Green, Simple, and Profitable: The Paradox of Failed Best Practices in University Building Maintenance

John L. Lyneis¹
John D. Sterman
MIT Sloan School of Management

Many green practices are widely understood and known to bring benefits beyond reduced energy use. Yet, organizations often fail to implement them. What explains these failures? Past theory suggests that adoption and implementation will be most likely to fail when practices are difficult to recognize given current competencies or organizational structures, require complex knowledge, or when the organization faces short term pressures that force it to abandon implementation early. Here, we present a case study of an organization that fails to adopt an important best practice despite the fact that the benefits and steps toward implementation are well understood and external short term pressures are minimal. We find that instead, short term pressures are created entirely internally by the structure of relations across organizational boundaries, causing individuals to misperceive the best practice as a cost that can be put off rather than an investment with positive future returns. Thus, even the simplest of innovations and improvements can be stymied by dynamics internal to an organization.

Keywords: Best Practice Implementation, Organizational Change, Capabilities, Organizational Boundaries, System Dynamics, Green Management

One of the central contributions that management research can make towards reducing the threat of global warming is to provide an understanding for when and how organizations adopt green best practices. Organizations are responsible for a large share of worldwide carbon dioxide emissions, and therefore any steps that organizations take to reduce emissions would go a long way towards putting our world on a path to sustainability. While some green practices may come at the expense of an organization’s financial or other interests, others are highly cost effective, widely beneficial, practiced in many places, and yet still are not implemented in the vast majority of organizations. What explains the failure of so many organizations to adopt these practices?

Organizational theory offers two main explanations for the failure of organizations to adopt and implement the best practices of others. First, failed adoption is often attributed to the difficulty that organizations have recognizing or interpreting practices. For example, practices may be ambiguous (Strang & Still, 2006), rooted in organizational cultures that are inherently hard to interpret (Barney, 1986), composed of information that crosses typical communication patterns and organizational boundaries (Henderson & Clark, 1990), adopted for legitimacy rather than technical reasons (Westphal et al., 1997) or may take on new and less effective forms when interpreted (Zbaracki, 1998).

¹ 30 Wadsworth St. E53-358, Cambridge, MA 02142, jlyneis@mit.edu
Second, implementation may stall if short term pressure is high and the benefits of a new practice are delayed. March (1991) points out the inherent tension that organizations face between maximizing returns today and sacrificing some fraction of those returns for investment in the future. More specifically, organizations may initiate efforts to learn and grow capabilities, but when delays between investment and return are long, these efforts will be easily abandoned in response to short term pressures, such as the need to maintain competitive position, satisfy customers today, or meet quarterly numbers (Repenning & Sterman, 2002). Thus, even when a practice is understood technically, implementation may stall.

Here, we test and extend existing theory through a case study of failed adoption of an important green best practice at a large, private research university. The practice that we choose is proactive building maintenance, which is the practice of fixing building equipment before rather than after breakdowns. Across a range of industries it is well known that proactive maintenance can lead to substantial savings and increases in reliability (cf. Carroll, Sterman & Marcus, 1997). In addition, proactive maintenance can bring important environmental benefits, as illustrated in Figure 1. The graph shows steam consumption in two different buildings at our field site over a 21 day period, plotted along with the outside air temperature. In Building B, the steam traps were replaced proactively, while in building A the old traps remain. (The purpose of a steam trap is to remove condensate from the system such that steam can flow efficiently to its destination.) In Building A, steam use is constant, whereas in Building B steam use responds to changes in outside temperature, as one would expect in an efficient heating system. Thus, the simple task of replacing the steam traps proactively can have a large impact on energy use and emissions. Maintenance to other aspects of building heating and cooling systems can be expected to have a similar effect.

Yet, despite the potential environmental benefit and cost savings, the university that we study – like many others – has failed to develop the capability to maintain its buildings proactively. Instead, maintenance work consists almost entirely of responses to expensive
emergencies and customer requests. The failure of this university to adopt proactive maintenance is surprising for two reasons. First, the idea behind proactive maintenance is extremely simple and widely known, suggesting that difficulties recognizing and interpreting the practice do not explain its absence. Second, unlike for-profit entities, universities are immune from many of the short term pressures that can cause managers to abandon improvement, as in Repenning & Sterman’s (2002) theory. Universities do face considerable financial pressure, but endowments are typically managed for the long term, and universities can uniquely afford to make investments that have a relatively certain but delayed return. Building maintenance is exactly such an investment: clear best practices are known that can lead to significant cost savings in the long run.

Here, we expand the theory of the capability trap and argue that failed adoption of effective best practices is possible even when practices are simple and widely understood, and even when organizations should have ample resources to invest. Interestingly, while a university should be immune from the short term pressures that stymie improvement elsewhere, for the maintenance department that we study, such pressures are in fact substantial. Over a period of many years, limited budgets and demanding customers have shifted resources away from proactive work, causing more breakdowns and still fewer hours on proactive work – a classic example of Repenning & Sterman’s (2002) capability trap. However, rather than being externally created, these pressures are a direct result of the internal organizational structure, and in particular, the structure of relations across organizational boundaries (Carlile, 2002; Galbraith, 1970). Budget pressures are a result of decisions made by senior university leadership, and in turn decisions to cut budgets are reinforced by the success that maintenance has had operating under conditions of scarcity. Thus, despite the fact that proactive maintenance is a known best practice and resources exist that might support an investment, implementation is stymied.

In the sections that follow, we review existing theory on the diffusion and implementation of best practices, present the background for our case, and provide evidence for the importance of the internal structure of organizations for diffusion. For the latter, we make use of both qualitative evidence and a simulation model, calibrated to match the data from our case. The model confirms the benefits of proactive maintenance, and also validates how actors constrained by the existing organizational structure and by poor communication across organizational boundaries may fail to appreciate these benefits.

