third essay would not have been possible without the help and support of many people at the research site. Thanks to Dave Lazor, Vic Leo, Frank Murdock, Ron Smith, and Bill Colwell, who have all played important roles. A special thanks to Tim Tiernan whose substantial investments, far above and beyond the call of duty, have kept the research project going. An extra special thanks to Laura Cranmer, who read countless drafts, provided detailed feedback, guided me through the strange language and customs of corporate America, and was always willing to give me much needed advice and counsel. I am very grateful for her help and support.

Thanks to Chris Griffiths, Walt Hecox, and Bill Weida for piquing my interest in the study of economics and management and encouraging me to go on. Special thanks to Esther Redmount for giving me an early push, and an extra special thanks to Mark Paich who introduced me to System Dynamics and MIT. Bob Pizzi also had a critical influence on my life, and, were it in my power, the only thing I would change about the entire experience would be to have him still with us. I am sure that he is now comfortably ensconced in a higher department and, having done a few favors for Saint Peter, is probably already a full professor. However, I miss him just the same.

The OM doctoral students were great. Geoff Parker, Nitin Joglekar, Julic Gomes, Luis Pedrosa and Sharon Novak all have provided valuable feedback and much needed encouragement. The SD/OM students were my social life support system. Anjali Sastry and I had many valuable conversations about System Dynamics, its relation to other fields, and whether or not our advisor had secretly been replaced by an alien life form (it made more sense at the time). Rogelio Oliva read drafts, provided timely and valuable feedback, and was the anchor in our Friday afternoon get togethers. Tom Fiddaman, Ed Anderson, and Scott Rockart have all been great. A special thanks to Elizabeth Kraemer who provided detailed, comma specific, feedback on every page of this thesis. An extra special thanks to Drew Jones who enthusiastically took over the Harley-Davidson research project so that I could finish.

Mom and Dad, besides just being Mom and Dad (jobs at which they excel), were always there to remind that this wasn't the first or last challenge that I would face in life. I hope this isn't a thesis that only a mother (or father) could love, but it is comforting to have that as a fallback position. To this day I smile when I remember my brothers accusing me of joining 'the enemy' when I told them I wanted to be a professor. Since then Win, John, and Jamie have continued to keep me in good spirits.

My wife-to-be Elizabeth Saltonstall makes a mean snickerdoodle (the world's greatest cookie), and was uncanny in her ability to provide a batch when I needed them most. She read every paper multiple times and always gave useful feedback. Elizabeth was also my 'personal trainer', telling me when to take a break, when to lighten up, and when it was time to go to the movies. I fear that she will need to continue in this role, but, fortunately, she is very good at it. I couldn't have done it without her, and even if I could have, it wouldn't have been any fun.

Finally, I dedicate this thesis to my grandmother, Rena Buck Robinson, as a token of appreciation for her fine example and generous support.

## Table of Contents

### Essay #1:

#### **Modeling the Failure of Productivity Improvement Programs**

1. Introduction.....	7
2. The Improvement Process.....	9
3. Supporting the Improvement Effort.....	16
4. Factor Acquisition and Pricing Under Unbalanced Improvement:.....	20
5. Firm Level Effects of Improvement.....	22
6. The Firm with Endogenous Improvement.....	25
7. Discussion.....	28
8. Conclusion.....	29
9. Tables.....	33
10. Figures.....	35

### Essay #2:

#### **Agency Problems in Process Improvement Efforts**

1. Introduction.....	56
2. When Do Improvements Yield Lay-offs?.....	59
3. The Model.....	61
4. Analysis.....	66
5. Collusion and Side Contracts.....	75
6. The Role of Fear.....	79
7. Discussion.....	82
8. Conclusion.....	85
9. Appendix.....	89
10. Figures.....	92

### Essay #3

#### **A Tale of Two Improvement Efforts:**

#### **Towards a Theory of Process Improvement and Redesign**

1. Introduction.....	94
2. Related Literature.....	98
3. An Expanded Framework.....	103
4. Two Improvement Efforts.....	112
5. Analysis and Discussion.....	128
6. Conclusions and Future Directions.....	136
7. Figures.....	144
8. Tables.....	155

## Essay #1

### Modeling the Failure of Productivity Improvement Programs

#### 1. Introduction-The Improvement Paradox

The tools and ideas embodied in the philosophy of Total Quality Management (TQM) have been widely accepted by both managers and academics. Companies such as Motorola and Xerox, as well as many Japanese firms, attribute much of their success to their on-going quality programs. Advocates of TQM and similar programs claim that, in the near future, the ability to improve continuously will be a prerequisite for any firm's success (Deming 1986; Stata 1989; Shiba, Walden, and Graham 1993). Easton and Jarrell (1995) find that firms making a long term commitment to TQM outperform their competitors in terms of both profitability and share price. Yet, despite the widely publicized success stories and the exhortations of TQM 'gurus', many firms do not receive significant pecuniary benefits from their TQM programs. A study by Ernst and Young (1991) concludes that few companies that use TQM experience a significant change in profitability, and another by Arthur D. Little reports that only a third of companies studied felt that their quality programs have had a significant impact on competitiveness (The Economist 1992). A study of Baldrige award finalists by the US General Accounting Office concluded that, while it did significantly improve quality and productivity, TQM did little to improve the returns on sales or assets (GAO 1991). TQM has been demonstrated to be a powerful tool for improving both quality and productivity, but, paradoxically, many firms abandon their programs due to lack of perceived impact on profitability.

To date, these issues have received little formal analysis (Kim and Burchill 1992 is a notable exception), and few hypotheses have emerged to explain why TQM can be so successful in some organizations but not in others. Some writers in the popular press suggest that techniques such as TQM are little more than new management fads (Harte 1992, Taylor 1992), while academic research has focused primarily on issues concerning implementation (Kaufman 1992). Each explanation is certainly true in some cases; even the most successful management innovations are likely to be used in inappropriate contexts, and poor training, inadequate support, and general organizational resistance can limit the effectiveness of any new program. However, neither the hypothesis that TQM does not work, nor that it does, but is often "implemented incorrectly", explains the paradoxical experience of companies whose TQM programs, while generating significant improvements in both quality and productivity, produce little improvement in profitability, and, as a result, are abandoned (see Kaplan 1990 for an example).

In this paper I present an alternative hypothesis: The implementation of a successful TQM program (1) increases the dynamic complexity of the decisions faced by both management and labor, and (2) invalidates many of the traditional decision making and performance evaluation heuristics used by both those in the firm and those in external capital markets. The consequence of these changes is that implementing TQM in an environment characterized by uncertainty, asymmetric and incomplete information, and decision makers with limited computational and cognitive capability produces unanticipated side effects which can limit the impact of an otherwise useful improvement program. Building on an earlier study which analyzed one company's experience with TQM in depth (Sterman, Repenning and Kofman 1994), I explore this hypothesis by developing a dynamic model to analyze the firm-level effects of productivity improvement and to highlight the interactions that may limit the program's positive impacts.

The adoption and use of a program like TQM is a dynamic process of learning, adjustment, and adaptation that involves interactions between many organizational levels within the firm. Capturing these interactions and their resulting influences on quality, productivity, and profitability requires an explicitly disequilibrium perspective. The model developed here describes a firm consisting of a production technology, a demand curve, and three conceptually distinct groups of stakeholders; senior management, the labor force, and a staff of quality improvement 'experts' or trainers. Management is assumed to be responsible for factor acquisition and price setting, the support staff determines which functional areas within the firm receive resources to support the improvement effort, and the general labor force determines the amount of time and effort actually dedicated to the improvement process. In the spirit of Forrester (1961), Morecroft (1985), and Cyert and March (1992), each of the representative actors responds dynamically to changes in the environment using decision rules, or policies, that use locally available information, and are consistent with the cognitive and computational limitations of the decision maker. The rules are consistent with Simon's (1976) principle of bounded rationality, current research on human decision making in dynamic environments (Paich and Sterman 1993, Sterman 1989a 1989b), and available knowledge concerning the historical performance of actual agents in similar situations (Sterman *et al.* 1994, Kaplan 1990). The resulting model, when simulated, suggests that the introduction of an improvement program in an environment of decentralized decision making may result in unexpected outcomes which, if misinterpreted, may induce actions that result in the demise of an otherwise successful improvement program.

The paper is organized as follows: Section 2 begins with a simple representation of the core improvement process, a modified version of the half-life model suggested by Schneiderman (1988), and develops a simple framework to understand the rate at which a particular process



improves. The framework is then augmented by an explicit representation of the diffusion of skills. In Section 3 the allocation of resources to support the improvement effort is considered. Section 4 analyzes a static model of a monopolistic firm which experiences exogenous improvement in its productive capabilities. A simple condition under which this improvement will generate excess capacity is presented. Section 5 presents a dynamic re-formulation of the model presented in section 4 with extensions to better represent the factor acquisition and pricing setting functions. Section 6 combines the models in Sections 3 and 5. Section 7 discusses the results and their implications, while section 8 presents final conclusions and possible directions for future research.

## 2. The Improvement Process

In the model the labor force is assumed to have primary responsibility for applying the appropriate improvement tools to the firm's various processes. The rate of improvement of each process depends upon (1) the inherent complexity of the process and (2) the willingness and ability of the workforce to use the appropriate tools for improvement. The inherent complexity of the process is represented using a model and supporting conceptual framework suggested by Schneiderman (1988). The model is then augmented with an explicit characterization of the workforce's skill level and their willingness to participate in the program.

### **2.1 Inherent Process Complexity- The Half-Life Model of Improvement**

The core improvement process is modeled using an empirical regularity documented by Schneiderman (1988):

...Any defect level, subjected to legitimate QIP [Quality Improvement Process] decreases at a constant (fractional) rate so that when plotted on semi-log paper, it falls on a straight line.

Here a defect is broadly defined as any measurable, undesirable, component in the process of bringing a product to market. Defective products, late deliveries, and long product development times are all considered defects. Schneiderman's observation, labeled the Half-Life Model since the time required for any defect measure to fall by 50% is constant, translates to a first order differential equation describing the time path of defect measure  $i$ .

$$\frac{dD_i}{dt} = -\phi_i D_i$$

The parameter  $\phi_i$  is specific to each defect generating process. Schneiderman (1992) suggests a framework to predict which processes will be easy to improve, implying a large value for  $\phi_i$  and a short half-life, and which processes will be difficult to improve, yielding a small  $\phi_i$  and a long

half-life. A defect generating process can be ranked along two dimensions; technical complexity and organizational complexity. Technical complexity refers to the state of knowledge concerning the particular process, while organizational complexity is a function of the number of organizational boundaries spanned by the particular process. When these two dimensions are represented graphically (Figure 2.1), processes with the shortest improvement half-lives will reside near the origin. These processes are technically well understood and cross few organizational boundaries. Standard TQM tools can be readily applied, experiments are easily performed and analyzed, and adjustments quickly made. Conversely, processes located to the northeast of the origin will be more difficult to improve. They are technically less well understood, so standard tools are more difficult to apply, and they cross numerous organizational boundaries, so any changes are time consuming to implement since they require input from numerous people.

Empirical evidence suggests that processes with short half-lives are more likely to be associated with direct manufacturing where the technology is relatively well understood and few, if any, organizational boundaries are crossed. Conversely, processes in areas such as product development and administration are likely to improve more slowly (Schneiderman 1988, Kaplan 1990) because they are either technically complex, as in basic research, or organizationally complex, as in sales and marketing, or both, as in product development. Further, the difference is compounded by the fact that much of the accumulated improvement experience is with processes in direct manufacturing. The applications of these tools to other areas is a relatively recent development.

## 2.2 Willingness and Ability to Use Improvement Tools- A Diffusion Model

The half-life model rests on the assumption that effort directed toward the improvement program is constant, and that the participating team's facility with the appropriate tools is fixed. A more complete model of improvement is developed here to capture the effects of learning and the workforce's changing beliefs concerning the benefit of continuing to participate in the improvement effort.

Assuming the firm has  $n$  distinct areas, each containing  $i$  defect generating processes, two additions are made to the half-life equation.

$$\frac{dD_i}{dt} = -\phi_i(D_i - D_{i_{Min}}) \cdot C_n, \quad 0 < C_n < 1 \quad (2.1)$$

First, a theoretical minimum defect level is explicitly defined for each process. The rate of defect reduction in process  $i$  is then proportional to the current level minus the minimum value. The formulation can now represent a broader class of improvement efforts, such as reducing product

development time, for which the theoretical minimum level is not zero. Second, the improvement rate is also assumed to be a multiplicative function of  $C_n$ , a measure of the effective usage of improvement techniques in area  $n$ . The variable  $C_n$  is restricted to the zero-one interval and is defined as the percentage of the full time equivalent workforce in area  $n$  that has acquired the skills appropriate to the particular improvement program and is committed to participating in the improvement effort. Thus  $\phi_i$  becomes the rate of defect reduction achievable by a fully committed workforce.

The effective usage of the improvement tools, represented by  $C_n$ , combines two distinct concepts; ability and willingness. The distinction is important in understanding the improvement effort since, in many cases, the knowledge required to improve a particular process resides only with those who have contact with that process on a daily basis. Management can provide support and training, but they cannot do the actual work of improving. Further it is difficult for management to enforce effective participation in an improvement program since it is impractical to supervise the daily efforts of every QIP team. As a result, the workforce is only likely to participate seriously if they believe it is in their best interest to do so.

Improvement does not begin immediately with the implementation of a given program; time is required to develop and disseminate the appropriate skills. The workforce also needs to develop confidence that the methods in question actually work and that it is in their interest to use them. Many TQM advocates suggest diffusion contains a 'push' from management and a 'pull' from results (Shiba, Walden, and Graham 1993). The model incorporates both premises as drivers of commitment.

$$\frac{dC_n}{dt} = \theta(C^* - C_n) + w_n C_n (1 - C_n), \quad 0 < \theta < 1, \quad 0 < \mu < 1 \quad (2.2)$$

The first term on the right hand side of (2.2) represents the push. Management sets a target,  $C^*$ , assumed here to be 100%, for firm-wide participation in the improvement program. Absent pull effects, the actual effective usage level approaches management's target via a first order exponential adjustment. The average delay,  $1/\theta$ , represents the time required for management to teach the workforce the new improvement tools and to enlist their participation in the program.

The second term of on the right hand side of (2.2) determines the strength and sign of the "pull" effects (Bass 1969, Homer 1987, Paich and Sterman 1993). Similar to its normal usage in the marketing literature, this process represents word of mouth—the spread of information through repeated contacts between those who have experience with the appropriate tools and those who do not. Early in a program's life, the population of committed users will be small, few contacts will

occur, and management's push will dominate. As committed usage increases, participants can evaluate the effectiveness of the program based on personal experience and the experience of colleagues with whom they make repeated contact.

However, word of mouth does not have to be positive. The workforce's preferences over whether or not the program should be continued are represented by the variable  $w_n$ , the 'sign and strength' of the word of mouth. If continuing the program is strongly preferred,  $w_n > 0$ , the committed portion of the workforce believes that the program is in their best interest and should be continued. If this is the case, additional contacts produce an increase in effective usage. If preferences are strongly against the program,  $w_n < 0$ , then those that are currently using the tools believe the program should be discontinued and additional contacts will produce a decline in effective usage.

Developing an accurate description for  $w_n$  is difficult because the process through which participants in an improvement program form beliefs about the expected benefit resulting from their continued participation is likely to be both complicated and highly subjective for each person involved. Reducing this process to an explicit mathematical representation is additionally complicated by the fact that much of the available information concerning the effect of changes in the surrounding environment on these beliefs is qualitative in nature. The strategy taken here is to first assume that  $w_n$  is determined by a linearly separable function (2.3) of three distinct pieces of information; the normalized rate of productivity improvement  $p_n$  ('does the program work?'), the current adequacy of resources to support the improvement effort  $a_n$  ('are our efforts being supported?; does this program increase my normal work level?'), and the perceived level of job security  $z$  ('will improvement cost me my job?'). The parameter  $\omega_n$  represents the intensity of communication in the particular area.

$$w_n = \omega_n [f_r\{p_n\} + f_a\{a_n\} + f_z\{z\}] \quad (2.3)$$

The assumption of linear separability allows the 'sign and strength' of word of mouth to be fully determined by simple inequalities when the three information streams are evaluated at the possible combinations of extreme values. Qualitative data (e.g. field studies and interviews) can then be used to determine the sign of each inequality, and upper and lower bounds for each function are chosen to satisfy the assumed relations. At intermediate points, each function, although separable from the other two, will, in general, be a complicated function of the given input. The choice of functional form is restricted to a class whose properties are consistent with the available qualitative information, and then scaled to the established bounds. Specifically, each function  $f_j\{\cdot\}$  is assumed to have the following general form:

$$f_j\{\cdot\} = (B_j^u - B_j^l)\varphi_j\{\cdot\} + B_j^l, \quad 0 \leq \varphi_j\{\cdot\} \leq 1$$

which, given the constraint on  $\varphi_j\{\cdot\}$ , is bounded from above and below by  $B_j^u$  and  $B_j^l$  respectively. These bounds are established using simple rank ordering arguments based on the beliefs generated by the three underlying information streams evaluated at their extreme values.

If a successful improvement program results in excess labor capacity, management may be tempted to cut costs by firing some portion of the workforce. Under the assumption that the workforce understands or fears that productivity improvements can result in excess capacity, effective usage of the improvement tools will be reduced if workers believe that further improvement will increase the probability of lay-offs or firings. The workforce is assumed to value job security above other factors since the loss of a job causes both financial and emotional difficulty, outweighing any positive benefits from support or results. As a result, when perceived job security is at its minimum, the lower bound  $B_z^l$ , it is assumed to dominate any positive effects of results or resource availability.

$$B_r^u + B_a^u + B_z^l < 0$$

Similarly, low resource availability is assumed to dominate any positive effects of results and/or perceived job security. If the work force believes it is not being supported both in terms of technical assistance and a reduction in normal work requirements so that they may participate in improvement activities, then any positive effects resulting from job security or results will be outweighed by frustration and the decline in utility resulting from an increased work requirement.

$$B_r^u + B_a^l + B_z^u < 0$$

Finally, results are assumed to be a necessary condition for positive word of mouth. Even if job security is high and resources are fully adequate, results play a key role in the formation of preferences over continuing the program. The workforce will not spend a significant amount of time pursuing an improvement program that has not demonstrated its usefulness. Thus job security and support cannot overcome the negative effect of poor results:

$$B_r^l + B_a^u + B_z^u < 0$$

If a program generates strong results, has adequate support, and job security is high, then word of mouth will be positive, so:

$$B_r^u + B_a^u + B_z^u > 0$$

With the extreme condition combinations established, the functions  $\varphi_r\{p_n\}$ ,  $\varphi_a\{a_n\}$ ,  $\varphi_z\{z\}$  need to be specified. A similar approach is used for each.

The rate of productivity improvement in area  $n$ ,  $P_n$ , is scaled by dividing by top management's goal for improvement  $P_n^*$ , assumed to be calculated using the simple half-life model. The model is specified so that the improvement rate never exceeds the prediction which implies that  $p_n$  is restricted to the zero-one interval.

$$p_n = \frac{P_n}{P_n^*} \quad (2.4)$$

The formulations is consistent with the "aspiration" concept of Cyert and March (1992) whereby performance is evaluated relative to an explicit goal or aspiration. The half-life concept was originally developed for the purpose of setting goals for quality improvement (Schneiderman 1988, Kaplan 1990). Qualitative information suggests that the effect of  $p_n$  on beliefs is monotonically increasing,  $f'_r\{p_n\} > 0$ . This gives an improvement program its true power—initial results demonstrate the validity of the approach and beget more results. Without this effect management would have a difficult time developing such program due to the substantial time required to individually enlist each member of the workforce in the program. In the neighborhood of  $p_n=1$  the second derivative is assumed to be strictly negative,  $f''_r\{1\} < 0$  since improvement measures are likely to be noisy and small deviations from the prediction will be discounted. These two conditions restrict the function to being either strictly concave and increasing or s-shaped and increasing. The s-shape is chosen,  $f'_r\{0\} > 0$ , and is represented by the logistic curve

$$f_r\{p_n\} = (B_r^u - B_r^l) \cdot \left( \frac{\exp(4\gamma(p_n - \delta_r))}{1 + \exp(4\gamma(p_n - \delta_r))} \right) + B_r^l \quad (2.5)$$

This specification takes the value  $B_r^u$  for  $p_n=1$  and the value  $B_r^l$  for  $p_n=0$ . The inflection point is at  $p_n=\delta$  and the slope at the inflection point is  $\gamma(B_r^u - B_r^l)$ . The inflection point is assumed to be at  $\delta=.5$ .

A similar procedure is used to specify the function that reflects the effect of resource availability. The total support resource requirement in area  $n$ ,  $r_n^*$ , is the product of the number of people in the area,  $L_n$ , the improvement resource requirement per person in area  $n$  assuming full participation,  $\rho_n$ , and the current commitment level in that area,  $C_n$ . The adequacy of resources,  $a_n$ , is the level of resources currently allocated to the area,  $r_n$ , divided by the resource requirement,  $r_n^*$ .

$$r_n^* = L_n \rho_n C_n \quad (2.6)$$

$$a_n = \frac{r_n}{r_n^*} \quad (2.7)$$

Workers, in order to participate effectively in an improvement program, require resources in the form of management's attention and a reduction in their current responsibilities. This suggests that the effect of resource adequacy on beliefs is monotonically increasing,  $f'_a\{a_n\} \geq 0$ . As

management increases its willingness to support the effort, workers become more committed. In the neighborhood of  $a_n=1$ , the second derivative is assumed to be negative,  $f''_a\{1\} < 0$ , implying a diminishing marginal return to additional support near the requirement level. These requirements restrict the functional form to being either strictly concave and increasing or s-shaped and increasing. Again, the s-shaped function is chosen,  $f'_a\{0\} > 0$ , and represented by the logistic curve with a similar parameterization.

$$f_a\{a_n\} = (B_a^u - B_a^l) \cdot \left( \frac{\exp(4\gamma(a_n - \delta_a))}{1 + \exp(4\gamma(a_n - \delta_a))} \right) + B_a^l \quad (2.8)$$

The inflection point is assumed to be at  $\delta=.5$ .

The workforce's belief in management's commitment to job security,  $z$ , is assumed to be solely a function of management's past actions. The effect of a change in perceived job security on beliefs is modeled as a function of the workforce's "memory" of the annual fractional lay-off rate,  $s$ .

$$z = 1 - s \quad (2.9)$$

A non-linear memory structure is used to represent the workforce's memory,  $s$ . When the annual percent lay-off rate,  $S$ , is greater than the workforce's current memory,  $s$ , the memory is updated very quickly, while when the converse is true, the memory is updated very slowly.<sup>1</sup>

$$\frac{ds}{dt} = \psi(S, s)(S - s) \quad (2.10)$$

$$\psi(S, s) = \begin{cases} \eta & \text{when } s > S \\ \nu & \text{when } s \leq S \end{cases}; \quad \nu \gg \eta \quad (2.11)$$

The effect of changes in job security on word of mouth is assumed to be monotonically increasing in perceived job security  $f'_z\{z\} \geq 0$ . In the neighborhood of  $z=1$ , the second derivative is assumed negative,  $f''_z\{0\} \leq 0$ , a decrease in job security reduces faith in management's commitment at an increasing rate. As in the two previous cases the s-shape is chosen with the inflection point at an annual lay-off rate of 20%,  $f'_z\{.8\} = 0$ . The logistic specification is chosen with properties identical to those selected previously.

$$f_z\{z\} = (B_z^u - B_z^l) \cdot \left( \frac{\exp(4\gamma(z - \delta_z))}{1 + \exp(4\gamma(z - \delta_z))} \right) + B_z^l \quad (2.12)$$

## 2.3 Partial Simulation

### Base Case

Assuming, momentarily, that there is a single defect generating process, that support resources are totally adequate, and job security is high, the introduction of an improvement program can be

---

<sup>1</sup>. This set-up should not be confused with the (s,S) policy from inventory theory.

simulated by initializing commitment to zero, and introducing, in month twelve, a unit step input to parameter  $C^*$ , management's target commitment level (see Table 2.4 for additional parametric assumptions). The basic stock and flow and feedback structure is shown in figure 2.2. The system has two levels, commitment and defects, both of which are governed by first order control (negative loops B1 and B2). The reinforcing nature of successful improvement is represented by the positive feedback loop R1: an increase in commitment increases the rate of defect reduction resulting in positive word-of-mouth and further increasing commitment. The simulation results are shown in figures 2.3, 2.4, and 2.5. The initial result is a rising commitment to improvement and a reduction in defects. Figure 2.3 shows that with this choice of parameters, management's initial push results in a small increase in commitment as initial worker skepticism retards diffusion. As the initial push begins to generate noticeable improvement, commitment increases. The reduction in defects further increases commitment in a positive feedback process. Eventually, the defect level approaches its minimum, and commitment declines as it becomes increasingly difficult to make additional improvements. Figures 2.4 and 2.5 show that defects in both areas fall and quickly approach the potential level.

### *Sensitivity*

A wide range of sensitivity tests have been performed in developing the model. A key parameter in this small system is the shape of the function  $f_r(\cdot)$  that determines the effect of results on commitment. Figure 2.6 shows some sample test inputs for this function. All functions are assumed to have the same left hand limit and the right hand limit,  $b$ , ranges from .5 to 4. The results of the six simulations are shown in figures 2.7. The slope of the curve and the right hand limit determine the gain of the positive feedback loop R1. For values greater than that of the base case, commitment rises more rapidly. For values less than the base case, commitment increases more slowly. For low values of  $b$ , commitment does not reach one, and for  $b=.5$ , the positive loop never overcomes the negative loops and commitment does not increase much beyond the initial level. Figure 2.8 shows the results from fifty monte carlo simulation in which both the left and right hand limits were drawn from uniform distributions on the  $[-1,0]$  and  $[1,4]$  intervals respectively. Although the inputs are uniformly distributed, the time path of commitment shows two modes. If the gain of the positive loop is strong, commitment rises rapidly, while if it is weak, the program never grows beyond the initial level of commitment engendered by management.

### 3. Supporting the Improvement Effort

The allocation of resources to different areas in the firm plays an important role in determining the final outcome of the improvement program. In this section a second organizational tier, the TQM



support staff, is added to the model. This group is composed of TQM ‘experts’ or trainers and is assumed to be primarily responsible for implementing and supporting the improvement effort. They are also assumed to be distinct from top management, who, although they set the target for commitment, do not provide training or support improvement activities.

### 3.1 The Allocation of Support Resources

The total amount of resources available to support the improvement effort,  $R$ , is assumed to be fixed at a level below that required to support the improvement effort in each area at 100% commitment. Issues surrounding changing the total amount of resources are not considered. The actions of those charged with supporting the improvement effort are represented by one autonomous decision; the fraction of the available resources allocated to each area, denoted  $x_n$ . The amount of resources allocated to area  $n$ ,  $r_n$ , is equal to the product of the resource constraint  $R$  and the allocation fraction.

$$r_n = x_n R \tag{3.1}$$

The allocation decision is based upon two pieces of information: the current rate of productivity improvement in each area,  $P_n$ , and the current resource requirement in each area,  $q_n$ , calculated as a percentage of the total resource requirement.

$$q_n = \frac{r_n^*}{\sum_n r_n^*} \tag{3.2}$$

The decision rule is specified using the  $US/(US+THEM)$  formulation (Kalish and Lillien 1986). The ‘attractiveness’ of each area is determined by weighting both the fractional resource requirement,  $q_n$ , and the rate of productivity improvement,  $P_n$ , by exponents. The fractional allocation to area  $n$  is then determined by calculating the ‘attractiveness’ of area  $n$  as a percentage of the attractiveness indices summed over all the areas.

$$x_n = \frac{P_n^\alpha q_n^\beta}{\sum_N (P_n^\alpha q_n^\beta)}, \quad \alpha, \beta > 0 \tag{3.3}$$

This type of equation has been used to model the formation of market share for products with multiple attributes. Recently variants have been used successfully to represent human decision making in various contexts (Arthur 1993). If  $\alpha=0$  and  $\beta=1$  then resources are allocated strictly according to need, if  $\alpha>0$  the allocation fraction is biased towards areas showing more rapid improvement, and conversely if  $\alpha<0$ . For the base simulation runs, the attractiveness parameters  $\alpha$  and  $\beta$  are chosen to represent a policy of allocating more support to areas with faster

improvement rates,  $\alpha, \beta > 0$ . This policy is consistent with the maxim a 'successful change program begins with results' (Schaffer and Thomson 1992).

Equations (3.2) and (3.3) add additional feedback loops which affect the dynamics of diffusion and performance in a firm with multiple areas engaged in an improvement effort. The new stock/flow and feedback structure is shown in figure 3.1. Three new feedback loops have been added to the system. First there is the negative loop B3. As commitment rises, support requirements are increased, and, holding the support allocated constant, the change in commitment is reduced. Second, there is the positive loop R3: as commitment rises more requests for support are made. More requests lead to more support being given to that area increasing commitment. Third, there is the positive loop R2. As commitment rises, the defect improvement rate increases causing the support staff to allocate more attention to that area, further increasing commitment.

### **3.2 Simulation Results**

#### *Base Case*

For the base simulation all other parametric assumptions are identical to those in the previous section (additional assumptions are listed in Table 3.1). As in the previous section, the start of an improvement program is simulated via the introduction of a unit step in the parameter  $C^*$  at month twelve. The initial results are similar to the single area, resource-unconstrained case: commitment to the improvement effort jumps initially and then begins to grow exponentially as the positive feedback of the diffusion process begins to dominate (see figure 3.2). Due to its faster improvement rate (see figure 3.3), Area One receives a greater share of the available improvement resources (figure 3.4). As support resources become inadequate, commitment in Area Two declines, further strengthening management's commitment to supporting Area One: the positive loop R2 begins to dominate. Eventually, as measurable improvements become more difficult to make, loop B1 begins to dominate and commitment declines in Area One. Management support is then re-focused to Area Two, which experiences a subsequent recovery in commitment. Area One significantly outperforms Area Two in terms of defect reduction. It is important to note that this difference is much larger than that predicted by the simple half-life model. The results suggest that localized decision making coupled with a policy of supporting a program with an early success may result in a wide differential between the improvement rates in areas characterized by simple processes and areas with complex processes. The gap is much wider than that predicted by Schneiderman's simple half-life model because the fast-improving area receives the lion's share of support, while the slow-improving area is starved for the resources needed to improve.

## Sensitivity

As with the earlier model, a wide range of sensitivity tests have been conducted on this version of the model. Three of these tests are of particular interest. First, the same test as conducted in the previous section was repeated. The model was re-run six times, once for each curve shown in figure 2.6. The results were almost identical for area one. However area two shows more interesting behavior (see figure 3.5). First, the simulation for  $b=4$  shows that if the  $f_a$  function is steep enough, then the resource constraint no longer matters: the positive effect of results outweighs any impact of slack resources: the positive loop R1 dominates. In this case workers are so motivated by results that they persist in their improvement efforts despite inadequate support. Even more interesting is the time path for  $b=1$ . In the earlier test, with  $b=1$  commitment in area one never reaches 100%. The same is true in this simulation. Since commitment never reaches 100% in area one, area two is given more resources with which to improve. With the extra resources, area two makes more improvement and the positive loop R2 kicks in, giving area two progressively more resources. Figure 3.6 shows that for this parameter, the area two begins to receive the higher proportion of resources earlier than in the base case simulation.

A similar test was performed with the function  $f_a(\cdot)$  that represents the effect of support on commitment. The family of test curves is shown in figure 3.7. Commitment in area one shows almost no change for any the inputs. This occurs because the TQM expert allocates the majority of her time to area one and, as a result, area one never experiences the effects of scarcity. For area two, the results are identical as for all inputs that are *steeper* than the base case ( $b \geq 2$ ) (see figure 3.8). For those inputs that are less steep, commitment declines less, since a flatter curve indicates that scarcity has a smaller effect on commitment. In addition, since resources matter less, the improvement rate is higher, implying that the expert allocates more resources to the area (figure 3.9). The analysis shows that there are few parameter combinations which lead to intermediate outcomes: either commitment grows in both areas, or one area begins to dominate. Figures 3.10 and 3.11 show the results from monte carlo simulations in which both functions were varied together. Figure 3.10 shows that there are two dominant behavior modes in area one; either commitment grows very rapidly and stays high, or else it never rises beyond 50% and then falls quickly. Figure 3.11 shows that area two displays three basic behaviors: 1) commitment grows rapidly, and stays high, 2) commitment grows more slowly and reaches its peak between months 60 and 90 but is always increasing, or 3) commitment rises, then collapses and recovers as in the base case. The sensitivity analysis shows that over a wide range of parameters, the model displays a small number of behavior modes. The model's base case will obtain as long as workforce commitment is sensitive to results and to resources. Commitment will be high in both areas if it is

sensitive to results but not to resources, and nothing will happen if commitment is not sensitive to results.

#### 4. Factor Acquisition and Pricing Under Unbalanced Improvement-Comparative Statics

In this section, the behavior of the final organizational tier, top management, is analyzed under the condition of unbalanced rates of improvement in productivity. The model presented here is static and assumes that the firm behaves optimally. In subsequent sections this framework will be adapted to a dynamic, boundedly rational setting. The change in the optimal demand for the factors of production under the assumption of unbalanced improvement in productivity is calculated. A simple result is presented that identifies the conditions under which unbalanced improvement will result in a reduction in the demand for the improving factor. The firm's optimal behavior, assuming that the excess capacity cannot be eliminated, is then analyzed.

##### **4.1 Optimal Factor Demand with Improvement**

The firm is assumed to be a monopolist who faces a constant elasticity demand curve, (4.1), with scale parameter A and price elasticity of demand  $\epsilon$ .

$$Q_D = AP^{-\epsilon} \quad (4.1)$$

This firm is endowed with technology described by a Leontief production function that requires two factors of production  $L_1$  and  $L_2$ .

$$Q_s = \text{Min}[\alpha_1 L_1, \alpha_2 L_2] \quad (4.2)$$

$L_1$  and  $L_2$  cost the firm  $p_1$  and  $p_2$  respectively. The monopolist's profit maximizing output is determined by setting marginal cost equal to marginal revenue, which, it can be shown, results in a constant mark-up rule for calculating the profit maximizing price,  $P^*$  (Varian 1992):

$$P^* = \left( \frac{p_1}{\alpha_1} + \frac{p_2}{\alpha_2} \right) \cdot \frac{1}{1 - \frac{1}{\epsilon}} \quad (4.3)$$

Assuming productivity improvements are being made for the first type of labor, it is possible to show that if  $\epsilon \cdot F_1 < 1$ , where the  $F_1$  is the fraction of cost accruing from the use of  $L_1$ , then the demand for  $L_1$  will decline as  $\alpha_1$  increases. To do this, first, substitute the optimal quantity demanded,  $Q^*$ , into the cost minimizing demand function for factor  $L_1$ .

$$L_1^* = \frac{A \left( \frac{1}{1 - \frac{1}{\epsilon}} \cdot \left( \frac{p_1}{\alpha_1} + \frac{p_2}{\alpha_2} \right) \right)^{-\epsilon}}{\alpha_1}$$

Then taking the derivative with respect to  $\alpha_1$  yields,

$$\frac{dl_1^*}{d\alpha_1} = \frac{A}{\alpha_1^2} \left( \frac{1}{1 - \frac{1}{\varepsilon}} \cdot \left( \frac{p_1}{\alpha_1} + \frac{p_2}{\alpha_2} \right) \right)^{-\varepsilon} \left( -1 + \varepsilon \cdot \frac{\frac{p_1}{\alpha_1}}{\frac{p_1}{\alpha_1} + \frac{p_2}{\alpha_2}} \right)$$

which is the desired result since all multiplying factors are positive and, for this technology,  $\frac{p_1}{\alpha_1} / \left( \frac{p_1}{\alpha_1} + \frac{p_2}{\alpha_2} \right)$  is the fraction of the total expenditure spent on factor  $L_1$ .

This condition has an intuitive interpretation. Under the model's assumptions, the fraction  $F_1$  measures factor  $L_1$ 's contribution to the output price  $P^*$ . The effect of any productivity improvements for factor 1 on output price is determined by this fraction, while the demand elasticity,  $\varepsilon$ , determines the effect of price on quantity demanded. The product of these two quantities is, then, the elasticity of demand with respect to changes in productivity. If it is less than one, then any change in productivity will result in a proportionally smaller change in demand, resulting in a decline in the demand for the improving factor.

#### 4.2 Pricing Behavior with Excess Capacity

Under the assumption of complete flexibility in factor acquisition, the profit maximizing firm will reduce its holdings of a particular factor if it is optimal to do so. However, if that factor is labor, significant reductions may not be possible due to existing contracts, or in the firm's long term interest because, as mentioned in Section 2, the perception of low job security may limit the possibility of future improvement. It should be obvious that given the firm's technology, if the firm is forced to hold an amount of factor 1 greater than the optimum, then the firm's marginal cost of selling an additional unit—evaluated at the optimum—falls to  $p_2/\alpha_2$ . Thus, if the firm is forced to hold too much of factor 1, its constrained optimal production quantity must rise.

In the context of this simple model, the consequences are also obvious. The firm's sales revenue will rise, and profitability will increase, but by a proportionally smaller amount. Thus, under the stated assumptions, the firm which undertakes a successful productivity improvement program that results in excess capacity, is likely to experience significant increases in units sold and sales revenue. However, this will be accompanied by a proportionally smaller increase in profitability, and a decline in profit margin. These effects will persist until the firm is able to reduce its holdings of the factor in question to the long-run optimal level.

## 5. Firm Level Effects of Improvement

In this section a dynamic extension of the model presented in Section 4 is developed. As in Section 4, the firm is assumed to be a monopolist facing a constant price elasticity demand curve, (4.1). The firm is again assumed to require two factor inputs, direct labor,  $L_1$ , and indirect labor,  $L_2$ , and to be endowed with productive technology represented by the Leontief production function (4.2). Direct labor,  $L_1$ , is associated with direct manufacturing and indirect labor,  $L_2$ , is assumed to perform tasks that are not directly associated with manufacturing. Three state variables are added to the model, the available stock of direct labor, the perceived capacity utilization, and the traditional mark-up ratio used for pricing.

### 5.1 Factor Acquisition

The current stock of direct labor,  $L_1$ , is equal to the time integral of hiring,  $L_h$ , attrition,  $L_a$ , and lay-offs  $L_d$ .

$$L_1 = \int_t (L_h - L_a - L_d) dt \quad (5.1)$$

Hiring, which is constrained to be positive, is the attrition rate plus a fractional correction for the difference between the current labor stock and the desired level,  $L^*$ .

$$L_h = \text{Max}[\zeta_h(L^* - L) + L_a, 0] \quad (5.2)$$

The desired stock of labor is set to the profit maximizing value, the long-run optimal quantity demanded,  $Q_{LR}^*$ , divided by the productivity parameter  $\alpha_1$ .

$$L_1^* = \frac{Q_{LR}^*}{\alpha_1} \quad (5.3)$$

The attrition rate is equal the current labor stock divided by the average career length,  $\tau_l$ .

$$L_a = \frac{L}{\tau_l} \quad (5.4)$$

The rate of induced work force reduction, or lay-offs, is formulated similarly to hiring with the addition of a multiplicative constant  $\lambda$ , which measures management's willingness to fire or lay-off excess labor.

$$L_d = \lambda \cdot \text{Max}[-\zeta_d(L^* - L), 0] \quad (5.5)$$

For the sake of simplicity, it is further assumed that the second type of labor can be immediately hired or fired with no additional consequences to the firm. As a result,  $L_2$  is always equal to the short-run optimal level, the profit maximizing short-run demand,  $Q_{SR}^*$ , divided by the productivity parameter  $\alpha_2$ .

$$L_2 = \frac{Q_{SR}^*}{\alpha_2} \quad (5.6)$$

## 5.2 Pricing

In an effort to more realistically represent the price setting operation, a more complex pricing structure is used than in Section 4. First, it is assumed the equilibrium contribution to marginal cost of the direct labor,  $\frac{PL_1}{\alpha_1}$ , is known with certainty. Second, when the firm's stock of direct labor is above the optimum level,  $L_1 > L_1^*$ , the marginal contribution, instead of being just the contribution of indirect labor as is the case in the static model, is scaled downward by a non-linear, decreasing, function of the perceived level of direct labor utilization  $f\{u\}$ . Perceived utilization is assumed to be an exponentially weighted average of past utilization since utilization is likely to be noisy, and managers will seek to filter out short term fluctuations. The adaptive smoothing procedure has been used widely to model the process of forming perceptions (Forrester 1961, Cyert and March 1992).

$$\frac{du}{dt} = \xi_u \left( \text{Min} \left( \frac{\alpha_1 L_1}{Q_d}, 1 \right) - u \right) \quad (5.7)$$

Third, it is assumed that the productivity of indirect labor is difficult to observe or calculate. The standard management accounting solution to this problem is to estimate a product's total cost based upon its direct labor content (Horngren and Foster 1992). The structure used here approximates this practice by calculating expenditure on direct labor as a fraction of the total expenditure, and then using this quantity to scale the direct marginal cost accordingly. However, since this adjustment is only made periodically, the actual cost adjustment,  $M$ , is assumed to be an exponentially weighted average of the adjustment indicated by the mark-up rule, with an average adjustment delay of  $\frac{1}{\xi_M}$ .

$$\frac{dM}{dt} = \xi_M \left( \frac{f\{u\}L_1 p_{l_1}}{L_2 p_{l_2} + f\{u\}L_1 p_{l_1}} - M \right) \quad (5.8)$$

The result of these assumptions is a rule for calculating marginal cost that takes the following form;

$$mc = \left( \frac{PL_1}{\alpha_1} \cdot f\{u\} \right) \cdot \frac{1}{M} \quad (5.9)$$

where marginal cost is equal to the equilibrium direct marginal cost adjusted for utilization and multiplied by a mark-up,  $1/M$ , to account for indirect costs. It should be noted the structure results in the correct calculation for marginal cost, and, as a result the profit maximizing price, when the model is in a steady state equilibrium.

### 5.3 Simulation Results

#### *Base Case*

It is now possible to simulate this portion of the model. The parametric assumptions are given in Table 5.1. For the purpose of this simulation, the productivity of direct labor, measured by  $\alpha_1$ , is assumed to rise approximately four-fold, while the productivity of indirect labor,  $\alpha_2$ , remains constant (Figure 5.1). The firm is also assumed to have a policy of no lay-offs. As the comparative statics results suggest, the result of this improvement is a substantial increase in unit sales and sales revenue. However, profit initially decreases (Figure 5.2), and then increases slowly, while profit margin actually decreases (Figure 5.4).

The initial decrease in profit results from the delays with which the accounting system recognizes a change in the ratio of direct to total costs. The mark-up factor  $1/M$  requires time to adjust to changes in the cost structure caused by unbalanced productivity improvement. Until the adjustment process is complete, the marginal cost of production is underestimated if the productivity of direct labor improves more quickly than that of indirect labor. The output price is below and the quantity demanded is above its optimal level, which, in turn, causes the desired level of direct labor to initially rise. As the estimated marginal cost begins to approach its true value, price is adjusted upward, demand downward, and the desired level of direct labor is reduced. As the actual stock of labor is also reduced, following the desired level, price is further increased causing sales revenue to decrease and profit to increase.