**THEORY – THE DIFFUSION OF BEST PRACTICES**

Why don’t organizations always successfully adopt and implement the best practices of others? Existing theory provides a number of possible answers to this question. At the level of the organization, theory emphasizes the inherent difficulty of organizational learning and change. According to March (1991), organizations naturally favor adaptations that have a quick and certain return over those with returns that are more distant or ambiguous. As a result, when the benefits of implementing an industry best practice are delayed, organizations facing throughput pressure may easily learn the wrong lesson and abandon change efforts (Repenning & Sterman, 2002). Organizations may also fall victim to “superstitious learning” (Levitt & March, 1988), in which initial successes are attributed to the wrong cause, thereby reinforcing behaviors that are in fact detrimental to improvement.
Diffusion of best practices may also be hindered by a failure to recognize the content or significance of new ideas. For example, Henderson & Clark (1990) argue that new practices may not be recognized when they are architectural in nature. Established functions and communication patterns within organizations cause individuals and groups to focus on certain kinds of information, leaving them less able to incorporate information that does not fit existing categories. In a similar vein, Cohen & Levinthal (1990) define an organization’s “absorptive capacity” as its ability to recognize and incorporate new information. Organizations must develop absorptive capacity over time, and if such capacity is lacking organizations may fail to successfully imitate best practices of competitors.

Even after an innovation or best practice is recognized and implemented, scholars have noted the difficulty inherent in maintaining both commitment and expertise in the execution of complex tasks. Maintaining expertise may require an ability to reflect and learn from events—an ability that Weick (1987) describes as essential to “high reliability” organizations. Crucially, learning must reflect not only the immediate causes of events, but also the social and cultural systems underlying organizational outcomes (Argyris, Putnam & Smith, 1985; Carroll, Rudolph & Hatakenaka, 2002). Such learning is inherently difficult and is often not practiced (Carroll, Rudolph & Hatakenaka, 2002).

The adoption of best practices has also been considered in the context of diffusion of ideas and practices among organizations. According to institutional theory (DiMaggio & Powell, 1983), once practices become established, organizations will adopt them out of a desire to appear legitimate. At the same time, institutional theory provides a number of possible explanations for why organizations may not successfully adopt the best practices of others. First, if practices are adopted to satisfy legitimacy rather than technical needs, organizations may not sufficiently customize practices, thereby contributing to implementation difficulties (Westphal et al., 1997). For example, Westphal et al (1997) show that late adopters of Total Quality Management among hospitals mimic peers rather than customize the innovation for their own needs. Thus, in settings where legitimacy is the main goal of organizational action, complex innovations may ultimately lack effectiveness.

A final subset of organizational theory that addresses the diffusion of practices and ideas across organizations is the literature on networks and organizations. A look at how organizations are connected provides one final explanation for why best practices might not always be adopted. In short, if an organization is not well connected with peers, the organization may face difficulty learning about and implementing new ideas. Simple ideas spread easily across both weak and strong ties, but the sharing of more complex practices both within and among organizations requires closer, more embedded connections (Centola & Macy, 2007; Hansen, 1997). Along these lines, Zuckerman & Sgourev (2006) argue that parallel peer networks among capitalists play a large role in both the sharing of best practices and in the creation of incentives (through peer pressure) to implement them. Even then, the influence of peers may be less effective when the innovation is ambiguous (Strang & Still, 2006). If implementation of an innovation requires complex knowledge, the quality of an organization’s connections to peers may influence both the chances for initial adoption and the success of any implementation effort.
How far does existing theory go towards explaining the paradox of university building maintenance? Despite the existence of a range of approaches that cross different levels of analysis and include several traditions within organizational theory, the failure of universities to adopt best practices in building maintenance remains a puzzle. As discussed, universities can afford to invest for the future and should be largely immune from the short term pressures that often stall implementation and meaningful organizational learning in for-profit entities (Repenning & Sterman, 2002). But other theories also fall short. Here, we consider the practice of proactive maintenance, a practice that is commonly understood by managers and mechanics to be far superior to the alternative of reactive maintenance. A number of maintenance best practice publications place significant emphasis on the importance of proactive work (for an example, see Levitt, 1997). As a result, failure to develop this practice cannot be attributed to a failure to recognize either its existence or its value. Furthermore, the idea behind proactive maintenance is not complex. For the most part, fixing or replacing a machine before it breaks requires the same set of skills that are required to fix a machine after it breaks. Mechanics are generally well trained to perform either task. Thus, failure to perform proactive maintenance cannot be attributed to a failed transfer of complex knowledge, or to a general inability to adapt or develop new competencies. Also, in the absence of direct competition, universities are willing to provide information for benchmarking studies, and peer networks are active (for example, the International Facility Management Association provides networking opportunities and publishes a highly regarded benchmarking study). Thus, proactive maintenance has many of the characteristics of a practice that one would expect to diffuse widely: its benefits are clear and widely known, and its implementation is unambiguous. Furthermore, universities, free of short term or competitive pressures and with endowments to invest, should be most able to act.

Instead, we find dynamics very similar to those that exist when short term pressures are high. For example, Carroll, Sterman and Marcus (1998) illustrate the challenges of implementation in field studies of maintenance practices at an oil refinery and a chemical plant. For-profit maintenance organizations face strong tendencies to avoid proactive work in the short run, including the need to cut costs and increase uptime so that output and profit are maximized. Avoiding proactive work, however, causes defects to accumulate, creating more breakdowns and even less time for proactive work in the future. This vicious cycle is a perfect example of the “capability trap” described by Repenning & Sterman (2002).

We argue that the picture painted by Carroll et al. and by Repenning & Sterman (2002) is only part of the story. The capability trap dynamics are exactly the dynamics that can so easily hinder organizational improvement when the benefits of improvement activities are delayed. However, we argue that Repenning & Sterman’s (2002) theory is even more general than originally described, due to the fact that pressures needed to create the trap can be created entirely internally, and thus can arise in almost any organization. Moreover, the capability trap can stymie improvement even when the practice or improvement initiative is decidedly non-complex. As a result, theories of the diffusion of best practices should pay more attention to the role of intra-organizational relations, and resist treating each organization as a single node subject to environmental influence. We next develop our theory in more detail, beginning with a presentation of research methods and the background to our case.
METHODS

Research Setting

To investigate the question of failed best practice implementation, we employ a mixed methods field study of maintenance practices at a single large research university, named Alpha for the purposes of this paper. We strategically choose a single extreme case that challenges existing theory (Flyvbjerg, 2006), following the practice of theoretical sampling (Glaser & Strauss, 1967). For all of the reasons discussed above, proactive maintenance in the university context is a very good “most likely” case (Flyvbjerg, 2006): that is, if best practice implementation is to be successful anywhere, we should especially expect it to be successful here. The benefits of proactive maintenance are well understood, and the university context is one where resources and commitment should exist to support implementation. The fact that proactive maintenance is not practiced in the university that we study provides a unique opportunity for inductive theory generation, following the tradition of management case study research (Eisenhardt, 1989).