#### *Sensitivity Analysis*

Additional simulations demonstrate that behavior is quite sensitive to the price elasticity of demand, the fraction of total cost that results from the use of direct labor, and the delay in perceiving changes in the ratio of direct to indirect cost. Table 5.2 shows the results of different assumptions on the cost structure and the price elasticity of demand. Figure 5.5 and 5.6 show the effect of changes in the time constant  $\xi_M$ . As the time constant is decreased below the base case value the decline in profit margin is smaller (figure 5.5). With a shorter delay, the error in estimating the correct price is smaller, and, as a result, the firm hires fewer extra direct laborers (figure 5.6). An interesting feature of the system is that the time required for profit margin to recover to its normal level is *not* affected by the time delays in the pricing process. Instead it is determined by the time required to reduce labor to the desired level, in the base case 240 months. Thus, although, delays in the pricing system initiate the decline in profit margin, they are sustained by the expansion of the labor force beyond the desired level.



## 6. The Firm with Endogenous Improvement

In this section the models discussed in Sections 3 and 5 are integrated to form a fully endogenous representation of the improvement process. The parametric assumptions are identical to those made in the previous sections (see Tables 3.2 and 5.1). Direct labor is assumed to work in area one, while indirect labor works in area two. The respective defect levels are translated to productivity parameters via the following equation, where  $A_i$  is the gross (i.e. defective units included) production per unit of labor type  $i$  ( $A_i$  is assumed to equal 133 units/month for subsequent simulations).

$$\alpha_i = 1 - \left( \frac{D_i}{A_i} \right) \quad (6.1)$$

To close the final feedback loop, one additional piece of structure is required. Autonomous reduction in direct labor,  $L_d$ , is a function of the difference between the desired and actual labor force, and a parameter  $\lambda$ , management's willingness to lay-off workers. While this has been set to zero in previous simulations, the results of two different simulation will be presented. In the first, as in the previous example, management maintains its commitment to no lay-offs, and in the second  $\lambda$  is assumed to equal to one, indicating management is willing to lay-off any excess labor.

### **6.1 Simulation Results**

#### *Base Case*

In both cases the profit margin begins to fall as productivity improves (Figure 6.1). Although profit is increasing, sales revenue is increasing at a faster rate (Figure 6.2). In the first case management reacts by laying-off excess labor (Figure 6.3). The resulting reduction in direct labor makes a very small improvement in profit (Figure 6.2), and a larger improvement in the profit margin. However, this short term gain comes at the expense of the firm's long run success. By resorting to lay-offs to improve profitability, management effectively ends the improvement program. Commitment in both areas falls quickly after the lay-off (Figure 6.4). Due to the nature of the memory process assumed, commitment recovers very slowly, as it takes a long time for management to regain the trust of the workforce. As a result, profit in the no lay-off case ultimately exceeds that of the case with lay-offs.

The long run performance of the firm is clearly better if commitment to job security is maintained. However, there are a number of reasons why the firm may resort to lay-offs to cut costs. First, these dynamics may not be well understood. A company embarks upon an improvement program in an effort to improve its competitiveness and profitability. Unless a manager understands the dynamics described above, she is likely to favor the faster improving areas in direct manufacturing

and generate excess capacity. If prices are cut to utilize this capacity and increase profit, the subsequent decline in profit margin may be misinterpreted as a sign that the program is not living up to its promise and costs are actually growing. Few people would suspect that a *decline* in profit margin might be a temporary consequence of a successful TQM program. It is possible that, were this to happen in a real company, top management would suspect that the TQM program was not living up to its promise and begin to consider other ways to cut costs, particularly if utilization was below normal.

Second, publicly held firms must face the scrutiny of financial analysts who are even less likely to appreciate the dynamic consequences of improvement. External capital markets frequently view profit margin as an indicator of a firm's ability to 'control' costs. Few analysts are likely to believe that a decline in profit margin might be a necessary consequence of a successful productivity improvement program. As a result, a firm may be forced to downsize in an effort to improve profitability and to demonstrate to the capital markets that they are "serious" about controlling costs.

### *Policy Analysis*

So far the model shows conditions under which a successful productivity improvement program can lead to declining profit margins and the possibility of lay-offs. In this section policies that can mitigate these dynamics are considered. The key control parameters in the model are  $\alpha$  and  $\beta$ , the weights in the expert's decision rule. These parameters represent the incentive scheme faced by the 'expert'. If the expert is rewarded for results, as is assumed in the base case,  $\beta$  will be large and positive. Two alternative policies are considered. First, in the *neutral* policy  $\beta$  is set to zero. Such a policy represents an incentive scheme for the expert that gives no reward for results. In such a situation the guru then allocates her attention in proportion to the number of requests received from each area. Setting  $\beta$  equal to zero eliminates the positive loop R2 from the system. Second, is the *balanced* policy in which  $\beta$  is set to a large negative number. The balanced policy represents an incentive scheme in which the 'expert' is actively encouraged to allocate more resources to areas showing less improvement. Setting  $\beta$  less than zero changes the polarity of loop R2, which becomes a negative loop that constantly seeks to balance the improvement rates. Implementing such a scheme in a real organization might be difficult. One method would be to reward the 'expert' based on the rate of improvement of the slowest area.

Figure 6.5 shows the results from these two policies and the base case. In each simulation it is assumed that the firm *cannot* commit to a policy of no lay-offs. Figures 6.5 and 6.6 show that both policies outperform the base case in terms of both profit and profit margin and avoid any

possible lay-offs. Interestingly, the neutral policy has the advantage in profit while the balanced policy does slightly better in terms of profit margin. The reason for the difference can be seen in figures 6.7 and 6.8. Under the neutral allocation policy commitment remains high in both areas, while under the balanced policy commitment in area one never reaches beyond 50%. Figure 6.9 shows that this occurs because under the neutral policy, the expert allocates her attention equally between the two areas, while under the balanced policy area two gets a much larger fraction of the attention. Under the balanced policy, the negative loop created when  $\beta < 0$  tries to equalize the improvement rates by allocating more resources to area two. Equalizing the improvement rates is not optimal given the large difference in potential improvement. Under the balanced policy, the expert allocates 'too much' attention to area 2. The neutral policy outperforms the balanced policy because it yields a better allocation of attention between the two which results in a higher combined level of commitment.

Figure 6.10 shows the results of Monte Carlo analysis with the two parameters. One thousand simulations were run with  $\alpha$  and  $\beta$  being drawn from uniform distributions on  $[-20,20]$  and  $[-200,200]$  respectively. The vertical axis shows the accumulated profit for the entire simulation. The response surface shows that the system's response to the parameter changes is highly non-linear. There are four distinct levels in the diagram that correspond roughly to four quadrants of the  $x$  and  $z$  axes. Figure 6.11 shows the average pay-off in each of the four quadrants. The negative orthant clearly dominates the other three. In this region the two positive loops, R2 and R3, become negative. The two mixed regions produce similar results. A negative value for sensitivity to resource requests does produce better results when paired with a positive value for sensitivity to the results than the opposite case. The positive orthant produces by far the lowest pay-off. In these cases both loops remain positive, excess capacity is quickly generated and lay-offs follow.

Figure 6.12 and 6.13 show select cross-sections of the surface in 6.11. These graphs show more clearly the desirable properties of the neutral policy. Viewed from both directions, the neutral policy provides a near optimal control policy. To check this intuition the optimal policy is calculated using a Powell search routine. The best policy found places a weight of .41 on resources and -4.76 on the improvement rate. The cumulative pay-off for this policy is 207411 compared to 207,074 for the neutral policy which is a less than a .5% difference. Clearly, any policy in the negative orthant is quite good, as the average for the entire region is 204,941 (approximately a 1% difference).

## 7. Discussion

The philosophy of continuous improvement represents a substantial increase in the dynamic complexity of the manager's task. Many managerial tasks are often conceptualized and modeled as static optimization problems, choosing the optimal price, setting the optimal incentive scheme, etc. Improvement adds a fundamentally dynamic element to almost every managerial problem and invalidates many the results developed in the static framework. For example, in the literature on experience and learning curves, Fine(1986) has shown that, in the presence of learning, optimal pricing and investment policies change significantly. Unfortunately, human performance in such environments is rarely optimal. In a wide array of tasks, ranging from fighting a simulated forest fire to managing a simple production and distribution chain, human subjects routinely perform well below even the simplest of decision rules (Sterman 1989a). Research has shown that in such situations, subject frequently rely on simple decision making heuristics derived from a static conception of the problem which ignore many of the important feedbacks within the system.

Managers that actively pursue TQM or similar process improvement efforts take actions that have delayed and uncertain consequences. In addition, these decisions can have multiple impacts due to the complex feedback structure of most organizations. The use of traditional decision making heuristics, such as constant mark-up pricing, based on a static conception of the environment can lead to highly undesirable in behavior. In the model presented in this paper a policy of favoring those areas that show higher rates of improvement leads to a number of unanticipated consequences. First, it starves slower improving areas of important resources and further increases the difference in improvement rates. Second, the differential improvement rates invalidate the traditional pricing heuristics. A policy of constant mark-up pricing that fails to recognize the changing ratio of direct o indirect cost, or recognizes the change with a delay, leads to the price being below optimal and the firm expanding capacity beyond the desired level. In addition, profit margin falls dramatically even though the firm is successfully *cutting* costs. Few managers would anticipate that a successful productivity improvement program might lead to *declining* profit margins. Third, the different rates of improvement can lead to excess capacity, which coupled with declining profit margins, may tempt managers into downsizing, and thus ending the effort long before its produced its full benefit.

In the context of a small model, it may appear as though the hypothesized decision rules represent 'stupid' managers. However, each of these actions was observed in the field study that was the catalyst for this model (Sterman *et al.* 1994) and taken individually each policy appears rational. For example numerous authors recommend a strategy of developing early results in a new

improvement effort ( Kotter 1995, Schaffer and Thomson 1992). Such a strategy makes intuitive sense – demonstrate the methods work and then propagate them to the rest of the organization. Unfortunately, as shown above, such a strategy can lead to a self-fulfilling prophecy in which those that show early results get support while others are neglected as managers come to believe that slow improving areas ‘can’t be improved’ or that participants ‘just don’t get it’. Similarly, constant mark-up pricing is a popular and useful heuristic for pricing products in an environment in which the ratio of direct to indirect cost remains constant .

The interaction between process improvement efforts and other parts of the firm has not received sufficient attention in the literature. A firm that actively pursues a productivity improvement program which results in unbalanced improvement in the productivity of its factors of production is faced with a classic “worse before better” situation that is frequently found in complex systems (Forrester 1969). Any short run improvement in profitability achieved through downsizing or lay-offs comes at the expense of long success. Analysis of the model in this paper suggests that simple strategies can mitigate many of these problems and lead to a solution that is very near optimal. For example, by evaluating the improvement expert on the basis of her service to her customers rather than on the improvement rate, an organization might be able to closely approximate the neutral policy described above. There is a clear need for further research in this area.

## 8. Conclusion

In an attempt to explain the failure of improvement programs such as Total Quality Management and Business Process Re-Engineering a model of a firm that attempts to implement such a program has been developed. As a result of the diffusion process used to model commitment and the policy of starting with early successes, the results presented suggest that improvement is likely to be unbalanced, with the relatively simple direct manufacturing processes improving quickly, while the more complex indirect processes, due to a lack of management attention, improve more slowly. If the percent change in productivity is larger than the resulting change in demand, unbalanced improvement will result in excess direct capacity. In the face of excess capacity, the profit maximizing firm will price below the long run equilibrium level, resulting in a substantial increase in unit sales, a small increase in profit, and a decline in profit margin. The decline in profit margin, if it is misinterpreted as poor cost control, may induce management to lay-off excess labor in an effort to improve profitability and demonstrate that it is "serious" about cost control. The lay-off effectively ends the improvement program. Analysis of the different control policies shows that improved results can be obtained using a neutral policy which allocates resources to different areas based on their requests rather than on their improvement rates. Such a policy could be

approximated in real organizations by evaluating improvement experts as a function of the quality of the service they provide rather than the results they produce.

Future research on this topic might be profitably focused on two areas. First, the analysis should be extended to models of the firm which contain more complex representations of technologies and market structures. The monopolistic model also should be extended to a competitive environment. Second, the model contains a number of hypotheses that could be tested empirically. Foremost among these, the analysis suggests that firms whose production costs are largely a function of direct labor content will benefit more from the current batch of improvement programs, than firms whose costs derive largely from indirect sources such as product development.

## References

- (1992). 'The Cracks in Quality', *The Economist*, 322, 67.
- Arthur, B. (1993). 'On designing economics agents that behave like human agents', *Journal of Evolutionary Economics*, 3:1-22.
- Bass, F. (1969). 'A New Product Growth Model for Consumer Durables', *Management Science*, Vol. 15, No.5, January.
- Cyert, R. and J. March (1992). *A Behavioral Theory of the Firm*, Cambridge, Ma., Blackwell Publishers.
- Deming, W. E. (1986). *Out of the Crisis*, Cambridge, MIT Press.
- Easton, G. and S. Jarrell (1995). 'The Effects of Total Quality Management on Corporate Performance: An Empirical Investigation'. Working Paper, University of Chicago, Chicago, Illinois, 60637.
- Ernst and Young (1991). "International Quality Study – Top Line Findings" and "International Quality Study – Best Practices Report" Ernst and Young/American Quality Foundation.
- Fine, C. (1986). 'Quality Improvement and Learning in Productive Systems', *Management Science*, 32(10).
- Forrester, J. (1961). *Industrial Dynamics*, Cambridge, The MIT Press.
- Forrester, J. (1969). *World Dynamics*, Cambridge, The MIT Press.
- General Accounting Office (1991). US companies improve performance through quality efforts. GAO/NSIAD-9-190 (2 May).
- Harte, S. (1992). 'Corporate Style', *Atlanta Journal & Constitution*, 11 October, R1.
- Homer, J. (1987). 'A Diffusion Model with Application to Evolving Medical Technologies', *Technological Forecasting and Social Change*, 31, 197-218.
- Hornigren, C. and G. Foster (1992). *Cost Accounting: A Managerial Emphasis*, New Jersey, Prentice Hall.
- Jacob, R. (1993). 'TQM: More than a dying fad?' *Fortune*, 18 October, 66-72.
- Kaufman, R. (1992). 'Why Operations Improvement Programs Fail: Four Managerial Contradictions', *Sloan Management Review*, Fall, 83-93.
- Kalish, S. and G. Lillien (1986), 'Applications of Innovations Diffusion Models in Marketing', in *Innovation Diffusion Models of New Product Acceptance*, Mahajan, V. and Y. Wind. Eds. Cambridge, Ma. Ballinger.
- Kaplan, R. (1990). *Analog Devices: The Half-Life System*, Case 9-191-061, Harvard Business School.

- Kim, D. and G. Burchill (1992). System Archetypes as a Diagnostic Tool: A Field Based Study of TQM Implementation. *Proceedings of the 1992 International System Dynamics Conference*. Utrecht: University of Utrecht, 311-320.
- Kotter, J.P. (1995). 'Leading Change: Why Transformation Efforts Fail', *Harvard Business Review*, March-April.
- Morecroft, J. (1985). 'Rationality in the Analysis of Behavioral Simulation Models', *Management Science* 31 (7): 900-916.
- Paich, M. and J. Sterman (1993). 'Boom, Bust, and Failures to Learn in Experimental Markets', *Management Science*, 39(12), 1439-1458.
- Schaffer, R. and H. Thomson (1992). 'Successful Change Programs Begin with Results', *Harvard Business Review*, Jan/Feb: 80-89.
- Schneiderman, A. (1988). 'Setting Quality Goals', *Quality Progress*, April, 55-57.
- Schneiderman, A. (1992). Personal Interview, April.
- Shiba, S, D., Walden, and A. Graham (1993). *A New American TQM. Four Practical Revolutions in Management*. Portland, OR., Productivity Press.
- Simon, H.A.(1976). *Administrative Behavior*, New York, NY., Free Press.
- Stata, R. (1989). 'Organizational Learning— The Key to Management Innovation', *Sloan Management Review*, 30(3) Spring, 63-74.
- Sterman, J., N. Repenning and F. Kofman (1994). Unanticipated Side Effects of Successful Quality Programs: Exploring a Paradox of Organizational Improvement, Working Paper #3667-94-MSA, Sloan School of Management, Cambridge, MA 02142.
- Sterman, J. D. (1989a). 'Misperceptions of Feedback in Dynamic Decision Making', *Organizational Behavior and Human Decision Processes*, 43 (3): 301-335.
- Sterman, J. D. (1989b). 'Modeling Managerial Behavior: Misperceptions of Feedback in a Dynamic Decision Making Experiment', *Management Science*, 35 (3): 321-339.
- Taylor, P. (1992). 'Such an Elusive Quality', *Financial Times*, 14 February, 9.
- Varian, H. (1992). *Microeconomic Analysis*. New York, N.Y., W.W. Norton & Co.



## Tables

### Table 2.4

Parameter	Value
$\phi$	.077 (1/months)
D Initial	100 (defects)
D Minimum	10 (defects)
$\theta$	.084 (1/months)
$\omega$	.5
C Initial	0
C*	0 until time=12, then 1
$\gamma$	2
$\delta$	.5
$B_r^u$	1.5
$B_r^l$	-.5

### Table 3.1

Parameter	Value
$\phi_1$	.077 (1/months, 6 month Half-Life)
$\phi_2$	.02 (1/months, 36 month Half)
D <sub>1</sub> ,D <sub>2</sub> Initial	100 (defects)
D <sub>1</sub> ,D <sub>2</sub> Minimum	10 (defects)
$\theta_1,\theta_2$	.084 (1/months)
$\omega_1,\omega_2$	.5
C Initial	0
C*	0 until time=12, then 1
L <sub>1</sub> ,L <sub>2</sub>	100 (people)
$\rho_1,\rho_2$	1 (resources/person/month)
$\alpha$	25
$\beta$	1
fs{.}	0
$\gamma_a$	
$\delta_a$	
$B_a^u$	0
$B_a^l$	-2

**Table 5.1**

<b>Parameter</b>	<b>Value</b>
A	25,600
$\epsilon$	2
$\alpha_1, \alpha_2$ Initial	.25 (1/units produced)
$\zeta_h$	.083 ( 1/months )
$\zeta_d$	.5 ( 1/months )
$T_L$	240 (months)
$\xi$	.083 (1/months)
PL1	.5 (dollars/person/month)
PL2	1.5 (dollars/person/month)

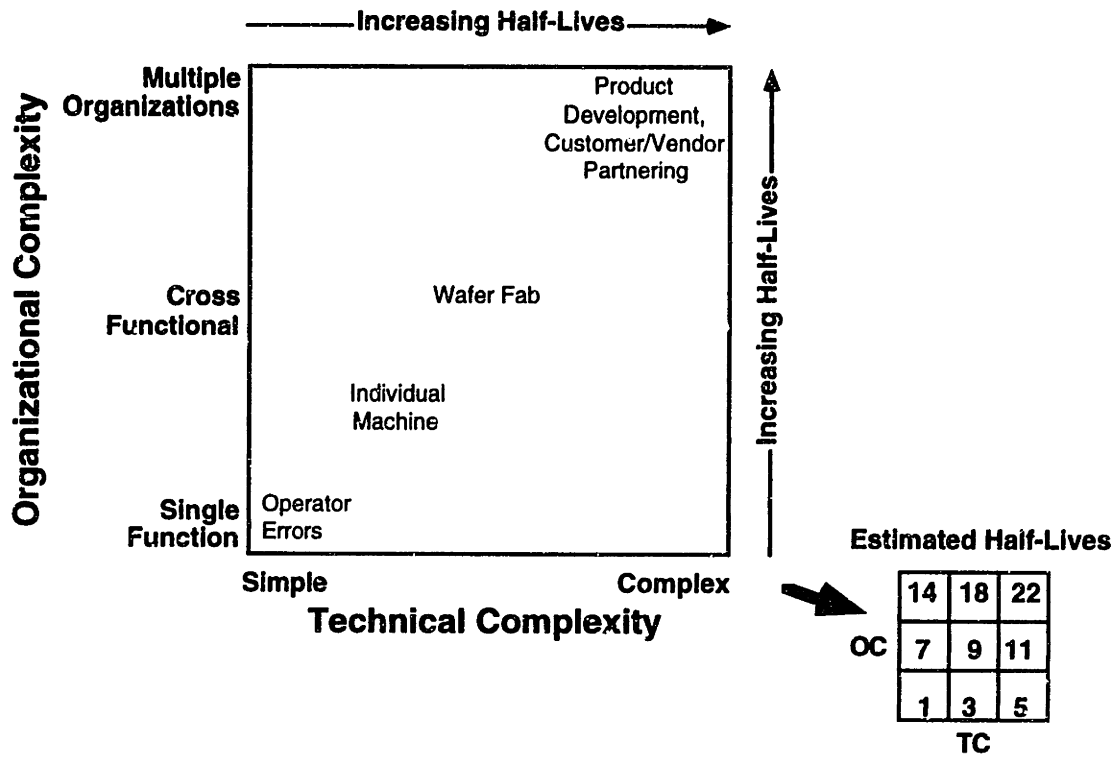
**Table 5.2**

% Change in Equilibrium Demand for Direct Labor due to Productivity Increase

% Direct Labor of Total Cost	Price Elasticity $\epsilon=1.5$	Price Elasticity $\epsilon=2$	Price Elasticity $\epsilon=3$	Price Elasticity $\epsilon=4$
25%	-63%	-59%	-50%	-39%
50%	-46%	-33%	+6%	+68%
75%	-11%	+33%	+194%	+553%

# Figures

## Figure 2.1



Source: Adapted from Schneiderman(1992)

Figure 2.2  
Basic Feedback Structure

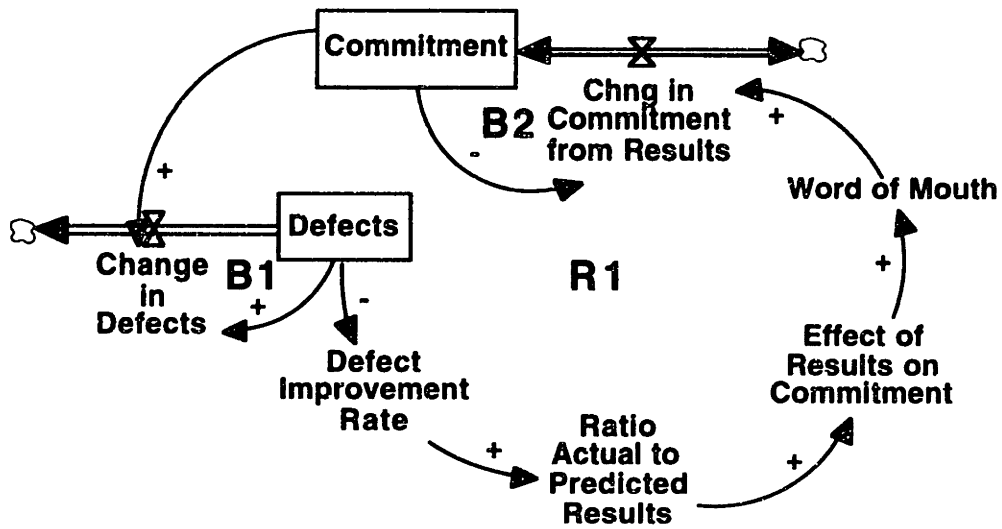
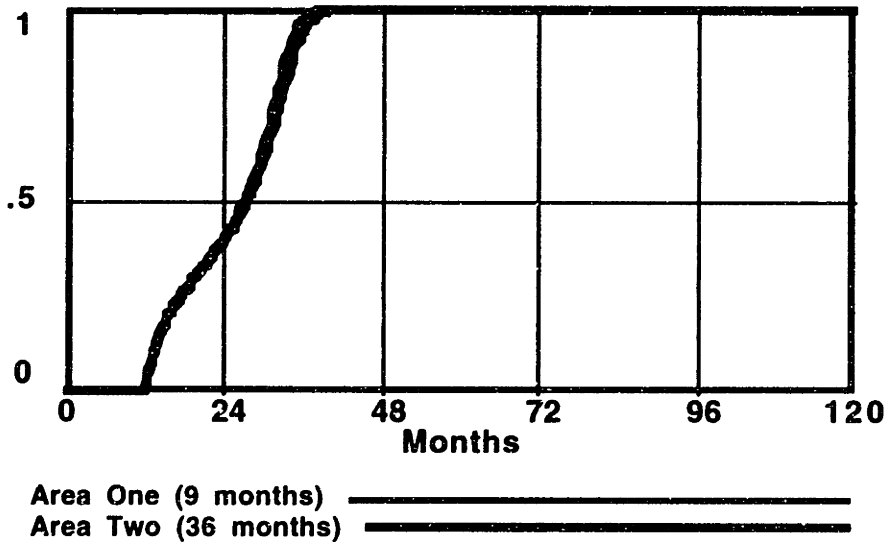
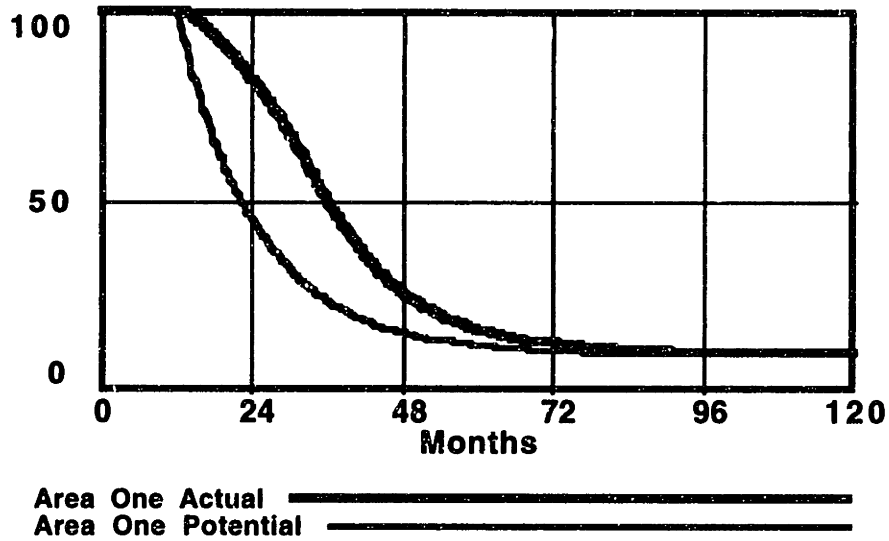


Figure 2.3  
Commitment to Improvement



**Figure 2.4**  
**Defects in Area One**



**Figure 2.5**  
**Defects in Area 2**

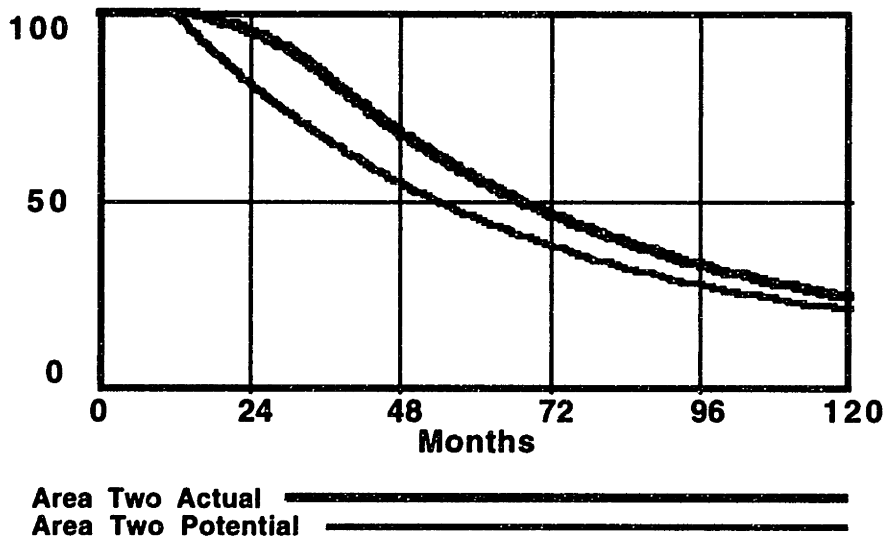


Figure 2.6  
**Test Inputs**

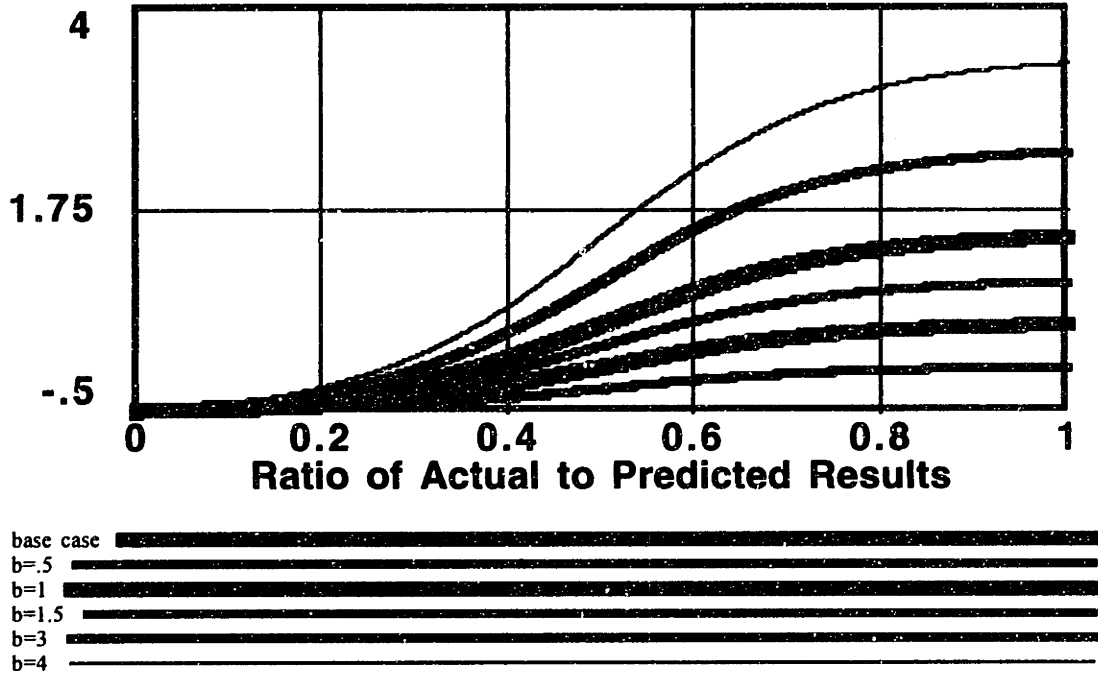
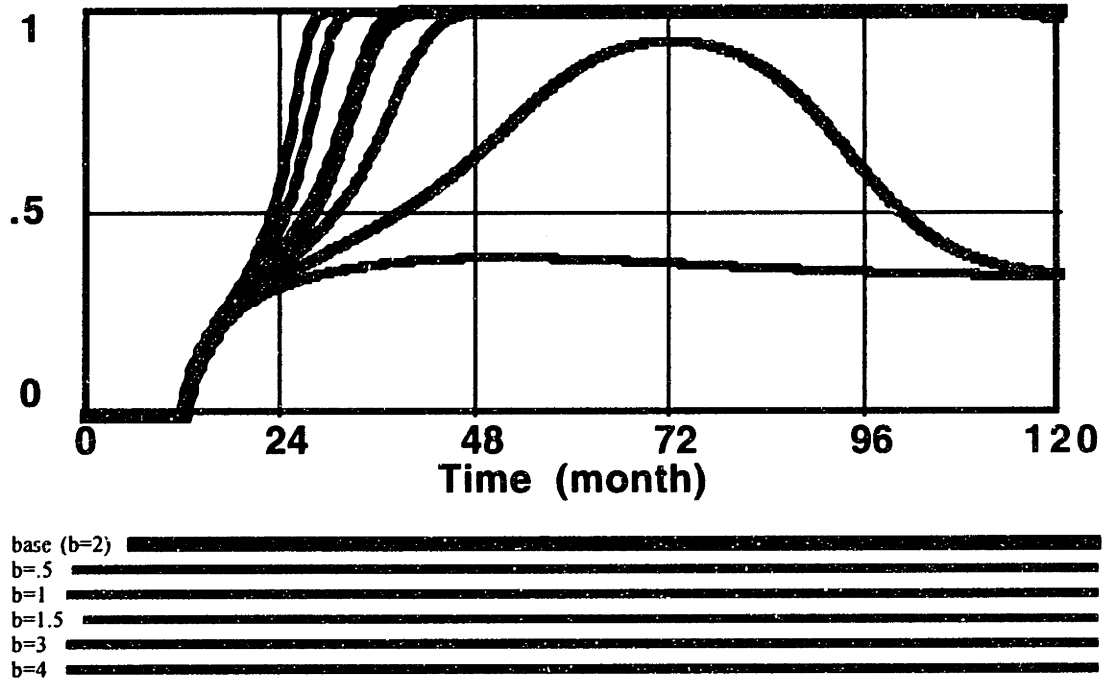


Figure 2.7  
**Commitment Area One**



**Figure 2.8**

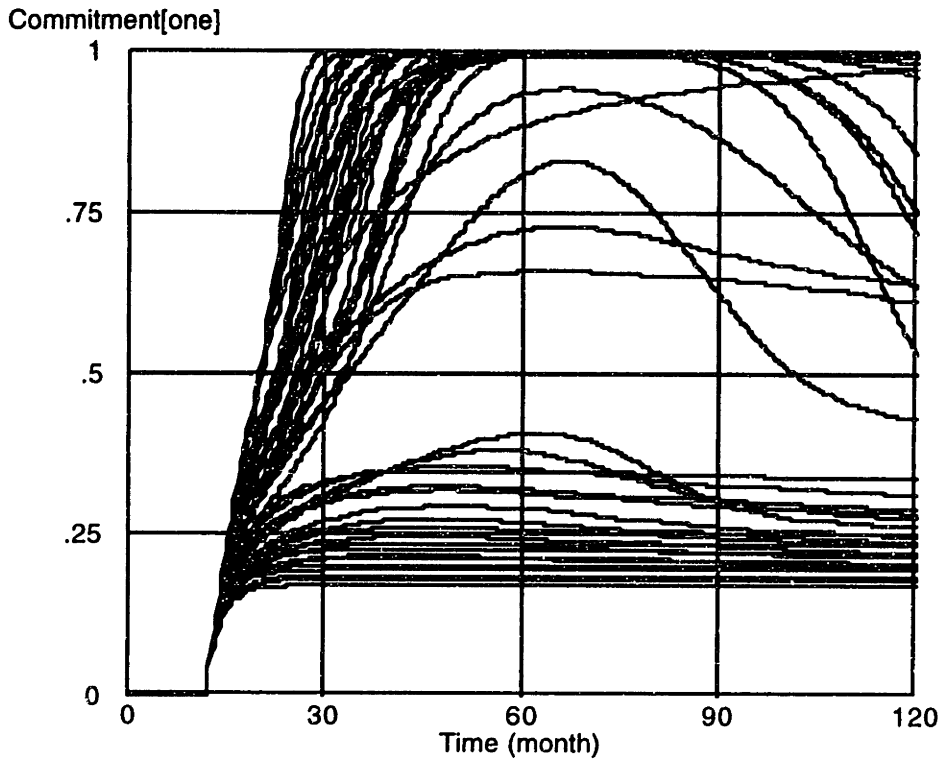
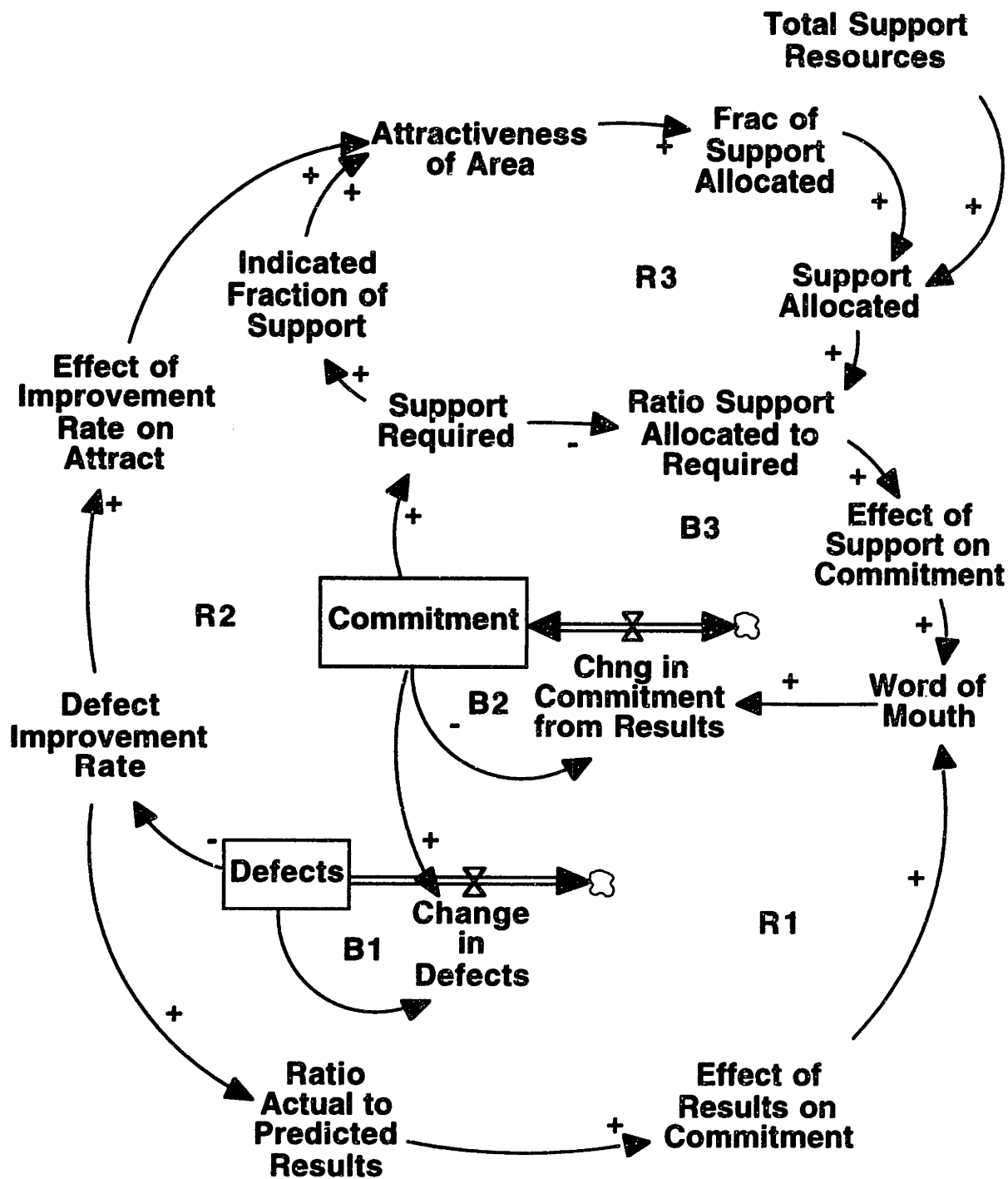
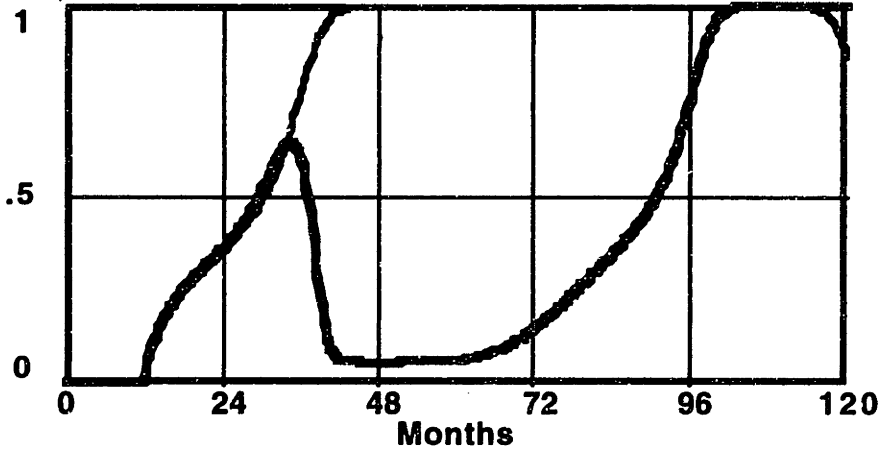




Figure 3.1



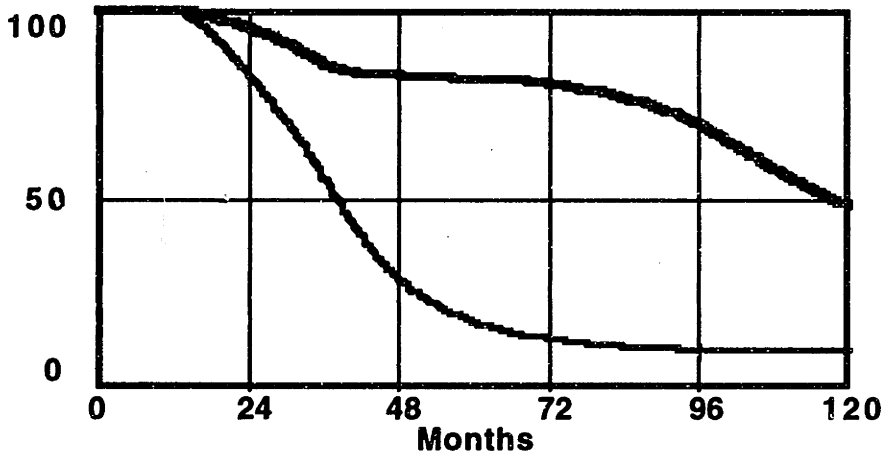


**Figure 3.2**  
**Commitment to Improvement**



Area One (9 months)   
 Area Two (36 months) 

**Figure 3.3**  
**Defects**





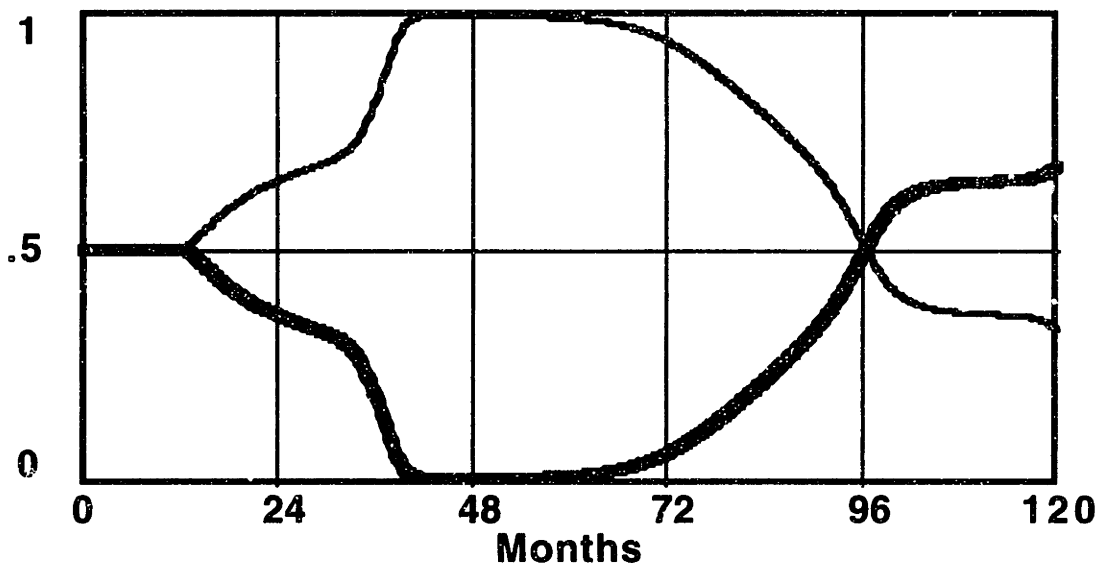
Area One (9 months)   
 Area Two (36 months) 