Several different sources of data were collected. First, semi-structured interviews were conducted with 30 different individuals over a period of 9 months, beginning in the summer of 2007. Interviews lasted between 45 and 120 minutes and were recorded. Individuals were asked to describe their history with the organization, their daily activities, and their opinions regarding the state of the campus and of the maintenance organization. Interviewees represented the broadest possible range of backgrounds and positions within the organization. Senior managers were interviewed along with hourly mechanics, engineers, departmental facilities liaisons, and analysts working on broader issues of facilities planning and strategy. Several interviewees were relatively new to the organization and were able to offer a comparative perspective on current operations, while others had been with the organization for upwards of 30 years. Most reported to a director of facilities through the same hierarchy, but others (the departmental liaisons) worked directly for academic departments and thus could offer a different perspective.

A substantial amount of quantitative data was also collected. Financial data on maintenance spending over the 8 year period from 2000-2008 was collected, along with detailed operational data on maintenance work orders from 2005 until the present. The latter data set includes a full listing of all work orders opened and closed each week, classified by priority, and including the number of labor hours required to complete each order. The fact that orders are classified by priority is especially helpful because it allows us to calculate exactly how many hours are spent each week on emergencies, routine work, and proactive work over a period of three years. Thus, we can directly observe adherence to the best practice of interest: namely, the fraction or amount of time spent on proactive as opposed to reactive maintenance.

Data Analysis

Both the qualitative and quantitative data are included in the analysis. To understand what is unique about the case of university building maintenance, we first employ qualitative research methods to gain an in-depth understanding of the phenomenon. Qualitative data were analyzed using standard methods for inductive fieldwork (eg., Miles & Huberman, 1984). Specifically, interviews were transcribed and then coded in search of common categories or
themes (Glaser & Strauss, 1967). Next, we incorporate themes into a dynamic theory following the method of causal loop diagramming, common in the field of System Dynamics (eg., Sterman, 2000) but also increasingly used in management research (eg., Perlow et al, 2002; Rudolph & Repenning, 2002). The resulting causal loop diagram provides a dynamic theory of adherence to proactive maintenance over time.

We next transform the causal loop diagram into a simulation model using the System Dynamics methodology (Sterman, 2000), and calibrate the model using quantitative data. The purpose of a calibrated System Dynamics model is to test the internal consistency of our theory (Davis et al, 2007; Sastry, 1997). Specifically, analysis of a simulation model shows why maintenance remains predominantly reactive despite the fact that maintenance managers know the benefits of proactive maintenance and understand the specific steps needed to implement it. In addition, a calibrated simulation model allows us to demonstrate why the culture of cost cutting that has emerged is so detrimental to future investments in proactive maintenance.

**CASE BACKGROUND – BUILDING MAINTENANCE AT ALPHA UNIVERSITY**

We start by providing a brief overview of the maintenance organization at Alpha University. Alpha is a research university with a large and diverse campus spanning many acres. As in many organizations, the maintenance function is part of a larger facilities group that is also responsible for renovations and new construction, custodial work, utilities, security, and other activities necessary for the general functioning of the university. Within this overall mission, the maintenance group is responsible specifically for keeping building and mechanical systems running. The group has approximately 100 employees organized into several trade groups including plumbers, electricians, HVAC mechanics, carpenters, general mechanics, and others. General maintenance groups are organized by geographic region of the campus and are supported by centralized groups of specialists. Each group has a supervisor who is responsible for assigning and monitoring work on a daily basis. The department also employs at least one engineer who advises tradesmen on technical matters and several senior managers who undertake longer range planning.

The daily activities of the maintenance department are heavily organized around an SAP work order system that has been in place since 2003. All work performed by mechanics is charged to a work order that is opened when a problem is reported, and closed when the work is completed. In addition, a number of proactive or “preventive” work orders are opened according to a pre-set schedule. Thus, operations are very much driven by the hundreds of customer calls that are received each week. Supervisors look at the list of open work orders in their area each morning and assign work so as to complete work by a target completion date when possible. In addition, work is also managed via the close relationships that supervisors have with representatives from each academic department. All told, supervisors thus have an extremely demanding job: in addition to managing their own backlog and monitoring their own workforce, they must stay in communication with customers regarding the status of work.

As we note above, one would expect Alpha to be a place where best practices in maintenance are followed. The benefits of performing maintenance in a planned, proactive manner are substantial, and include both lower operating costs and higher reliability. Yet,
maintenance at Alpha is around 90% reactive, indicating huge inefficiencies. Figure 2 shows the fraction of work that is proactive vs. reactive over a 2 ½ year period beginning in late 2005. The emphasis on reactive work is also clear from the above description of operations. Customer calls drive the vast majority of work, and keeping customers satisfied is seen as a major priority. One individual described the situation as follows:

“You know, we’re a customer service organization. It’s almost like we’re afraid to commit completely to the behind the scenes stuff, because we want to get to the visible stuff so quickly. That’s not spoken, but I think that’s – having the resources available – the customer doesn’t care if a belt is flapping in a fan. It might not matter for a year down the road, but to us it might be in January in the middle of the night that the fan shuts down – to us it’s important, but to the customer it’s not, so our resources go to what the customer wants, for the most part.” – Facilities employee

The above quote illustrates precisely why reactive work is so inefficient: by neglecting “behind the scenes stuff” today, the organization is forced to instead address the same issues later as emergencies, when repairs are likely to be far more costly. The speaker is clearly emphasizing these inefficiencies when he speaks of a breakdown occurring in the middle of the night in January – a time when overtime and a host of other charges might be incurred. A second interviewee offers a very similar assessment of current operations:

“It’s a fire drill… it’s who’s screaming right now. So your priorities change on an hourly basis, probably a half-hourly basis during the day, and it’s kind of – it’s basically a constant fire drill. [So you] kind of have a tendency to leave it once you get to a point where no one is complaining.” – Facilities employee

Firefighting is a major theme that emerges from all of our interviews. Workers describe having to often leave work to attend to emergencies, and supervisors speak of the constant need to respond to customer demands and the steady stream of calls that require attention.

Thus, a major theme from our qualitative field work is the large distance that exists between maintenance at Alpha and industry best practice. What explains this gap? Carroll et al. (1998) attribute the poor maintenance performance in the plants that they study to limitations in the mental models of employees throughout the organization. However, one of the more
puzzling findings in our interviews is just how much knowledge employees have regarding proactive maintenance, and how seriously they take the small amount of proactive maintenance that is completed. Over the past five years, beginning with the change in maintenance management in 2003, the department has slowly increased its emphasis on proactive maintenance. Figure 2 indeed shows a slight increase in the fraction of labor hours spent on proactive work over the past two plus years. “Preventive maintenance” work orders are now created according to a regular schedule, and employees are expected to complete preventive work along with reactive work. In contrast, previously preventive work orders were printed from a separate database on different color paper, and were placed at the bottom of the pile, or often were not completed at all. Today, preventive maintenance is given a target completion date along with all other work, and is especially emphasized for the equipment most essential to the university’s research mission, for example in science laboratories. Furthermore, not only is the current preventive maintenance regime well adhered to, but most employees that we talked with could clearly articulate the theory behind proactive work:

“If you do more predictive and preventive maintenance, you do less unplanned maintenance, and you have less breakdowns. If you do that, then you’re doing your work in the most efficient and effective way you can. I know that… [but] it’s just a matter of making that happen, of turning that corner.” - Facilities employee

As this quote suggests, employees also understand that the current level of preventive maintenance is not adequate. The current regime includes brief checks of equipment and replacing of belts and filters; most acknowledge that the frequency of checks could be increased and some equipment replaced entirely. Along these lines, the department keeps a wish list of desired proactive capital replacements in case resources ever become available. Thus, implementing proactive maintenance, at least initially, would not pose much of a challenge: managers can list specific areas where resources would be used and can articulate the benefits that proactive maintenance would bring.

Why then, aren’t these steps taken? Two additional themes dominate our interviews and provide a partial answer to this question: the poor condition of campus buildings, and the limited resources available to the maintenance department. Maintenance professionals describe a campus where the size of the deferred maintenance backlog is substantial. In fact, an independent engineering firm recently assessed this backlog and documented potential improvements that dramatically exceed the annual budget of the maintenance department. Over a period of many years, mechanical systems survived renovations and had their lives extended due to the skilled work of mechanics, leaving mechanics today with an even more difficult job. Given such a large stock of latent defects and a lack of resources to fix identified problems proactively, it is not surprising that the department faces a high volume of breakdowns. In fact, breakdowns are sometimes the only way that equipment gets fixed, as one employee noted somewhat sarcastically: “If it breaks, then you’re lucky. Then, you have to have the money.” Lately the pressure has only increased due to increases in the price of materials and in the size of the campus served by the maintenance department.

In sum, maintenance at Alpha is far below industry best practice. Work is mainly reactive, creating a tremendous opportunity for improvement, as many acknowledge. To get a
better sense of these savings and to test our contention that managers are currently constrained, we next develop a calibrated simulation model of the maintenance organization at Alpha.

A MODEL OF BUILDING MAINTENANCE

Calibrating a Simple Model

We start by formulating and calibrating a detailed System Dynamics (Sterman, 2000) model to match the quantitative time series data on proactive vs. reactive maintenance. Building maintenance is best represented as a labor capacitated process (Sterman, 2000), and we form the core of the model by adapting the structure developed by Sterman (2000) and Oliva & Sterman (2001), shown below in Figure 3. The causal loop diagramming convention is used to provide an overview of the main relationships; these diagrams capture most but not all of the complete model equations.

Work orders are opened and accumulated in a backlog until they are closed. Two kinds of work orders are considered: reactive maintenance work orders and proactive (planned) maintenance work orders. The rate of closed work orders, is governed by two key feedback loops: “work faster” and “work longer.” As the work order backlog rises, the desired completion rate increases, creating more work pressure given a fixed capacity (represented as the “standard completion rate”). Work pressure, in turn causes productivity and hours worked to increase, increasing the completion rate and thus decreasing the backlog, all else equal.

Given a backlog of both types of work orders, the desired completion rate for each type is determined by:

\[
\text{Desired Completion Rate} = \frac{\text{Backlog}}{\text{Desired Time to Complete Work Orders}} \quad (1)
\]
The desired time to complete work orders is a management policy, set to be an average of 2 weeks for reactive work and 4 weeks for planned work. (A fraction of emergency work orders are completed much more quickly; however, for a simulation with a time horizon of several years it is appropriate to aggregate all reactive work.) Given a desired completion rate for each type of work, capacity is allocated between the two types according to a logit choice model (McFadden, 1974), where the variable $A$ is the “Attractiveness by Priority” variable shown in Figure 3:

\[
\text{Fraction of Work}_i = A_i / \sum_j A_j
\]  

\[
A_i = \exp(\alpha_i \times \text{Desired Completion Rate}_i)
\]  

Next, we estimate the relationship between work pressure and productivity and the relationship between work pressure and total labor hours worked from the data. Work pressure is defined as:

\[
\text{Work Pressure} = \sum \text{Desired Completion Rate}_i / \text{Standard Completion Rate}
\]  

Here, the “Standard Completion Rate” takes into account staff levels, the average work week, and the average productivity of workers over the 2 ½ year time period where data exists. The relationship between work pressure and productivity is given by a nonlinear increasing function whose slope around an operating point is determined by estimating the parameter $s$ from (5) using log linear regression:

\[
\text{Productivity} = \text{Work Pressure}^s \times \text{Nominal Productivity}
\]  

Nominal productivity is again the average productivity over the 2 ½ year period. A similar procedure is used to formulate the relationship between work pressure and total labor hours. Results give a statistically significant relationship between work pressure and productivity ($s=.14$, t value = 2.71), but not between work pressure and the total hours worked per week. The latter result is consistent with both qualitative and quantitative data regarding the use of overtime; overtime is rare, and is more often used for scheduled shutdowns than to catch up on work.