Figure 3.4

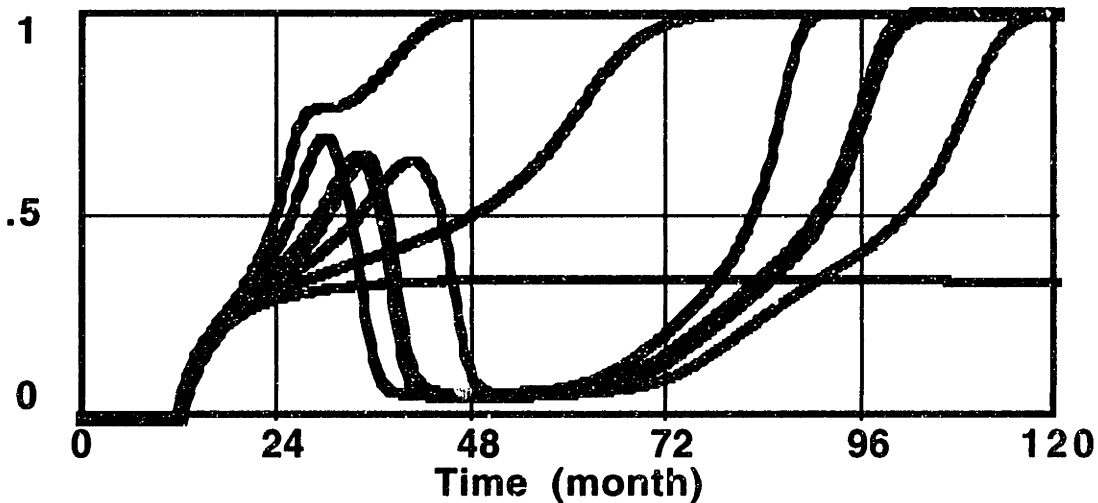
### Fraction of Support Allocated



Area One (9 months)   
Area Two (36 months) 

Figure 3.5

### Commitment: Area Two



base ( $b=2$ )   
 $b=.5$    
 $b=1$    
 $b=1.5$    
 $b=3$    
 $b=4$  

Figure 3.6

### Fraction Support Allocated to Area One

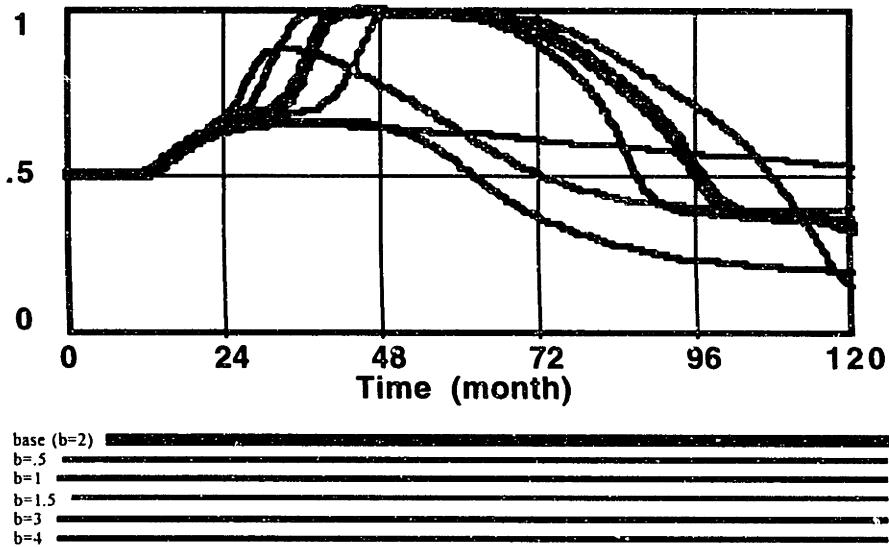


Figure 3.7

### Effect of Resources on Commitment

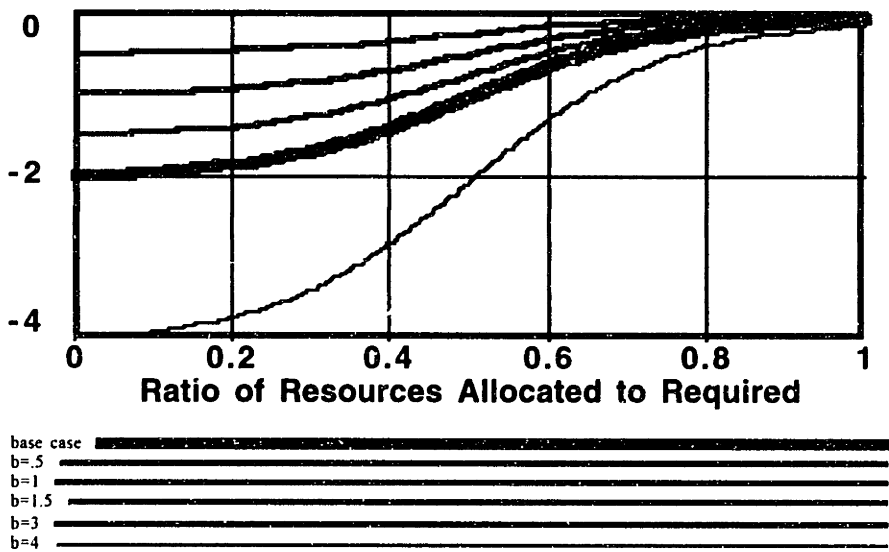


Figure 3.8  
Commitment: Area Two

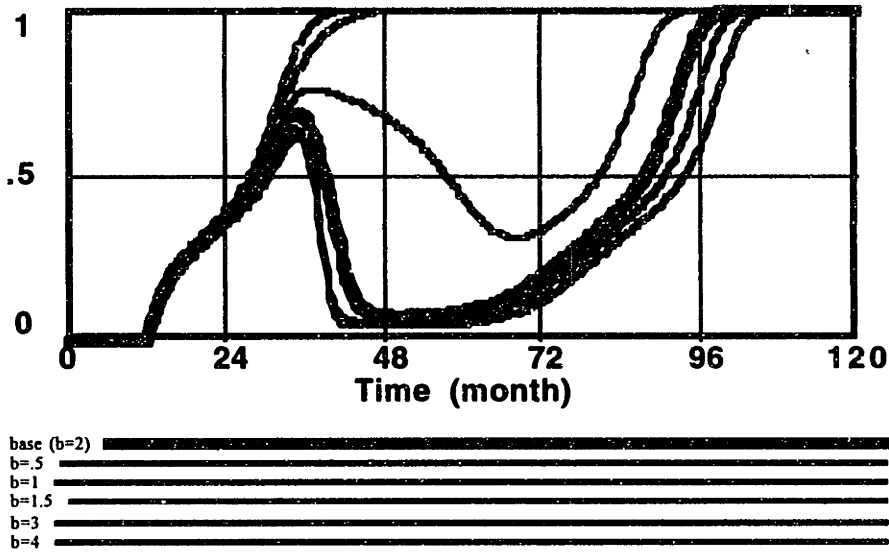
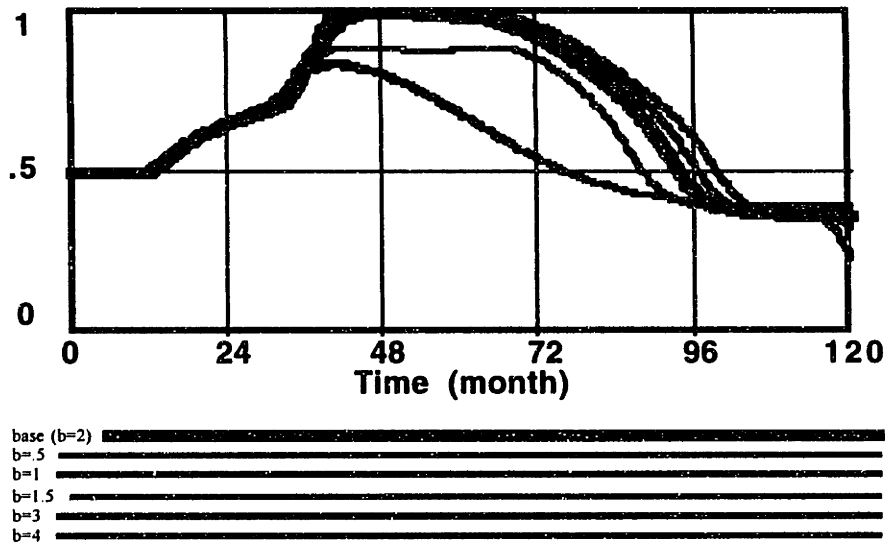
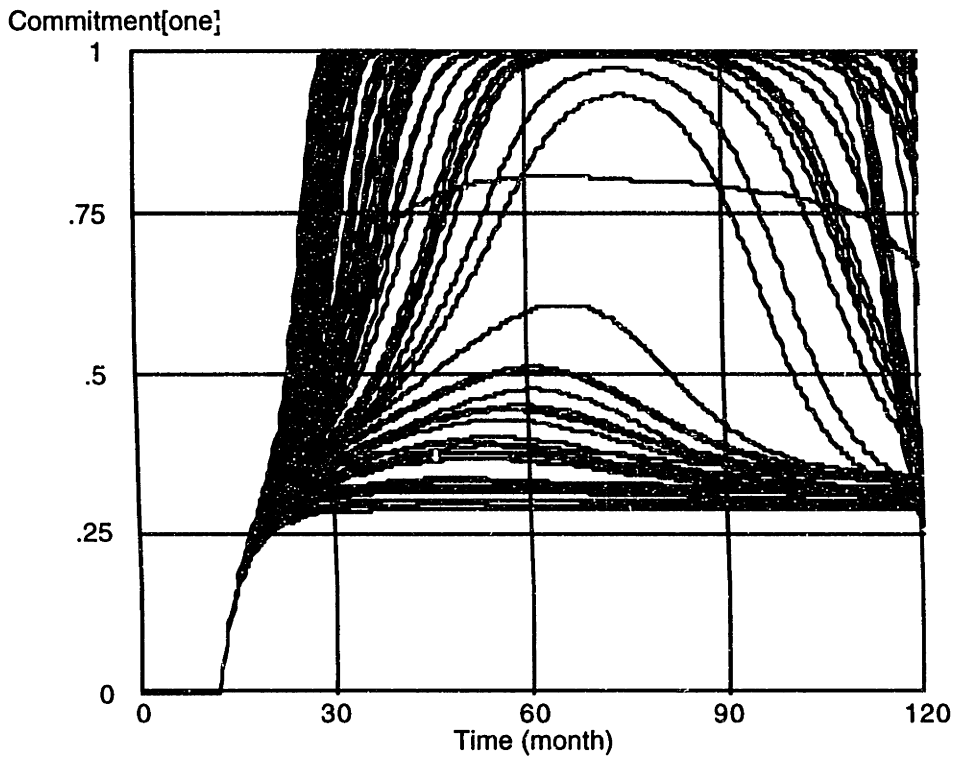


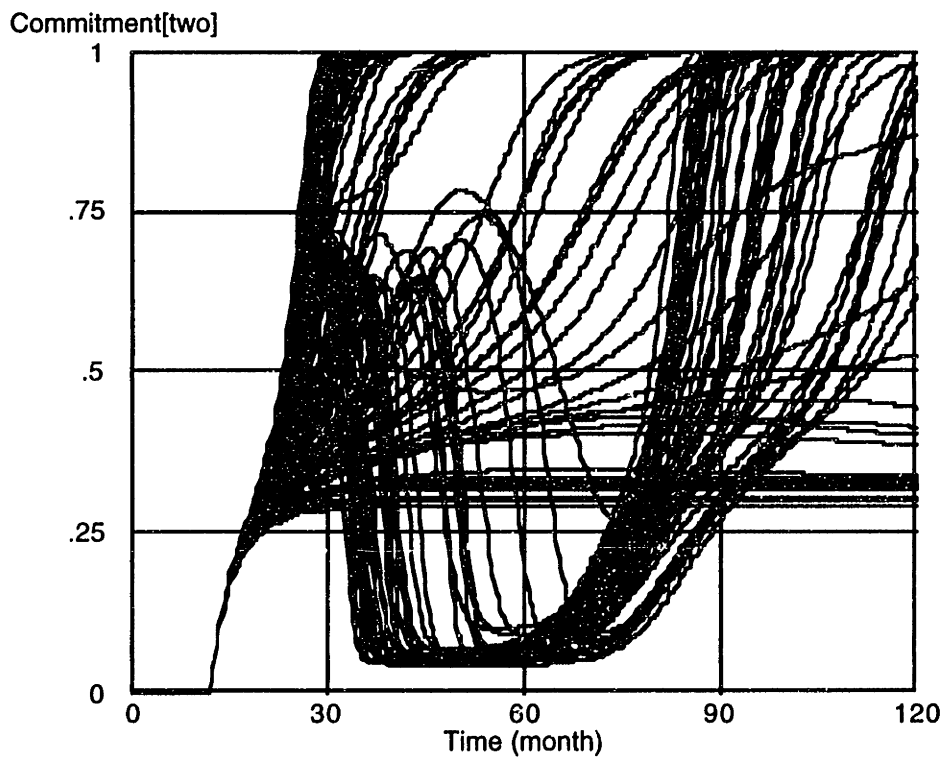
Figure 3.9  
Fraction Support Allocated to Area One



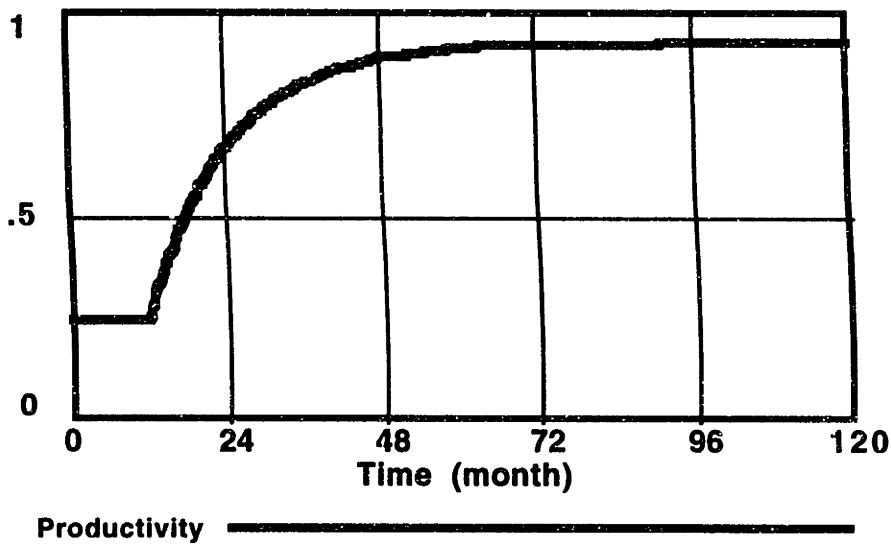
**Figure 3.10**



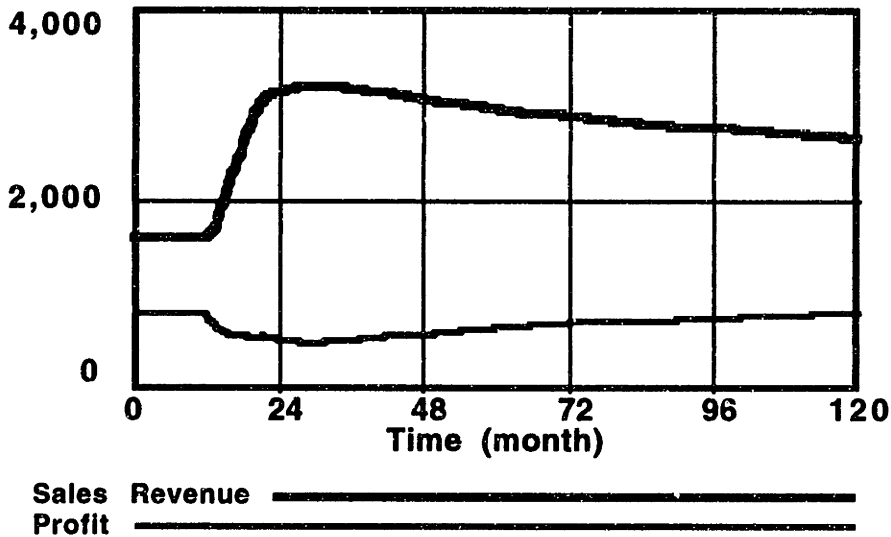
**Figure 3.11**



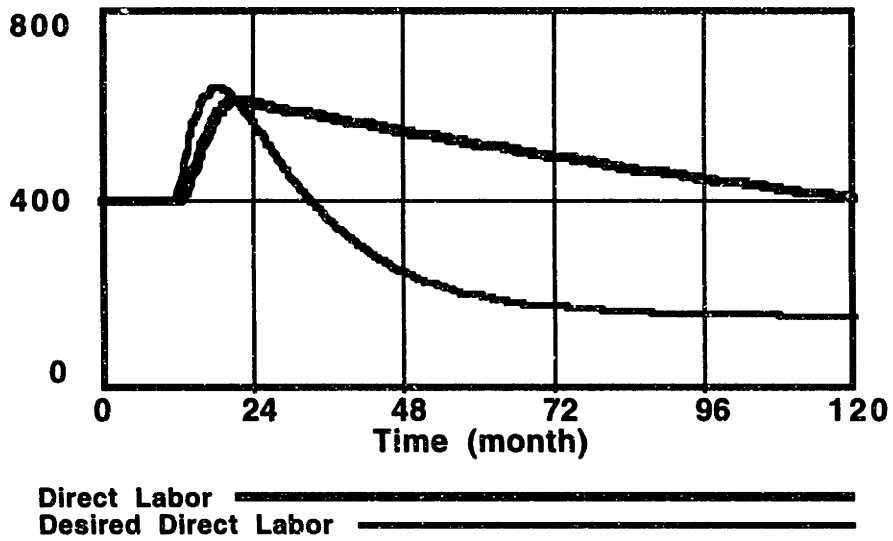
**Figure 5.1**  
**Productivity of Direct Labor**



**Figure 5.2**  
**Sales Revenue and Profit**



**Figure 5.3**  
**Direct labor**



**Figure 5.4**

**Profit Margin**

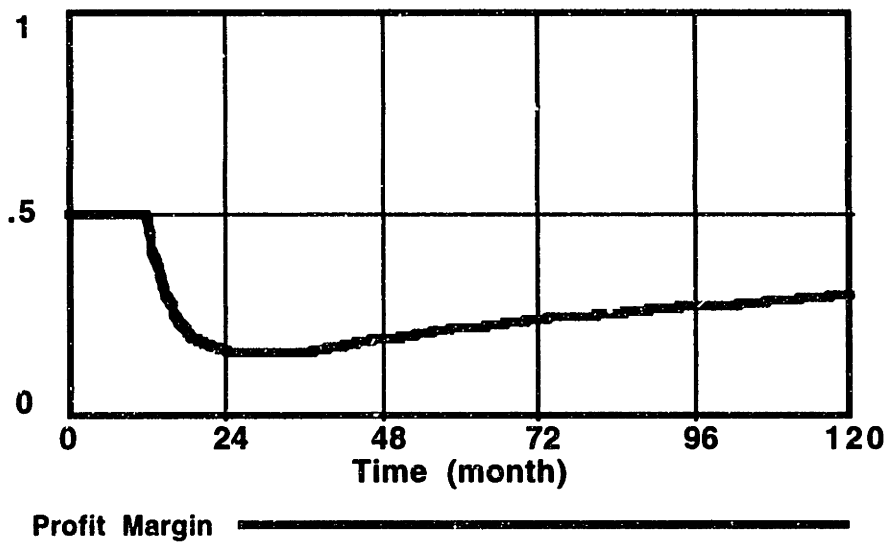


Figure 5.5

### Profit Margin

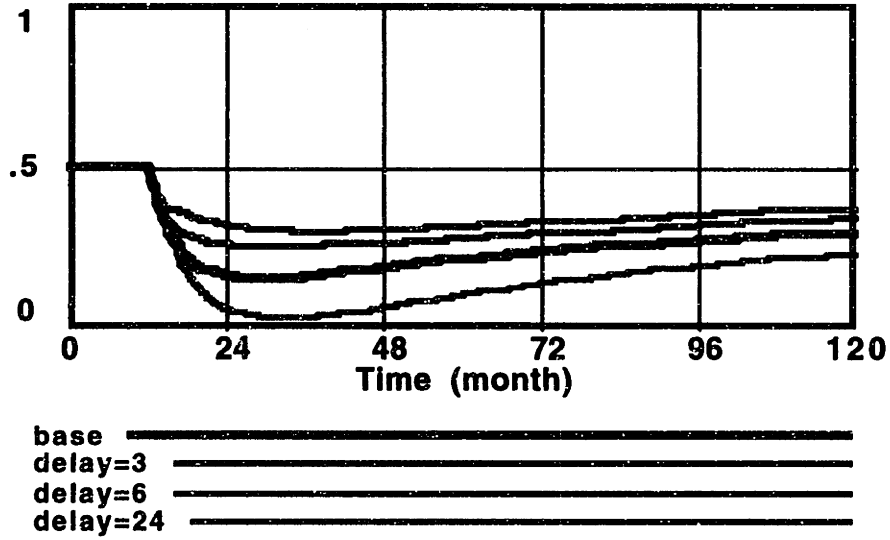


Figure 5.6

### Direct Labor

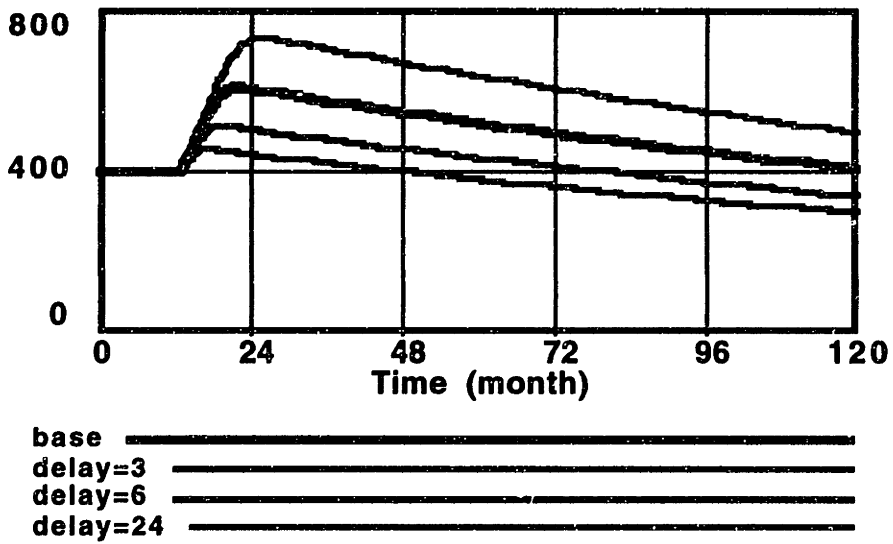




Figure 6.1

### Profit Margin

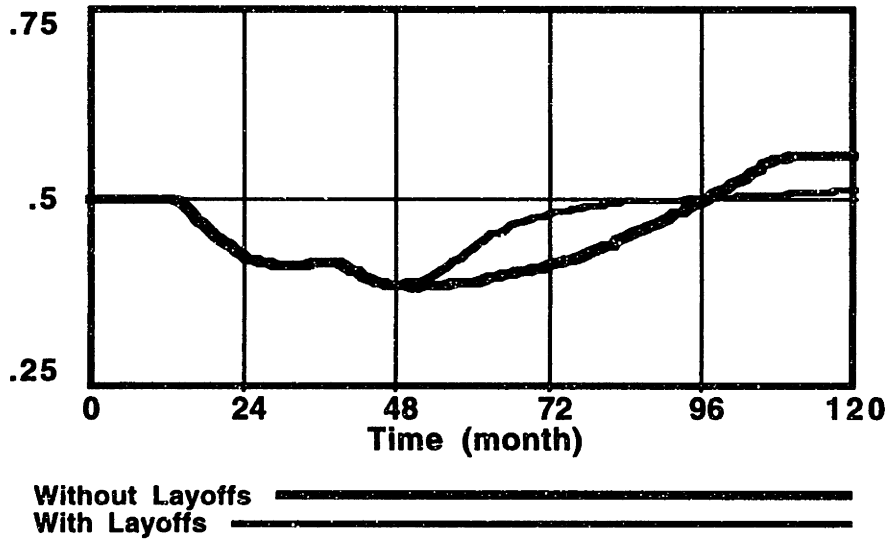


Figure 6.2

### Profit

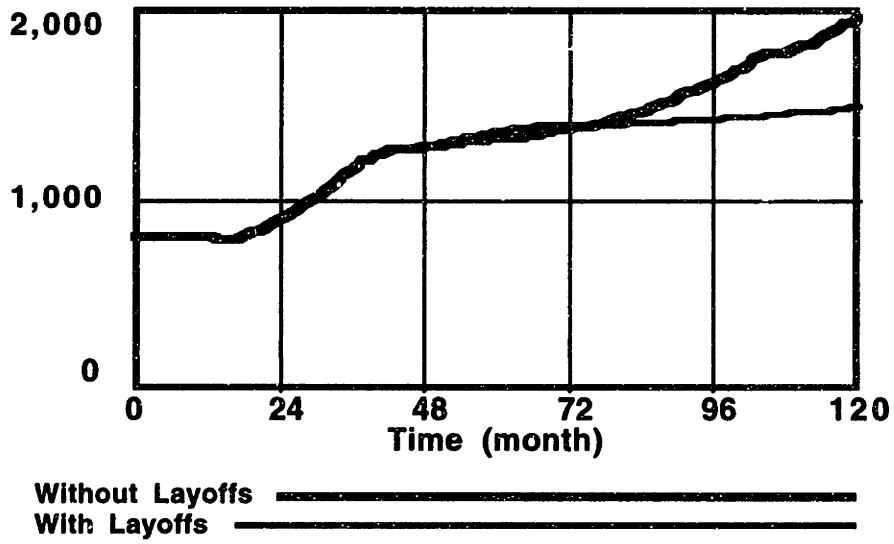


Figure 6.3

### Direct Labor

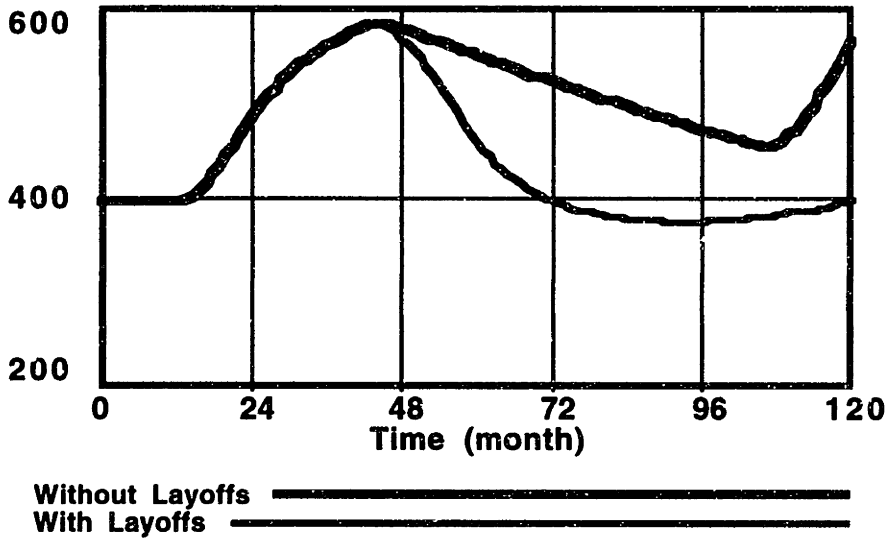


Figure 6.4

### Commitment

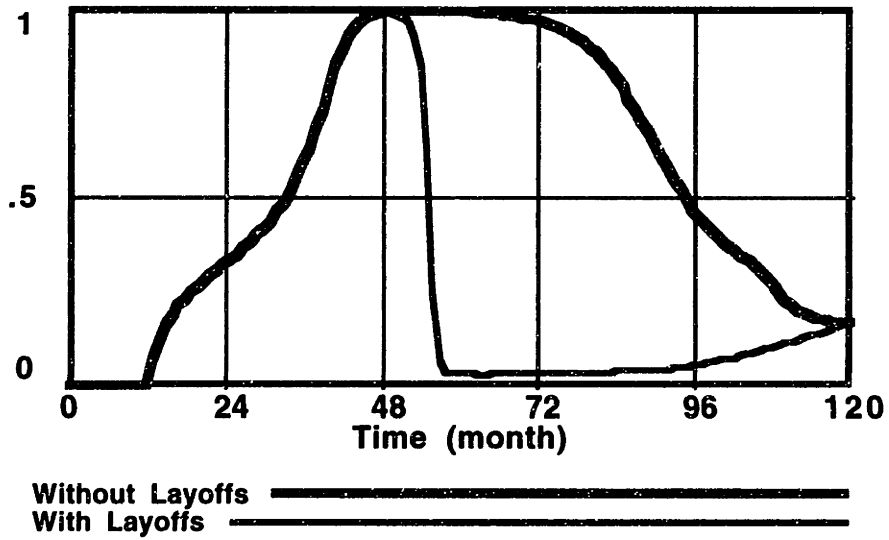


Figure 6.5

### Profit

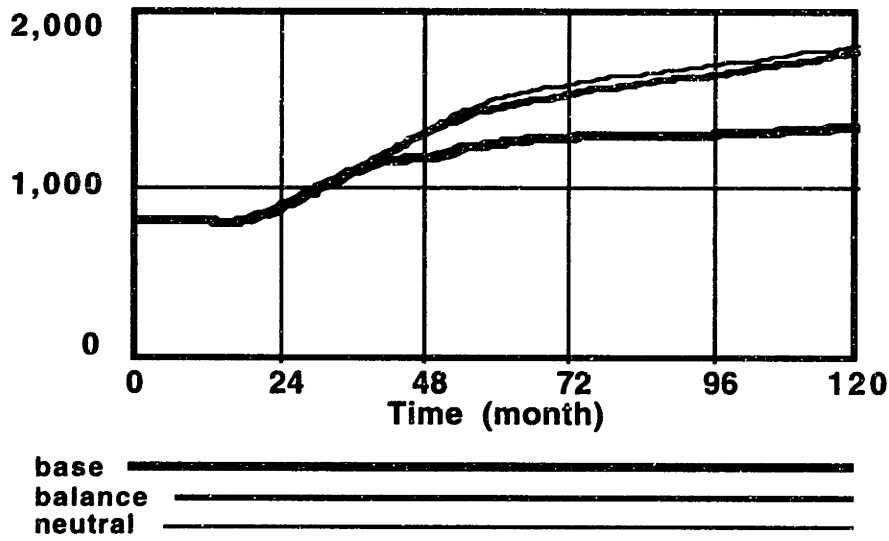


Figure 6.6

### Profit Margin

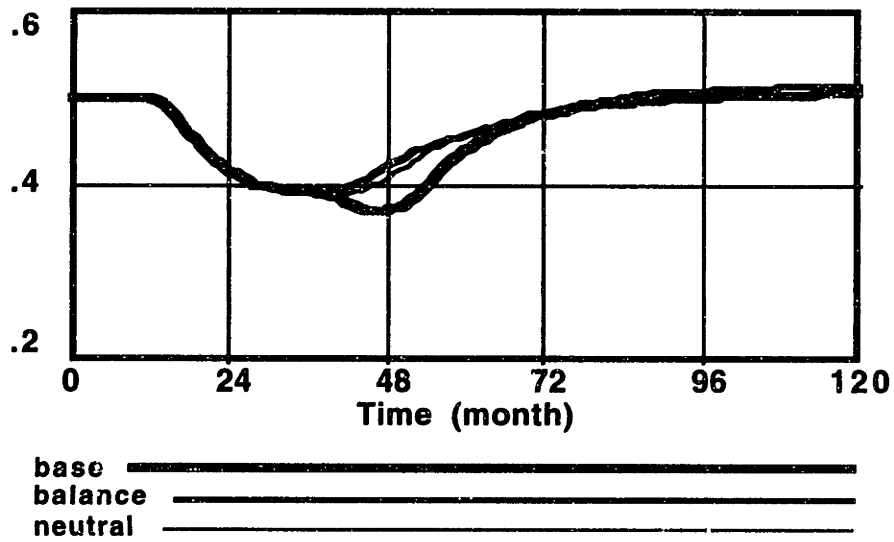


Figure 6.7

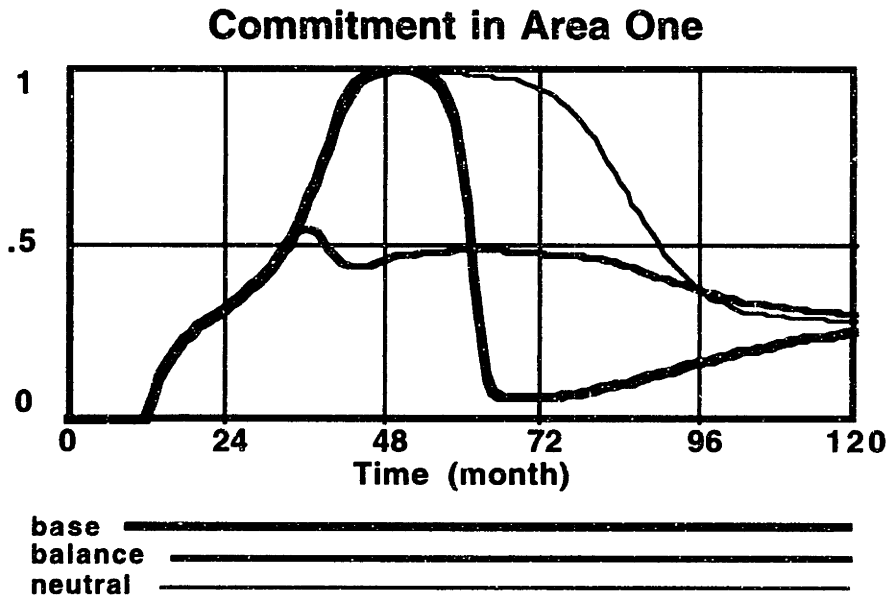


Figure 6.8

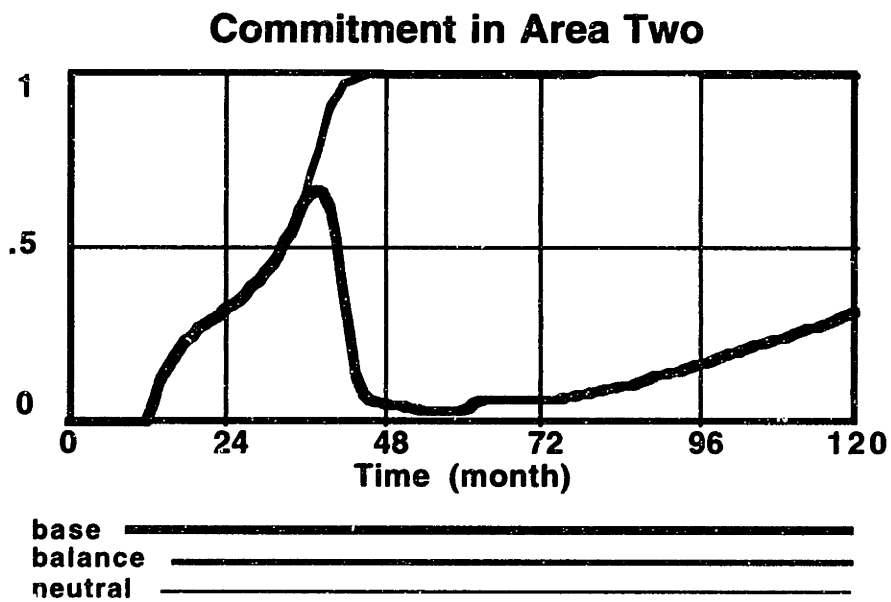


Figure 6.9

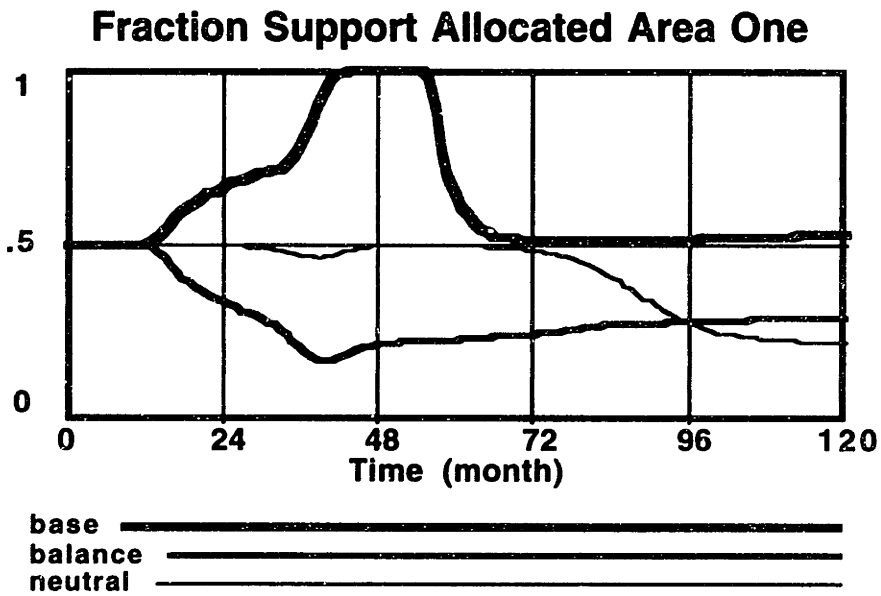


Figure 6.10

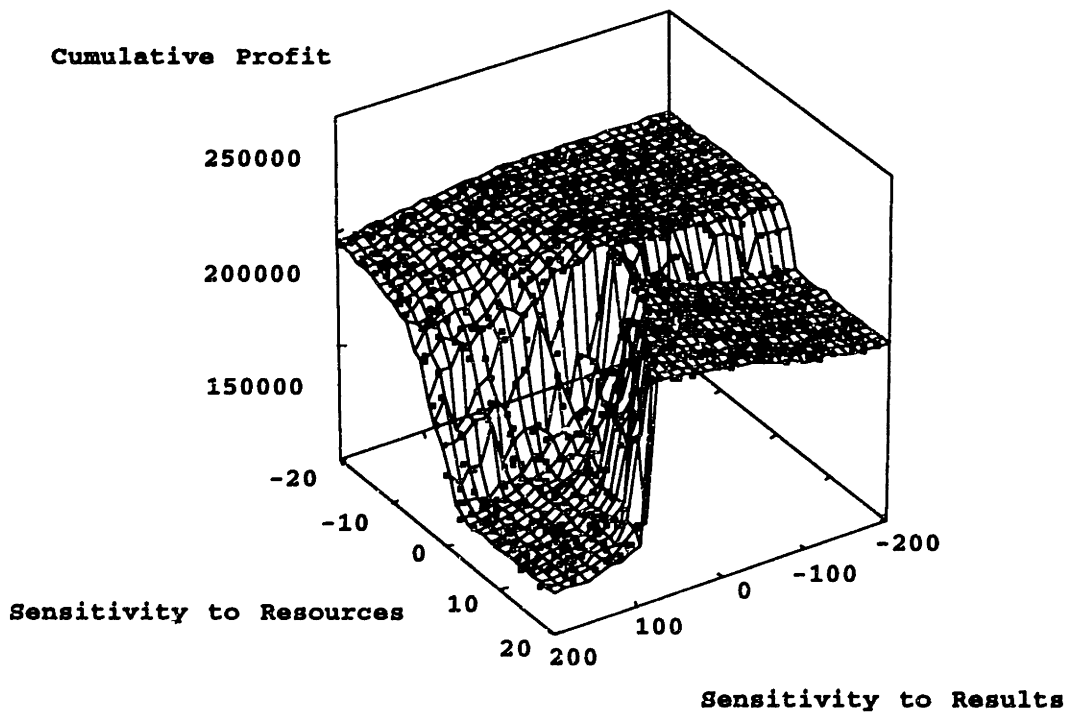


Figure 6.11

	$\beta < 0$	$\beta > 0$
$\alpha < 0$	204,941	193,566
$\alpha > 0$	184,677	143,370

Figure 6.12

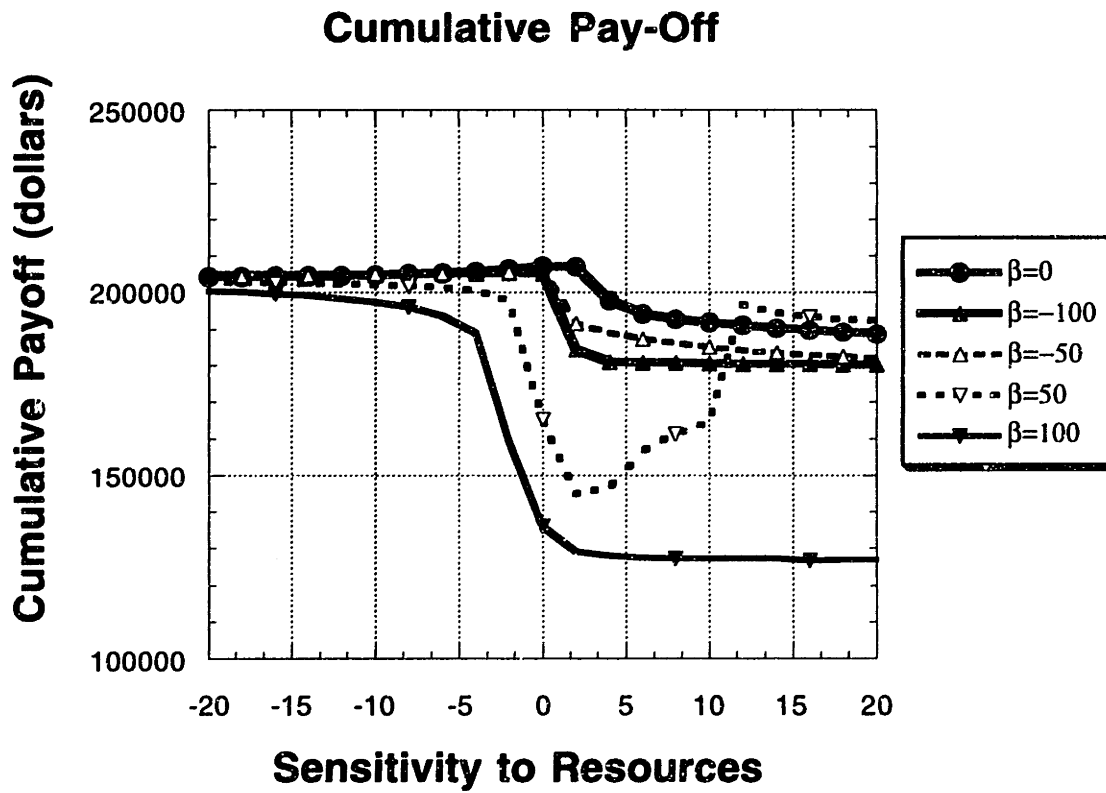
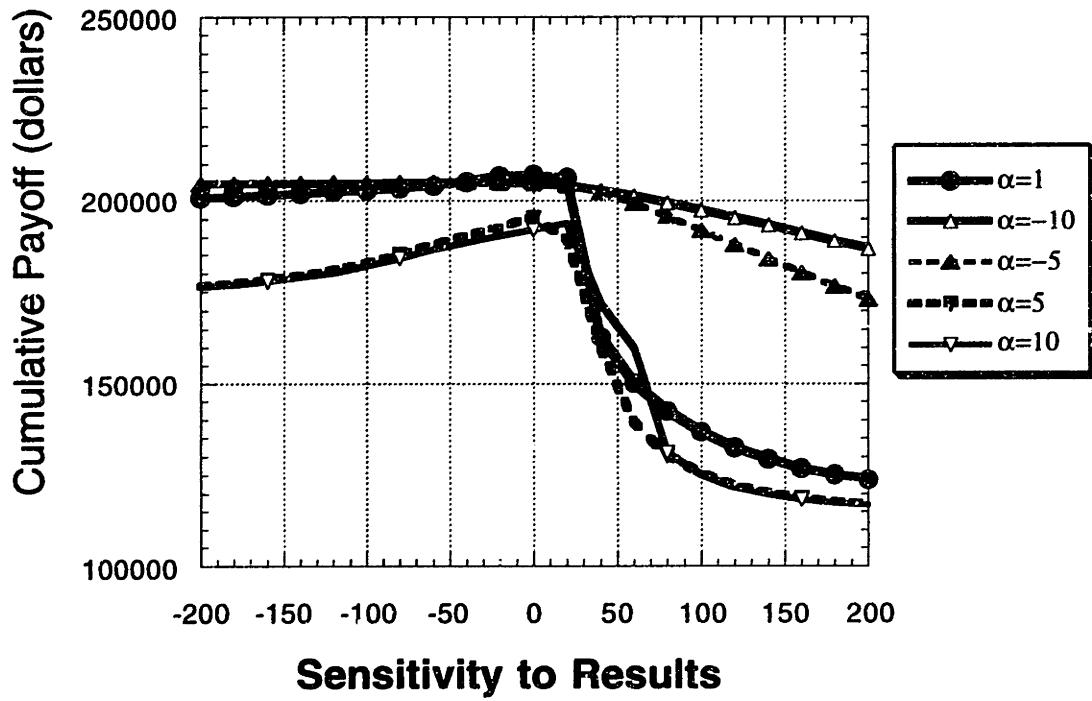


Figure 6.13

### Cumulative Pay-Off



## Essay #2

### Agency Problems in Process Improvement Efforts

#### 1. Introduction

##### **a. Motivation**

Many popular quality and productivity improvement techniques rely on the general workforce to do much of the actual 'improving'. For example, Total Quality Management (TQM) dictates that, on the factory floor, machine operators are responsible for collecting data, designing experiments, and making changes to improve quality and productivity. When firms achieve high levels of participation in programs like TQM, the results can be impressive. For example, Analog Devices, a major semiconductor manufacturer was, by internal estimates, able to double its effective production capacity in less than three years using TQM and related techniques (Sternan, Repenning, and Kofman, 1994). During that period they were also able to improve product quality and delivery reliability.

The ability of such techniques to produce dramatic improvements in productivity does, however, create an important dynamic that can limit their impact: taking advantage of productivity improvements may require firing some portion of the workforce. If a lay-off or 'downsizing' is required, employees are presented with the possibility of 'improving themselves out of a job' and thus may not wish to participate in the program. After making such dramatic improvements, Analog Devices laid off over 10% of its workforce and then saw its quality and productivity measures deteriorate for two subsequent years as participation in improvement efforts declined. Both TQM proponents and scholars have recognized this dilemma. Through his famous dictum 'Drive out Fear', W. Edwards Deming argues that a firm must assure the security of its workforce if it is to make the transformational changes required to become a 'quality' organization (Deming 1986). Levine and Tyson (1990) also study the effect job security has on the firm's ability to make productivity improvements.

While the argument that people will not be willing to 'improve themselves out of a job' is compelling, it is not totally supported by available data. In contrast to the experience of Analog Devices, McPherson (1995) reports that a division of AT&T laid off over 6,000 workers – more than half its workforce – on its way to winning the Malcolm Baldrige National Quality Award and achieving record quality and productivity improvements. Reid (1990) reports that the Harley-Davidson Motorcycle Company laid off almost half its



workforce while adopting TQM and was able to make significant improvements in quality, productivity, and profitability. Thus the link between the possibility of down-sizing and the ability of firms to implement productivity enhancing innovations such as TQM remains an open and important question for both managers and scholars.

In this paper I study the problem faced by a firm that tries to induce its workforce to reveal information that leads to productivity improvements. I develop a model in which the firm attempts to extract productivity-improving information from its workforce by providing monetary incentives for such revelations. The analysis begins by establishing the conditions under which productivity improvements are likely to lead to lay-offs. The impact of different contractual and institutional assumptions on the firm's ability to implement such programs is then investigated. There are two main results of the analysis. First, the employees' ability to collude or participate in binding side agreements – to write contracts with each other or to join a union – is a critical determinant of the firm's cost of implementing new programs. Second, the program's perceived impact on the firm's survival influences the firm's cost and the ability of employees to profitably collude. These results allow me to explain the differing experiences of the companies mentioned above, and to generate some insight about the effect that a firm's financial health has on its ability to implement programs like TQM. My theory is also consistent with the empirical study of Easton and Jarrell (1994) which finds no correlation between the adoption of TQM leading to downsizing and improved profitability.

## **b. Related Work**

My model is based on the view, first advanced by Jensen and Meckling (1976), of the firm as a 'nexus of contracts' (also discussed in Holmstrom and Tirole (1989)): the firm induces the workforce to take costly actions by offering contracts to provide incentives. Jensen and Meckling largely focus on the relationship between the owners of the firm – the shareholders – and the firm's managers. My model differs from theirs because it focuses on the relationship between managers and workers. The idea that the firm and its workers could write explicit contracts for efforts to improve productivity seems somewhat suspect – such arrangements are rarely observed. Thus, one might restrict attention to self-enforcing or implicit contracts (Hart and Holmstrom 1986). I analyze both possibilities.

Beyond the methodological approach of viewing the firm as a combination of contractual arrangements, a number of different papers are relevant to the topic discussed here. The idea that programs like TQM allow the workforce to accumulate private information

valuable to the firm has been recognized by Wruck and Jensen (1994). They argue that TQM increases the *specific* knowledge of its users – knowledge that is costly to transfer. Proponents of TQM have written extensively about implementation and the effect job security has on improvement (Deming 1986 and Shiba, Walden and Graham 1993 are just two examples).

The idea that firms face difficulties in getting employees to reveal important information is not new. The information economics literature has analyzed extensively the general problems faced by principals trying to induce agents to take costly actions (moral hazard) or reveal costly information (adverse selection). Hart and Holmstrom 1986 provide a survey. My approach is somewhat different because all actions are observable, unlike moral hazard models, and information is only generated if employees take observable actions, unlike adverse selection models. The model presented here is in many ways simpler, however it still yields distortions. These result from the institutional structure in which the contracting relationship takes place rather than the informational structure. Also consistent with the ‘nexus of contracts approach’ I allow the possibility of side contracting and collusion following the methodology of Tirole (1986).

Finally, there is a strand of literature that discusses lay-offs, quits, and the impact of unionized workforces. Grossman (1983) and Weiss (1985) argue that unions maximize the utility of their senior members, perhaps at the expense of more junior ones. My model gives an alternative interpretation: union-like arrangements may arise endogenously to better distribute the gains from productivity improvements between the firm and its employees.

### **c. Outline of Paper**

The paper is organized as follows: In Section 2, the conditions under which productivity improvements are likely to lead to lay-offs are discussed. In Section 3 the main model is presented. In section 4, the problem of inducing the workforce to reveal information that leads to possible lay-offs is analyzed. In section 5, the possibility of collusion is introduced, and in section 6, I show how the workforce’s fear of firm failure influences the results. Section 7 contains discussion, and section 8 presents concluding thoughts, and future directions.

## 2. When Do Improvements Yield Lay-offs?

Before proceeding to the main focus of the paper, I address the question of when productivity improvements are likely to lead to lay-offs by developing a simple model of the firm that yields a closed form solution for the elasticity of demand for labor with respect to changes in productivity. Let the firm be a monopolist producing a homogeneous good. The firm's production technology requires labor,  $L$ , and raw materials. For simplicity, raw materials are never a production constraint and their content within the firm's product is constant and results in a material cost of  $\rho$  dollars per unit produced. Each laborer can produce  $\alpha$  units of output per period without any additional constraint, and earns a wage  $w$  each period. The production and cost functions for the firm are:

$$Q = \alpha L \quad (2.1)$$

$$C(Q) = wL + \rho Q \quad (2.2)$$

The demand curve has constant elasticity and is represented by:

$$Q = P^{-\varepsilon} \quad (2.3)$$

where  $P$  is the price of output.

Using standard arguments (Varian 1992), the firm sets its optimal price  $P^*$ , output quantity  $Q^*$ , and demand for labor  $L^*$ , as:

$$P^* = \frac{[\frac{w}{\alpha} + \rho]}{1 - 1/\varepsilon}$$

$$Q^* = (P^*)^{-\varepsilon}$$

$$L^* = \frac{Q^*}{\alpha}$$

To analyze the effect of a change in productivity on the demand for labor, differentiate  $L^*$  with respect to  $\alpha$ :

$$\frac{dL^*}{d\alpha} = -\frac{L^*}{\alpha} + \frac{1}{\alpha} \cdot \frac{dQ^*}{dP^*} \cdot \frac{dP^*}{d\alpha}$$

The change in labor demand caused by changes in productivity,  $dL^*/d\alpha$ , can then be decomposed into two pieces. The first piece,  $-L^*/\alpha$ , represents the ‘direct’ effect – as productivity increases, current output can be produced with fewer workers. The second piece represents the change in the demand for labor caused by the increase in the demand for the firm’s product. Demand for output increases because the increase in productivity allows the firm to lower its price. Expanding the second component further using the definitions for  $Q^*$  and  $P^*$  yields:

$$\frac{dL^*}{d\alpha} = -\frac{L^*}{\alpha} + \frac{Q^*}{\alpha^2} \cdot \varepsilon \cdot \left( \frac{\frac{w}{\alpha}}{\frac{w}{\alpha} + \rho} \right)$$

which reduces to:

$$\frac{dL^*}{d\alpha} \cdot \frac{\alpha}{L^*} = (\varepsilon \cdot f_1 - 1)$$

where  $f_1$  equals  $\frac{w}{\alpha} / \left( \frac{w}{\alpha} + \rho \right)$ , the fraction of total cost resulting from expenditure on labor.

The term  $(\varepsilon \cdot f_1 - 1)$  is the elasticity of demand for labor with respect to productivity. The intuition is that  $\varepsilon$  is the price elasticity of demand, while  $f_1$ , given the optimal pricing rule, is the fraction of the price determined by the cost of labor. If  $\varepsilon \cdot f_1$  is greater than one, making labor more productive increases the demand for labor, while if  $\varepsilon \cdot f_1$  is less than one, making L more productive decreases the demand for labor.<sup>1</sup>

The analysis suggests that two factors will determine whether a successful improvement effort will generate excess labor capacity. First, if demand is very inelastic, there will be little opportunity to use the excess capacity created by productivity growth. Second, if labor is a small portion of total cost, as in high-tech industries like semi-conductors, improving productivity will have a small impact on price and thus do little to stimulate demand even if it is price sensitive.

---

<sup>1</sup> . A similar analysis has been done with the more general class of CES production functions.

### 3. The Model

The environment being modeled is one in which the firm is periodically presented with a managerial innovation that, if adopted, might increase the productivity of its workforce. The critical assumption is that the nature of these innovations is such that the firm cannot adopt them directly. Instead, the firm must provide some incentive scheme that induces the workforce to do the adopting. The types of innovations under consideration are thus not purely technical in nature – the firm cannot just purchase them. Instead, these innovations involve a combination of organizational methods and technical tools that allow workers to utilize their accumulated experience more effectively (organizing technologies in the language of Wruck and Jensen (1994)). A key feature of these tools is that once they have been used, the new knowledge generated becomes available to all in the firm. In this model, once a particular improvement has been made, it becomes a permanent part of the technology set available to the firm and does not disappear even if the worker who made that improvement is subsequently fired.

The archetypal example of such an organizational/managerial innovation is TQM and the Plan-Do-Check-Act cycle of W. Edwards Deming. As pointed out by Wruck and Jensen (1994), TQM represents a new way of organizing productive activities that allows the firm to take better advantage of the accumulated knowledge of its workforce. By focusing efforts on the elimination of the root causes of defects, the TQM methodology allows firms to make permanent improvements in their productive processes. For example, by applying the TQM method in a manufacturing setting, workers on an improvement team might discover a better way to set-up a particular job so that product defects are permanently reduced. Once implemented, the new procedure remains in place even if those workers subsequently leave. The key feature of this method is that quality is the responsibility of those who do the productive work. The firm cannot ‘do’ TQM directly, rather it has to find some way to induce its workforce to use the method.

#### **a. Set-Up**

The model to be developed is a stochastic game composed of an infinitely lived firm and the labor force it must hire to produce output. The model takes place in discrete periods. The length of a period corresponds to the frequency with which the firm is able to change its labor hiring – the length of the basic labor contract.

## Workforce Actions and Preferences

Each member of the workforce can engage in two different activities: normal work and improvement work. Normal work represents activities related to production and improvement work is focused on improving the productivity of that production process. Normal work is governed by a pre-existing contract that pays each worker a wage  $w$  in return for making productive efforts with a dis-utility of  $e$ . Workers are risk averse and have a utility function  $U(w,e)$  that is quasi-linear in income and effort,  $U(w,e)=u(w)-e$ .  $u(\cdot)$ , the utility over income, satisfies the usual restrictions,  $u'>0$ , and  $u''<0$ . Critically, I further assume that  $u(w)>e$  – the equilibrium wage is such that people strictly prefer employment within the firm to the best available alternative. There are numerous theories that justify such an assumption; they include costly search, efficiency wage theories, and the equilibrium unemployment theory of Shapiro and Stiglitz (1984). Those fired in period  $t$  are assumed to earn 0 utility in all future periods. Since it plays no role in the analysis, I normalize the cost of normal efforts,  $e$ , to zero.

## Adoption and Contracts

Each period the firm learns of a new managerial innovation that has the potential to improve productivity. *Ex ante* the firm does not know whether the innovation is legitimate. With probability  $p$ , the innovation yields improvement if adopted, and with probability  $1-p$ , productivity remains the same regardless of improvement efforts. Given a legitimate innovation, the firm, if it is to reap the benefits, must induce each member of the workforce to engage in improvement work and *adopt* the innovation. The adoption decision is modeled by assuming there is a set,  $I$ , of workers, and each worker,  $i \in I$ , makes a binary choice,  $a_i \in A_i = \{0, 1\}$ , to adopt,  $a_i=1$ , or not adopt  $a_i=0$ . Let  $\mathbf{a}=(a_1, \dots, a_I)$ . If a worker adopts, she incurs a private cost  $c>0$ , representing the dis-utility of obtaining the required training and making any extra efforts required to use the new tools. The cost of adopting enters the utility function linearly,  $U(w,e,a) = u(w)-c(a)$ .

There are two basic contracting frameworks used in the analysis. In the first, a worker's improvement effort,  $a_i$ , is both observable and verifiable – contracts are complete. The firm can offer a contract  $\{(0,0),(1,b)\}$  that pays a bonus,  $b$ , to each worker if she is observed to have adopted the innovation and 0 otherwise. If there is no possibility of lay-off then the firm, to induce adoption, must set  $b$  such that the incentive compatibility constraint (IC) is satisfied:

$$u(w+b)-c \geq u(w) \quad (\text{IC})$$

Let the value of  $b$  that satisfies (IC) with equality be  $b^*$ . The main implication of the verifiability assumption is that such a contract is enforceable by an outside party and hence the firm cannot renege on its commitment to pay  $b$  even if it turns out that the innovation was not legitimate or if that worker was subsequently fired. Under such an arrangement any innovation whose benefit exceeds its cost ( $c$  times the number of workers) gets implemented.

In contrast to the observable and verifiable case, the situation in which workers' adoption decisions are not verifiable and cannot be the basis for enforceable contracts is also considered. A solution to the lack of enforceability of contracts is for workers and the firm to agree to an *implicit contract*. In such a scenario, the firm agrees to pay workers for their efforts even if the innovation does not turn out to be legitimate. Such an arrangement is not feasible in a one period game because the firm will always wish to renege on such an agreement *ex post*. However, in a multi-period setting, an implicit contract may be possible since the firm is better off in the long run if it has the reputation for honoring its implicit agreements.

To model this arrangement, assume that each worker uses a trigger strategy that dictates cooperation as long as the firm honored its commitment to that worker in the previous period and no cooperation if the firm has failed to honor its commitment in any previous period. Using such strategies, and assuming the firm honored its commitments in previous periods, workers will again adopt the innovation if  $b \geq b^*$ . If the firm has reneged on its commitment in any previous period, workers never accept a future offer.

For the moment, let  $\pi$  be the incremental increase in profit generated by an additional adoption and assume that the firm discounts future benefits with the discount rate  $\delta$ . To sustain a cooperative agreement the firm's present discounted value of future benefits from the cooperative regime must exceed the pay-off it could receive from not honoring the agreement, not paying  $b$ , and then playing the non-cooperative outcome for every period afterwards. Thus define the reneging constraint as:

$$(p_s \cdot \pi - c) / \delta \geq p_s \cdot \pi \quad (\text{RG})$$

Which reduces to:

$$p_s \cdot \pi (1 - \delta) \geq c$$

An implicit contracting relationship of the type outlined will exist as long as the firm is sufficiently patient – if  $\delta$  is close enough to 0. Such an arrangement, if feasible, is efficient since the risk neutral firm now bears all the risk, and any innovation whose expected benefit exceed its cost again gets implemented. For the remainder of the analysis I assume that the required conditions are satisfied and that an implicit contract of this type is feasible. It is important to note that in this model workers only look at their own past experience to determine whether or not to trust the firm in the future.

## b. Productivity and Lay-offs

Standard principal agent models assume the agent has the opportunity to engage in some activity that generates profit for the principal without specifying the exact nature of that activity. Here the agent's efforts have a specific impact: they improve the productivity of the agent's normal work. Thus, additional structure connecting the workers' actions with the realized productivity and *ex post* labor requirements is required.

### *Productivity*

Improvements manifest themselves through changes in the parameter  $\alpha$ , the productivity of labor. If the innovation is legitimate, then the realized productivity in the period following the adoption of the innovation is a function,  $\alpha_{t+1}(\cdot)$ , that maps from the space of adoptions,  $A = \prod_{i=1}^I A_i$ , to the real line,  $\alpha_{t+1}:A \rightarrow \mathbb{R}$ . It is easiest, although not necessary, to think of  $\alpha$  as being simply a function of the sum of the  $a_i$ 's:  $\alpha_{t+1} = \alpha(\sum a_i)$ . If all workers in  $I$  adopt, then the innovation produces the maximum gain in productivity,  $\alpha' = \alpha_{t+1}(I)$ . Realized productivity is assumed to be increasing in the number of adopters,  $\alpha(n) \geq \alpha(n-1)$ , and the first differences are assumed to be decreasing,  $\alpha(n) - \alpha(n-1) \geq \alpha(n+1) - \alpha(n)$ , representing a diminishing marginal improvement.<sup>2</sup> Let  $L^*(\alpha_{t+1})$  be the optimal demand for labor in period  $t+1$  after the productivity improvements in period  $t$  are realized. For clarity I add the

---

<sup>2</sup> . Realistically, the curve is probably S-shaped rather than concave. Most of the results presented do not depend critically on this assumption although in some cases, if the s-shape were used, an assumption on the location of the point of inflection would be required.



assumption that an additional adopter saves the firm at most one job. Thus,  $L^*(\alpha(n-1)) - L^*(\alpha(n)) \leq 1$ .<sup>3</sup>

### *Demand for Labor*

The state of the game,  $L_t$ , is the number of workers required in period  $t$ . The firm's demand for labor in period  $t+1$ ,  $L_{t+1}^*$ , is a function of two variables, the realized productivity,  $\alpha_{t+1}$ , (determined in period  $t$ ) and the growth in demand from the previous period,  $g_t - L_{t+1}^* = L^*(\alpha_{t+1}, g_t)$ .<sup>4</sup> For simplicity, there are two possible growth rates: with probability  $p_g$ , growth will be  $g$ , and with probability  $1-p_g$ , the growth rate will equal 0. Given that innovations are legitimate with probability  $p_s$ , there are four possible state transitions. With probability  $p_s p_g$  the innovation is both legitimate and growth is  $g$ . In this state the growth rate is sufficiently large that, even if the innovation produces its maximum improvement in productivity, there is no possibility of lay-offs,  $L_{t+1}^*(\alpha(I), g) \geq L_t^*$ . With probability  $(1-p_s)p_g$  growth is positive and the innovation is not legitimate, in this case the firm will increase its hiring for the following period,  $L_{t+1}^*(\alpha_t, g) > L_t^*$ . With probability  $(1-p_s)(1-p_g)$  nothing happens and the state remains the same through the period, and with probability  $p_s(1-p_g)$  the innovation is legitimate and there is no growth,  $L_{t+1}^*(\alpha_{t+1}(I), 0) < L_t^*$ . Let  $p_d = p_s(1-p_g)$ . It is in this state that productivity improvements have the potential to lead to lay-offs.

### *Timing*

Within each period the model proceeds through four stages (see figure 1). At the beginning of the period the firm learns of a new innovation and all parties learn whether or not the growth rate  $g$  is positive or zero. The firm then offers its workers some type of incentive scheme or contract to induce them to adopt. After seeing the firm's offer, but before learning whether or not the innovation is legitimate, each member of the workforce decides whether or not to adopt. Following the adoption decision, any improvement is realized and then the firm makes its hiring/firing decisions for the following period.

<sup>3</sup>. Clearly this must be true for all  $n$  greater than some threshold, otherwise labor requirements would be negative if the innovation was fully adopted. Strengthening this requirement to all  $n$  simplifies many of the arguments that follow but is not necessary.

<sup>4</sup>. For simplicity one can think of this set-up as a firm that makes product to order so that orders received in period  $t$  are not actually produced until period  $t+1$ .

## 4. Analysis

### **a. Statement of the Problem**

The focus of the analysis is the problem faced by the firm in periods in which lay-offs are possible, when  $L_{t+1}^*(\alpha_{t+1}(I),0) < L_t^*$ . The game is analyzed using the behavioral assumption that players' strategies conform to a sub-game perfect Nash equilibrium (Fudenberg and Tirole 1992). Working backwards, at the end of period  $t$  the firm observes the realized improvement for that period and makes the firing decision for the following period. Formally, the firm's strategy is a map from the observed adoption decisions and the realized productivity to the probability that each worker gets laid off. Thus

$P = \times_{i=1}^I P_i$ ,  $P_i \in [0,1]$  and  $p: A \times [\alpha, \alpha'] \rightarrow P$  where  $p_i(\mathbf{a})$  is the probability that worker  $i$  gets

fired given the adoption vector  $\mathbf{a}$ .<sup>5</sup> Given the requirement of sub-game perfection, after improvements have been realized the firm will always lay-off as many workers as is necessary to reach the optimum level. I assume that the firm cannot credibly commit to a 'no lay-off' policy and, as a result, will always choose  $\mathbf{p}$  such that:

$$\sum_i (1-p_i) = L^*(\alpha_{t+1}(\mathbf{a}), 0) \quad (4.1)$$

Given (4.1) the firm still has a large number of options. For example it may choose to lay-off all workers with equal probability or it may choose to keep some workers with probability one and fire others for sure.

Let  $(a_i, a_{-i})$  denote  $i$ 's choice of action holding the choices of other players constant and assume that workers discount future benefits at rate  $r$ . The probability that a worker gets fired and does not continue working for the firm is  $(1-p_i \cdot p_i(a_i, a_{-i}))$ . In period  $t$ , the effective discount rate that workers apply to period  $t+1$  is  $(1-r) \cdot (1-p_i \cdot p_i(a_i, a_{-i}))$ . Let  $V_i(a_i, a_{-i})$  represent a worker's expected pay-off in the continuation game if that worker is not fired.

The decision problem faced by members of the workforce depends on the contracting arrangement. Under the complete contracts regime, the firm can credibly commit to paying the bonus  $b$  even to those that it will fire. In equilibrium each worker's conjecture about  $p_i$  given an adoption vector  $\mathbf{a}$  is correct, thus, if she is purely self-interested, player  $i$  will choose  $a_i$  to maximize her expected utility:

---

<sup>5</sup> . The variable  $P$  is redefined here as the space of probabilities and no longer represents price.

$$a_i \in \arg \max u(w + b(a_i)) - c(a_i) + (1 - r)(1 - p_s \cdot p_i(a_i, a_{-i})) \cdot V_i(a_i, a_{-i}) \quad (\text{IC-CC})$$

$b(a_i)$  is the incentive scheme offered by the firm in stage one. If the contract is implicit the firm cannot credibly commit to paying  $b$  to those it will fire. Thus workers solve:

$$a_i \in \arg \max (1 - p_s \cdot p_i(a_i, a_{-i})) (u(w + b(a_i)) + (1 - r)V_i(a_i, a_{-i})) + p_s \cdot p_i(a_i, a_{-i}) \cdot u(w) - c(a_i) \quad (\text{IC-IC})$$

Finally, given its actions in step three, and the actions of the workers in step two, in the first step the firm must choose  $b(a_i)$ , the incentive scheme it offers to each member of the workforce, to maximize *ex ante* profits. The firm solves:

$$\begin{aligned} \max_{\mathbf{p}(\mathbf{a}), \mathbf{b}(\mathbf{a})} & p_s \pi(\alpha_{t+1}(\mathbf{a})) + (1 - p_s) \pi_{t-1} - \sum_I b_i(a_i) \\ \text{s.t.} & \sum_I (1 - p_i(\mathbf{a})) = L^*(\alpha_{t+1}) \\ & b_i(a_i) \geq 0 \forall i \in I \\ & (\text{IC - CC}) \text{ or } (\text{IC - IC}) \text{ depending on contracting scheme} \end{aligned}$$

The firm must select an incentive scheme,  $\mathbf{b}(\mathbf{a})$ , and a lay-off strategy (one lay-off vector  $\mathbf{p}(\cdot)$  for each adoption vector  $\mathbf{a}$ ) to maximize expected profits given that each member of the workforce will maximize her expected utility, and that the firm cannot, by assumption, credibly commit to job security.

### b. Pure Strategy Equilibria

Consider first the case in which the firm cannot use randomized lay-off schemes –

$p_i \in \{0, 1\}$ . This may be true for a number of reasons. First, the firm may not control both the size and the composition of the lay-off. Via seniority, union rules, or otherwise, the firm may be forced to lay-off workers in some previously established order. Second, absent an explicit ordering for lay-offs, there may exist an implicit ordering that is common knowledge. For example, firms often lay-off workers with the most accumulated time on the job because they earn the highest salaries. For the purpose of the analysis in this section, I assume that there is an order on the set of workers known to all ( $i > j$  means  $i$  gets laid off before  $j$ ).

To find an equilibrium when the firm is restricted to pure strategies, it is helpful to define two subsets of  $I$ ,  $\Gamma(\alpha^*)$  and  $\Omega(\alpha^*)$ . Let  $\Gamma(\alpha^*)$  be the set of workers who adopt in equilibrium and for whom the firm sets  $p_i=0$  if the realized productivity is  $\alpha^*$ , and let  $\Omega(\alpha^*)$  be the set of workers who adopt in equilibrium and for whom the firm sets  $p_i=1$  if the realized productivity is  $\alpha^*$ . For members of  $\Gamma(\alpha^*)$  the decision to adopt only affects their continuation pay-offs,  $V_i$ , since they will not be fired in the current period. Conversely, for members of  $\Omega(\alpha^*)$ , their decision to adopt can only affect their pay-offs in the current period since their continuation pay-offs are zero.

First consider the decision problem faced by members of  $\Gamma(\alpha^*)$ . By construction their decision to adopt does not lead to their being laid off in the current period. They can be induced to adopt with a scheme that pays:

$$b^{cc}=b^*+(1-r)(V_i(0,a_i)-V_i(1,a_i)) \quad (4.2)$$

The decision problem faced by members of  $\Omega(\alpha^*)$  depends on the contracting scheme.

### *Complete Contracts*

If contracts are complete, then the firm can credibly commit to paying  $b(a_i)$  to members of  $\Omega(\alpha^*)$  if they adopt, and, because they have a continuation pay-off of zero, they can be induced to adopt if the firm offers  $b^*$  in return for their adopting. Given the incentive schemes that induce members  $\Gamma(\alpha^*)$  and  $\Omega(\alpha^*)$  to adopt there exists an equilibrium in which the set of non adopters,  $(\Gamma(\alpha^*) \cup \Omega(\alpha^*))^c$ , is small (in a sense to be made precise in below). The key insight is that, no matter how much productivity improves, the firm will never lay-off all its workforce, and, thus, the set of adopters is non-empty,  $\Gamma(\alpha^*) \supseteq \Gamma(\alpha')$ .

Since some adopt, productivity rises, and some are guaranteed to be fired –  $\Omega(\alpha^*)$  is not empty. Those who will be fired regardless of their actions also adopt. The basic argument is that each set of adopters creates a set that will be fired for sure and, thus, further enlarges the set of adopters. The line of reasoning can be continued, and I show in the following proposition that, if contracts are complete, in equilibrium the innovation can be *almost fully* implemented.

**Proposition I:** If contracts are complete and of the type discussed above, then there exists a sub-game perfect Nash equilibrium in which the set of non-adopters has at most  $(1/(1-\beta))$  members (only one member for all  $\beta \leq 1/2$ ) where  $\beta$  is the reduction in labor requirement caused by the adoption of the lowest ranking player in  $\Gamma(\alpha)^c$  (the last person to be laid off).

**Proof:** see appendix

For intuition, consider the following example: The firm starts with 100 workers and learns of an innovation with  $p_s=1$  that, if fully adopted, could reduce labor requirements by 50%. Assume further that the marginal improvement in productivity is constant so that each adoption saves one half of a worker. In this case the first 50 are assured job security regardless of the realized productivity, hence they adopt given the scheme in (4.2). Because the first 50 adopt, at least 25 jobs will be eliminated – workers 76 through 100 will be fired – hence at least 75 adopt. This process can be continued. However, does the last worker to be fired, number 51, adopt? Perhaps not. If she does not, then, *ex post*, the firm has a labor requirement of 50.5. Without further assumptions one does not know whether the firm is better off with 50 or 51 workers. Assume it prefers 51, then player 51 can save her job by not adopting. Can #52 save her job? No. If 51 and 52 do not adopt then the *ex post* labor requirement is still 51 and player 52 gets fired. Thus in this example the innovation gets implemented *almost fully*, meaning all adopt except those few who can unilaterally save their jobs by not adopting.

### *Incomplete Contracts*

Now consider the case in which contracts are incomplete and enforced via reputation. The key feature is that the firm *cannot* credibly commit to the incentive scheme for those it plans to fire. The period  $t$  pay-offs for members of  $\Gamma(\alpha^*)$  do not change – their decisions only affect their continuation pay-offs – and they can again be induced to adopt with  $b \geq b^{cc}$ . Others, however, risk investing  $c$  and not getting compensated if they are fired. To induce this group to adopt, the firm must set  $b$  such that:

$$(1-p_s)(u(w+b)+(1-r) \cdot V_i(1, a_{.i})) + p_s \cdot u(w) - c \geq u(w) + (1-p_s)(1-r) \cdot V_i(0, a_{.i})$$

Which implies the firm must set  $b$  such that:

$$u(w+b) - u(w) \geq c / (1-p_s) \tag{4.3}$$

The expected continuation pay-off,  $V$ , for members of this group is equal in both states of the world. However, to induce them to adopt the firm must offer an extra insurance payment above and beyond the cost of adopting. If the firm chooses to offer this payment its cost of inducing adoption rises. As  $p_s$  approaches 1 it becomes prohibitively costly to induce those that will be fired to adopt, and if  $p_s=1$ , obviously, there is no payment that can satisfy (4.3).

There are, then, at least two possible equilibria. If  $p_s$  is sufficiently small, the firm offers  $b^*$  to members of  $\Omega(\alpha^*)$  and  $b^{cc}$  to members of  $\Gamma(\alpha^*)$ . The innovation is almost fully adopted as in the complete contracts case. A second possibility, if  $p_s$  is large enough to make the first scheme unprofitable, is to offer  $b^{cc}$  to all. Those not fired adopt, and those laid off do not. The innovation is then adopted by only a fraction of the workforce. Here it is most clear to proceed graphically. Again in equilibrium those guaranteed job security adopt, implying that there must be some increase in productivity, which, in turn, means that some are guaranteed to be fired. However, as shown in figure 2, in contrast to the previous case, the members of  $\Omega(\alpha^*)$  do not adopt which means that  $\alpha'$  will never be realized in equilibrium, thus implying that  $\Gamma(\alpha^*)$  is larger than  $\Gamma(\alpha')$ , which, in turn, implies that  $\Omega(\alpha^*)$  is smaller than  $\Omega(\alpha')$ . This process also can be continued and, as I formalize in the following proposition, there is an equilibrium in which the innovation is adopted by some fraction of the workforce and the realized productivity is less than the potential of the innovation.

**Proposition II:** If the contract is implicit and the firm offers  $\{(0,0),(1,b^{cc})\}$  then:

- i) there exists an equilibrium such that  $\Gamma(\alpha^*) \supset \Gamma(\alpha')$  and  $\Omega(\alpha') \supset \Omega(\alpha^*)$
- ii) those that are not fired adopt and those that are fired do not
- iii) if the marginal benefit of an additional adoption is constant, then the number of adopters is equal to largest integer less than  $I/(1+\beta)$  where  $\beta$  is the reduction in labor requirements that results from an additional adoption.

Proof: see appendix

Returning to the previous example with 100 people and each adoption reducing *ex post* requirements by 1/2, at least 66 people adopt and 33 do not and are fired.

### *Cost of Adoption*

Members of  $\Omega(\alpha^*)$  are paid either  $b^*$  or 0. Members of  $\Gamma(\alpha^*)$ , however, require a bonus sufficiently large to compensate them for their impact on their expected continuation pay-offs,  $(1-r)(V_i(0,1)-V_i(1,1))$ . A necessary condition for a member of  $\Gamma(\alpha^*)$  to have a profitable deviation is that, in some future sequence of plays, that player is the lowest ranking person in a lay-off. If we let a player  $j$  be the lowest ranking person in a given lay-off then a deviation by any player  $k > j$  improves  $j$ 's pay-off not  $k$ 's. However, every player could be the last fired in *some* sequence of growth states and lay-off states. To approximate this cost I add more restrictive assumptions. Specifically, assume that growth and productivity improvements (in equilibrium) are exactly offsetting. That is, in the growth state labor requirements increase by a factor of  $\beta$ , and a fully adopted productivity improvement decreases them by  $1/\beta$ . If both occur simultaneously, labor requirements are unchanged. Also assume that only  $n$  consecutive lay-offs are possible before the firm reaches a minimum efficient scale. With these restrictions the only members in  $\Gamma(\alpha^*)$  who have profitable deviations are the  $n$  players who have the lowest rank in the  $n$  consecutive lay-offs. The firm can ensure adoption in one of the two possible equilibria if it can induce these players to adopt. The firm must pay each of these players a *blocking* payment of  $p_s \cdot (1-r)^n (V_i(0,1) - V_i(1,1))$ . The blocking payment is at most  $p_s \cdot (1-r)^n u(w)/r$ , since, at best, a deviation would prevent that player from ever being laid off. Thus the firm's total extra cost of implementation is at most:

$$p_s \cdot \sum_n (1-r)^n u(w)/r$$

where  $n$  is the number of consecutive lay-offs required for the firm to reach its minimum efficient scale.

Returning to the earlier example, can any player in the first 50 (those not laid off in period  $t$ ) improve her continuation pay-off by a deviation? Any deviation by someone in the first 50 causes the *ex post* labor requirement to be at most 52 players. If another innovation comes along that again saves 1/2 a worker per adopter, who benefits from the previous deviation? If all others adopt the *ex post* requirement will be at most 26 players, and player 26 might have saved her job by deviating. The firm must pay 26 an extra *blocking* payment to maintain full adoption in equilibrium. How much will that payment be? Clearly it is bounded above by  $(1-r)u(w)/r$ . A similar situation obtains for player 13. In principle

this process could continue for every subsequent division of the existing workforce by two. However, given a minimum efficient scale for the firm, this does not appear to be a critical issue. If only three consecutive lay-offs are possible, the firm's total excess cost of implementation is at most  $3 \cdot u(w)/r$  even though the total loss in workforce utility is  $50 \cdot u(w)/r$ .

Thus, if it can commit to  $b(a_i)$  – if contracts are complete – the firm can implement the innovation at close to minimum cost. If contracts are implicit, the firm's cost of implementing is still close to the minimum, but the innovation is only partially adopted. The main intuition developed is that adopting is a weakly dominant strategy unless a player can save her job unilaterally by not adopting, and, in general this is not possible because a player's decision to adopt generally affects someone else's pay-offs. The 'mismatch' between an individual's actions and their consequences allows the firm to implement the innovation at a low cost and reap most of the benefit it generates.

### c. Mixed Strategy Equilibria

In contrast to the previous section, lay-offs may also have a random component. To analyze this I consider a policy in which the firm lays off workers with equal probability,  $p_L[a]$ . The efficacy of this policy relies on the firm not announcing in advance who it wishes to lay-off and ensuring that employees cannot make additional investments – 'influence activities' in the language of Milgrom (1988) – to tip the scales in their favor. In fact, Meyer, Milgrom, and Roberts (1992) argue that a firm undergoing down-sizing is more prone to such problems than more stable organizations. One method for doing this is to draw names randomly once the size of the lay-off has been determined. Given such a mechanism, the firm must offer the bonus  $b$  to all workers (otherwise revealing who will be fired). Again the issue of the firm's ability to commit to the scheme  $b$  plays a role, although in this case its implications are less dramatic than in the ordered lay-off case.

In the third step, after productivity is realized, the firm determines the size of the lay-off ( $L^*(\alpha_t) - L^*(\alpha_{t+1}^*)$ ), and then, using some randomization device, lays-off each worker with probability  $p_L = (L^*(\alpha_t) - L^*(\alpha_{t+1}^*)) / L^*(\alpha_t)$ . If the scheme  $b$  is governed by a complete contract, the firm must set  $b$  such that the expected pay-off to adopting:

$$u(w + b) + (1 - r)(1 - p_S \cdot p_L(1, a_{-i})) \cdot V(1, a_{-i}) - c$$



is greater than that from not adopting:

$$u(w) + (1 - r)(1 - p_s \cdot p_L(0, a_{-i})) \cdot V(0, a_{-i}).$$

The scheme  $b(a_i)$  will produce a Nash equilibrium in which everybody adopts the innovation as long as  $b(0)=0$  and:

$$u(w + b) - u(w) \geq c + (1 - r)((1 - p_s p_L(0, a_{-i}))V(0, a_{-i}) - (1 - p_s p_L(1, a_{-i}))V(1, a_{-i})) \quad (4.4)$$

The right-hand side of the inequality can be re-written as:

$$c + (1 - r)((1 - p_s p_L(0, a_{-i}))(V(0, a_{-i}) - V(1, a_{-i})) + p_s \cdot V(1, a_{-i})(p_L(1, a_{-i}) - p_L(0, a_{-i})))$$

The firm must provide a bonus sufficient to compensate the worker for her impact on her own pay-offs and, as the above relation shows, that impact can be decomposed into two pieces. The first represents the change in continuation pay-offs resulting from a deviation; the second represents the change in the probability of receiving that continuation pay-off.

### *Incomplete Contracts*

If the scheme is only enforceable through reputation the result is similar. The firm must set  $b$  such that the expected pay-off from adopting is greater than that from not adopting:

$$(1 - p_s \cdot p_L(1, a_{-i}))(u(w + b) + (1 - r)V(1, a_{-i})) + p_s \cdot p_L(1, a_{-i})u(w) - c \geq (1 - p_s \cdot p_L(0, a_{-i}))(u(w) + (1 - r)V(0, a_{-i})) + p_s \cdot p_L(0, a_{-i})u(w)$$

There exists an equilibrium in which everyone adopts as long as  $b(1)$  satisfies:

$$u(w + b) - u(w) \geq \frac{c + (1 - r)((1 - p_s p_L(0, a_{-i}))(V(0, a_{-i}) - V(1, a_{-i})) + p_s \cdot V(1, a_{-i})(p_L(0, a_{-i}) - p_L(1, a_{-i})))}{1 - p_s \cdot p_L(1, 1)} \quad (4.5)$$

The firm must offer a larger bonus payment, but the payment only goes to those who adopt and are not fired. If workers were risk neutral, then the firm's total expenditure on implementing the innovation is identical under both schemes. With risk aversion, the firm's cost is higher under the implicit contract because workers bear the additional risk of investing  $c$  and not getting compensated.

### *Cost of Implementation under Random Lay-offs*

To approximate the firm's cost of implementation under random lay-offs, I again add the additional assumption that the innovation, if fully adopted, produces the same fractional reduction in the workforce each period. With this assumption, in a full adoption equilibrium the probability of being fired is constant each period, and hence the continuation pay-offs are unchanged by a deviation (the pay-off relevant state is the null state). In the current period, a unilateral deviation increases the *ex post* labor requirement by at most one job. The probability of lay-off in the current period changes from  $(L^*(a) - L^*(\alpha'))/L^*(\alpha)$  to  $(L^*(\alpha) - (L^*(\alpha') + 1))/L^*(\alpha)$  – it is decreased by at most  $1/L^*(a)$ . Thus the maximum payment above the cost of training required to induce each worker to adopt is  $(p_s/L^*(\alpha)) \cdot V(1,1)$ . The firm's total cost of implementing the innovation in excess of the training cost,  $c \cdot L^*$ , is then equal to the expected loss in utility of *one* worker who is fired –  $p_s \cdot V(1,1)$ .

As an example consider again the case with 100 players where each adoption saves 1/2 of a worker. A deviation means that only 99 adopt instead of 100. The *ex post* labor requirement is at most 51 workers changing the probability of lay-off this period from 1/2 to, at best, 49/100. Does this change the continuation pay-off of those that are not fired? No: if lay-offs are required in the following period the probability of lay-off is still 1/2. Thus, although the firm is required to compensate each worker for the probability that she might increase the possibility of her own firing, the compensation is relatively small since each worker takes no account of the costs her actions impose on others.

### **e. Summary**

While it is difficult to quantify the exact payment required to implement the innovation in equilibrium, the main message is clear: If the firm is able to implement equilibria in which workers do not engage in side contracts or other types of collusion, the firm can induce the workforce to adopt and only compensate them for a small fraction of their total loss in expected utility. This occurs because in all cases described, players do not account for the negative effect of their actions on the utility of others. In the ordered lay-off equilibrium those to be fired reduce the continuation pay-offs of those that are not fired. In the random lay-off equilibrium all players face the same decision problem, but still do not account for the negative effect of their actions on the utility of others. As a result, the firm has to compensate players for a very small portion of their loss in expected utility.