Starting from the above parameterization, we next determine values for alpha in the logit formulation by calibrating the model against time series data for both backlog and the rate of closed work orders, using the Vensim modeling software. The rate of new work orders is taken as exogenous. The resulting model gives a good fit to the data, as illustrated in Figure 4.

A calibrated model of work order backlog and completion confirms several themes from our qualitative data. First, the maintenance organization is indeed overstretched. The volume of incoming reactive work is enough to consume the vast majority of the organization’s capacity, assuming a very realistic average completion goal of two weeks. We also see why pressure from customers is so significant; an average completion time of two weeks - while acceptable for much of the department’s work - does leave room for improvement. Second, the department does make an effort to complete some essential planned and preventive work. On average 216
proactive work orders are opened each week, and for the most part these are closed to meet a completion goal of 30 days. The majority of work is reactive, but decision rules do favor placing some emphasis on preventive work ahead of reducing the completion time on reactive work still further. Thus, the model confirms a major paradox of our case: that Alpha remains predominately reactive despite widespread understanding among maintenance professionals of the benefits of proactive maintenance.

FIGURE 4
Results of Model Calibration

Expanding the Model Boundary

While the above service delivery model provides some insight into the maintenance operation, to understand why reactive work is so dominant we must also consider the source of reactive work orders. Figure 5 shows an expanded model where the rate of new work orders is now endogenous. We expand this model by drawing on existing literature (Repenning & Sterman, 2002; Sterman, 2000) and themes from our qualitative research.

As we note, a major theme from our interviews is the amount of deferred maintenance that exists. Interviewees describe a long list of equipment that is running past the recommendation of manufacturers and that will soon require repair or replacement. We represent deferred maintenance in our model as a stock of defects that rises and falls over time, analogous to the backlog of work orders. The stock of defects rises through defect creation, and falls through defect resolution. Some defects are known to mechanics, while others are hidden. Above all, defects are the source of reactive work orders: the higher the number of defects, the higher the breakdown rate, as indicated by the positive arrow in Figure 5 between the stock of defects and the rate of opened work orders.
Consistent with the descriptions of interviewees, defects can be resolved through one of two balancing feedback loops: proactive maintenance or reactive maintenance (shown in Figure 5). Defect resolution through proactive maintenance occurs when resources are available to fix defects before they become breakdowns. As the stock of defects grows, more defects will be identified, leading to more proactive repairs when resources are available. In practice, maintenance managers follow exactly such a process: a list of known problems is kept and updated regularly, and when resources are available items are addressed from this list. Defects that are not resolved through proactive maintenance must instead be resolved through reactive maintenance. A rise in the stock of defects causes the breakdown rate to increase, leading to more reactive work orders and more repairs, through the feedbacks in the service delivery model presented above.

**FIGURE 5**
Causal Diagram for the Full Simulation Model

The introduction of the planned and reactive maintenance balancing loops creates a third reinforcing feedback loop that is crucial to the resulting dynamics. As the stock of defects grows, breakdowns, work pressure and the time and money spent on repair also increase. Resources spent on repair, however reduce the amount of resources that are available for planned work, resulting in less planned defect elimination. The stock of defects continues to rise, forming a reinforcing loop.

Together, the expanded model provides important insight into how Alpha may have fallen into a state of such low performance, despite the best intentions of managers. Given the feedback structure of the system, the organization is highly vulnerable to Repenning & Sterman’s (2002) capability trap. If at any time resources become stretched, attention to proactive maintenance must decrease, leading to more defects, more reactive work, and still higher work pressure. Given Alpha’s current high stock of defects and low allocation towards proactive work and improvement, it is clear that the organization is stuck in exactly such a trap.
Maintenance managers, moreover, have little latitude for action once stuck in the capability trap. Faced with a stream of reactive work orders and a fixed budget, managers control only the detailed service delivery portion of the system introduced above.

ANALYSIS AND RESULTS

Organizational Boundaries and the Failure to Invest

When the larger feedback structure is considered, however, it is abundantly clear where investments might be made to improve the performance of the system. Figure 6 shows the behavior of the system when a substantial investment is made in proactive maintenance. The diagram shows two simulation runs: in the first there is no investment, and in the second an investment is made in the form of an increase in the maintenance budget and workforce. Decision rules of managers remain as developed above. Following a period of several years, capacity is allowed to readjust to only what is needed by the organization (both proactive and reactive).

Simulation results confirm the benefits of a large up front investment in proactive maintenance. The investment immediately increases the fraction of work that is proactive, but more importantly, the new emphasis on proactive work is sustained even as capacity falls. Feedback through the stock of defects is responsible for the improved performance: the initial investment begins to reduce the stock of defects immediately, freeing resources for still more investment in proactive work. Over a period of 10 years, the return is substantial in the form of far fewer defects (indicative of a better functioning campus) and lower annual maintenance costs. The up front investment pays a clear return, especially as the time horizon is extended.

Why then, hasn’t Alpha made an investment in proactive work? On the one hand, Repenning & Sterman’s (2002) “capability trap” theory could provide a partial explanation. Simulations confirm that the benefits of an investment in proactive work are highly delayed: as shown in Figure 6, managers do not realize net cost savings until as many as six years after the start of the investment – an eternity for managers facing yearly budget pressure. In contrast, if resources are shifted from proactive to reactive work, the organization will experience an immediate increase in the response time to customers - a desirable outcome given the importance of customer satisfaction to the organization. Meanwhile, the cost of a drop in proactive work will not be apparent until much later, when equipment that is not fixed proactively begins to break. Thus, managers may easily learn to avoid proactive work.