## 5. Collusion and Side Contracts

So far, the analysis suggests that productivity improving innovations can be adopted at a relatively low cost to the firm. This stands in contrast to much of the empirical evidence: many firms struggle to implement such programs, and, as mentioned previously, many programs fail due to lack of commitment from the workforce. In this section I analyze the effect that collusion and side contracting have on the firm's cost of achieving full adoption. To simplify the analysis and economize on notation, I assume for the remainder of the paper that workers are risk neutral –  $U(w,b,c)$  now equals  $w+b-c$ .

### **a. Side Contracts Under Pure Strategies**

The possibility for collusion and/or side contracting between members of the workforce exists because members of  $\Omega(\alpha^*)$  forego  $(1-r) \cdot V_i$  in the collusion-free equilibrium and might be willing to pay to prevent the innovation from being adopted. Further, the members of  $\Gamma(\alpha^*)$  suffer a loss in utility due to changes in their continuation pay-offs.

There are two possible agreements in the ordered lay-off case. First, players could simply use trigger strategies that dictate cooperation (not adopting) as long as cooperation has been sustained in previous periods. However, since this equilibrium is similar to that discussed below with random lay-offs, a second arrangement, one in which members of  $\Omega(\alpha^*)$  actually pay members of  $\Gamma(\alpha^*)$  to not adopt, is discussed here.

Construction of an equilibrium with side contracts is complicated by two issues: (1) the relative bargaining strengths of the players involved and (2) the process through which members arrive at such side agreements. To deal with the first I follow a methodology similar to that outlined by Tirole (1986): the issue of bargaining strength is ignored and the focus of the analysis is on the contract the firm has to offer in order to ensure that side agreements are not profitable. Dealing with the second issue is more complicated because, in contrast to Tirole's model, my model has many players and a large number of potential side agreements.

Consider again the situation of the last person to be laid off, player  $j$ . To engage in a profitable side agreement  $j$  needs to induce some set of other players not to adopt. Since her job is the first to be saved, she can contract with any set of adopters as long as she

prevents enough adoption to increase the *ex post* labor requirement. The situation is more complicated for player  $j+1$ . Her efforts to side contract depend on what player  $j$  does. If  $j$  has already entered into an agreement with some players and  $j+1$  can observe this, then she needs to find another group to save her own job. If  $j+1$  cannot observe  $j$ 's agreement, then she must induce a larger set of people not to adopt since she must save  $j$ 's job before her own.

To simplify the analysis, construct the new games as follows: The firm moves first and offers  $b(a)$ . The players then move sequentially with the first mover being the highest ranking player in  $\Omega(\alpha^*)$  – the last person to be laid off. The first player makes a take or leave it offer to a group of adopters sufficiently large to save his job. If he reaches any agreement, it is observable to all other workers. Then the next highest ranking player moves and so on until all have moved. This set-up is essentially a two player game between the firm and worker  $j$  because, if the firm can prevent the first worker from reaching a profitable side agreement, then it has effectively prevented all others from doing the same since each subsequent player has to convince more workers not to adopt. I also assume that both sides can only commit to an agreement that is one period in length, since members of  $\Gamma(\alpha^*)$  will wish to re-negotiate a long term contract each and every period. Thus if player  $j$  negotiates a side agreement in period  $t$ , she only prevents adoption in that period.

Given a potential one period side contract, members of  $\Omega(\alpha^*)$  will be willing to pay up to  $p_s \cdot (1-r) \cdot w$  (the expected net benefit to postponing the lay-off one period) to a group of adopters if it prevents them from adopting the innovation, and prevents the lay-off. In general, the size of the payment would be a function of the relative bargaining power of the two groups. However, this set-up has a special feature: the preferences of the firm are exactly opposite those of members of  $\Omega(\alpha^*)$  – the firm wishes to see the innovation adopted 100%, those facing a lay-off wish to see no adoption whatsoever. To ensure that side contracting does not prevent adoption, the firm must set  $b(1)$  sufficiently high that each worker prefers adopting to the best available side contract. Consider again player  $j$ .  $j$  must convince some group of  $m$  players not to adopt the innovation. Player  $j$  must offer at least  $s \geq b-c$  to each member of the group. Player  $j$  can offer up to  $s \leq p_s \cdot (1-r) \cdot w$  to prevent adoption, so each of the  $m$  players can earn up to  $s/m$  in such an agreement. If the firm wishes to implement an equilibrium in which no player reaches a side agreement it must set

$b \geq c + p_s \cdot (1-r) \cdot w/m$  where  $m$  is the number of participants in the best possible side agreement – the fewest number of adopters that could save one job. If  $m$  is constant for all side agreements then the firm's cost of implementing the innovation rises by  $p_s \cdot (1-r) \cdot w \cdot (L(\alpha_t) - L(\alpha_{t+1}^*))$  each period – the expected utility lost by all those who face lay-off.

To continue the earlier example with 100 players,  $p_s=1$ , and each adoption saving  $1/2$  a worker, in equilibrium an adopter could earn at most  $1/2 \cdot (1-r) \cdot w$  from a side contract. To prevent side contracting the firm must set  $b \geq c + 1/2 \cdot (1-r) \cdot w$ . Thus the total cost to the firm increases by  $50 \cdot (1-r) \cdot w$  over the side contract-free equilibrium. In addition there is redistribution of the surplus. Those in  $\Gamma(\alpha^*)$  are able to earn more because they are indifferent between adopting and not adopting – they will never be laid off – and as a result are able to sell their adoption decision to whomever will pay the most.

If the scheme  $b$  is not fully enforceable the situation does not change much. Again the highest ranking player in  $\Omega(\alpha^*)$  moves first and the firm must offer  $b$  sufficiently high to prevent any members of  $\Gamma(\alpha^*)$  from negotiating with him. The firm must set  $b \geq c + (p_s/m) \cdot (1-r) \cdot w$  where  $m$  is the fewest number of players that can save one job. The number of players needed to save one job,  $m$ , may be smaller in this case since in equilibrium few players adopt and the marginal improvement is decreasing in the number of adopters. The main insight, however is the same: the firm's cost increases when side contracts are allowed.

### **b. Random Lay-offs**

If the firm resorts to random lay-offs the situation is more straightforward. Under a random lay-off scheme the joint welfare of the workforce is clearly higher if it colludes to prevent adoption. Since the relationship is a repeated one collusion can be sustained. The set-up is similar to the infinitely repeated prisoner's dilemma. Assume again that players use simple trigger strategies that dictate cooperation if the cooperative outcome occurred last period and otherwise play the non-cooperative equilibrium described earlier. Clearly in a one shot version of this game each player would defect from the cooperative equilibrium as long as  $b > b^*$ . However, in the repeated game environment, cooperation can be sustained as long as the discounted benefit of cooperating exceeds the pay-off that one can achieve

from defecting. Sustained cooperation implies that the probability of lay-off is zero in every period and that the total pay-off is equal to  $w/r$ .

If a player deviates, she earns  $w+b-c$  in the period of the deviation and then the continuation pay-off to the non-cooperative equilibrium discussed earlier,  $V(1,1)$ . Cooperation can occur in equilibrium as long as:

$$w + b - c + (1 - r) \cdot (1 - p_s p_L(1,1)) \cdot V(1,1) \leq \frac{w}{r}$$

To prevent collusion and achieve full adoption the firm must set  $b$  such that:

$$b \geq c + (1 - r) \left( \frac{w}{r} - (1 - p_s p_L(1,1)) \cdot V \right) \quad (5.1)$$

To approximate the difference between the two pay-offs assume again that in the non-cooperative equilibrium the probability of lay-off is constant in each period in which they are possible. The continuation pay-off is then:<sup>6</sup>

$$V(1,1) = w + b - c + (1 - r)(1 - p_d p_L) V(1,1)$$

Which implies that:

$$V(1,1) = (w + b - c) / (1 - (1 - r)(1 - p_d p_L))$$

Substituting into (5.1) and some manipulation yields:

$$b \geq c + \left( \frac{1 - r}{r} \right) \left( \frac{((1 - r)p_d + r \cdot p_s) p_L}{1 - (1 - r)p_L(p_s - p_d)} \right)$$

Adding the additional assumption that  $p_s = p_d$ , implying that the potential for lay-offs exists every period, reduces the expression to:

$$b \geq c + (1 - r) \left( \frac{w \cdot p_s \cdot p_L}{r} \right)$$

Thus the firm's total payment is approximated by all the expected utility lost from switching from the cooperative to the non-cooperative equilibrium.

<sup>6</sup>. This implicitly assumes that the innovation produces the same proportional reduction in the workforce every period.

### **c. Summary**

The results in both cases change significantly when the possibility of side contracting or collusion is allowed. The main insight is the same in both cases. With collusion or side contracts the firm must compensate its workforce for a much larger portion of their loss in expected utility (all of it if the marginal benefit of adopting is constant). This occurs because, under these regimes, workers now account for the impact of their decisions on the joint welfare of the workforce. Institutions like unions or worker federations that facilitate side contracting or collusion would, thus, substantially increase the cost to the firm of implementing such participatory improvement techniques.

## **6. The Role of Fear**

With the addition of collusion, the model offers an explanation for the difficulty that firms like Analog Device experience in trying to implement programs like TQM. However, it does not offer a convincing explanation for the cases of AT&T and Harley-Davidson. Both firms had unionized workforces yet managed to implement successful TQM programs even though those programs included substantial lay-offs. In the analysis so far, the firm's survival has not been an issue. However, an important feature of both the AT&T and Harley cases was that the organization's survival was in jeopardy. A complete model of this phenomenon would include competitors and link the probability that a firm fails to its profitability, capital expenditure, etc. I take a more stylized approach here. I assume that although the firm cannot commit to lay-offs *ex ante*, it can commit to shutting down the firm if it does not achieve some level of adoption. This is justified based on the assumption that adoption correlates with profitability; if the firm cannot achieve a given level of adoption, it could liquidate the enterprise and invest the money elsewhere. Let the probability that the firm is shut down at the end of period  $t$  be represented by  $p_t$  and assume that  $p_t$  is function of the adoption vector  $\mathbf{a}$ . Thus  $p_t: A \rightarrow [0,1]$ . It is most useful to think of  $p_t$  as being determined by the accumulated number of adoptions, thus the probability of shut down equals  $p_t(\sum a_i)$ .

### **a. Ordered Lay-Off**

In the ordered lay-off scheme, without the possibility of side contracting, fear of shut down has a relatively small impact. It causes no change in the behavior of those in  $\Omega(\alpha^*)$ , they

lose their jobs anyway. Fear of shut down does, however, change the continuation pay-offs of those in the set  $\Gamma(\alpha^*)$ . The payment required to induce this group to adopt is now:

$$b \geq b^* + (1-r) \left( (1-p_t(0,1)) \cdot V(0,1) - (1-p_t(1,1)) \cdot V(1,1) \right)$$

Which can be re-written as:

$$b \geq b^* + (1-r) \left( (1-p_t(0,1)) \cdot (V(0,1) - V(1,1)) - (p_t(0,1) - p_t(1,1)) \cdot V(1,1) \right)$$

The change is modest however, because, again, players do not take into account the positive effect of their actions on the utility of others.

The impact of fear is larger when side contracting is allowed. The timing of the game is similar: the firm moves first and offers an incentive scheme  $\mathbf{b}(\mathbf{a})$ . The players then move sequentially, the highest ranking member of  $\Omega(\alpha^*)$  moving first. Each is again allowed to strike a side contract with any subset of adopters. Workers then adopt, any productivity improvements are realized, and the firm makes the next period's hiring/firing decisions or is shut down.

The player moving first can offer, in expectation, up to  $p_s \cdot (1-p_t(I-m)) \cdot (1-r) \cdot w$  to induce the required  $m$  players not to adopt. Players that enter such an agreement thus can earn up to  $p_s \cdot (1-p_t(I-m)) \cdot (1/m) \cdot (1-r) \cdot w$  in side payments. However, their decisions to not adopt reduce their expected regular incomes by at least  $(p_t(I) - p_t(I-m)) \cdot (1-r) \cdot V$  since they have increased the probability of shut down. To prevent side contracting the firm needs to set  $b$  such that:

$$w + b - c + (1-r) \left( 1 - p_t(I) \right) \cdot V \geq w + (1-r) \left( 1 - p_t(I-m) \right) \cdot (p_s \cdot w/m + V)$$

Which reduces to:

$$b \geq c + (1-r) \left( (1-p_t(I-m)) \cdot (p_s \cdot w/m) - V \cdot (p_t(I-m) - p_t(I)) \right) \quad (6.1)$$

The cost to the firm falls for two reasons. First, those to be fired have less to offer since there is some probability that the firm will be shut down. Second, participating in a side agreement reduces the continuation pay-offs of those that would not be fired otherwise since it raises the probability of shutdown.

## **b. Random Lay-offs**

Under the random lay-off scheme, without the possibility of collusion, workers now compare the pay-off to adopting:



$$(w + b) + (1 - p_s \cdot p_L(1, a_{-i}))(1 - p_t(1, a_{-i})) \cdot V(1, a_{-i}) - c$$

with the pay-off from not adopting:

$$w + (1 - p_s \cdot p_L(0, a_{-i}))(1 - p_t(0, a_{-i})) \cdot V(0, a_{-i}).$$

Thus, to maintain full adoption in equilibrium, the firm must set  $b$  such that:

$$b \geq c + (1 - p_s p_L(0, 1))(1 - p_t(0, 1))(V(0, a_{-i}) - V(1, a_{-i})) \\ + p_s \cdot V(1, a_{-i})((1 - p_s p_L(0, 1))(1 - p_t(0, 1)) - (1 - p_s p_L(1, 1))(1 - p_t(1, 1)))$$

The payment required to sustain full adoption falls since not adopting increases the probability of shutdown. However, the incentive effects are modest as workers do not account for the effect of their actions on others.

With collusion, the effect is more dramatic. Previously, to prevent cooperation the firm chose  $b$  to satisfy:

$$b \geq c + (1 - r) \left( \frac{w}{r} - (1 - p_s p_L(1, 1)) \cdot V \right)$$

Now, assuming the shutdown strategy is applied at the end of every period,  $b(1)$  needs to satisfy:

$$w + b - c + (1 - r)(1 - p_t(1, 1))(1 - p_s p_L(1, 1)) \cdot V \geq \frac{w}{1 - (1 - r)(1 - p_t(0, 0))} \quad (6.2)$$

Using the approximation for  $V(1, 1)$  from the previous section and assuming that  $p_s = p_d$ :

$$V \approx \frac{w + b - c}{1 - (1 - r)(1 - p_s p_L(1, 1))(1 - p_t(1, 1))}$$

which yields:

$$\frac{w + b - c}{1 - (1 - r)(1 - p_s p_L(1, 1))(1 - p_t(1, 1))} \geq \frac{w}{1 - (1 - r)(1 - p_t(0, 0))}$$

Thus the firm must set  $b$  such that:

$$b \geq c + (1 - r) \cdot w \cdot \left( \frac{p_L(1 - p_t(1, 1)) - (p_t(0, 0) - p_t(1, 1))}{1 - (1 - r)(1 - p_t(0, 0))} \right)$$

If we let  $p_i(0,0)=1$  and  $p_i(1,1)=0$  the condition reduces to:

$$b \geq c - (1 - r) \cdot w \cdot \left( \frac{1 - p_L}{r} \right)$$

Here, clearly  $b$  is at most  $c$ , and, in fact, there exists the possibility that collusion will lower the firm's cost below the non-cooperative equilibrium since everybody is better off if the innovation is fully adopted.

### c. Summary

In all cases fear of shutdown reduces the firm's cost of implementing the innovation. The change is most dramatic when collusion or side contracting is allowed. In both cases the cost is reduced because the possibility of the firm's failure has increased the pay-off to adopting – the chance of the firm surviving is increased – and reduced the pay-off from participating in outside agreements. Thus *if* the firm can credibly communicate that its survival depends upon the innovation being adopted, it may be able to reduce its cost of implementation. An obvious information problem arises in this setting: managers have a strong incentive to overstate the company's probability of failure, since they are then able to implement innovations at a lower cost. Institutions that allow managers to credibly communicate the firm's health thus play an important role in implementing programs like TQM. In particular, if managers develop the reputation for manipulating internal reports workers are likely to discount this information and look to other sources.

## 7. Discussion

### *Drive Out Fear*

Many scholars, managers, and TQM proponents have argued that long term job security is the key to implementing successful employee based improvement initiatives (Deming 1984, Levine and Tyson 1990, Bluestone and Bluestone 1992). A key factor identified in the Serman *et al.* (1994) study of Analog Devices was that members of the labor force began to view TQM as contributing to lay-offs. Analog's then Vice-President for Quality stated, "People didn't want to improve so much that their jobs would be eliminated." The decline in commitment to TQM was sufficiently large that Analog went from being number two on a key customer's 'ten best supplier' list to being number one on that same customer's 'ten worst supplier' list in two years. The important role of job security in successful participatory improvement effort is further supported by an empirical study by Easton and Jarrell (1995) which finds no evidence to support the hypothesis that the gains attributed to

TQM derive from down-sizing. In fact, they find that "...there is evidence of a positive association between the implementation of TQM and firms that do not down-size." These findings are all consistent with the model developed here, once allowances for collusion are made, since I show that the firm's implementation cost is high if downsizing is required.

Although it can contribute to a successful improvement effort, committing to job security is not always a feasible strategy, or, in low growth environments, in the firm's best interest. Firms that are growing slowly may never need the excess capacity created by productivity improvements. Further, keeping excess labor that could be laid off is, at best, an uncertain investment predicated on there being substantial demand growth in the future. Even if the firm expects substantial growth it still may be difficult for managers to resist the temptation to improve short term profits through lay-offs. Publicly owned companies are frequently penalized in the form of low share prices for not taking advantage of potential cost reductions. Analog Devices was forced into its lay-off in large part because it needed to demonstrate to the external capital markets that it was 'serious' about cost control.

#### *Drive In Fear*

In cases in which job security is not a 'feasible' strategy, the results of the analysis are consistent with another theme in the organizational change literature: major crises are useful for precipitating change. For example, Kotter (1995) discusses the case of a CEO who, as part of a successful change effort, "...deliberately engineered the largest accounting loss in the company's history, creating huge pressure from Wall Street in the process." The analysis shows that the firm can implement change efforts more cheaply *if* the workforce believes the firm's survival is in jeopardy. However, the firm needs a credible way of communicating this information since it has a strong incentive to "drive in fear" and overstate the probability that it will fail. External sources of information play an important role. If the firm appears to be profitable and competitive, management may have a difficult time convincing its workforce that the initiative and the consequent downsizing are really needed.

Both AT&T and Harley-Davidson had credible signals of financial distress. AT&T was losing business to external competitors at such a high rate that it was experiencing excess capacity and lay-offs even without improvement efforts. At Harley-Davidson, as reported by Reid (1990), the union leadership required management to show them detailed financial statements concerning the company's profitability and cash flow before they would agree to the lay-off and accept TQM. Harley was also forced to publicly request, and subsequently

received, a bail-out from the Federal Government in the form of protection against Japanese motorcycle imports. Managers at Harley believe that they were able to achieve full TQM implementation even with a substantial workforce reduction because employees realized that the company was at significant risk of bankruptcy (Gelb 1995). One manager who shepherded the company through the turn around said, "...it was easy to get people to change then. Everybody knew if we didn't make changes we were going under."<sup>7</sup>

These two theories indicate that either job security or fear of shutdown is a necessary condition for success. They suggest that participatory improvement programs will be easier to implement when the firm is either, growing very quickly and can absorb the excess capacity generated by productivity improvements (and can, thus, commit to job security), or is doing very poorly and may be forced to shut down. However, such programs are difficult to implement, when the firm is profitable but growing slowly. Current prescriptions for change offer little help to these firms since they can neither commit to job security – they may never be able to use the excess labor – or convince the workforce that the risk of shut down is significant.

### *Planned Improvements*

For firms caught between these two extreme points, the study yields an unexpected insight not captured in current theories of improvement. It suggests that the timing and pace of an improvement effort are important. In the model, the firm's cost of implementing an innovation depends on the growth rate of demand. If growth is zero then, with collusion or side contracts, the cost can rise considerably, whereas if growth is high the firm does not incur these costs. It is easy to construct examples in which the firm – given the option – is better off postponing an innovation until growth is again high (let  $\delta \cong 1$  and assume that the lost utility of workers is greater than the incremental increase in per period profit). Thus, it may in fact be possible to improve productivity too quickly or start a program too soon. The firm may be able to minimize its long run cost by matching the rate of productivity improvement with the rates of natural attrition and demand growth. By doing this, the firm never raises the possibility of lay-offs and thus does not incur any of the extra costs discussed above. In effect, the firm credibly commits to job security by never giving itself the opportunity to lay people off. Further, if the firm can deploy some resources towards generating additional demand instead of extra productivity, it may be able to further mitigate many of these effects.

---

<sup>7</sup>. Tom Gelb personal interview 3 April 1995.

The strategy of balancing growth in demand and productivity is, however, rarely observed. Many firms undertake such programs precisely because they are experiencing low rates of growth in demand and are looking for new ways to improve profitability. Quality and productivity improvement programs are generally given less attention when demand is growing faster than the firm's ability to produce product. It appears that firms are most likely to undertake improvement efforts during periods when the external conditions – low growth and employee turnover rates – are the least favorable for a program's success. The feasibility and desirability of the balanced strategy requires more study.

## 8. Conclusion

In this paper I have developed a model to help explain the wide range of experience with participatory improvement efforts like TQM. I model such programs as an attempt by the firm to induce its workforce to reveal information that increases productivity and, as a result, may lead the firm to downsize. There are two main results of my analysis. First, the employees' ability to collude or participate in binding side agreements is a critical determinant of the firm's cost of implementation. Second, the program's perceived impact on the firm's survival strongly influences the firm's cost and the ability of employees to profitably collude. These results allow me to explain the different experiences of firms described in case studies and are consistent with the available large sample empirical study.

For managers trying to implement such programs, the analysis gives some insight into the decision process used by participants in such programs. It matters critically whose job is lost when the innovation is adopted and whether adopters and those getting fired have any opportunity for collusion. The framework presented also partially confirms the conventional wisdom concerning the implementation of improvement efforts: job security can reduce the cost of implementing a program, and, if job security is not feasible, fear of failure can contribute to successful implementation. The analysis also suggests a new strategy for those firms that cannot commit to job security and are not in danger of failure: balance the rate of productivity improvement with the natural rates of demand growth and attrition.

For scholars, the analysis suggests that future empirical studies need to consider the context in which improvement programs are implemented. The growth rate of demand, macroeconomic conditions, employee turnover, and the firm's financial health may all

contribute to the firm's ability to implement and reap the full benefit of a program like TQM. More data are needed on both the case specific and aggregate levels. There are also a number of directions in which the model might profitably be extended. For example, the analysis ignores competitive dynamics. Many firms adopt such techniques because competitors use them. Firms that cut costs through productivity improvement and lay-offs may force their competitors to do the same or risk ceding valuable market share. The development of industry level models of the adoption of new improvement techniques appears to be an important area for future modeling efforts.

## References

- Bluestone B. and I. Bluestone (1992). 'Workers (and Managers) of the World Unite', *Technology Review*, November/December.
- Deming, W. E. (1986). *Out of the Crisis*, Cambridge: MIT Center for Advanced Engineering Study, Cambridge, MA.
- Easton, G. and S. Jarrell (1994). 'The Effects of Total Quality Management on Corporate Performance: An Empirical Investigation'. Working Paper, University of Chicago, Chicago, Illinois, 60637.
- Ernst and Young (1991). 'International Quality Study – Top Line Findings' and 'International Quality Study – Best Practices Report', Ernst and Young/American Quality Foundation, Milwaukee, WI.
- Fudenberg, D. and J. Tirole (1992). *Game Theory*, Cambridge MA., MIT Press.
- General Accounting Office (1991). 'US companies improve performance through quality efforts'. GAO/NSIAD-9-190 (2 May).
- Gelb, T. (1995). Personal Interview.
- Grossman, G.M. (1983). 'Union Wages, Temporary Lay-offs, and Seniority', *American Economic Review*, June, pp:276-290.
- Hart, O. and B. Holmstrom (1986). 'The Theory of Contracts', *Advances in Economic Theory*, Cambridge University Press.
- Holmstrom, B. and J. Tirole (1989). 'The Theory of the Firm', *Handbook of Industrial Organization*, Volume I, R Schmalensee and R.D. Willing Eds. Elsevier.
- Kaplan, R. (1990a). Analog Devices: The Half-Life System, Case 9-191-061, Harvard Business School.
- Kaplan, R. (1990b). Analog Devices: The Half-Life System, Teaching Note 5-191-103, Harvard Business School.
- Kotter, J.P. (1995). 'Leading Change: Why Transformation Efforts Fail', *Harvard Business Review*, March-April.
- McPherson, A. (1995). Total Quality Management at AT&T. Unpublished MS thesis, MIT Sloan School of Management.
- Meyer, M., P. Milgrom, and J. Roberts (1992). 'Organizational Prospects, Influence Costs, and Ownership Changes', *Journal of Economics and Management Strategy*. 1(1), Spring.
- Milgrom, P. (1988). 'Employment Contracts, Influence Activities, and Efficient Organization Design', *Journal of Political Economy*, 96(11).

- Levine, D. and L.D. Tyson (1990). Participation, Productivity, and the Firm's Environment, in A. Blinder ed., *Paying for Productivity*, Brookings Institution, Washington, D.C., 183-244.
- Reid, P. (1990). *Well Made in America: Lessons from Harley-Davidson on Being the Best*, New York, McGraw Hill.
- Shapiro, C., and J. Stiglitz (1984). 'Equilibrium Unemployment as a Discipline Device', *The American Economic Review*, June: 433-444.
- Shiba, S, D. Walden, A. Graham (1993). *A New American TQM. Four Practical Revolutions in Management*. Portland, OR: Productivity Press.
- Sterman, J., N. Repenning, and F. Kofman (1994). Unanticipated Side Effects of Successful Quality Programs: Exploring a Paradox of Organizational Improvement. Working Paper #3667-94-MSA, Sloan School of Management, Cambridge, MA.
- Tirole, J. (1986). 'Hierarchies and Bureaucracies: On the Role of Collusion in Organizations', *Journal of Law Economics and Organization*., 2(2), Fall.
- Varian, H. (1992). *Microeconomic Analysis*., New York, N.Y.: W.W. Norton & Co.
- Weiss, Y. (1985). 'The Effect of Labor Unions on Investment in Training: A Dynamic Model', *Journal of Political Economy*. October, pp:994-1007.
- Wruck, K.H., and M.C. Jensen (1994). 'Science, Specific Knowledge, and Total Quality Management', *Journal of Accounting and Economics*, 18, 247-287.



## Appendix

**Proposition I:** If contracts are complete and of the type discussed above, then there exists a sub-game perfect Nash equilibrium in which the set of non-adopters has at most  $(1/(1-\beta))$  members (only one member for all  $\beta \leq 1/2$ ) where  $\beta$  is the reduction in labor requirement caused by the adoption of the lowest ranking player in  $\Gamma(\alpha')^c$  (the last person to be laid off).

**Proof:** Clearly, if the firm could commit to lay-offs *ex ante*, it could achieve complete adoption in equilibrium since it would commit to firing all those in excess of  $L^*(\alpha')$  and thus the union of  $\Gamma(\alpha^*)$  and  $\Omega(\alpha^*)$  would be exactly equal to  $L^*(\alpha)$ .

However, without such a commitment device this equilibrium may not be sub-game perfect; if some player deviated and no longer adopted, it may not be in the firm's *ex post* interest to fire them.

To establish the claim, first check if any members  $\Omega(\alpha^*)$  have a profitable deviation. Since sub-game perfection is required, the firm's firing strategy is contingent on the realized productivity  $\alpha^*$ . Consider a candidate equilibrium in which everyone adopts. There are two conditions that must be satisfied for a member of  $\Omega(\alpha^*)$  to have a profitable deviation. First, the player's adoption decision must change the firm's *ex post* firing decision, thus the deviation must satisfy:

$$L^*(\alpha(1,1)) < L^*(\alpha(0,1)) \text{ and } L^* \in \text{integers} \quad (*)$$

The difference in productivity caused by the player's adoption choice must be sufficiently large that the firm's optimal labor demand changes by enough that it prefers one additional worker. Second, contingent on the first condition being satisfied, the deviator must also be in the correct position in the order to take advantage of her deviation. So, for a deviation to be profitable for player  $j$ ,  $j$ 's adoption decision must satisfy (\*) and  $j$  must also be sufficiently high in the ordering that if she changes the labor requirement it will be her job that is in fact saved.

I now characterize the set,  $\Theta$ , of those who do not adopt in this equilibrium . Index the person with the lowest rank who is a member of the set  $\Gamma(\alpha^*)^c$  as  $j$ .  $j$  is the last person to be laid off if adoption is 100%. If  $j$  does adopt, can any other player  $k > j$  profitably deviate? No, since by assumption  $k$ 's impact on productivity is at most sufficient to save one job, which in this case would be  $j$ 's. Now, assume that  $j$ 's impact on productivity is sufficiently large that if she does not adopt, she does not get fired – adoption is a dominated strategy – thus  $j$  is an element of  $\Theta$ . If  $j$  does not adopt then player  $j+1$  may also be better off not adopting if  $j$  does not adopt and so on. Thus the question is, can this process continue through all the workers? No, for some integer  $k$ , adopting is a dominant strategy for the  $j+k$ th player regardless of the actions of  $j$  through  $j+k-1$ . The reasoning is as follows: By assumption the marginal improvement in productivity is decreasing in the number of adopters, and each adoption saves less than one job. Assume player  $j$  by not adopting increases *ex post* labor requirements by some amount  $\beta < 1$ , and make the *a fortiori* assumption that if  $L^*(\alpha)$  is not an integer the firm always chooses  $\lfloor L^*(\alpha) \rfloor$  as the next largest integer. Adopting becomes a dominant strategy for the  $j+k$ th player when  $k \cdot \beta \leq (k-1)$  – when the number of jobs saved is fewer than the number of deviators. Thus  $\Theta$  contains at most  $(1/(1-\beta))$  players and contains only one player for all  $\beta \leq 1/2$ . Thus in equilibrium the set of non-adopters  $\Theta$  has at most  $(1/(1-\beta))$  members.

Proposition II: If the contract is implicit and the firm offers  $\{(1, b^* + V(0,1) - V(1,1)), (0,0)\}$  then:

- i) there exists an equilibrium such that  $\Gamma(\alpha^*) \supset \Gamma(\alpha')$  and  $\Omega(\alpha') \supset \Omega(\alpha^*)$
- ii) those that are not fired adopt and those that are fired do not
- iii) if the marginal benefit of an additional cost is constant, then the number of adopters is equal to largest integer less than  $I/(1+\beta)$  where  $\beta$  is the reduction in labor requirements that results from an additional adoption.

Proof: Part (ii) of the claim is trivial. In equilibrium those that are not fired are indifferent between adopting and not and hence accept the contract. Those that are fired do not earn a bonus and adopting costs them  $-c$ , hence they are better off not adopting.

To establish the i): Full adoption is not an equilibrium since some who adopt will be fired and thus lose  $c$ . Conversely, zero adoption cannot be an equilibrium since some are guaranteed not to be fired hence they will adopt. Thus if an equilibrium exists it must be at some intermediate point. Thus I need to show that there exists a partition of  $I$  into two disjoint sets  $\Omega$  and  $\Gamma$  such that members of  $\Omega$  do not adopt, members of  $\Gamma$  do adopt, and nobody can do better by deviating. Assume such a partition does not exist. This implies that for every partition at least one player  $j$  can deviate and do better. It is not possible that members of  $\Omega$  and  $\Gamma$  could both have profitable deviations simultaneously since the strict ordering assumption would be violated. Thus there must exist two partitions of  $\Omega$  and  $\Gamma$  that are identical except for the residence of player  $j$  such that  $j$  has a profitable deviation in both candidate equilibria. This is a contradiction since it implies that  $j$ 's decision to adopt leads to his being fired in one equilibrium but not in the other even though they are identical in all respects.

For part (iii), the number of adopters,  $y$ , is less than or equal to the number of people not fired in equilibrium, thus  $y \leq I - y\beta$  which implies that  $y \leq I/(1+\beta)$ .

Figure 1

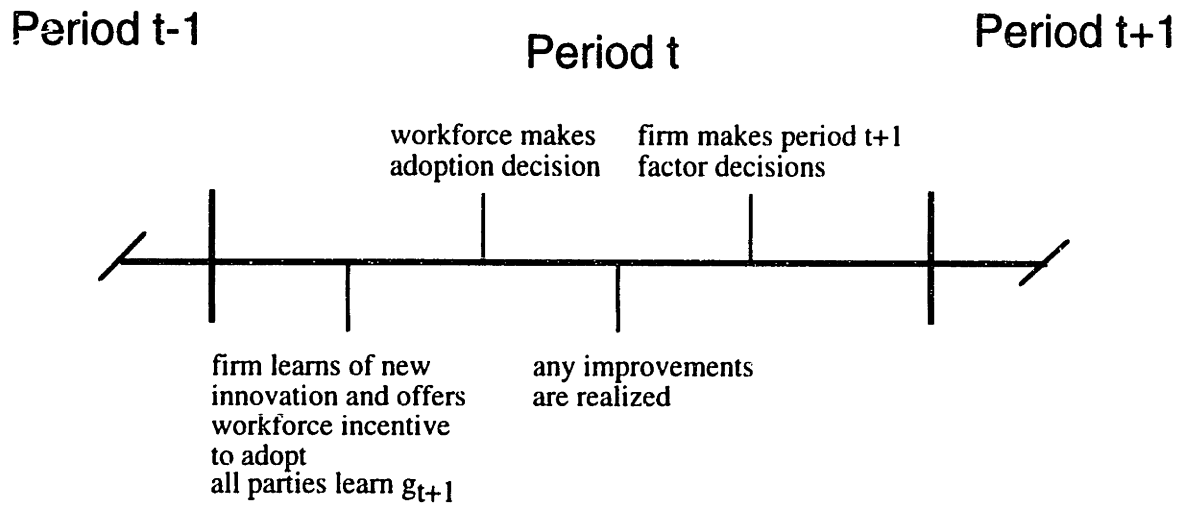


Figure 2

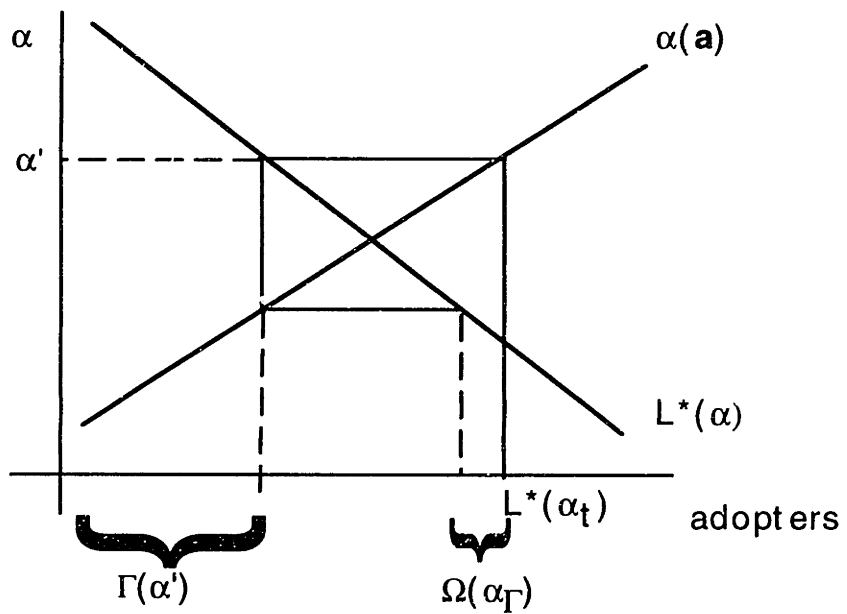
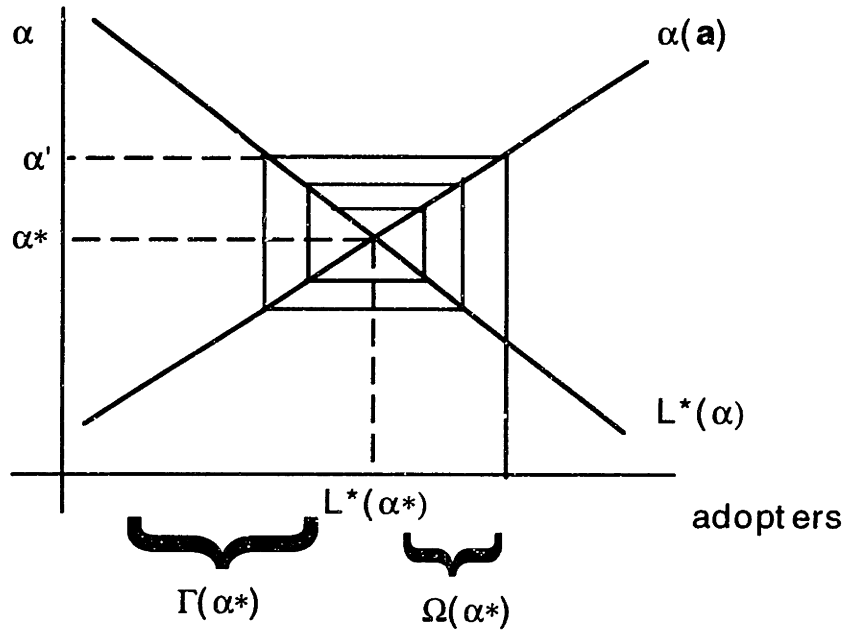


Figure 3



## Essay #3

### **A Tale of Two Improvement Efforts: Towards a Theory of Process Improvement and Redesign**

*“The prototypic question in organizations becomes: given the blue print, what recipe will produce it?”*

- Karl Weick, 1979

#### 1. Introduction

##### **a. Motivation: The Paradox of Process Improvement**

Managers, consultants, and scholars have increasingly begun to recognize the value of considering an organization's activities in terms of processes rather than functions. The current popularity of the 'process approach' stems from its ability to drive improvement within organizations (Garvin 1995b). Starting with Total Quality Management (TQM) (Deming 1986), and continuing with business process re-engineering (BPR) (Hammer and Champy 1993), many recent trends in management focus on the process, rather than the function, as the critical unit of analysis for improvement. The popularity of these approaches is one testament to the benefit of the process view; another is the data. There are numerous examples of firms that have made significant improvements in quality and productivity using TQM and related techniques. Easton and Jarrell (1995) find that firms who make a long term commitment to TQM outperform their competitors in both profitability and stock returns. There has yet to be a large sample study concerning re-engineering; however, there are reported examples of it producing substantial improvement (Hammer and Champy 1993).

Designing, executing and improving business processes is, however, not easy. Paradoxically, for every successful process improvement effort, there are many more that fail. Even though some firms make improvements using TQM, numerous studies indicate that most quality efforts fail to produce significant improvements in profitability and are eventually abandoned (Ernst and Young 1991, USGAO 1990). Although there have yet to be large sample studies, the results appear similar for re-engineering; even its proponents claim that the majority of re-engineering efforts fail to produce significant improvement (Hammer and Champy 1993). Scholars and managers alike have long realized the

difficulty of making fundamental changes to the technology, processes, and structures of organizations, and the process focus does not appear to mitigate those difficulties. While they can suggest new and valuable improvement opportunities, process-focused improvement techniques still fall prey to all the barriers that traditionally limit other organizational change and improvement efforts.

Unifying and synthesizing the literatures related to process improvement and redesign presents a challenge because, like processes themselves, studies on and theories about process improvement and redesign cross a number of distinct literatures. The physical design of manufacturing and service processes has traditionally been in the domain of industrial engineering, operations research, and operations management (Chase and Aquilano 1989). Tools for improving these processes have come from a wide range of areas. TQM has its origin in the field of statistics (Shewhart 1939, Deming 1986), while re-engineering has its roots in information technology and computer science (Hammer and Champy 1993). Theories from these areas generally focus on modifying the physical environment and give less attention to the concomitant and consequent organizational changes required to produce improved performance.

In parallel to the development of the physical and technological improvement tools, consultants and scholars have invested substantial effort in the study of improving and redesigning organizations (Huber and Glick 1993; Kanter, Jick and Stein 1992). However, the current theory on improving and redesigning organizational processes is not fully developed. Many academic studies focus on predicting when organizations are likely to undergo significant change and identifying what those changes may be (see Huber and Glick 1993 for examples). These theories do not, in general, offer a comprehensive framework that identifies factors separating successful change efforts from those that fail, and offer little insight to managers trying to make changes to the processes within their organizations. In addition, this literature largely ignores the physical structure of the environment in which the changes are taking place.

In contrast to the academic literature, authors writing for the practitioner audience focus almost exclusively on strategies for successful change (Kotter 1995; Kanter, Jick, and Stein 1992). However, many of these authors are promoting their own particular change methodology and tool set. Proponents of re-engineering, for example, offer a different change methodology than proponents of TQM (Deming 1986, Hammer and Champy 1993). The literature is virtually silent on the question of which techniques are more

appropriate in different situations. For example, should a manager choose the incremental approach favored by advocates of TQM, or the radical re-design approach favored by proponents of re-engineering? Current frameworks have little to offer in answering these questions.

The purpose of this paper is to articulate a theory of process improvement and redesign that accounts for both the physical *and* the behavioral components of the environment in which improvement is taking place. The aim is to develop a framework useful to both managers that engage in such activities and to scholars trying to explain the varied experiences of the many organizations that undertake process improvement efforts. The main tools for theory development are intensive case study research (Eisenhardt 1989), and the development of stock and flow feedback diagrams (Richardson 1992, Senge 1990, Mausch 1985, Weick 1979, Forrester 1961). The main thrust of the argument is that, contrary to the popular conception, TQM and re-engineering are complementary activities, and that a more general improvement and redesign methodology can be developed using precepts from each theory. TQM offers an organizational structure and decision making methodology. Re-engineering is re-conceptualized as a tool for challenging the dominant mental models that guide the organization.

The remainder of the paper is organized as follows: In the rest of this section I offer some basic definitions of processes and discuss the factors that make improvement and redesign efforts difficult. Section two presents related literature, and section three develops the building blocks of the basic theory. In section four a short summary of two case studies performed in a major U.S. automobile manufacturer is presented; the first field study deals with a very successful process improvement and redesign effort, the second with one that was less successful. In section five I map the results from the field studies into some of the existing frameworks on organizational change and transformation and analyze the case using the theory developed in section two. Section six presents concluding thoughts and implications for future research and practice.

### **b. Definitions**

While the term *process* is widely used, it is worth discussing its precise definition. Garvin (1995a) provides a comprehensive survey of its use in the study of organizations. He writes "...[the process focus] starts with a simple but powerful idea: organizations accomplish their work through linked chains of activities cutting across departments and



functional groups” (Garvin 1995a:3). Garvin identifies three broad categories: work processes, behavioral processes, and change processes. Work processes are “...a sequence of linked, interdependent activities, that taken together, transform inputs into outputs.” The process improvement activities considered here are those focused on improving work processes. Behavioral processes are “...underlying behavior patterns...so deeply embedded and recurrent that they are displayed by most members of the organization.” and include decision making, communication, and learning processes. They have “...no independent existence apart from the work processes in which they appear” but at the same time they “...profoundly affect the form, substance and character of activities by shaping...how they are carried out.”

### **c. From Blueprints to Recipes: The Challenge of Process Design**

Simon (1962) makes a distinction between the notions of blueprints and recipes. In the words of Weick (1979:46) “...[Simon] equates recipes with process descriptions and he contrasts these with state descriptions or blueprints....Recipes provide the means to generate structures that have the characteristics you want.” Thus the challenge of process redesign and improvement is twofold. First, one must develop the blue print of the desired process, including designing both the work process (the lay-out of machines on the factory floor) and the supporting behavioral processes (decision rules for production scheduling). Second, a recipe must be created that constructs the hypothetical process created in step one. A theory of process improvement and redesign must answer Weick’s question “given the blue print, what recipe will produce it?”

Three features of processes make the transition from blueprints to recipes difficult. First, processes are *technically* complex. Generating blueprints that lead to efficient, flexible work processes is a technical challenge. Management scientists and industrial engineers have invested substantial time and energy in developing better lay-out, scheduling and coordination algorithms to aid in the design and execution of manufacturing processes. Further, with these blueprints in hand, implementing such systems presents a further technical challenge. As the boundaries of a process are re-defined, new and different technologies become a part of that process.

Second, processes and process improvement efforts are *organizationally* complex. Glick and Huber (1993:5) define three elements of organizational complexity: numerosity, diversity, and interdependence. By definition, a process is more complex on each of these

dimensions than a given functional activity. Making a process rather than a function the subject of an improvement effort increases the number of people involved. The diversity of those people is increased since a process involves multiple functions. Those charged with improving a process must account for a higher degree of interdependence relative to those trying to improve a particular function or activity.

Third, process improvement efforts are *dynamically* complex. Processes take place over time. Those charged with managing and improving processes must deal with a dynamically complex environment characterized by a high degree of coupling and information that is imperfect and received only with a substantial time delay. There is a substantial body of research that shows learning in such complex environments is difficult (Sterman 1994). People often make decisions which yield highly undesirable behavior in these environments (Paich and Sterman 1994, Sterman 1989a, 1989b). The process view is a powerful perspective from which to improve the organization's effectiveness, but it does come with a cost. Those charged with managing and improving processes must deal with higher levels of technical, organizational and dynamic complexity, all of which may thwart process improvement and redesign activities.

## 2. Related Literature

### **a. Process Improvement and Redesign**

Process improvement and redesign have been addressed by a wide range of authors. Authors generally fall into one of two camps regarding the appropriate methods for improvement: incremental improvement and radical redesign (Garvin 1995b). These two groups are represented by proponents of TQM and BPR, respectively. Each of these methods suggests a tool set and modifications to the physical environment and also contains – at least implicitly – an organizational component.

#### *Total Quality Management*

Many of the ideas embodied in TQM and other continuous improvement methodologies were first articulated by W. Edwards Deming (Deming 1986). Subsequent authors have made important contributions to both the tools and philosophies underlying continuous improvement (Juran 1988; Feigenbaum 1983; Ishikawa 1985; Garvin 1988; Shiba, Walden and Graham 1993). At TQM's core is the idea that, rather than being trade-offs, quality and cost move together (Deming 1986:1). In his famous "fourteen points", Deming summarizes a recipe for improvement that contains three basic elements: 1) effort is shifted

from defect correction to defect prevention using statistical process control; 2) the goal of the operation is to "...improve constantly and forever the system of production and service," (Deming 1986:23), and 3) improvement is the job of everyone within the organization. Under such a system responsibility is dispersed, and those that work in the process are given the primary responsibility for quality improvements. The role of supervisors is to "...help people and machines and gadgets do a better job." (Deming 1986:23).

Although it has its roots in statistics, TQM has an organizational dimension. Wruck and Jensen (1994) argue that the implementation of TQM requires a fundamental change in the "...critical organizational rules of the game, namely the organization's systems for allocating decision rights, its performance measurement system, and its reward and punishment systems." (Wruck and Jensen 1994:249). Changes in the organizational structure are critical to reaping the benefits of the tools suggested by Deming and others. The location of decision rights is particularly important, as "The TQM process temporarily transfers decision rights from their location in the hierarchy to problem solving teams, and often permanently reassigns them based on the outcomes of the problem solving process" (Wruck and Jensen 1994:248).

### *Re-engineering*

Re-engineering offers an alternative view. Its proponents suggest that many work processes are not worth improving, but instead have grown so inefficient that they should be discarded and rebuilt. Re-engineering relies heavily on the concept of process as the driver of improvement (Hammer and Champy 1992, Garvin 1995a). Re-design activities are centered around creating business processes that are fast and efficient. Implementations tend to utilize new information technology to aid in the development. In contrast to TQM, re-engineering suggests more centralized control. The radical redesign of processes requires the commitment and effort of senior managers. Proponents spend less time discussing the role of those that will work in the process once it has been implemented.

### *Summary*

Simon's blueprint/recipe distinction provides an interesting perspective on the contrast between TQM and re-engineering. Advocates of TQM provide recipes for improvement: the PDCA cycle, seven step improvement methods, fish bone diagrams, etc. TQM has very little to say about the structure of manufacturing, service, or development processes.

Re-engineering, on the other hand, is focused almost solely on the creation of better blueprints. Its proponents offer very little on implementation, and nothing resembling a recipe. Clearly either situation can produce problems. Many TQM programs go off track as participants begin to apply the recipe to areas that are not in need of improvement. Florida Power and Light, winner of the prestigious Deming Prize given by the Japanese Union of Scientists and Engineers, “dismantled much of its quality program” soon after winning the award (Lee 1994). In another company, managers reported that people were using TQM methods in such decisions as where to move the water cooler (Krahmer and Oliva 1995). Conversely, the high failure rate of re-engineering efforts are testament to the difficulty of implementing blueprints without recipes.

### **b. Organizational Change and Design**

In their survey of the organizational change literature Van de Ven and Poole (1995) reviewed over 200,000 titles. I make no attempt to summarize this literature here. Instead, in this section I discuss two theories that relate to TQM/Re-engineering contrast discussed above: Weick’s theory of organizational design as improvisation (Weick 1993), and Orlikowski’s theory of technological structuration (Orlikowski 1993).

#### *Weick’s Theory of Organization Design as Improvisation*

Weick’s central contention is that the traditional model of designing and implementing an organizational structure – derived from an architectural metaphor– relies on the assumption “...of ideal conditions and focuses on structures rather than processes.” He suggests an alternative metaphor, theatrical improvisation, in which members of an organization respond to uncertainty in the environment by trying to agree to a set of shared meanings or interpretations that allows for coordinated action, but is sufficiently small that “...people retain the capability to make individual adjustment to local irregularities.” The act of designing is to “...notice sequences of action that are improvements, call attention to them, label them, repeat them, disseminate them, and legitimize them.”(Weick 1993:375). Weick’s theory provides a interesting lens through which to interpret the incremental and radical approaches to process improvement. Although they start from very different premises, many of his suggestions bear a striking resemblance to those of proponents of TQM .

Weick develops his metaphor by contrasting its assumptions with those standard in the study of organizational design. The conventional assumptions are taken from the precepts of the CODE study (Glick, Huber, Miller, Doty, and Sutcliffe 1990), and Weick supplies

the alternative. Three of these contrasts are particularly relevant to the improvement and redesign of processes

1.      **Conventional Assumption:**    Design creates planned change.  
          **Alternative Assumption:**      Design codifies unplanned change after the fact.

In contrast to the view of organizational redesign as planned change, Weick views the act of design as reinforcing changes that have already occurred. The design of a process emerges as the result of experience. The main mechanism posited is the response of members to uncertainties. Participants improvise as their daily experience presents them with unexpected and ambiguous events. The act of designing is to choose those improvisations that were perceived as desirable and repeat them. Weick's conception is consistent with TQM, the focus on continuous improvement, and Deming's charge to "...constantly improve the system of production." In contrast, re-engineering is more consistent with the conventional assumption of design creating planned change.

2.      **Conventional Assumption:**    Proper organizational design reduces current crises and inefficiencies.  
          **Alternative Assumption:**      Proper organizational design exploits crises and inefficiencies.

Weick's contention that organizational design should exploit mistakes and crises is similar to the TQM maxim that "a defect is a treasure". Continuous improvement relies on identifying undesirable outcomes and taking actions to correct them. Conversely, re-engineering focuses on eliminating past mistakes and takes little account of future imperfections.

3.      **Conventional Assumption:**    Activities are the object of control.  
          **Alternative Assumption:**      Ideas are the object of control.

If organizational design is a sequence of unplanned and unforeseeable changes, traditional methods of controls have little impact. Perrow (1986) distinguishes between three modes of control in organizations. First order control – orders, rules, and direct surveillance – second order control – bureaucracy, specialization standardization – and third order control – control by ideas. Weick suggests that control in an improvised design is third order. Weick posits these as justifications for actions or "...the socially acceptable reasons people

give themselves for doing something irrevocable.” Control is exerted primarily through the organization’s system of interpretation and legitimation.

### *Orlikowski’s Theory of Technological Structuration*

In her model of Technological Structuration, Orlikowski (1992) attempts to synthesize and reconcile two competing views of technological adaptation -- the technological imperative model and the strategic choice model. The technological imperative model posits technology as an exogenous feature of the environment and analyzes its effect on human behavior. The strategic choice model views human behavior as fixed and examines the evolution of technology through its interactions with humans. Orlikowski’s structuration model combines both views by recognizing the “dual nature of technology”. She states,

“...technology is the product of human actions, it is physically constructed by actors working in a social context, and it is socially constructed by actors through different meanings that are attached to it...”(Orlikowski 1992:406)

but, she continues,

“however,...once developed and deployed, technology tends to become reified and institutionalized, losing its connections to human agents that constructed it or gave it meaning, and it appears to be part of the objective structural properties of the organization.(Orlikowski 1992:406)

Three elements underlie this conception of technology: human agents, technology, and the institutional properties of the organization (Orlikowski 1992:409). Four relationships are posited: 1) technology is the product of human action, 2) technology facilitates and constrains human action, 3) institutional conditions influence how people interact with technology, and 4) institutional norms are influenced by people’s interactions with technology. Structuration is, then, a “...complex, recursive, process of mutual adaptation” between each of these factors.

Such a conception has important implications for the study and improvement of processes. Processes evolve and change over time, and changes to a process can have multiple and delayed impacts. Consider, for example, the consequence of changing the performance measurement scheme within a manufacturing facility. Such a change will affect the behavior of those that work within the process. Those that work within the process may then make physical changes to manufacturing equipment to enhance their performance on the new metrics. These changes then become institutionalized as new social norms and cause managers to make further adjustments to the measurement system. Structuration suggests that redesigning and improving a process is more difficult than simply developing

a better blueprint. Although managers may intend to implement the blueprint, the multiple adaptive processes involved in structuration may produce an entirely different outcome.

### 3. An Expanded Framework

Proponents of TQM, re-engineering, and other process improvement techniques offer methodologies that suggest changes to both the physical and the behavioral components of conversion processes. However, the theory embodied in such techniques is not explicitly stated. In contrast, organizational theories are rigorously stated but generally do not explicitly consider process improvement. The aim of this section is to develop a framework that explicitly links the physical and organizational dimensions of process improvement and redesign.

#### **a. The Physical Structure of Improvement**

##### *First and Second Order Improvement*

A fundamental contribution of TQM was to recognize the critical distinction between preventing defects and correcting them. Though it was originally applied to manufacturing, the idea can be broadened to include any work process. *Defects* will be used as a generic term for any undesirable outcome of a conversion process (Schneiderman 1988). *Process problems* are the features of the process, either physical or behavioral, that generate defects (the TQM literature has generally referred to these problems as *root causes* (Ishikawa 1985)).

*First order* improvements are actions targeted at identifying and correcting defects. These actions are obvious with relation to product quality: first order actions include quality assurance, inspection, and rework. First order actions are less obvious in other domains. For example, expediting in a manufacturing system can be considered a first order response to a manufacturing process that is slow and inflexible. In product development, defects take the form of longer-than-anticipated development times, products that don't meet customer requirements, or final products that are too costly relative to initial plans; first order improvements include overtime to get the project back on schedule, schedule slippage, rework, and changes in specifications .

The physical structure of first order improvement is shown in figure one. Defects are divided into two categories, known and unknown, and are represented as accumulations or

stocks. The stock of unknown defects is increased by defect introductions. The stock of known defects is increased by discovery, and depleted by corrections. The link between known and unknown defects is determined by both physical and behavioral processes. A defect is a matter of definition, and remains undiscovered if it has yet to be defined as such. Once defined, defects still may not be discovered immediately after introduction. For example, in manufacturing operations there are often delays between defective parts being produced and being identified by test equipment. In product development, defective designs may not be identified until assembled into a prototype. The level of known defects determines the through-put of the process, where through-put is defined as the number of items produced that are defect free. The level of unknown defects is *not* connected to process through-put because through-put represents an assessment by management, and, by definition, unknown defects cannot enter into that assessment. First order improvement is represented by the negative feedback loop, or goal seeking process, labeled B1. The current process through-put gets compared to the desired through-put and generates a gap between the desired and current state of the process. If the gap is positive, resources are allocated to reduce the stock of known defects. These efforts increase the outflow of defects, closing the gap between the existing stock and the desired level.

Second order improvement targets the causes of defects. In manufacturing second order actions include making adjustments to machinery to improve yield, shortening cycle times to increase flexibility, and changing the lay-out of the factory floor. Second order actions focus on prevention. The dynamic structure of second order improvement is similar to that of first order changes. Process problems – those features of the process that generate defects – are represented as accumulations and are either known or unknown. However, unlike defects, process problems are not directly observable. Instead, their existence is inferred via measurements of the process. The inference process is represented by the arrow between defect discovery and process problem identification. The link represents both a measurement and an assessment. Defects are defined and measured, then process problems are inferred.

Second order improvements are represented by the balancing loop B2. The existence of known defects increases the assessed process capability gap. As in first order improvement, the gap leads to an increase in the resources dedicated to eliminating process problems. Reducing the stock of process problems decreases the rate of defect introduction, and decreases the *inflow* to the stock of known defects. Two features of this loop deserve comment. First, second order improvement does nothing to reduce the



existing stock of defects – the defects that have already been created. Second, the B2 process often works more slowly than the B1 process. There are time delays in both correcting problems and in perceiving the benefits.

A key structural relationship between first and second order improvement is shown in figure three. The stock of available resources is explicitly represented and a negative link is added between resources to correct defects and resources for process improvement. The structure implies that any extra allocation of resources to correction comes at the expense of efforts aimed at prevention. The link represents the assumption that improvement resources are finite. The new link creates the positive loop, R1. The loop represents either a virtuous or vicious circle depending on the state of the system (Mausch 1985). If the organization is able to make improvements in process capability, then the defect introduction rate decreases, less effort is required for defect correction and even more effort can be dedicated to process improvement. Conversely, if the defect level increases, more time and resources are required for correction, reducing the resources that can be dedicated to improvement. Positive loops generally add instability to dynamic systems. In the case of improvement, the loop R1 gives the system a tendency to move to either extreme. Organizations able to make initial process improvements should experience a high degree of success. Organizations that are not able to make the initial improvements will have the opposite experience: more of their time will be dedicated to correction efforts and they will experience little improvement in process capability. The reinforcing nature of improvement helps explain one important feature of improvement efforts: they tend to be very successful or quickly fail with fewer cases in between (Ernst and Young 1992).

### **b. The Organizational Structure of Improvement**

The physical structure of improvement offers an explanation for why some improvement efforts succeed and others fail with very few falling in between. It does not, however, explain what factors contribute to successful improvement efforts. The high level of generality in organizational theories, such as those of Weick and Orlikowski, make them difficult to apply in specific situations. Knowing that processes are the product of a complex adaptive process is valuable in understanding why they are difficult to redesign and improve, but proves less useful in developing improvement and redesign strategies that overcome those difficulties.

### *Constructs*

Two specific types of human agents are considered: those who oversee the conversion process (managers) and those who work in the process on a day-to-day basis (workers). The distinction is important because typically only those who work in a process have firsthand experience with its inner workings. In contrast, managers experience the process largely through the performance measurement and evaluation system, and are subject to all its filters, distortions, and biases. Technology is construed as the process specified by managers. Technology includes the physical artifacts of the process, such as machines on the factory floor, and the rules explicitly defined by management for their use – standard operating procedures. Technology also includes the explicit process control structures and measurement and evaluation systems. Institutional properties are the behavioral processes that workers use in the process. In some cases, these may overlap with the process as specified – workers follow the process to the letter – and in other cases there may be a wide gap between the process as specified and the process as practiced. Structuration theory does not posit a driver for the adaptation process, and although agents, technology, and institutional properties are constantly changing, no catalyst for that change is identified. Here the perceived difference between the capability of the current process and that which is desired drives the structuration process.

### *Building Blocks*

In this section a number of behavioral processes are added to the physical structure discussed above. These processes are ideas presented in the literature mapped to the stock/flow feedback format. Each piece is a different building block that will be used subsequently to develop a theory of process improvement.

The first role of managers is to assess the process capability gap and to determine what constitutes a defect. The goal setting process is represented by the balancing loop B3. The current through-put is compared to the desired level to determine the capability gap. The capability gap leads to an assessment of the maximum process capability which in turn influences the desired level. The balancing loop represents an ‘eroding goals’ process in which the target adjusts towards the actual over time (Forrester 1968, Forrester 1969). The goal formation and adjustment process is consistent with the aspiration adjustment processes discussed by Cyert and March (1992) and the empirical study of Lant (1992). Loop B4 represents the adjustments made over time to the definition of a defect. As the capability gap rises, the definition of a defect becomes more strict – fewer outcomes are

considered defects – and the organization's tolerance for defects rises. B4 also represents an eroding goals structure. Sterman (1994: 299-302) discusses a number of examples in which such a reinterpretation process led to defects or inconsistencies being ignored. The two loops are consistent with the premises of structuration: managers constantly reinterpret their environment in ways that minimize the process capability gap. These processes can lead to diminished expectations and the acceptance of sub-standard performance. The only structure that prevents these loops from pulling the system towards equilibrium is the desired improvement rate, which represents a bias that keeps the capability gap from reaching zero (Cyert and March 1992).

To counteract the forces of diminished expectations, managers must engage in and support improvement efforts. An expansion of the balancing loop B2 is shown in figure five. Managers can take a number of actions to correct process problems. First, they can change the physical process directly, loop B2.1. To be successful, such changes must modify the physical process to correct the problem and make the appropriate changes in the supporting behavioral processes. Such a strategy is consistent with the philosophy of re-engineering. Alternatively, managers can encourage experimentation by those that work within the process. Experimentation requires a number of features. First training is required to make the experiments effective learning tools (represented by B2.4). Second, to participate in experiments and analyze the results, workers need release time from their normal responsibilities. Third, managers must accept the short term reductions in through-put caused by experiments. Each of these concepts is aggregated into the general construct, resources for improvement.

Besides supporting improvement, managers also exert control over the activities of the workforce. Control structures include restrictions on allowable activities (e.g. only so many coffee breaks) and pressure to reach through-put goals (see figure six). Pressures include penalties for low outputs and incentives for high outputs. Managers always have the option of strengthening these structures based on the assumption that there is some slack in the behavioral processes of the work force. Slack includes low effort, inattention to detail, or lack of discipline in following the specified process. The link between the capability gap and the belief in low efforts by the workforce represents an inference made by managers. Such an inference may be correct, workers may not be putting in a full effort. However, such an inference may also be incorrect. The mis-attribution of undesirable events to attitudes and dispositions rather than to systemic or environmental causes has been widely documented in psychology (Plous 1992). The problem has been

so persistent that it has been labeled the “Fundamental Attribution Error.” The inference loop B5 represents the attempt, by managers, to close the capability gap by increasing the control they exert over those that work within the process. Loop B6 represents the use of production pressure to close the capability gap.

Loops B1-B6 represent the behavioral processes of managers. The expectations, preferences and beliefs of workers also play an important role in process improvement. Figure seven shows another important construct, the commitment of the workforce to improvement. Commitment is determined by four variables. First, it is positively related to the rate of observed improvement. Workers need to see change in order to remain committed (Sterman *et al.* 1994, Schaffer and Thomson 1992). The link between improvement and commitment adds a second reinforcing loop, R2. Improvement leads to an increase in commitment and commitment leads to an increase in improvement. Second, support resources are important. Process improvement efforts require support in the form of training, release time from normal responsibilities, and a reduction in through-put objectives. Third, commitment requires a stock of known problems on which to work, creating the negative loop B7. The link between known problems and commitment is not discussed in the improvement literature, but emerged as an important dynamic in the field study to be discussed below. Finally, job security plays an important role (loop B8). Earlier studies indicate that commitment to improvement can be significantly reduced if workers believe that further improvement may lead to lay-offs (Sterman *et al.* 1994, Reppenning 1996c).

Besides participating in improvement efforts, workers are still responsible for the execution of their day-to-day responsibilities. These are left largely implicit in the model, except when they conflict with improvement objectives. Weick’s improvisational metaphor and structuration theory suggest that process participants are continually making changes to the process. Many of these changes deviate from the standard operating procedures and work rules set by managers. The organizational literature contains many examples in which process participants depart from the specified process; examples range from simple ‘work arounds’ on the manufacturing floor (Orlikowski and Tyre 1992) to changing the standards for O-ring tolerance on the space shuttle (Wynne 1988). One driver for these changes in the context of process improvement and redesign efforts is a conflict between the pressure to produce through-put and the capability of the process as specified.

If managers react to low performance by increasing production pressure and tightening the control over worker activity, such actions may be incompatible. There is no guarantee that stronger controls will help workers accomplish their objectives. If not, then workers are forced to make *ad hoc* changes to the process in order to reach their through-put objectives. Such modifications include changes to the process, either physical or behavioral, and manipulation of the performance measurement system. Adding these links to the diagram creates two more negative loops, B9 and B10, that act to close the process capability gap.

### **c. The Integrated Theory: Failure Modes**

The feedback and stock/flow structures discussed above provide the building blocks for a theory of process improvement and redesign. In this section the basic building blocks are used to identify some basic failure modes in improvement efforts. Much of the analysis centers on the identification of positive feedback loops that result from the combination of the processes discussed above. In a system composed solely of negative loops the outcome depends critically on the strengths or gains of the various loops. In contrast, positive loops tend to push the system towards extremes. By identifying positive loops, one hopes to gain sharper predictions of the system's behavior. The positive loops identified in this system are those that can lead to the failure of improvement and redesign efforts.

#### *Self Confirming Attribution Errors*

The first set of dynamics focuses on the results of managers incorrectly attributing low performance to the sub-standard efforts of the work force. Consider the situation shown in figure nine. Managers observe a gap between the desired and current through-put. The gap is attributed to low workforce effort and managers react by increasing the strength of controls they impose on the workforce. Stronger controls increase through-put forming the negative loop B5. However, increasing the level of control has an additional effect: it limits the ability of workers to experiment and learn. If the ability to experiment and learn is decreased, then the effectiveness of improvement efforts declines leading to a *higher* level of defects and a larger through-put gap. These links create the positive loop R3 shown by the bold arrows. The key feature of this loop is that managers' attribution of low effort, correct or incorrect, is a self-fulfilling: managers infer slack from low performance, they react by increasing controls, and increased controls limit learning and further reduces performance. Upon seeing the decline in performance, the initial attribution of slack is confirmed and managers conclude that they did not increase control enough. Thus, although the attribution may have been initially correct, as management purges slack from

the system by increasing control, they also limit learning, lower performance, and create the situation they are trying to correct. This dynamic has been observed in other studies, including maintenance in the chemical and nuclear power industries (Carroll *et al.* forthcoming).

A similar loop is created by considering loop B6 in the same context (see figure ten). Upon seeing a capability gap, managers increase the pressure to produce. Pressure leads to extra effort and higher through-put levels. However, the extra production effort comes at the expense of other tasks, in this case improvement efforts. The negative link between production effort and resources for process improvement creates the positive loop R4. Again, management's attribution of low effort is self confirming.

It is not hard to picture the type of environment produced by loops R3 and R4. Management continually increases both its control over the activities of the workforce and the pressure to produce. Workers feel increasingly under stress to produce while having less freedom to work within the process. These conflicts lead to a contentious and antagonistic relationship between managers and workers. Over time, managers come to view workers as avoiding work at all costs, while workers come to view managers as ruthless and willing to resort to any means necessary to increase productivity. In the meantime, performance is sub-standard, revenue and profit fall, and the size of the 'pie' the organization divides between the two groups becomes smaller, further increasing the adversarial relationship.

In both cases, managers may also try to promote prevention efforts through loop B2. However, these efforts will produce few results if they are accompanied by stronger controls and increased production pressure. There is a delay between successfully eliminating process problems and experiencing increased through-put. Managers that implement an improvement effort may underestimate this delay and increase production pressure before the improvements have been realized. In doing so, loops R3 and R4 may negate any early progress, pushing the organization back to 'business as usual.'

#### *The Ad Hoc Change Process Loop*

Another key failure mode in process improvement and redesign centers around the need for workers to make *ad hoc* changes to the process. The environment described above continues to evolve towards higher levels of production pressure and stronger controls over the activities of the work force. These controls include the specificity of standard operating

procedures and the required level of documentation for work activities. Loops B9 and B10 represent the need for workers to make *ad hoc* departures from the rules as rising production pressure and stricter controls begin to conflict. In figure eleven the dynamic is represented by a positive link between *ad hoc* changes and the introduction of new process problems. The link creates another positive feedback loop, R5. In reaction to the capability gap, managers increase the strength of controls and/or the pressure to produce. These changes increase the conflict between the objectives faced by the workforce. The strong control structure suggests one course of action. However, highly standardized processes are likely to be, in the short run, less efficient. Workers have a better chance of hitting their production targets by making *ad hoc* changes not specified by the control structure. These changes increase through-put and close the capability gap. The changes also introduce process problems, which eventually cause the capability gap to increase.

#### **d. Summary**

Structuration theory suggests that processes evolve over time as managers and participants reinterpret and reconstruct their environment. Weick suggests that flexible and efficient organizations emerge from allowing participants to improvise and experiment. However, these ideas are presented at a high, abstract level, and must be grounded in mid-range theory supported by empirical examples. Using available literature, these processes have been mapped into a stock/flow feedback structure. The analysis has identified two related dynamics that can limit the success of improvement and redesign efforts. If managers attribute low performance to workers and react by increasing through-put pressure and their control over the process, improvement will be difficult. These problems are exacerbated if workers are further forced to make *ad hoc* departures from the process in order to achieve their through-put goals. These dynamics have implications for improvement and redesign.

Designing a process off-line requires the designer to anticipate all the physical outcomes of the new process and any conflicts between the process and the performance evaluation system. Otherwise, workers will be forced to make *ad hoc* adjustments to the process that may produce undesirable behavior in the long run. Decision making research indicates that human ability to mentally simulate higher order dynamic systems is limited (Sterman 1994, Sterman 1989a, 1989b). The possibility of designing a process and supporting measurement system that could account for all the feedback relationships discussed above seems beyond the limits of human cognition. TQM offers a partial solution to this dilemma. By emphasizing experimentation and a scientific approach to decision making, TQM can improve the quality of the *ad hoc* changes made to the process. However, TQM does not

suggest which changes should be made, or provide any solution to the reinforcing behavior in loops R3-R5.

#### 4. Two Improvement Efforts

The field research was performed within one division of a major American automobile manufacturer. The division manufactures electronic components that are integrated into the vehicle at the company's main assembly facilities. The division is quite large with over two billion dollars in annual sales and has many major manufacturing facilities. Two process improvement initiatives were studied. The first was targeted at reducing the cycle time of the manufacturing process – the Manufacturing Cycle Time (MCT) initiative – and the second was designed to improve the efficiency, speed, and reliability of the product development process – the Product Development Process (PDP) initiative.

##### **a. Methodology**

The research was retrospective. Both initiatives were completed at the time the research was undertaken. While the company has undergone numerous change initiatives in the past fifteen years, the MCT and PDP initiatives were chosen for several important reasons. The MCT initiative was very successful. During the course of the effort, the division was able to reduce its average cycle time from more than 15 days to approximately one day. Further, the division's experience with MCT continues to influence how other improvement efforts are implemented and managed throughout the company. The PDP initiative was selected because it was heavily influenced by the success of MCT. In particular, the same senior executive launched both initiatives and viewed PDP as a logical extension of the success of MCT, and tried to use many of the same strategies to initiate and manage the PDP effort that had been so successful in the MCT initiative. The two initiatives represent a rare opportunity to 'control' for the effect of senior leadership.

The primary data collection method was semi-structured interviews. Over sixty interviews were conducted with participants in the two initiatives. All levels within the organization were represented, from the general manager of the division to development and operations engineers who do actual product engineering or run production lines. The researcher visited two different manufacturing facilities and the product development headquarters. Interviews lasted between 45 and 90 minutes and were all recorded. Each interview began with the subject describing his or her background with the organization and any relevant



previous experience. Participants were then asked to give a detailed historical description of their experience with the initiative. Once the description was completed, subjects were asked to assess the key successes and failures of the initiative and to give any personal hypotheses for their causes. Finally, subjects were asked to describe any lessons learned and to speculate on what they would do differently if they were to participate in a similar initiative in the future.

The interviews were supplemented with extensive review of available archival data. The researcher was given access to a wide range of promotional and training material associated with each initiative including pamphlets, newsletters, instructional books, and video and audio tapes. The historical performance data were also reviewed. In the case of the MCT effort, extensive data on actual cycle times, product quality, productivity and other operational variables were available. Less data was available for the PDP effort (the reason for this will be discussed in the analysis).

The data were summarized in the form of two detailed case studies (Repenning 1996a, 1996b). The case documents describe the history of the initiatives with emphasis on both the available quantitative and archival data and the recollections of participants. Both cases make significant use of quotations taken from the recorded interviews. The case documents were provided to participants for their feedback; participants were asked to review their quotations for accuracy but were not allowed to change the content. Participants were also asked to review the entire case for accuracy. The case documents are available from the author upon request.

The research was also supported and enhanced by a team of company people formed specifically for this study. Participants were drawn from multiple levels, and played a number of important roles in the study. First, they provided the researcher access to key players in each of the initiatives. Second, they provided valuable assistance in explaining and interpreting the organization's unique language. Finally, the team met with the researcher on a regular basis to review the case documents for factual content and completeness and to assess the relevancy of the theory being developed. While it is not possible to describe both cases in any detail, in what follows I try to highlight the main phases of each.

## **b. Manufacturing Cycle Time (MCT)**

### *State of the System Prior to the Initiative*

Prior to the MCT initiative in 1988, the division's manufacturing facilities were operated in a manner similar to that of other companies whose business requires substantial capital investment and labor expense. Line supervisors were charged with keeping each piece of equipment and each laborer fully utilized, and the division used a traditional performance measurement and evaluation system that emphasized direct labor performance (roughly defined as the number of units produced per person). The focus on utilization gave supervisors strong incentives to keep high levels of work-in-process inventory (WIP) to ensure that breakdowns and quality problems at upstream machines would not force downstream machines to be shut down. Over time the manufacturing system evolved to the point where a large portion of each plant's floor space was dedicated to holding WIP. An operations manager summarized the environment,

Before [MCT] if you were to walk out onto the floor and ask a supervisor how things were going, he would say "Great, all my machines are running" and you would see tons of WIP sitting around.

This mode of operation was problematic for a number of reasons. First, between sixty and eighty percent of the division's total costs derived from purchased components, so holding high levels of WIP was expensive. Second, high levels of WIP delayed quality feedback – a machine could produce a large batch of defective parts before the defect would be discovered by a downstream operation. Third, since the average cycle time was so long, it was very difficult for the manufacturing facilities to change the production schedule at short notice. Last minute changes were usually accommodated through expediting – *ad hoc* changes in the production schedule – which were very destabilizing to the production floor. Prior to the start of the MCT initiative, a number of attempts to improve the manufacturing process had been made, but none of them had led to reduction in the manufacturing cycle time or the level of WIP.

### *Launching the Initiative*

The beginning of the MCT initiative can be traced to the arrival of a new general manufacturing manager, JD. JD's previous employer – a major computer manufacturer – had managed its operations by minimizing inventory and cycle time, and he believed that the division's manufacturing operations could be improved if they were similarly focused. Although he had been hired into a senior position, JD believed he could not dictate that the division's operations focus on these measures. Instead, his first step was to analyze the

path of a typical product as it traveled through a division manufacturing facility. He recalled,

We analyzed [for a sample product] the time elapsed between when a part came in the back dock until the time it left the shop floor, and asked the questions “How long did it take?”, and “What was the value added?”. We found out [for this product] it took 18 days to make the product and we were adding value to the product 0.5% of the time. When I laid this out for everybody...they were astonished.

The simple presentation of the data played an important role in stimulating interest in the initiative. JD set only one official requirement for each plant: that they calculate and report manufacturing cycle time and value added time. Interestingly, he did not specify how these metrics should be calculated. Instead, he challenged the plants to develop their own definitions. One plant manager said, “JD didn’t give us a lot of the details...he wanted us to take a fresh look.”

JD then spent the vast majority of his time visiting the division’s manufacturing facilities. These visits were focused on providing concrete examples of how the notions of cycle time and value added percentage could lead to improvements in the manufacturing process. JD recalls his trips,

They [people in the plants] wanted to give me presentations in the conference room, and I would say “no, let’s go out to the floor”... I wanted to show them examples of what I was talking about. I might look at the shipping labels in the warehouse. If it were May, I would usually find parts that had been received the previous August, and I would ask, “if you aren’t using this stuff until May, why has it been sitting here since last August?”

JD also used these trips to augment the division’s traditional objective setting process. For example, at one facility, a tent had been constructed to hold the extra work-in-process inventory. Although originally planned as temporary, the tent had become a permanent part of the plant’s lay-out. JD challenged that plant to reduce inventory to a sufficiently low level that the tent would no longer be needed. As one participant said, “JD told us, ‘I’ll know your plant is running well when I come to visit and the tent is gone’”.

### *Early Measurements and Experiments*

JD’s charge to measure and reduce cycle time was received with differing levels of enthusiasm in the division. Many plants initially perceived the effort as just another fad being promoted by the corporate staff that would soon be replaced by a new ‘flavor of the month.’ One facility, alpha, took a different view. Alpha’s plant manager recalls,

The concept of cycle time and value added were not unheard of here....however, JD had a very different vision about how those measurements could be used to really drive the operation and improve productivity... we [the plant staff] sat down together and talked about this idea and decided it looked like it did have merit. We then, very quickly, put together some pilots to try out these concepts.

Alpha undertook an intense period of experimentation that lasted for approximately two years. Early efforts focused on developing a measurement system that could capture cycle time and value add percentage. Improvement began almost immediately. As the plant manager recalls,

...in the first year we started with simple counts at different times during the day, and we started to plot them and to try and understand what was happening. Very quickly our creative engineering personnel came up with clever ways to control the buffers that helped make big improvements.

In the first year, the plant was able to reduce cycle time by more than fifty percent.

In the second year of the effort Alpha still did not focus on creating a standard improvement methodology. Instead, participants were actively encouraged to experiment. The plant manager recalls, "...if somebody had a better idea about how to manage the buffer, they could try it." These experiments did come with a cost. Reducing WIP buffers and experimenting with new scheduling systems inevitably resulted in lines being shut down as they were starved for parts. Shutdowns reduced machine utilization as well as put the plant at increased risk of missing its production schedule. However, Alpha's plant manager was willing to accept these problems in the hope of making future improvement. He recalls,

...the best thing we did was that we didn't kill anybody when they shut down the line, and that happened a lot during this period of time as we experimented with new buffer management systems. We certainly shut it down more than we would have otherwise, but we were willing to do this in order to make more improvements.

The results of their efforts were significant: Alpha reduced its average cycle time from somewhere between 10 and 20 days to less than three days within the first two years.<sup>1</sup>

### *Manufacturing Cycle Efficiency*

The first step towards adding a formal improvement methodology for the division came in the middle of the second year when it created a four-person group to promote the initiative

---

<sup>1</sup> . An exact measurement of manufacturing cycle time prior to the initiative is difficult to obtain. Prior to the initiative cycle time had not been measured in a standard fashion. Further, as the measurement program was being developed many substantial improvements were being made simultaneously.

throughout all the plants. The group's first step was to institutionalize the method of value added analysis pioneered at alpha by having each plant calculate a metric called Manufacturing Cycle Efficiency (MCE). MCE was defined as the ratio of value add time (time in which function or feature was being added to the product) to total manufacturing cycle time. The results early were not encouraging, as another plant manager recalled, "...when we first started to calculate MCE, the numbers were so low [less than 1%] we really wondered how relevant they were."

The *process* of calculating the metric, however, proved valuable. A staff member recalled,

...you had to walk through the shop floor and ask the question, "Is this value added?" for every step in the process. By the time you were finished you had flow-charted the entire process and really highlighted all the value add stations....After calculating MCE, we really started to understand the process flow of our products. We knew where value was being added, and, more importantly, where value was not being added.

In the past, production steps set by engineering had been taken as given. After calculating the MCE measure plants began to question the length of time the products needed to spend in certain areas and why certain steps needed to be included in the production process at all. The division staff spent the next year helping the plants evaluate every step in the process based on its ability to add value. Many time-consuming operations in the process were re-evaluated: some were reduced; others eliminated. The group's leader recalls,

The process made us challenge specifications and engineering requirements that we had previously taken as given. Why, for example, did we need to protect a circuit board from the outside environment when it sits in the passenger compartment of the car? We finally decided after much thought and experimentation that we didn't, so we eliminated it [thus saving twelve hours].

Within a year, the MCE efforts helped cut the average cycle time for the division to less than five days (a better than fifty percent reduction).

### *Theory of Constraints*

Two years into the initiative, with the MCE effort well underway in most facilities, the corporate staff and others began to look for a new methodology to support further improvements. The group focused on shop floor management as the next opportunity for reducing MCT. The MCE effort had focused on eliminating non-value added operations and identifying unneeded buffer inventories. These improvements changed the structure of the manufacturing process. To make further reductions in cycle time, the team believed they needed a better way to *manage* the process, which presented two challenges. First,

the manufacturing processes were very complex and scheduling them was difficult. The division already employed a group of dedicated simulation specialists to develop scheduling and coordination strategies. Second, better management of the process required the participation of machine operators and material handlers. Thus, the problem was more than developing a better scheduling technique – itself no easy task – but also included training everybody within the manufacturing operation to use that technique.

A supervisor recalls,

...at the time people thought “this is important because it’s important to the general manufacturing manager” but they didn’t necessarily feel in their gut that it was important because they didn’t understand what was behind it...We needed more than just a definition of MCT or MCE, people needed a better understanding of how the shop floor really worked.

The corporate office started their search for a new shop floor management technique by interviewing consultants who offered methodology and training. They became interested in the offerings of the Goldratt Institute which taught the shop floor management philosophy Theory of Constraints (TOC) developed by its founder Eli Goldratt (Goldratt and Cox 1986). The attraction of the Goldratt group was twofold. First, they offered a scheduling and coordination methodology, but second, and more importantly, they offered a training program focused on developing intuition through hands on experience with a computer simulator. The director of the manufacturing simulation group saw it as an integrated way of teaching things that his group had learned in their computer studies of production lines:

I called it ‘Shop Floor Scheduling and Coordination Awareness 101’. If you wanted to concentrate in three days everything you would want to understand about the dynamics of the shop floor and how to keep the line running, this was it.

Within six months of the initial contact almost every manufacturing engineer and supervisor within the division had participated in a two day TOC class. In the following year, TOC training was given to almost every operator and material handler within the division. Participants viewed the extensive roll-out of the training program to all levels of the organization as one key to the program’s success.

The extensive roll-out of training made a big difference. Nobody in this plant could say that they didn’t have the opportunity to learn about TOC. I thought, at the time, it was a waste to train 1,500 hourly workers, but it really helped.

During this time management also worked to overcome the reliance on the traditional labor and overhead performance evaluation system. Many plant managers eliminated the labor and overhead standards and encouraged their staff to focus on inventory reduction.

Participants also cited the change in the evaluation system as a key contributor to success.

One operations engineer recalled,

They were willing to accept that we couldn't have our cake and eat it too. One day we were evaluated on labor efficiency and the next day we were evaluated on cycle time. It changed our entire focus.

### *Results*

Through the MCT effort, the division accomplished a remarkable transformation. Between 1988 and 1995 the average manufacturing cycle time was reduced from approximately fifteen days to less than one day, average inventory holdings were reduced by over fifty percent, the quality of finished products was improved, and sales revenue, profit, and cash flow all increased significantly. Further, the manufacturing process became less elaborate, more flexible, and more adaptable. For example, many facilities have discarded complex automated inventory storage and retrieval systems that are no longer needed in the low inventory environment. In addition, many facilities are able to change their production schedule on a daily basis, something that was impossible before the MCT effort. Finally, the reduction in WIP created enough extra floor space within the existing manufacturing facilities that two of the five planned new facilities were not needed.

### **c. Product Development Process (PDP)**

The second initiative, focused on improving the division's product development process, was initiated in large part due to the success of the MCT initiative. After three years as general manufacturing manager, JD was promoted to general manager of the division. He believed that a new, faster, more efficient product development process was important for the continued growth and success of the division, and many within the division shared this sentiment. An often-cited problem with the division's development process was the lack of standardization and discipline. A chief engineer describes the period preceding the initiative,

We went through a period where we had so little discipline that we really had the 'process du jour'. Get the job done and how you did it was up to you....It allowed many of the engineering activities to go off on their own and as long as they hit the key milestones, how they got there wasn't that important.

JD launched the initiative by forming a dedicated task force to design and implement a new development process. He describes his instructions to the group,

"I want a development process that is fast, that will give me a 50% increase in throughput in two years, and I want everyone to follow the same process."

### *Developing a New Process*

The team assembled to develop the new process included representatives from all the major stake-holders within the organization. The group focused on three main activities to design the new process. First, they hired an outside consultant to provide basic methodology and to provide an outside 'check' on the polices and processes the team would propose. Second, they benchmarked other companies, and, third, they spent time documenting the current process and determining how many of the problems that occurred repeatedly had come to be part of the process. A team member summarizes the process,

We spent a substantial amount of time looking at what other people did. How they structured their processes and the problems they had. We looked at...the current state of our process and tried to net out a process that had all the things we wanted and...allowed us to do things much more quickly.

### *The New Product Development Process*

The team consolidated learning from the benchmarking efforts, lessons from internal analysis, and the input of numerous people throughout the company into a new design process for the division. The process is quite detailed, and within each phase there are a large number of steps. However, three key elements distinguished this process from those the division had used in the past.

First, PDP was designed as a 'one pass process'. Historically, the division had created a larger number of physical prototypes in the course of a development project. Developing multiple prototypes was time consuming and expensive. To break this cycle, PDP required detailed documentation of the customer's requirements for the product before the design process was initiated. Historically, projects were initiated with ambiguous requirements which led to large amounts of re-work. When the requirements were established, engineers were then supposed to do the vast majority of the design work using computer engineering and design tools rather than developing physical prototypes. The group believed that a substantial improvement in efficiency could be made if a "one pass process" could be implemented.