At the same time, the capability trap theory does not entirely explain the case of Alpha. First, as we note in the introduction, the theory assumes the existence of short term pressures that need not exist in the university context. Where do short term pressures originate if not from an external competitive environment? Second, Repenning & Sterman’s theory – like many other theories of organizational improvement – begins with the assumption that the organization has identified and at least attempted to implement an “improvement initiative.” Yet, in Alpha’s case, a focused and adequately financed effort to introduce proactive maintenance on a large scale is never even initiated. Interestingly, proactive maintenance is viewed by maintenance workers as a simple idea and never as an “innovation” or improvement “initiative.” Thus, the question is
not only why a novel idea might be abandoned once begun, but also why a simple and known best practice is never even considered for serious adoption and implementation.

**FIGURE 6**
**Investing in Proactive Maintenance**
(1= No Investment; 2= Investment in Proactive Maintenance)

<table>
<thead>
<tr>
<th>Year</th>
<th>Work Capacity</th>
<th>Defects</th>
</tr>
</thead>
<tbody>
<tr>
<td>2006</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2009</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2012</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2015</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2018</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

One of the major themes from our qualitative field notes is the importance of the relationship between the maintenance department and senior university administration. First, from the perspective of maintenance managers and mechanics, senior leadership is a major source of the tremendous work pressure that constrains operations. As noted above, maintenance employees are highly aware of the amount of deferred maintenance that exists, and support making the many needed repairs in order to prevent future breakdowns and emergencies. However, the department is constrained by a yearly budget that is barely enough to attend to the high volume of reactive work. Short term pressure leaves the organization vulnerable to the afore-mentioned capability trap: in order to keep customer satisfaction high and remain under budget, employees face a strong incentive to neglect proactive work. Equipment replacement and known defect elimination are routinely postponed because the money does not exist. Long tenured employees describe a “constant budget crunch… since back around the mid–eighties,” a sentiment that is borne out in the now substantial stock of deferred maintenance.

Faced with such long standing budget pressure, maintenance employees have developed a mindset of scarcity. Justification for limited budgets is generally not provided, and so individuals are left to speculate. According to one individual:

They don’t give us excuses, but someone could say “well, we just don’t have the money,” or they could say “well, we really don’t think you’re as efficient as you should be, and
therefore I’m reluctant to give you more money because I don’t think you’re using what you have effectively.” – Facilities Employee

Others see limited budgets as an inevitable consequence of the low status of maintenance:

“If there’s a [need] to cut, this is one of the first places to get cut… at the bottom tier. We’re a necessary evil. Nobody wants us, but we’ve got to be here.” – Facilities Employee

Either way, maintenance employees have come to accept that maintenance is not a top priority for the university.

The attributions that maintenance employees have formed regarding work and budget pressures greatly influence department actions and priorities. First, knowing that budgets will remain limited and suspecting that the administration may expect more out of them, managers emphasize efficiency and cost cutting as their most important accomplishments and future policy goals. Several interviewees described recent improvements in productivity that have resulted from increased supervision and accountability, better tracking of work orders, and more attention to how reactive work is planned. Cost savings have been generated through reduced use of overtime and more selective use of vendors. Several years ago, the department began tracking and publishing performance metrics, including work order backlog, labor hours per work order, and average number of days that work orders remain open – all metrics that encourage speed and efficiency before quality. Looking ahead, the department expects that further improvements in productivity will be needed.

Second, the customer service bias that we describe above may also be a response to feelings of inferiority generated by years of budget cuts. Ashforth & Kreiner (1999) theorize that members of occupations who perform “dirty work” (Hughes, 1951) will seek to improve their self-worth by “recalibrating” the work that they perform to over-emphasize those aspects of the job that are most desirable. To maintenance employees, “serving customers” by responding to emergencies is exactly the aspect of the job that is most desirable, because it is challenging, of immediate importance, and noticed by all. According to one individual “a lot of guys would rather do the troubleshooting than the [proactive work] - that’s just a natural thing.” Because budget pressure reinforces the low status of maintenance relative to other priorities, it’s not surprising that maintenance mechanics and managers respond by favoring work that yields immediate and noticeable value. Customer satisfaction requires responding to reactive work quickly, and so once again efficiency on short term tasks is valued above quality and attention to proactive work. Thus, although they understand the value of proactive work, low status prevents and discourages individuals from acting on that knowledge.

Given the clear need, why don’t senior leaders increase funding and actively encourage a longer term view of maintenance? Are the attributions formed by maintenance employees correct? The perspective of the university administration is equally important to the story of proactive maintenance at Alpha. Without question, senior leaders are well aware of the amount of deferred maintenance at Alpha; as noted, the university very recently commissioned an independent assessment of the condition of the campus. However, to senior leaders maintenance
is but one of a large number of competing priorities for funds. The endowment is also typically used to fund academic scholarships, salaries, specific research initiatives, and new building construction – all activities crucial to the university’s educational mission and supported by important stakeholders.

Faced with large competing priorities, university leadership naturally views maintenance as a cost to be minimized. While acknowledging the large stock of deferred maintenance, leadership is content to address known issues as needed or as funds become available. Perversely, such an approach is made possible precisely because the maintenance department has been so successful at keeping the campus running despite limited budgets and the large stock of deferred maintenance. The successes that managers describe related to cost cutting and productivity improvements are real, and in addition customers report being satisfied with the work that maintenance mechanics do. Our interviewees included several building managers not affiliated with maintenance, and as one put it: “They’re doing a good job. A guy who comes in, he’ll try to finish his job... I get what I need. If something fell behind, you’d know who to contact – and it will get done.” Notice once more, however, that while the customer is clearly satisfied, his measure of success is responsiveness to reactive work. Confirming such sentiments, Figure 7 illustrates the gains that the department made in responsiveness over a recent three year period. The graph shows the percentage of work orders that are closed within a target number of days, as determined by the work order’s priority.