A second goal of the PDP process was to increase discipline. The development process was divided into six major phases, and at the end of each phase the development team was required to undergo a 'phase exit quality review' before they could proceed to the next step. These reviews were conducted by senior managers and required the development teams to assemble detailed documentation on the state of the project. In between those reviews, the



PDP process relied on each project being run using standard project management techniques such as developing work plans, creating Gantt charts, and using project management software. By using project management tools, engineers would be more efficient and better able to meet critical milestones in the development of a given product.

A third goal of PDP was to propagate learning through the use of the 'bookshelf.' The bookshelf was an engineering library of technologies, modules, and subsystems: Every time a new technology was used it was the user's responsibility to 'bookshelf' that technology by fully documenting its uses, capabilities, and limitations and placing it in the library. Historically the division did not share technological learning well and substantial effort was duplicated in learning about new technologies. Because of this, the bookshelf was seen as critical to improving the efficiency of the development process.

The final component of the process was a set of metrics and performance objectives. Developing metrics to measure the performance of the product development process turned out to be very difficult. A special committee was formed to undertake this task. Its chair recalls

We [the committee] developed an entire set of metrics for PDP....what we came up with had everything from what [the general manager] needed to look at down to what the engineers should be watching.

However, although they knew what they *wanted* to measure, they did not know what they could measure. The information collection and reporting systems within the division did not report much of the data the team felt was needed, as the team's leader said "...the infrastructure to provide the information we needed simply did not exist." This problem was compounded by considerable controversy over what in fact should be measured. The group never did reach agreement on the measurement system.

### *Pilot Development Projects*

After designing the new development process, the team tested the process on a number of pilot projects. The team hoped that the pilots would serve two purposes. First, the pilots would provide an opportunity for the team to identify and correct problems in the process as designed. Second, if the pilots were successful, they could be used as examples to drive the process through the organization. In many cases, the second concern dominated the first. The first pilot project chosen was a very high profile vehicle that used a number of new and unproven technologies. As the first test of the new process, engineers could not draw on the bookshelf which did not yet exist. As one engineer said "...we crashed right through the wall of innovation and didn't look back."

The project suffered further since much of the support infrastructure required for the new tools were not in place. Engineers did not have powerful enough computers to use the new CAD/CAE/CAM software, and once the computers had been obtained, the rest of the organization was not prepared to accept their output due to software incompatibility. In addition, learning how to use the tools imposed a substantial burden on the already overworked engineers.

...I had some background in CAD/CAE from my master's program and I still stayed at work until midnight every night for a month learning how to use the tools and trying to figure out how to get my work done...some of the older engineers, even with training, they just have a [computer] sitting on their desks gathering dust

...the value of the tools was way overestimated...we never had time to take the courses and get the equipment we needed to really make this stuff work....it was really exhausting trying to learn how to use the tools and do the design at the same time.

The effect of these problems on the morale of the participating engineers was significant. Every interviewee reported being frustrated with the process. Many felt that management had defined a development process and then immediately gave the engineering staff a project and time line that could not be accomplished using the process. There was also a substantial additional workload as engineers tried to teach themselves how to use the new tools while trying to accomplish their normal work requirements. As a result, many of the engineers working on the pilots were forced to abandon much of the methodology to meet the project's schedule and specifications. A common sentiment was expressed by one engineer that said, "...I believe PDP is a good process. Some day I'd really like to work on a project that actually follows it." The overwork and frustration further limited the usefulness of the process.

The use of the pilot projects to generate enthusiasm for PDP throughout the organization was hampered by its lack of success. The dynamic was exacerbated because some engineers believed that management was not willing to recognize the problems in the process. One engineer recalls,

...PDP had a big budget at the beginning and it seemed like it was being used to impress us [the engineers]. They had PDP magazines, fliers, slogans etc. Early on one of the members of our team got interviewed for one of these and she wanted to mention some of the problems, the tools weren't ready, stuff like that. They took what she said and totally sugar coated it....we kept seeing these magazines that said our project was signed

up for PDP and doing really well...we were just shaking our heads...everybody in the division knew we were having problems.

The credibility of those promoting PDP was hurt by the conflict between the information received through formal channels such as newsletters, and informal channels such as word-of-mouth. The informal communication networks among engineers indicated the pilot projects were struggling while the formal media stated that the projects were doing well.

### *Rolling out the Initiative*

The roll out strategy had three components: A high level awareness campaign designed to show senior management's support; a middle level effort to create interest in the actual process; and intensive training that would give supervisors and engineers detailed working knowledge of the process. Components one and two were very successful. JD and other senior managers were highly visible with respect to the project. A number of promotional documents were created including an audio tape and a high gloss brochure. There was also a PDP newsletter throughout the initiative that kept people updated on the effort. In many ways, the promotional campaign was too successful. The effort became so popular that the ability of the organization to provide the detailed training was totally overwhelmed:

The team that wrote the handbook made it clear...that they felt nobody should get the handbook directly. They should get it in a training session....The whole idea was to set the stage for this is a vision....However the initial demand was so strong that...Executive engineers started coming to the PDP office asking for books for their people. Executive engineers are the third rank in the division -- general manager, director, executive engineer -- and when they say "I need 500 books," it's not easy to say no....of the books we published, less than 15% were received with the appropriate training.

### *Results*

Evaluating the success of the PDP initiative is difficult. The time delays are sufficiently long that, as of the Fall of 1995, only the first pilots have reached the launch phase. Further, the difficulty that the PDP team experienced with the metrics continues. There is little quantitative data to evaluate the success of the initiative. However, problems with data aside, many people involved developed some strong feelings as to the successes and failures of the effort. Everybody believed that the process as designed was good, but that the division as a whole does not follow it. JD rates the effort as a fifty percent success. The executive in charge of the initiative believes that they achieved eighty to ninety percent of their objective on the use the new tools and less than twenty percent of their objectives in

documentation of customer requirements, using project management, and developing a more rigorous and repeatable process. Members of the design team also believe that the effort failed to achieve its objectives, but hope that the effort will provide a catalyst for future improvements. Among the engineers interviewed, not one believed that the initiative had materially influenced his or her job.

What accounts for the lack of success? A number of common explanations were offered. First and foremost, many of the planned improvements in efficiency did not materialize. The bookshelf never became a reality. Engineers simply did not have the extra time to rigorously document new technologies. In addition, no improvements were made in the documentation of customer requirements. The executive in charge comments,

...we said, "...we're going to get this clear specification and understanding for what the customer wants before we start designing." It turns out that the customers didn't know. ...nobody had ever asked them before ...they said, "Gee beats us, what do you think?"

Since efficiency did not increase, the engineering workload remained constant and there was no extra time available for engineers to dedicate to the disciplines of PDP.

A second major problem cited was conflict over the use of project management and the increased amount of required documentation. Many managers believed that engineers resisted it because it was another tool with which managers kept tabs on engineers. One manager said,

A lot of the engineers felt that it was no value add and that they should have spent all their time doing engineering and not filling out project worksheets. It's brushed off as bureaucratic...

Interestingly, the engineers had a different view of project management. Many believed in it, but felt management had not given them sufficient slack in resources to do it properly.

As one engineer said,

...under this new system the engineer was responsible for the doing the physical design work using the new tools...however, none of our old tasks went away, so the new workload was all increase...in some cases your workload could have doubled...many times you were forced to choose between doing the physical design and doing the project and administrative work. To be successful you had to do the design work first, but the system still required all this extra stuff...There just weren't enough hours in the day, and the work wasn't going to wait.

#### **d. Summary: Contrasts**

JD was ultimately responsible for both initiatives, and differences in leadership style cannot totally account for the differing outcomes of the MCT and PDP programs. In addition, JD

was in an even more senior position during the second initiative and had more authority and resources at his disposal. JD also used similar strategies to promote the initiatives. Both efforts were focused on the concept of speed and cycle time and he gave both organizations a substantial amount of freedom to accomplish those objectives. He was further willing to allocate a substantial amount of resources to training and promotion in each initiative. Yet for these high level similarities, the initiatives produced very different results: MCT produced big improvements and continues to influence the division today, PDP has faded and has little impact on the division.

### *The Physical Process*

The first difference between the two efforts is the physical process being improved. Using the three dimensions of process complexity discussed above (technical, organizational, and dynamic complexity) product development is more complex than manufacturing. Many interviewees felt that the difference in outcomes between the two initiatives could be explained solely by differences in the physical environments. One interviewee, who played a key role in PDP, said “....manufacturing is easy, what we were doing [improving the PD process] was really hard.” The physical structure in figure two suggests that a key difference between MCT and PDP is the delay between correcting process problems and observing the results. At its worst manufacturing cycle time was less than a month. The product development cycle time was between three and four years. In addition there is more pressure in the development environment to allocate effort to defect correction. Designs that are done incorrectly can not be scrapped, otherwise they would stall the launch of a new vehicle. Instead rework, particularly for new vehicles, takes a very high priority, drawing resources away from process improvement.

### *The Kick-off*

The second difference was in the launch of each initiative. In the first case, JD was new to the division and did not have the credibility or the resources to initiate a high profile effort. Instead he challenged the conventional understanding of the manufacturing process, then spent substantial time personally explaining the initiative to people in the plants. In contrast, the PDP effort was kicked-off with substantial fanfare and was immediately given a high priority within the division. However, there was no event analogous to the presentation of the cycle time data. While people believed that the PD area needed improvement, the initiative did not begin with a substantial challenge to the dominant mental model of the product development process.

### *Redesign Mode*

A third key difference between the initiatives was the mode of redesign. Nowhere in the MCT effort did a group of people form to 'design' the new manufacturing process. There was never a conscious effort to create a new blueprint. Instead, the process was highly experimental, changes were made incrementally on an as needed basis, and, although a world class manufacturing process emerged at the end of the initiative, it was never 'designed' in the conventional sense. Further, the process involved two distinct phases. The first phase focused on the physical *structure* of the process. Via MCE analysis the team analyzed each step in the work process in terms of its contribution to total cycle time. The focus on cycle time and value add redefined what constituted a defect and a process problem. Many people reported the experience of 'discovering' new buffer inventories. Although these inventories were never actually hidden from view, they were not discovered until the definition of a process problem was changed to include WIP inventory. The second phase of the MCT effort was focused on the *behavioral* processes that supported the work process. Via TOC and other tools, the division developed a better method to manage the manufacturing process that relied on changing and improving the decision processes of the operators and material handlers. Interestingly, participants in the effort found that new members of the organization who tried to participate in phase two without having participated in phase one were not able to achieve the same level of improvements. One plant manager said,

...the thing that TOC was missing was it didn't do what we had done...it didn't spend all the time on understanding the cycle time concepts and value added versus non value added....we made a conscious decision to go back and be sure that people who came to the division spent some time understanding MCT and MCE fundamentals before getting into TOC.

In contrast to MCT, the PDP design team spent almost two years designing the development process. There was no experimentation with new methods. Instead the team scanned the environment for new ideas that worked in *other* places. Most of the new concepts came from outside consultants and benchmarking efforts and were not individually tested before being integrated within the PDP process. There was no temporal separation between process structure and process management, and the planned changes in the behavioral processes were predicated on the structural changes having been made.

### *Metrics, Measurements, and Objectives*

A fourth key difference was in the measurement and evaluation process. In the MCT effort, JD dictated only that the plants measure cycle time without specifying how it was to be measured. As the initiative progressed, the cycle time metric became more standardized by joint agreement of the plant managers based on what had proven useful in practice. In contrast the PDP development team formed an entire committee to develop metrics and had little input from the design engineers themselves. The team never reached strong consensus on what should be measured and many believed the PD process could not be measured because of its complexity. A high level manager who participated in both efforts discusses the dilemma,

The metrics process was frustrating....However there was a reciprocal frustration on the manufacturing side. Finally we just decided to do it. We weren't sure that the metrics were the right ones, but we just decided to do the best we could, and it helped...It was also very difficult in manufacturing, but in manufacturing the goal was made operational immediately....we never really got to that point in PDP.

In addition, in the MCT initiative the metrics were changed rather than augmented. Labor utilization standards were dropped in favor of cycle time, and, as one engineer said, "This was a big change...normally they add new measurables without removing old ones and tell you to do well on both, even if some are directly in conflict." In PDP, the new metrics and requirements were added to the old without any reduction in schedule pressure or increase in resources.

### *Promotion and Dissemination*

A fifth difference was in the mode of promotion. PDP was promoted through a wide array of media including pamphlets, audio tapes, brochures, and newsletters. As the initiative started to flounder the informal communication networks provided different information and undermined the credibility of managers promoting the effort. As the conflict became more pronounced, engineers began to question the motives and credibility of management. In contrast, MCT had very little formal promotion. MCT was promoted primarily through JD's personal visits to the plants and later through meetings of the plant staff. As one participant in the research team pointed out, "...there was never an MCT newsletter."

## *Training*

A sixth difference was in the mode and content of the training. MCT training was done on an as needed basis, and emphasized developing intuition for the structure and dynamics of the process. An operations manager discusses his efforts,

We started by teaching each of the work teams how to manage their line using TOC...the classes were useful, but the real learning came from working with them on their lines on the floor. I would coach them through making actual decisions. I'd let them make the decisions and then we would talk about the results.

In contrast to the recipe format of the MCT training, PDP training consisted of a sequential description of the process. PDP was presented as a process that should be followed without exception. There was no discussion of the *current* structure of the development process or how the division might move from the current state to that which was desired. In addition, since the MCT effort unfolded over the course of five years, training was eventually provided to almost every member of the manufacturing organization. In contrast, detailed PDP training was never received by the majority of the engineering staff.

## 5. Analysis and Discussion

The MCT effort was largely consistent with the alternative assumptions developed by Weick. The design of the manufacturing process emerged from a sequence of local adaptations that exploited existing inefficiencies. Further, the effort was largely controlled by ideas. JD drove the effort with a simple idea: faster is better. In contrast the PDP process was designed ahead of time and off-line. It started with the idea that faster is better, but quickly degenerated into a focus on controlling the activities of the engineering staff. Further, the MCT effort made no temporal or organizational distinction between the design and use of the process. Thus, the improvisational metaphor and structuration theory offer a way to explain the differences between the two initiatives. However, an important question remains unanswered: If MCT was successful, why didn't JD simply follow the same strategy for PDP?

### **a. Failure Modes**

#### *Self Confirming Attributions*

Prior to the MCT effort, both systems shared some important characteristics. In many cases, participants reported taking actions that they knew would hurt the process in the long run, but were necessary to hit the short term objectives. As an example, consider the



following two quotes taken from interviews with operations engineers at *different* manufacturing facilities. In the first case the supervisor discusses the difficulty of finding time for preventative maintenance:

...supervisors never had time to make improvements or do preventative maintenance on their lines...they had to spend all their time just trying to keep the line going, but this meant it was always in a state of flux, which in turn, caused them to want to hold lots of protective inventory, because everything was so unpredictable. It was a kind of snowball effect that just kept getting worse.

Second, a manager at a different plant discusses the inability of operators to stop the line to make improvements that would increase yield.

In the minds of the [operations team leaders] they had to hit their pack counts. This meant if you were having a bad day and your yield had fallen ... you had to run like crazy to hit your target. You could say “you are making 20% garbage, stop the line and fix the problem”, and they would say, “I can’t hit my pack count without running like crazy.” They could never get ahead of the game.

In both cases, operators ran their machines even though it was not in the best interest of the plant. Management dictated high utilization levels by setting high pack counts. Machines had quality problems so they were run constantly to hit the count leaving no time for preventative maintenance or continuous improvement. Lack of preventative maintenance and improvement effort led to an unreliable production environment that required a high level of utilization to achieve through-put targets: The system that emphasized machine up-time became a self fulfilling prophecy. These examples can be mapped to the feedback structure discussed in section three (see figure 12). Prevention is represented by the balancing loop B2, and requires that machines be taken off line. The corrective action (loop B1) represents running the machine to increase through-put rather than taking time for preventative maintenance or continuous improvement. The positive loop, R1, gives the system the tendency to move towards one or the other extreme.

Why did the system move towards low quality and high production pressure? One hypothesis is shown in figure 13. As the process capability gap rises, managers make the attribution that some of the gap is due to low effort by those working in the process. Managers then respond by increasing the pressure to hit the day’s pack count. The addition of these forces creates another negative feedback loop, B6, which works to close the gap between the desired and actual process throughput. Although such actions may lead to a short term increase in through-put, they are eventually self-defeating. Additional

production pressure reduces the willingness of operators to shut down machines for preventive maintenance or continuous improvement, which ultimately leads to more machine breakdowns or product defects and creates the reinforcing loop, R4. Managers in the MCT effort believe that breaking this self confirming cycle was one key to the success of the effort. One manager said,

There are two theories, one says 'there's a problem let's fix it', the other says 'we have a problem, someone is screwing up, let's go beat them up'. To make improvement we could no longer embrace the second theory, we had to use the first.

The basic structure of product development is similar (see figure 14). Engineers can do process improvement work (loop B2) including learning how to use the new development tools, contributing designs to the bookshelf, and doing up-front improvement work like failure mode and effects analysis (FMEA). However, they are also responsible for rectifying past problems (loop B1) including redoing past designs that were discovered to be incorrect. As in the previous example, the balancing loop R1 pushes the system towards the extreme conditions. In PDP, correction efforts dominated. Engineers reported having almost no time for improvement related activities and spent much of their time 'catching up'.

PDP was not successful in reversing this trend. In fact, it may have made it worse. Managers in the PDP effort frequently mentioned that engineers were 'undisciplined' and resisted following a standard process. Many also mentioned that engineers did not 'want to be measured'. The response of those designing PDP is a good example of the self confirming attribution process discussed above. The PDP process was designed to 'add discipline' back into the process by very specifically laying out the tasks required of an engineer. In the meantime, the division accepted new customers under the assumption that PDP would lead to improved efficiency. Thus, the engineering workload was also increased. As one manager said,

We took on more work than there were people to do it...We'd do anything for anybody, anytime...[and]...we assumed we'd fund it [the extra business] with efficiencies from PDP.

The diagram shows how these actions were self defeating. As the capability gap increases, managers attributed some of the difficulties to the 'undisciplined nature' of the engineering staff. They respond by designing a new process that adds discipline back into product development. The regulation of engineering activities creates a balancing loop, B5; increased regulation of engineering activity leads to increased effort. The workload also

increased, creating loop B6. Just as in manufacturing, these actions may have led to short term improvements, but in the long run they help explain the limited impact of PDP.

By increasing the workload, managers reduced the time that engineers could dedicate to learning how to use the new engineering tools, placing designs on the bookshelf, and properly following the development process. In addition, by regulating engineering activities through project management, managers limited the ability of engineers to experiment with new techniques, thus thwarting continuous improvement. These relationships – between workload and improvement time, and regulation and willingness to experiment – create two reinforcing loops. Managers, believing engineers are undisciplined, increase the workload and nothing changes or the situation gets worse. Seeing this, adds further support to the contention that engineers were lazy or ‘lacked discipline’. A member of the PDP development team recalls the reaction of engineers to the PDP process,

The problem with [the process] was that sometimes management chose to adhere to it, and sometimes they chose not to .....when we set out the disciplines of PDP we said “there it is, it’s a very disciplined, rigid program, go follow it.” Then in the very next breath we would say, “I want you to ignore all that and bring this project home in half the time.” That just didn’t go down very well.

### *Self Defeating Changes*

The dynamics described above are further exacerbated by the combined impact of increasing regulation and increasing workload. In both cases those working in the process were eventually forced to make *ad hoc* departures from the regulations to meet their objectives. In the theory of structuration, technology is the product of human actions, but it also constrains and facilitates them. Although a process may be designed by managers, those that work within it are constantly making changes to both its physical and behavioral components. In the two studies a key driver of these changes was the conflict between the process as dictated by management and the performance objectives. Prior to the MCT effort, there were other attempts to reduce inventory, each of which ultimately failed. The suggested reason for these failures is that the initiative conflicted with the overriding philosophy of “Keep All Your Machines Running.” Operators circumvented these objectives by holding ‘secret’ caches of inventory that could be used to buffer against disruptions. These actions were driven by the need to achieve the desired labor efficiency. For example, one operations engineer recalls,

It didn't take long for them [line supervisors] to develop a buffer in front of their line so that if the schedule called for 700 and their line was fully utilized at 800, they could still run 800 units every day, and still make their labor performance

Operators and supervisors did this because they knew the penalty for missing the objective was substantial. One manager recalls "... supervisors who missed their targets knew they were going to get 'beat up' by their managers."

The dynamic structure of this example is shown in figure 16. Managers react to a gap in the through-put by strengthening their control over WIP inventory and by increasing the pressure to hit pack counts. Those working on the production line then experience a conflict in their objectives. Reducing WIP inventory makes it harder to hit production goals. Workers react to the conflict by departing from the process and holding 'secret' caches of WIP. The extra WIP allows them to satisfy their objectives and temporarily increases process through-put. However, increased WIP reduces the overall manufacturing cycle time and reduces the total through-put of the manufacturing process causing management to further tighten controls and increase production pressure. These connections create a positive feedback loop that constantly pushes the manufacturing system to high levels of WIP inventory and high levels of production pressure. The pressures manifest themselves as increased tension between workers and managers, a feeling of helplessness on the part of both workers and managers, and a chaotic manufacturing environment that seems to behave in an unpredictable manner.

In addition, the inventory does not remain secret. Management's goals erode and the once *ad hoc* changes and procedures become an accepted part of the system. For example, as mentioned earlier, one facility built a tent as a temporary storehouse for inventory. Over time, inventory goals eroded and the tent became a permanent part of the manufacturing technology. Thus, as suggested by structuration theory *ad hoc* and temporary changes to technology can eventually come to be perceived as permanent. As Weick observed, "A little structure goes a long way(Weick 1993)."

A similar dynamic existed in the PDP effort. For example, engineers frequently felt the need to depart from the PDP process in an attempt to finish their work on schedule. These departures took the form of neglecting documentation, not placing technologies on the bookshelf, or not filling out a detailed work plan. One manager recalls,

An engineer might not take the time to document her steps or put the results of a simulation on the bookshelf and because of that she saved engineering time and did her project more efficiently. But in the long run it prevented us from being able to deploy the reusability concepts that we were looking for....there was a lot of push back when it came to following a process that people could look at and say, "Hey, I can do that more efficient by not doing some of these interim steps."

### **c. Breaking the Cycle**

The self-confirming processes discussed above can explain the failure of PDP and previous efforts to improve manufacturing. However, they give little insight into how MCT was able to overcome these dynamics. How did the manufacturing organization overcome the dynamics discussed above and harness the virtuous cycle of improvement? How did the organization break the adversarial relationship between managers and those that worked on the production floor? In this section I advance some hypotheses based on the MCT and PDP experience.

#### *Redefining Defects and Process Problems Creates Improvement Opportunities*

JD initiated the MCT effort by arguing that the manufacturing system would be more efficient if it was focused on cycle time rather than machine utilization. Such a statement challenged the conventional wisdom – mainly that the system was not optimized by running each machine at 100% utilization. He supported his statement with data, and thereby exposed an important inconsistency in the dominant mental model. In contrast, PDP, for all its promotion, offered few new ideas.

Many successful change techniques start by challenging the conventional wisdom. For example, TQM is predicated on reversing the conventional logic that quality and cost are trade-offs (Deming 1986:1). In developing its famous production system, Toyota challenged the conventional belief that cost and manufacturing flexibility were inversely related (Womack, Jones, and Roos 1991). Ideas as the catalyst for change presents quite a different picture than other literature and practitioner oriented change theories. Many authors argue that the catalyst for change is a lack of fit with the environment which results in some type of crisis (Tushman and Rommanelli 1985). In contrast, the catalyst for change described here is a challenge to the dominant understanding of the inner workings of the organization. The two conceptions are related, organizational crises may increase the likelihood that participants will challenge the dominant model.

The challenge to the dominant mental model is important in the context of the improvement structure outlined above where a defect is a matter of definition, and a process problem is a matter of inference. JD's cycle time argument redefined what constituted a defect and led to new inferences about process problems. The redefinition was critical to setting the positive loops R1 and R2 working in a favorable direction. Redefining what constituted a defect greatly increased the stock of known defects. In the process of calculating cycle times and value add percentages, those working in the process quickly realized how excess inventory and redundant production steps created problems, and many were easily eliminated. If, as in Weick's conception, processes are designed to exploit existing inefficiencies, then a successful effort must start with the identification of such opportunities. PDP had no similar process. There was no change to the dominant mental model of the engineers or the supervisors and few new problems were identified.

Discovering a large number of new defects and inferring many new process problems was a great benefit to the MCT effort. Having a large stock of process problems makes early improvements easier (Schneiderman 1988). For example, Alpha's plant manager recalls being encouraged when the first MCE analysis showed that the value add percentage was less than 1%: A very low value add percentage meant there was substantial room for improvement. In the first year of the MCT effort many participants reported that making improvement was easy. Simply identifying WIP as undesirable allowed them to make improvements. The early and rapid improvements were critical to initiating the reinforcing cycle of commitment. The early results demonstrated to skeptics that the methods worked and allowed loop R2 to work favorably.

#### *Early Results Create Slack for Subsequent Improvement*

A common feature of successful improvement efforts is that they produce early results. Many authors argue that early results are important to stimulate interest and stimulate commitment (Kotter 1994, Schaffer and Thomson 1992). However, the structure discussed above suggests that early results have another important positive effect that is not well recognized in the literature: they create slack. The Alpha facility took advantage of the slack created by early improvement by undertaking an intense period of experimentation that led to still more improvement. The early improvements were substantial enough to allow the plants to undertake experiments that would yield further improvement thus initiating the reinforcing process represented by loop R1.

PDP did not experience a similar phenomenon for a number of reasons. They were not able to generate early results quickly. However, they should have experienced at least some improvement in efficiency. They did, for example, eventually achieve wide acceptance of the computer tools. Although this happened slowly, the improved efficiency still could have created slack. However, unlike in MCT, even though efficiency was improving, the workload was growing more rapidly. Rather than dedicating the slack resources to improvement, the division accepted even higher levels of business. The substantial increase in workload overwhelmed the early improvements. What caused managers to do this?

The stock/flow structure of improvement suggests that there is a delay between the correction of process problems and an improvement in through-put. Empirical research shows that human decision making in such environment is well below optimal (Paich and Sterman 1994, Sterman 1989a, 1989b). Managers within the PDP effort are likely to have both underestimated the delay between implementation and reaping the benefits and the impact of production pressure on improvement effort. Further, when those benefits were not observed, the lack of results was attributed to the undisciplined nature of the engineering staff. Managers responded by increasing their control over the process and further increasing production pressure. Few managers anticipated that providing a more detailed process actually forced engineers to be less disciplined as they sought ways to meet their deadlines while still appearing to follow the process.

In contrast, during the MCT effort, control was substantially reduced. For example JD allowed the plants to define, calculate, and report their own cycle time metrics. By allowing people to define and measure their own environment, the MCT effort eliminated the conflict between production pressure and the process controls. Instead of making *ad hoc* changes hidden from managers, participants performed experiments that were observable to all. The difference between *ad hoc* changes and experiments is important. Wruck and Jensen (1994) point out that TQM encourages science-based decision making. Experiments add rigor to the process and improve the chances of making a favorable change. In addition, by making the results observable, rather than actively hiding them, the organization is better equipped to quickly adopt the benefits of new learning.

To summarize, a successful improvement effort has the following characteristics. It begins with a challenge to the dominant mental model of the organization. The challenge takes the form of a new definition for what constitutes a defect. The new definition leads

to the identification of a large number of process problems. Such problems present early improvement opportunities. Early results create a reinforcing cycle of commitment and create slack that the organization can dedicate to further experimentation. A key component of this process is that managers relinquish control over the process and allow participants to make changes to the process. The changes, once allowed, can be guided by TQM or other methodologies that add rigor to the process.

## 6. Conclusions and Future Directions

### **a. Summary**

Process improvement and redesign efforts have both physical and behavioral dimensions yet past scholarly work has tended to focus on one component at the expense of the other. In contrast, practitioners of TQM and re-engineering offer both physical and organizational tools, but are divided on whether the appropriate mode of change is incremental or radical. The purpose of this paper is develop a grounded theory of improvement and redesign that captures both the physical and organizational dimensions of the issue, as well as reconciles the incremental/radical change dilemma. Through the development of stock/flow and feedback diagrams, a representation of both the physical and organizational structures of improvement is developed. The processes identified are used as building blocks to identify the common modes of failure in redesign and improvement efforts.

The failure modes discussed result from an error of attribution made by managers in assessing the cause of low process through-put. Specifically, if managers attribute low performance to the attitudes and disposition of those that work in the process, they react in a manner that makes such an attribution a self-fulfilling prophecy. The dynamic created leads to an environment characterized by increasing levels of production pressure, a high degree of managerial control over the process, and workers who are forced to make *ad hoc* and secret modifications to the process to achieve their through-put goals. Improvement is difficult in such an environment since a large portion of the available resources are dedicated to correction efforts and there is too much production pressure to allow the type of experimentation and adaptation needed for improvement.

The results from two case studies on improvement and redesign efforts were presented to support and explicate the theory. One initiative was successful, the other less so. Both processes suffered from the problems discussed above. The relationship between those that work within the process and those that manage the process was contentious, and each



group blamed the low performance of the process on the other. The successful effort overcame these difficulties by challenging the dominant mental model of those that ran the process and then allowing process participants to physically and socially reconstruct the process based upon that challenge. Such a mechanism has a number of desirable properties. The new causal statement redefined what constituted a defect and led to new inferences about process problems. Early improvements based on these definitions created slack that was then dedicated to further experimentation and improvement. In addition managerial control over the process was reduced. In contrast, in the unsuccessful effort, further attempts were made to increase control over those that worked within the process and any initial slack created was dedicated to more work rather than more improvement. The antagonistic relationship between managers and process participants was never resolved because the dominant mental models were never sufficiently challenged and then reconstructed through the improvement process.

The results of the analysis lead to the assertion that TQM and BPR are complementary activities. TQM is an improvement and redesign recipe that dictates science based decision making and allocates authority to those that are most able to make use of it (Wruck and Jensen 1994). Re-engineering normally focuses on creating new blueprints while giving less attention to the recipes required to create them. Here re-engineering is re-conceptualized as a tool for challenging the dominant mental model. By constructing a hypothetical process that is faster and more efficient, re-engineering challenges people's understanding of their organization. The flaw in current thinking on re-engineering is the assumption that the hypothetical blueprint can be immediately constructed in the real organization. Research on decision making shows that human ability to anticipate the dynamics of higher order complex systems is very limited. Planning an ideal process that can be feasibly implemented in advance requires a higher degree of foresight and cognitive capability than is normally displayed. In contrast to the implementation philosophy of re-engineering, TQM offers a recipe that can produce efficient and flexible recipes. Traditionally, however, these efforts have been limited to quality improvement.

The process described here is as follows: Successful improvement begins with a challenge to the dominant mental model of the organization. The challenge can take the form of new data, observation or benchmarking of a process that produces a similar output much more efficiently, or the construction of a hypothetical process with desirable properties. Process participants are then encouraged to reconstruct and reinterpret their process in light of the new challenge. Managers guide the process by reducing the level of control and providing

the appropriate training and slack to allow an experiment guided search for improvements to the process.

## **b. Implications for Research and Practice**

### *Researchers*

The analysis and conclusions have important implications for researchers. For operations researchers and management scientists, it suggests that traditional work on designing better systems of conversion should be complemented with research on implementing such systems. The development of models and tools that enhance real-world intuition about these systems would be very valuable. Processes are rarely implemented as designed and will be subjected to a myriad of adjustments and adaptations by those that work within the process. Such actions place a premium on developing better intuition in the minds of those making such changes. Operations research and management science can contribute much in this area. Early efforts in this area, including the development of 'conceptual models', simulation games, and management flight simulators, are promising. For example, in an effort to reduce maintenance cost, the DuPont corporation developed a board game to teach basic lessons about the dynamics of preventative maintenance with very favorable results (Carroll *et al.* forthcoming; Sterman, Banaghan, and Gorman 1992). For organizational scientists, the analysis suggests that future studies on organizational change and design need to explicitly consider the feedback structure created by the physical environment in which the change is taking place. Time delays, feedback processes, and interconnections all play an important role in determining the outcome of a change effort.

### *Practitioners*

The study also has implications for practitioners. Although TQM and re-engineering have normally been viewed as separate options, here they are viewed as complementary and related activities. In contrast to the conventional view, re-engineering is viewed as a means to challenge the dominant mental model rather than to design the new process. In fact it is argued the designing a process off-line is difficult since it is hard to anticipate both its dynamics and the large number of *ad hoc* changes and adjustments that will be made by those within the process. The role of management is then to challenge the dominant thinking and then, through methods similar to those used in TQM, provide the tools and support necessary for process participants to reconstruct a new and better process based on the challenge. Such an effort still involves a large number of localized changes, but in

contrast to the conventional model, these changes are made in an open and observable manner using rigorous methods such as experimentation and simulation.

Practitioner oriented change theories often emphasize the role of a vision and compare the leader to the architect (Kotter 1994, Senge 1990). The framework developed here offers a different conception. As pointed out by Weick, the architectural metaphor assumes ideal structures and process. The laws governing the domain of architecture are much better known and understood than the laws governing a complex organization with both physical and behavioral components. In such a setting the role of leadership is perhaps better compared to that of a teacher or the manager of a research project. Managers must challenge the conventional understanding of the project and they must supply the tools and support to those that work within the process to allow them to make localized changes in the most efficient and productive way possible.

## References

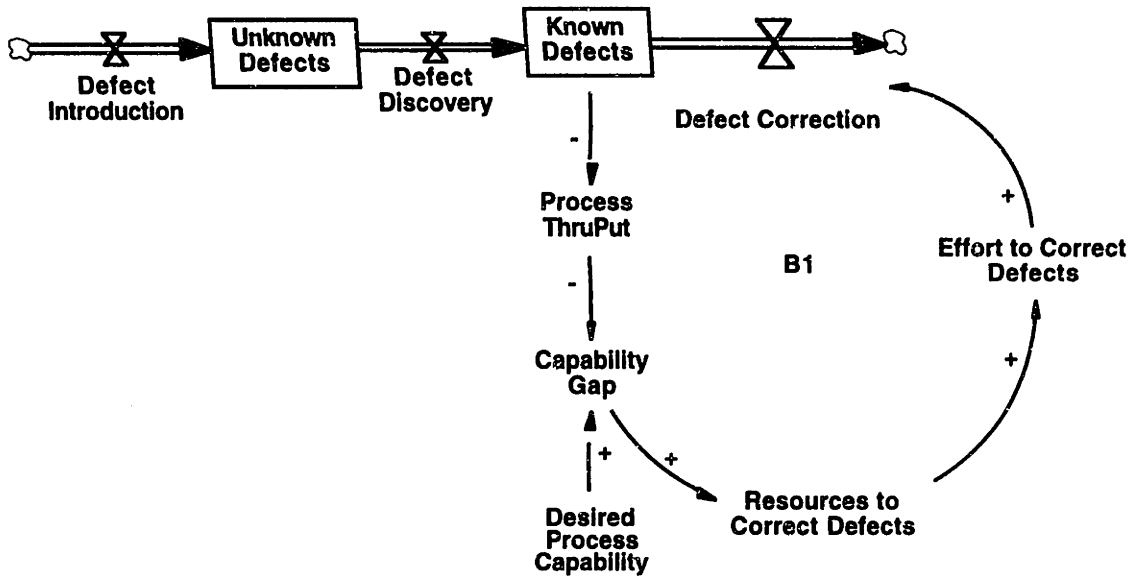
- Carroll, J., J. Sterman, and A. Marcus (forthcoming). 'Playing the Maintenance Game: How Mental Models Drive Organization Decisions', to appear in, R. Stern and J. Halpern (eds.) *Nonrational Elements of Organizational Decision Making*. Ithaca, NY, ILR Press.
- Chase, R.B. and N.J. Aquilano (1989). *Production and Operations Management*, Fifth Edition, Homewood, IL. Irwin.
- Cyert, R. and J. March (1992). *A Behavioral Theory of the Firm*, Cambridge Ma., Blackwell Publishers.
- Deming, W. E. (1986). *Out of the Crisis*. MIT Center for Advanced Engineering Study, Cambridge, MA.
- Easton, G. and S. Jarrell (1994). 'The Effects of Total Quality Management on Corporate Performance: An Empirical Investigation'. Working Paper, University of Chicago, Chicago, Illinois, 60637.
- Eisenhardt, K.M. (1989). 'Building Theories from Case Study Research', *Academy of Management Review*, Vol. 14, No. 4, 532-550.
- Ernst and Young (1991). 'International Quality Study – Top Line Findings' and "International Quality Study – Best Practices Report" Ernst and Young/American Quality Foundation, Milwaukee, WI.
- Feigenbaum, A. V. (1983). *Total Quality Control*, New York, McGraw Hill.
- Forrester, J.W. (1968). 'Market Growth as Influenced by Capital Investment,' *Industrial Management Review*, 9, No.2: 83-105.
- Forrester, J.W. (1969). *Urban Dynamics*, MIT Press.
- Forrester, J. W. (1971). 'Counterintuitive Behavior of Social Systems', *Technology Review*, 73(3), 52-68.
- Garvin, D. A. (1988) *Managing Quality*, New York, The Free Press.
- Garvin, D. (1995a). The Process of Organization and Management, Working Paper #94-084, Harvard Business School, Boston MA02163.
- Garvin, D. (1995b). 'Leveraging Processes for Strategic Advantage', *Harvard Business Review*, September-October, pp. 77-90.
- General Accounting Office (1991). 'US companies improve performance through quality efforts', GAO/NSIAD-9-190 (2 May).
- Glick W., G. Huber, C. Miller, D. Doty, and K. Sutcliffe (1990). 'Studying Changes in Organizational Design and Effectiveness: Retrospective Event Histories and Periodic Assessments', *Organization Science*, 1(3):293-312.
- Goldratt, E. and J. Cox (1986). *The Goal: A Process of Ongoing Improvement*, Croton-on-Hudson, NY., North River Press.
- Hammer, H. and J. Champy (1993). *Re-engineering the Corporation*, New York, N.Y., Harper Collins.
- Huber, G.P. and W.H. Glick (1993), *Organizational Change and Redesign: Ideas and Insights for Improving Performance*. New York, Oxford University Press.

- Ishikawa, K. (1985) *What is Total Quality Control?*, Englewood Cliffs, NJ, Prentice Hall.
- Juran, J. M. (1988) *Juran on Planning for Quality*, New York, The Free Press.
- Kanter, R.M., T.D. Jick, and R.A. Stein (1992). *The Challenge of Organizational Change*. New York, Free Press.
- Kaplan, R. (1990a). Analog Devices: The Half-Life System, Case 9-191-061, Harvard Business School.
- Kaplan, R. (1990b). Analog Devices: The Half-Life System, Teaching Note 5-191-103, Harvard Business School.
- Kaufman, R. (1992). 'Why Operations Improvement Programs Fail: Four Managerial Contradictions,' *Sloan Management Review*, Fall, 83-93.
- Krahmer, E. & R. Oliva (1995). Improving Product Development Interval at AT&T Merrimack Valley Works. Case history available from author, MIT Sloan School of Management, Cambridge, MA 02142.
- Kotter, J.P. (1995). 'Leading Change: Why Transformation Efforts Fail', *Harvard Business Review*, March-April.
- Lant, T. (1992). 'Aspiration Level Adaptation: An Empirical Exploration', *Management Science*. 38(5), 623-644.
- Leonard-Barton, D. (1988). 'Implementation as Mutual Adaptation of Technology and Organization', *Research Policy*, 17, 251-267.
- Mausch, M. (1985). 'Vicious Cycles in Organizations', *Administrative Science Quarterly*, 30:14-33.
- Orlikowski, W.J. (1992). 'The Duality of Technology: Rethinking the Concept of Technology in Organizations', *Organization Science*, 3(3).
- Orlikowski, W.J. and M.J. Tyre (1994). 'Windows of Opportunity: Temporal Patterns of Technological Adaptation', *Organization Science*, 5(1).
- Orlikowski, W.J. and D.C. Gash (1994). 'Technological Frames: Making Sense of Information Technology in Organizations', *ACM Transactions on Information Systems*, 12(2).
- Orlikowski, W.J. (1994). 'Improvising Organizational Transformation over Time: A Situated Change Perspective', forthcoming in *Information Systems Research*.
- Paich, M. and Sterman, J. (1993). 'Boom, Bust, and Failures to Learn in Experimental Market's. *Management Science*, 39(12), 1439-1458.
- Perrow, C. (1986). *Complex Organizations: A Critical Essay*. Third Edition. New York, Random House.
- Plous, S. (1993). *The Psychology of Judgment and Decision Making*, New York, McGraw-Hill.
- Repenning, N. (1996a). Reducing Manufacturing Cycle Time at MidWest Electronics, Case Study.
- Repenning, N. (1996b). Reducing Product Development Time at Mid West Electronics, Case Study.
- Repenning, N. (1996c). Agency Problems in Process Improvement Efforts. Working Paper.

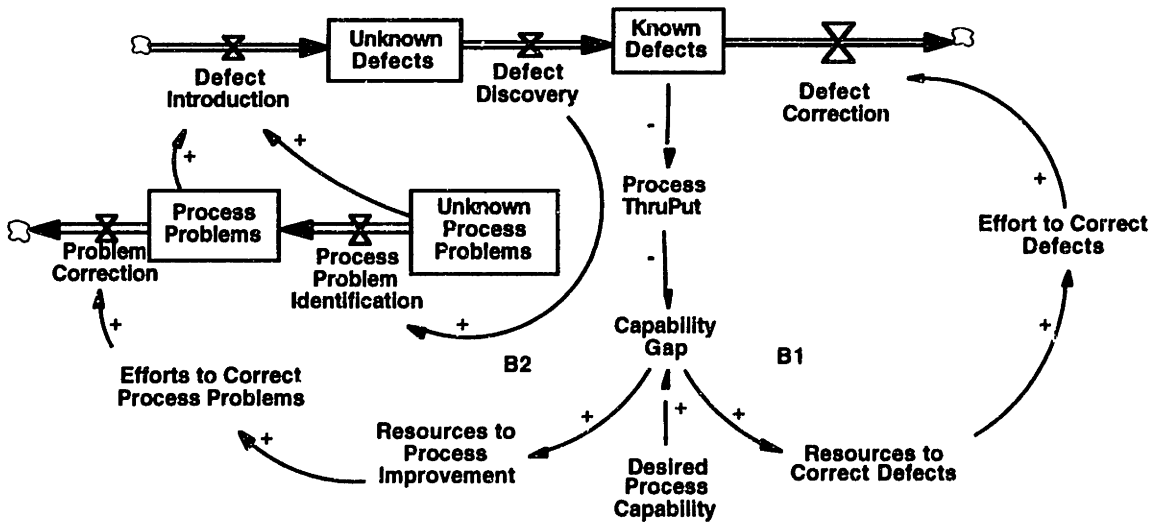
- Richardson, G. P. (1991). *Feedback Thought in Social Science and Systems Theory*. Philadelphia, University of Pennsylvania Press.
- Sastry, M. (1995). *Time and Tide in Organizations: Simulating Change Processes in Adaptive, Punctuated, and Ecological Theories of Organizational Evolution*. Unpublished doctoral dissertation, Sloan School of Management, MIT, Cambridge, MA.
- Schaffer, R. and H. Thomson (1992). 'Successful Change Programs Begin with Results', *Harvard Business Review*, Jan/Feb. 80-89.
- Schein, E.H. (1972). *Professional Education*. New York, McGraw-Hill.
- Shewhart, W. (1939). *Statistical method from the viewpoint of quality control*. Washington, DC: US Department of Agriculture.
- Schneiderman, A. (1988). 'Setting Quality Goals', *Quality Progress*, April, 55-57.
- Shiba, S., D. Walden, and A. Graham, (1993). *A New American TQM. Four Practical Revolutions in Management*. Portland, OR: Productivity Press.
- Senge, P.M. (1990). *The Fifth Discipline: The Art and Practice of the Learning Organization*, New York, Double Day.
- Simon, H.A. (1962). 'The Architecture of Complexity', *Proceedings of the American Philosophical Society*, 106(6), 467-82.
- Sterman, J.D. (1994). 'Learning in and about Complex Systems', *System Dynamics Review*, 10(2-3):291-330.
- Sterman, J.D., E. Banaghan, and E. Gorman (1992) Learning to Stitch in Time: Building a Proactive Maintenance Culture at DuPont. Case Study, Sloan School of Management, MIT.
- Sterman, J. D. (1989a). 'Misperceptions of Feedback in Dynamic Decision Making', *Organizational Behavior and Human Decision Processes* 43 (3): 301-335.
- Sterman, J. D. (1989b). 'Modeling Managerial Behavior: Misperceptions of Feedback in a Dynamic Decision Making Experiment', *Management Science* 35 (3): 321-339.
- Sterman, J., N. Repenning, and F. Kofman (1994). Unanticipated Side Effects of Successful Quality Programs: Exploring a Paradox of Organizational Improvement. Working Paper #3667-94-MSA, Sloan School of Management, Cambridge, MA.
- Tushman, M.L. and E. Romanelli (1985). 'Organizational Evolution: A Metamorphosis Model of Convergence and Reorientation', In B.M. Staw and L.L. Cummings (Eds.) *Research in Organizational Behavior*, Vol. 7:171-222. Greenwich, CT. JAI Press.
- Van de Ven, A., and M.S. Poole (1995). 'Explaining Development and Change and Organizations', *Academy of Management Review*, V20(3), 510-540.
- Weick, K.E. (1979). *The Social Psychology of Organizing*, Second Edition, New York, Random House.
- Weick, K. E. (1993). 'Organizational Redesign as Improvisation' in Huber, G.P. and W.H. Glick (eds.) *Organizational Change and Redesign.*, New York, Oxford University Press.
- Womack, J., D. Jones, and D. Roos (1991). *The Machine that Changed the World: The Story of Lean Production*, New York, Rawson and Associates.

- Wruck, K.H., and M.C. Jensen (1994). 'Science, Specific Knowledge, and Total Quality Management', *Journal of Accounting and Economics*, 18, 247-287.
- Wynne, B. (1988). 'Unruly Technology: Practical Rules, Impractical Discourse and Public Understanding', *Social Studies of Science*, 18, 147-167.

**Figure 1**  
**First Order Improvement**

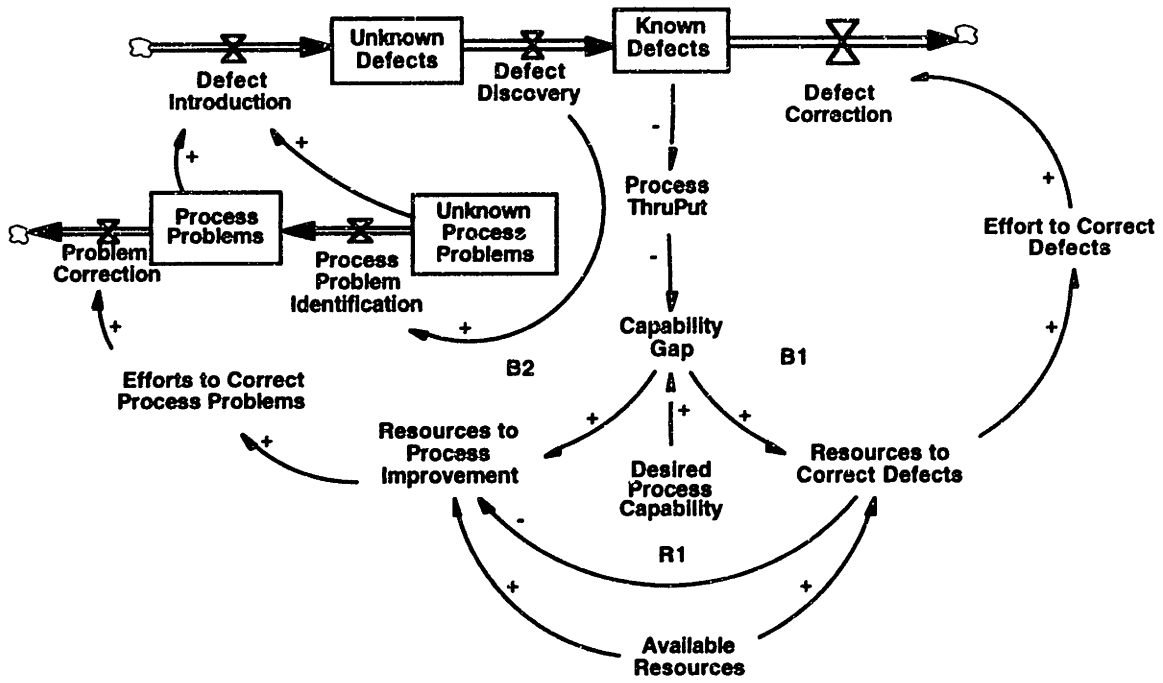


**Figure 2**  
**First and Second Order Improvement**

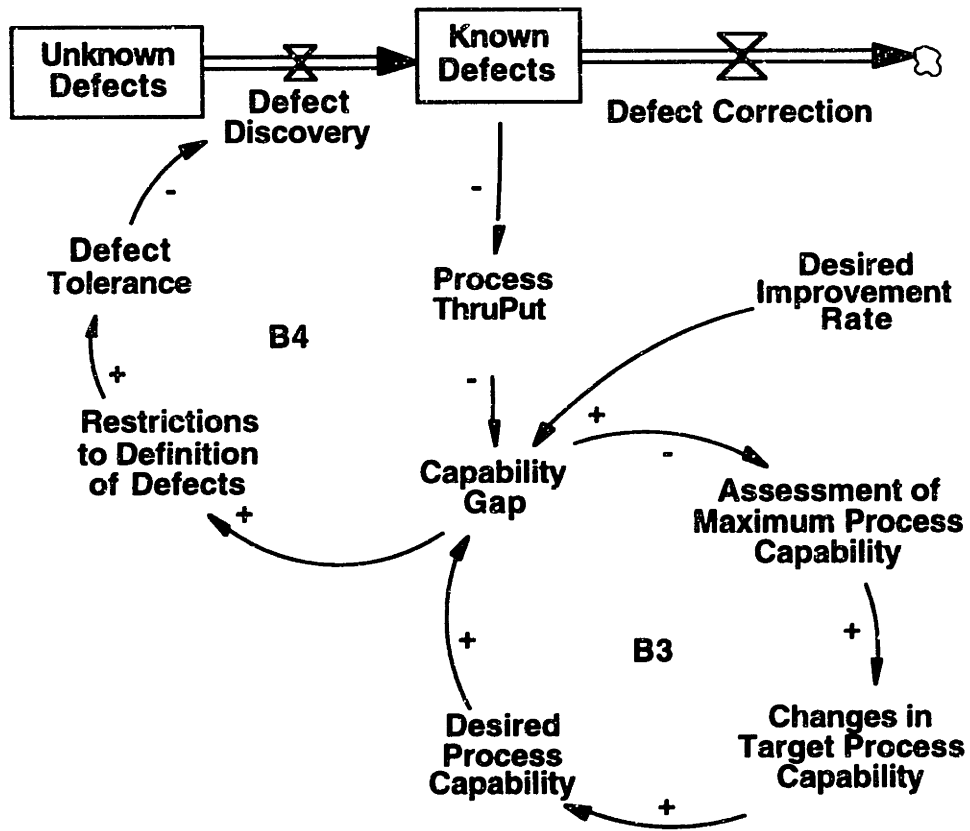




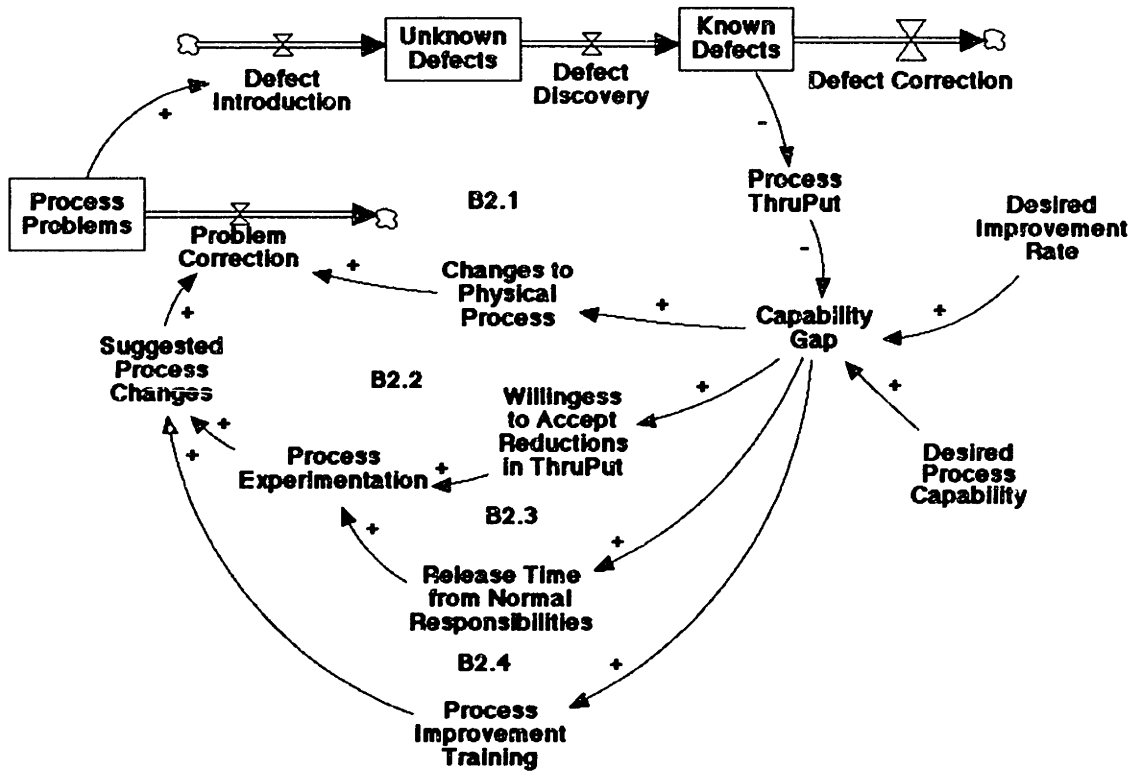
**Figure 3**  
**The Reinforcing Nature of Successful Improvement**



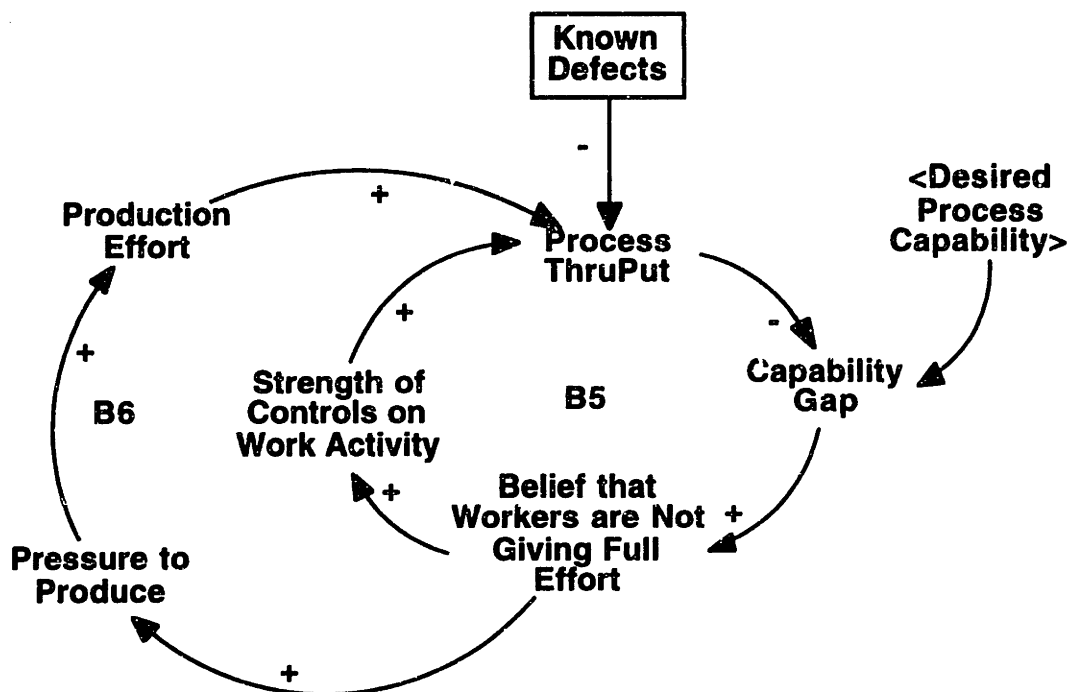
**Figure 4**  
**Managers Set Objects and Define Defects**



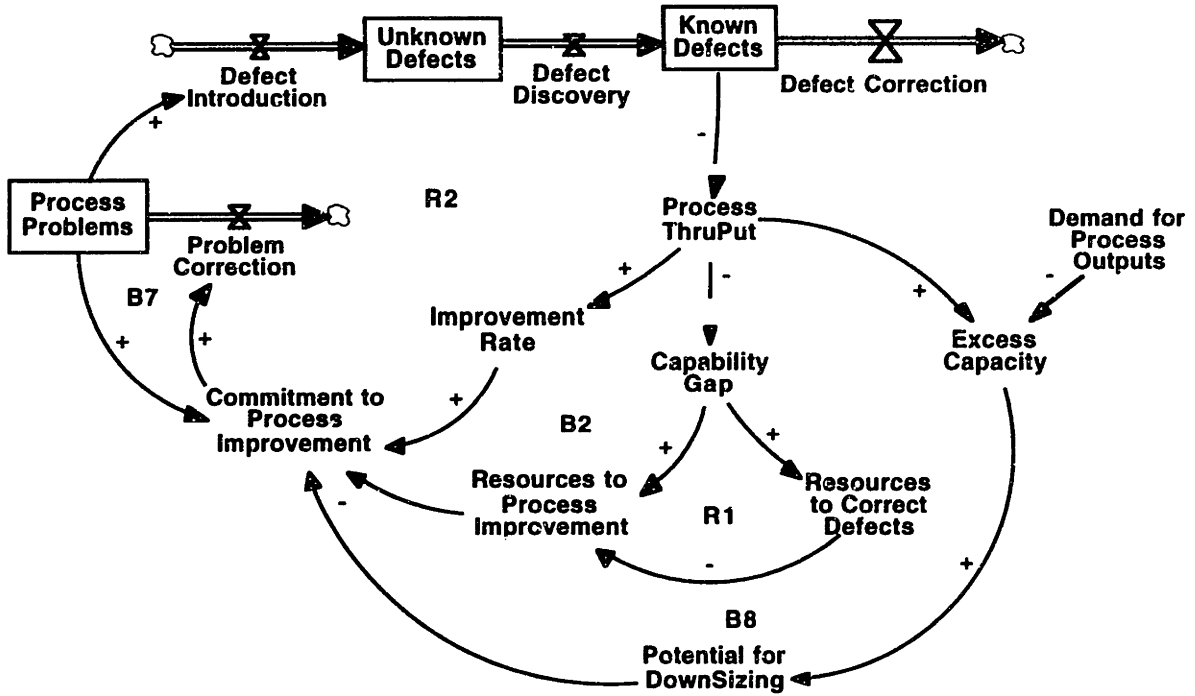
**Figure 5**  
**Managers Facilitate Improvement**



**Figure 6**  
**Managers Control Activities and Apply Production Pressure**



**Figure 7**  
**Worker Commitment to Improvement**



**Figure 8**  
**Workers Reaction to Conflicting Goals**

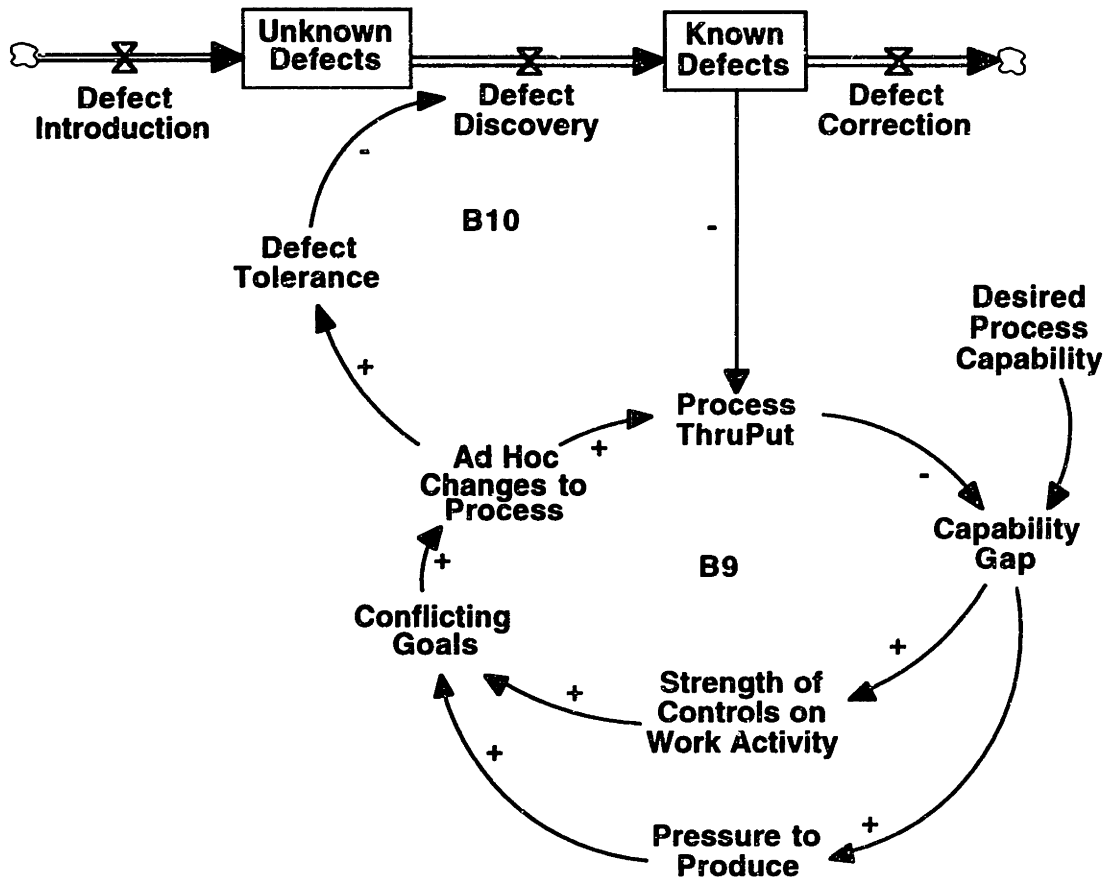


Figure 9

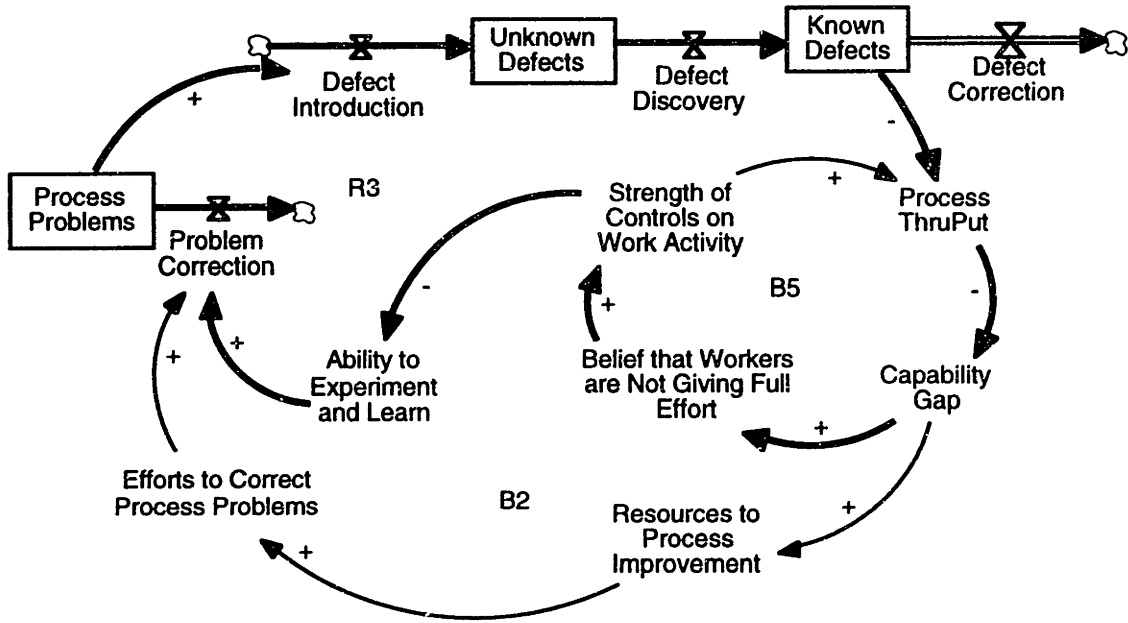


Figure 10

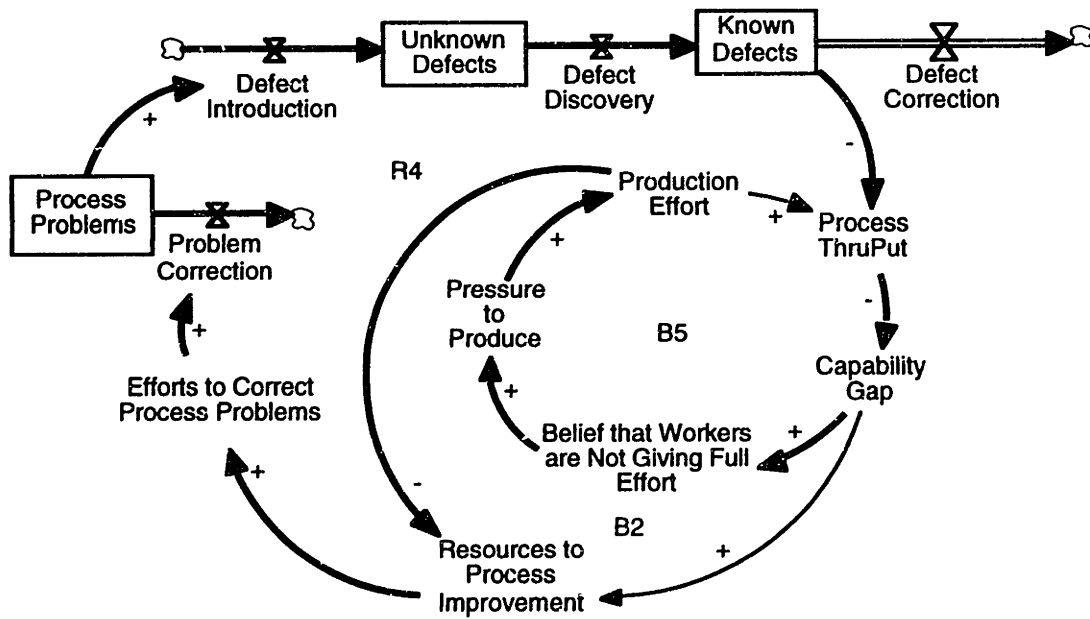


Figure 11

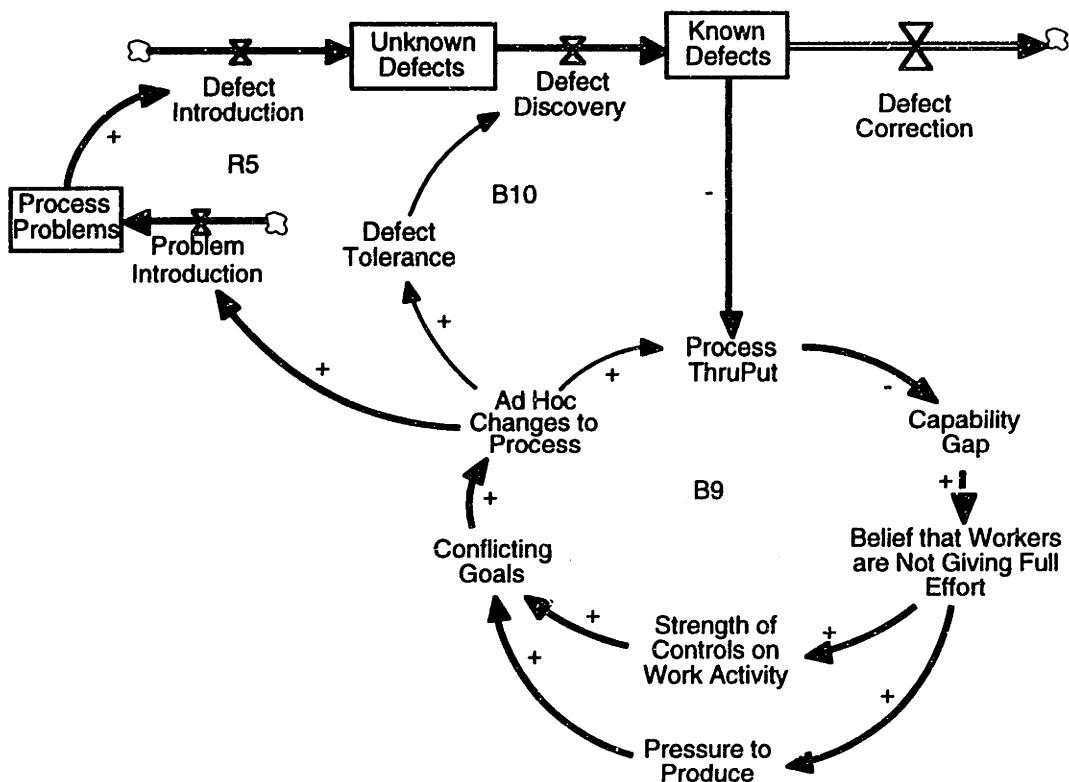


Figure 12

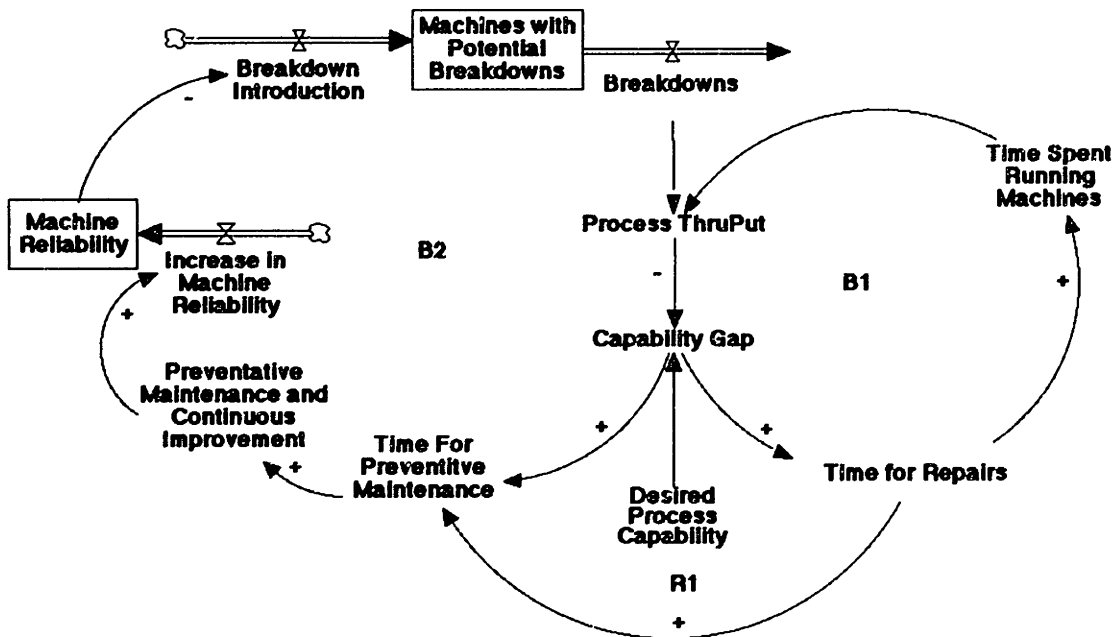




Figure 13

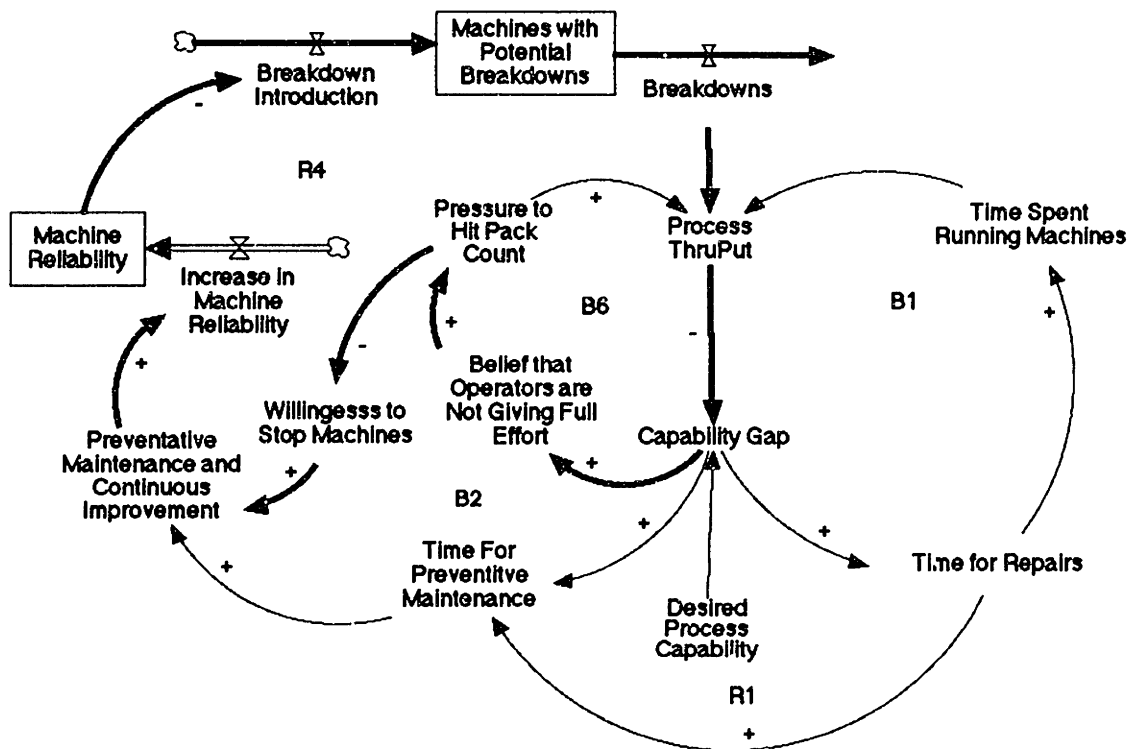


Figure 14

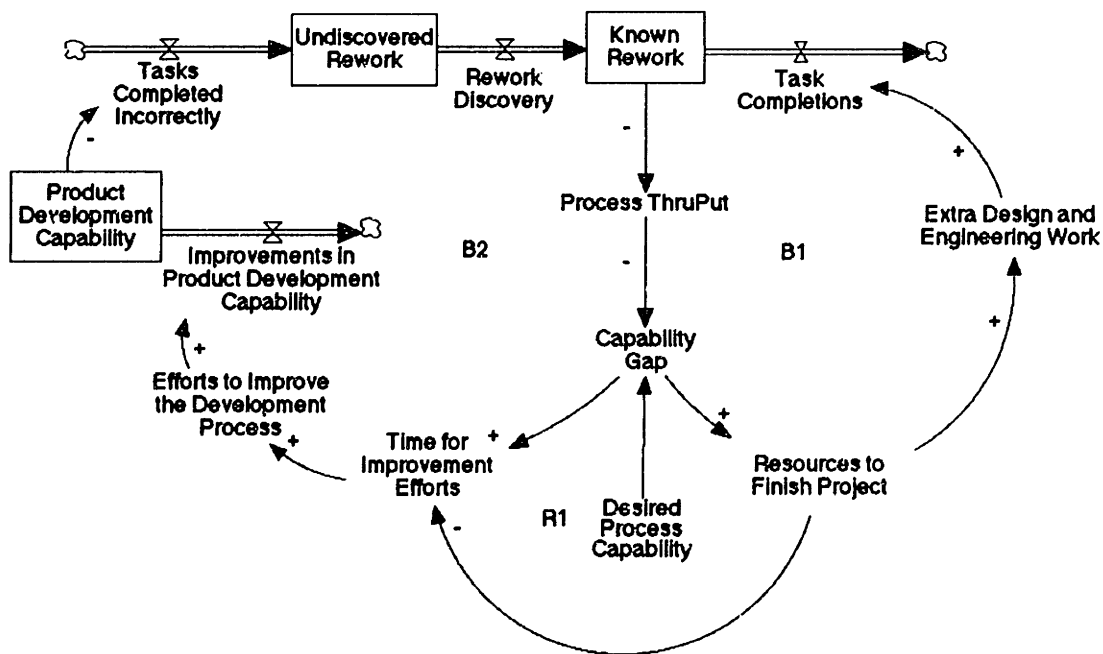


Figure 15

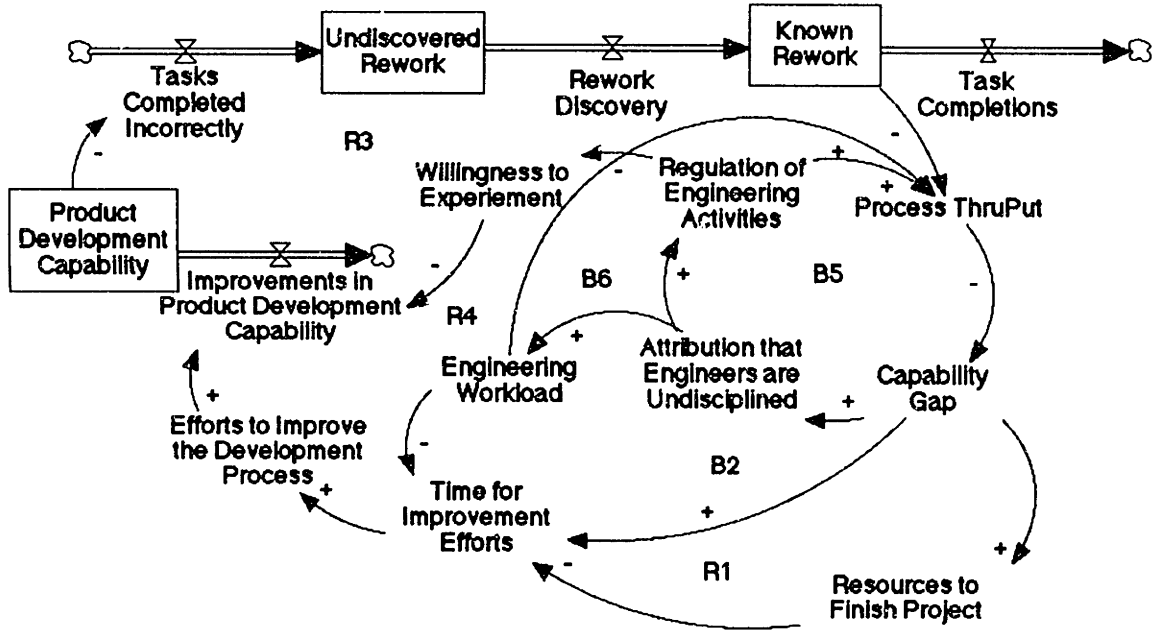
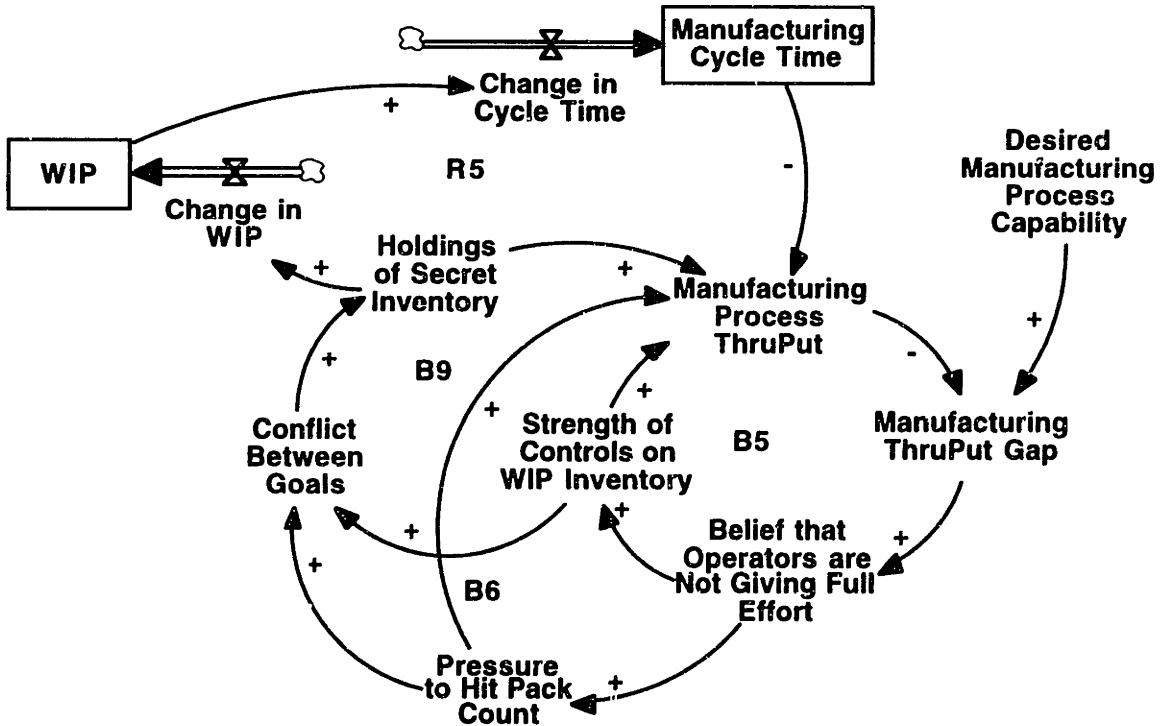


Figure 16



**Table 1**

<u>Loop Number</u>	<u>Process</u>	<u>Actors/Actions</u>	<u>Theoretical/Empirical Support</u>
B1	First Order Improvement. Closes the gap between desired and actual process through-put by correcting defects	Managers allocate resources to correction efforts	
B2	Second Order Improvement Closes the gap between desired and actual process through-put by eliminating the causes of defects.	Managers allocate resources to prevention efforts	TQM (Deming 1986)
B2.1	Physical Changes	Managers change the process technology.	Re-engineering (Hammer and Champy 1993)
B2.2	Allow Experimentation	Managers reduce through-put goals to allow process participants to experiment with new techniques	TQM (Deming 1986) & Theory of Improvisation (Weick 1992)
B2.3	Release Time	Managers reduce normal work responsibilities to allow workers time for experimentation and learning	organizational slack, (Cyert and March 1992)
B2.4	Training	Managers Allocate Resources to Training	Shiba <i>et al.</i> 1993
B3	Aspiration Adaptation.	Managers adjust goals based on past experience with process	Eroding Goals (Forrester 1968, 1969; Cyert and March 1992; Lant 1992)
B4	Defect Reinterpretation	Managers redefine what constitutes a defect based on past experience	Structuration: (Orlikowski 1992) Serman 1994
B5	Strength of Control	Managers attribute gap to worker slack and strengthen their control of process to induce extra production effort	Fundamental Attribution Error: Plous 1992; Carroll <i>et al.</i> forthcoming;
B6	Production Pressure	Managers attribute gap to worker slack increase production pressure to induce extra production effort	Fundamental Attribution Error: Plous 1992; Carroll <i>et al.</i> forthcoming;
B7	Diminishing Returns	As the stock of process problems is depleted, commitment to improvement declines	Schneiderman 1988, Serman <i>et al.</i> 1994
B8	Potential Down-sizing	Improvement can create excess capacity, which can lead to lay-offs and reduce commitment to improvement	Serman <i>et al.</i> 1994, Repenning 1996c

B9	Ad Hoc Process Changes	Workers react to conflict between controls and production pressure by making ad hoc process changes to close the through-put gap.	Structuration (Orlikowski 1992, Wynne 1988)
B10	Manipulate the Metrics	Workers make ad hoc changes to the measurement system to reduce through-put gap.	Structuration (Orlikowski 1992, Wynne 1988)
R1	The Reinforcing Nature of Improvement	As defects decline, fewer resources are needed for correction. More resources can be focused on prevention, further reducing the level of defect.	'Quality is Free'
R2	Commitment to Improvement	An increase in commitment increases the rate of improvement, further increasing commitment.	'Successful Change Begins with Results' Schaffer and Thomson (1992) Serman <i>et al.</i> 1994

**Table Two**  
Key Differences

	<u>MCT</u>	<u>PDP</u>
physical process	<ul style="list-style-type: none"> <li>- fast cycle times (days)</li> <li>- contained within one organization</li> </ul>	<ul style="list-style-type: none"> <li>- slow cycle times (years)</li> <li>- cross multiple organizations</li> </ul>
kick-off	<ul style="list-style-type: none"> <li>- low key</li> <li>- challenged mental models</li> <li>- new leadership with low credibility</li> </ul>	<ul style="list-style-type: none"> <li>- high profile</li> <li>- confirmed mental model</li> <li>- established leadership with high credibility</li> </ul>
design mode	<ul style="list-style-type: none"> <li>- emergent and flexible</li> <li>- focus on continuous improvement</li> <li>- emphasis on experimentation</li> <li>- selected a wide array of possible techniques</li> <li>- early focus on process structure</li> <li>- later focus on process management</li> </ul>	<ul style="list-style-type: none"> <li>- pre-planned, rigid</li> <li>- focus on creating ideal process</li> <li>- no experimentation</li> <li>- simultaneous changes in process structure and process management</li> </ul>
measurements and objectives	<ul style="list-style-type: none"> <li>- process participants defined own metrics</li> <li>- metrics evolved</li> </ul>	<ul style="list-style-type: none"> <li>- metrics by committee</li> <li>- metrics disappeared</li> </ul>
promotion	<ul style="list-style-type: none"> <li>- personal</li> <li>- word-of-mouth</li> <li>- "...there was never an MCT newsletter."</li> </ul>	<ul style="list-style-type: none"> <li>- impersonal</li> <li>- print, audio, video</li> <li>- heavy investment</li> <li>- conflict with informal communication</li> </ul>
training	<ul style="list-style-type: none"> <li>- recipes</li> <li>- hands-on with simulators and real experiments</li> <li>- "how the system works"</li> <li>- provided to almost every operator and material handler in the division</li> </ul>	<ul style="list-style-type: none"> <li>- blueprints</li> <li>- class room with overheads and books</li> <li>- "how things should be done"</li> <li>- received by a small portion of the total engineering staff</li> </ul>