FIGURE 7
Responsiveness to Reactive Work

Although the university administration is not a ‘customer’ in the same sense as a building manager, the general sentiment that maintenance is responsive to needs confirms leadership’s view that introducing more proactive maintenance is not an immediate necessity. The list of deferred maintenance might be long, but the maintenance process as it exists currently is sound. Thus, deferred maintenance remains a cost to be addressed later, rather than an opportunity for investment. Proactive work may bring benefits, but only at some expense to the organization. For entirely different reasons, both university leadership and the maintenance department hold mental models of maintenance as a cost to be minimized. Moreover, the structure of relations across organizational boundaries continually reinforces such a view: because maintenance is successful, the administration sees no reason to increase budgets, thereby forcing maintenance managers to renew their focus on efficiency rather than undertake real investment in best practices.
Misperception of Feedback

By itself, a focus on efficiency should not preclude both managers and administrators also understanding the long term value of proactive maintenance. An important puzzle remains: even if cost reduction and efficiency are necessary today, why doesn’t the university express a desire to invest in proactive maintenance in the future, especially given that proactive maintenance is a known best practice? We next argue that a mental model focused on efficiency and cost reduction not only encourages certain behaviors today, but also causes decision makers to underestimate the potential gains from proactive work into the future.

Research has shown how mental models of decision makers are often limited (e.g., Sterman, 1989), including in the maintenance context (Carroll et al, 1997). In particular, actors often misperceive how the implications of their decisions feedback through the environment to influence future decisions (Sterman 1989; 2000). Here, our interviews suggest that both managers and senior leaders underestimate the importance of reinforcing feedback in the full causal model developed above (Figure 5). The reinforcing “reinvestment” loop is crucial to the returns from proactive maintenance; when this loop is active, the majority of deferred work is more than paid for by reinvesting the gains from initial proactive work. Proactive work reduces the stock of defects, lowering the breakdown rate and thereby freeing resources for still more reactive work. Yet, given a culture of scarcity and many competing priorities for funds, individuals at Alpha instead expect that any savings will not be theirs to keep. In fact, such behavior is commonly observed in improvement initiatives of all kinds: organizations, when faced with a choice between cutting costs or reinvesting savings, often choose to cut costs. Thus, the absence of reinvestment feedback is an accurate portrayal of the mental model currently held by both managers and senior leadership.

Failure to fully account for positive feedback through the reinvestment of resources has dramatic implications for managerial decision making. To show the impact, we use the method of partial model testing (Morecroft, 1985), in which one or more feedback loops in a simulation model is deactivated. Figure 8 shows several simulation runs in which positive “reinvestment” feedback is no longer active. Here, the reduction in the stock of defects comes only from planned investment, and not from the reinvestment of freed resources; excess resources are instead saved or spent outside of the department. As such, all proactive work comes at a cost, matching the current mental model of maintenance.
Figure 8 shows three different simulation runs. The first, as in Figure 6, shows a base run in which no investment in proactive work is made, while the second and third runs show the effect of a small and large investment respectively. As shown in the graph of work capacity, both investments follow a similar pattern in which a one year ramp up gives way to a gradual decrease back to the base capacity needed to sustain operations. Clearly, the absence of the reinvestment feedback has a large influence on model behavior. Proactive work causes an initial fall in the stock of defects as before, but once extra resources are removed, the system now falls back into a mode of rising costs and declining quality. In direct contrast to full model results shown in Figure 6, a shift to a regime of sustainable proactive work never occurs, as more and more resources are appropriated to breakdowns as the stock of defects begins to rise again. Costs do fall below what they would be otherwise, but the return on investment is small and does not become positive until well into the future. Thus, without positive feedback through reinvestment, proactive maintenance is a cost rather than an investment. Senior leaders and managers who neglect such feedback – whether due to incomplete mental models or due to an expectation that savings will not be reinvested – will therefore not promote an investment in proactive maintenance, unless they are willing to finance the improvement at a loss. As such, failure to invest is a boundedly rational strategy (March & Simon, 1958) for senior leaders whose only interest is generating a return on investment.

Positive feedback through reinvestment is necessary to ensure a significant return on investment. Is it also sufficient? We next use the model to gain a more general understanding of
the conditions under which Alpha could transition from a reactive to a proactive maintenance organization. Figure 9 again shows three model runs: a base case, a small investment, and a large investment where feedback through reinvestment is active in all cases. We see that in addition to the presence of feedback through reinvestment, the size of the investment is also important. If only a small investment is made, the organization reverts to a reactive mode as above. When a larger investment is made, however, proactive maintenance becomes sustainable even as extra resources are pulled away. The consequences for the potential return on investment are dramatic; in the latter case, yearly spending continues to fall well into the future.

The sudden shift in qualitative behavior occurs as the organization crosses a tipping point. Tipping points are common in complex systems and are not new to management theory (eg., Repenning, 2002; Rudolph & Repenning, 2002). Here, the system tips from a reactive to a proactive regime when the reinforcing “reinvestment” loop becomes dominant. Reinvestment becomes dominant when the total defect elimination rate, net of any extra resources invested initially, exceeds the defect introduction rate. As long as the defect elimination rate is greater than the defect introduction rate, the stock of defects will fall, decreasing breakdowns, allowing more resources to shift to proactive work, and causing the defect elimination rate to rise still further (due to the fact that proactive work is always more productive than reactive work). Thus begins a virtuous cycle whereby proactive work becomes more and more dominant. Figure 5 shows the full causal chain.

**FIGURE 9**

Comparison of a Small and Large Investment in Proactive Maintenance
On the other hand, if the initial investment in proactive work is sufficiently small, the defect elimination rate will exceed the introduction rate only briefly. At first, extra resources cause the outflow of defects to exceed the inflow, reducing the stock of defects (as shown in Figure 8). However, the decline in the stock of defects is not large enough to cause the system to tip. Breakdowns fall and proactive work increases, but the increase in proactive work is not substantial enough to keep the defect elimination rate larger than the defect introduction rate once extra resources are scaled back. As a result, the stock of defects begins rising again, negating any previous gains and sending the system back into a reactive mode. Thus, a transition from a reactive to a proactive mode requires not only reinvestment of resources, but also a sustained initial investment that is large enough to push the organization past the tipping point.

The above simulation results reinforce our contention that maintenance at Alpha does not follow best practice. A switch to proactive maintenance both reduces yearly maintenance costs and increases reliability. At the same time, simulation results also provide insight into why the organization has chosen not to invest in proactive maintenance. In particular, if positive feedback is neglected – either due to cognitive limitations in mental models or due to an expectation that cost savings will not be available for reinvestment – proactive maintenance does not provide a significant return on investment. Given the many competing priorities for university dollars, it is understandable that improving maintenance may not be the highest priority as an expense. Only when proactive maintenance is seen as an investment does the case become clear.

**DISCUSSION AND CONCLUSION**

Taken together, simulation results and qualitative data provide a clear picture of the absence of proactive maintenance at Alpha University. Interviews illustrate that maintenance professionals do understand the importance of proactive work, and the list of outstanding deferred maintenance provides an easy starting point for a comprehensive implementation. In addition, a calibrated simulation model of the maintenance operation confirms that proactive work is valued in decision rules related to work allocation, and the fraction of labor to proactive maintenance has risen over the past three years, even if only slightly. Thus, the absence of proactive maintenance presents a challenge to existing theories of failed best practice adoption that emphasize the complexity of practices or the difficulty that organizations have recognizing and interpreting them. Furthermore, the university setting is one where short term pressures should not be dominant.

To help resolve this apparent paradox, we develop an expanded model of the maintenance system. We find that the structure of the maintenance system leaves Alpha highly vulnerable to the capability trap dynamics described by Repenning & Sterman (2002). Over a period of many years, the stock of deferred maintenance has grown, increasing breakdowns, decreasing proactive work, and therefore leading to still more breakdowns later on. Data confirms that short term pressures are in fact substantial, in the form of the large volume of breakdowns and demanding customers.

The key insight of our analysis is that short term pressures that inhibit improvement can be generated entirely internally. In the case of Alpha, relations across organizational boundaries
create a shared belief that resources are scarce and maintenance is a cost to be minimized. Thus, introducing proactive maintenance becomes an expense that can be put off, rather than an opportunity for growth and investment. This mindset is a natural consequence of the culture of scarcity: simulation results show that as long as actors misperceive the opportunity for reinvestment or believe that savings will not be theirs to keep, they will not view proactive maintenance as a source for return on investment. Only with a fuller shared understanding that allows for reinvestment of returns does proactive maintenance become profitable for the organization.

The results presented here validate Repenning & Sterman’s (2002) theory of stalled organizational improvement. We see that the capability trap dynamics describe the maintenance system very well: while the benefits of proactive maintenance are delayed, the organization can immediately improve response time to customers by instead focusing on reactive work. Yet, the context of university maintenance allows us to extend the capability trap theory in important directions. First, the fact that pressures are internally generated suggests that the trap can emerge in almost any organization, even in the absence of the external influences that often attract blame from managers. Second, intra-organizational misunderstandings can derail implementation of even the simplest and most widely understood of practices. The key characteristic of proactive maintenance is that its benefits are delayed, and our results suggest that implementation may be difficult for any practice – no matter how simple or widely adopted – as long as that is the case.

Together, these ideas suggest a different metaphor for the diffusion of best practices than the one that is commonly used. Typically, best practices are conceived of as novel innovations that are discovered by one or several organizations and then imitated by many others, following a contagion metaphor like the spread of a disease (eg. Strang & Macy, 2001). Imitation then either fails or succeeds. If successful, practices become part of an institutional logic that compels all remaining and future organizations to follow suit (DiMaggio & Powell, 1983). Alternatively, if practices are sufficiently complex or organizations sufficiently inert (Hannan & Freeman, 1984), imitation may fail and the practice may continue as a source of sustainable competitive advantage for its originator. For example, Barney (1986) argues that organizational cultures are especially difficult to observe, describe and imitate. In a similar vein, Zbaracki (1998) shows how a practice like TQM can take on many forms when it is imitated by organizations, thus leading to broader and sometimes less effective definitions.

Our results suggest a somewhat different metaphor. Of course, organizations do generate novel, complex innovations that others then attempt to imitate. At the same time, however, practices that drive competitive advantage can also be surprisingly simple and based on theories and practical knowledge that have existed for many years. Proactive maintenance is an example of such a practice. The idea can be described in specific, concrete terms, and organizations like Alpha already do some of it. In addition, as simulations show, proactive maintenance can have a dramatic influence on the effectiveness of the organization. Rather than a novel idea that either diffuses or fails to diffuse, best practices are sometimes ideas that originally spread long ago, and that now exist in some form almost everywhere.

As such, the main impediments to diffusion are not always the relationships between organizations, but the relationships within them. Much organizational theory is concerned with
how environments shape organizations, including most notably the resource dependence and institutional perspectives. In addition, diffusion research often emphasizes the quality of ties, the number of ties (e.g., Centola & Macy, 2008), or the structure of the network. We argue instead that the spread of simple ideas may be hindered also by properties endogenous to each organization.

Finally, our results should bring hope to practicing managers. Rather than the bleak view offered by existing theories of implementation and competitive advantage, we suggest that organizations can make huge improvements by acting on the simple ideas that they already know can work. Such is the case especially for practices that reduce emissions, including proactive maintenance, installing more efficient light bulbs, or changing behaviors to reduce energy use. Making these simple changes, however, requires overcoming the challenges that exist when the benefits of improvement activities are delayed. We suggest two requirements for overcoming these challenges. First, mental models must be rich enough such that the long term benefits of a new practice, and the nature of the delays before gains are realized, are understood. Second, these models must be shared and reinforced by strong norms of communication both within functional groups and across organizational boundaries. Without a shared understanding, short term pressures may well emerge between individuals and groups and prevent dedicated action towards improvement. As March observed, “what is good in the long run is not always good in the short run (1991; pg. 73).” This simple statement perhaps encapsulates much of why achieving high performance is so difficult for organizations.
References


