Principal-Agent Alignment and Relational Incentive Contracts in High-Performance Service Operations

by

Ryan Garrett Doss

B.S.E., University of Michigan, Ann Arbor, 2008
M.S.E., University of Michigan, Ann Arbor, 2009

Submitted to the Department of Aeronautics and Astronautics and the MIT Sloan School of Management in partial fulfillment of the requirements for the degrees of

Master of Science in Aeronautics and Astronautics

and

Master of Business Administration

in conjunction with the Leaders for Global Operations Program at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2014

Signature redacted

Author ....
Department of Aeronautics and Astronautics and the MIT Sloan School of Management
May 9, 2014

Signature redacted

Certified by ....
Deborah Nightingale, Thesis Supervisor
Professor of the Practice, Aeronautics and Astronautics and Engineering Systems Division

Signature redacted

Certified by ....
Georgia Perakis, Thesis Supervisor
William F. Pounds Professor of Management Science, MIT Sloan School of Management

Signature redacted

Approved by ....
Paulo C. Lozano
Chair, Aeronautics and Astronautics Graduate Program Committee

Signature redacted

Approved by ....
Maura Herson
Director, MBA Program, MIT Sloan School of Management

Principal-Agent Alignment and Relational Incentive Contracts in High-Performance Service Operations

by

Ryan Garrett Doss

Submitted to the Department of Aeronautics and Astronautics and the MIT Sloan School of Management on May 9, 2014, in partial fulfillment of the requirements for the degrees of
Master of Science in Aeronautics and Astronautics
and
Master of Business Administration

Abstract

This thesis focuses on the creation of a high-performance service operations organization. As organizations increasingly compete on service quality, increased attention has been given to measuring, tracking, and improving customer satisfaction. This thesis 1) provides a novel framework for service quality improvement and 2) explores concepts in game theory, relational contracts, and incentive mechanism design that impact service quality in the modern organization. The framework introduced in this thesis is comprised of four distinct steps. In the first step, service quality is quantitatively measured and drivers of service quality are determined both through qualitative methods and through statistical analysis on a customer-by-customer basis. In the second step, key drivers of service quality are addressed through process redesign and operational improvement. In the third step, the alignment of service operations incentive mechanisms with employee behavior consistent with high service quality is analyzed and considered in the context of building a high-performance service organization. Finally, the role of organizational learning and the relational contracts that may help to sustain a culture of experimentation, learning, and improvement are considered. These concepts are applied to a host organization, Atlantic Energy, by way of case study throughout this thesis; this acts to provide a concrete example of the application of these concepts and shows an example of the effectiveness of the framework when compared to traditional methods in service operations improvement.

Thesis Supervisor: Deborah Nightingale
Title: Professor of the Practice, Aeronautics and Astronautics and Engineering Systems Division

Thesis Supervisor: Georgia Perakis
Title: William F. Pounds Professor of Management Science, MIT Sloan School of Management

Acknowledgments

I would like to thank Atlantic Energy for their sponsorship of this project and for all of their support in the development of this work. Specifically, I would like to thank Jorge Calzada, Nancy Concemi, Chris Starr, Diane Rivera, Megan Piccaretto, and Mallik Angalakudati; without the generous help of these individuals, the project behind this thesis would not have been possible.

I would also like to thank my thesis advisors, Georgia Perakis and Deborah Nightingale. Professor Perakis's technical expertise and guidance in operations management was incredibly helpful throughout this project, and Professor Nightingale's insight and intuition into the high-level, systems-focused considerations of an organization were invaluable during this project.

I also wish to thank the LGO program and all fourty-seven of my classmates, whose support and friendship have made the last two years the great experience it has been.

Finally, I want to thank my wife for her incredible support and love over the last two years, as I am incredibly blessed to have had her as a partner in this two-year adventure, and I want to continually thank God for His grace and love in my life.

Contents

1 Introduction **12**

1.1 Motivation and Problem Statement 12

1.2 Prior Work in Service Quality Improvement 13

1.3 Hypothesis and Overview of Thesis Framework 15

1.4 Contributions and Structure of Thesis 17

2 Measuring and Improving Service Quality **20**

2.1 Defining an Appropriate Measurement of Service Quality 21

2.2 Qualitative Determination of Customer Satisfaction Drivers and Development of a Quality/Process Data Set 23

2.3 Analytical Determination of Customer Satisfaction Drivers 24

2.3.1 Quantitative analysis using ordinal service quality measures 25

2.3.2 Using the results of the regression analysis 29

2.4 Conclusion 30

3 Updating Process to Reflect Drivers of Service Quality and Cost **32**

3.1 Existing Tools in Operations Excellence 33

3.1.1 Lean Thinking 33

3.1.2 High Velocity Organization 35

3.1.3 Other related works in service quality 37

3.2 Developing Operational Changes for Improved Service Quality 38

3.2.1 Defining and addressing value from drivers of service quality 38

3.2.2 Reducing waste via the "funnel" concept 39

3.2.3 Improvement and learning through experimentation 41
3.3 Managing the Tradeoff between Quality and Cost 42
3.3.1 Quality / cost alignment from eliminating *muda* 43
3.3.2 Quality / cost alignment from the impact of quality on cost 44
3.3.3 First contact resolution trade-off with average handle time 45
3.4 Conclusions and Path Forward 46

4 Case Study I: Goal Definition and Operational Architecting at a Multinational Energy Utility 48
4.1 Introduction 49
4.1.1 Company context 49
4.1.2 Objective and motivation 50
4.2 Measuring and Improving Service Quality 51
4.2.1 Measuring service quality via customer satisfaction surveys 51
4.2.2 Qualitative assessment through employee interviews 54
4.2.3 Qualitative assessment through customer surveys 55
4.2.4 Quantitative assessment through customer satisfaction scores and process metrics 57
4.3 Updating Process for Improved Operations 62
4.3.1 CSAT Lever #1: Increasing First Contact Resolution (FCR) 63
4.3.2 CSAT Lever #2: Reducing errors through feedback loops 67
4.3.3 CSAT Lever #3: Implementing a priority queue for customers without power 68
4.4 Conclusion 69

5 Aligning Agent Behavior with High-Performance Operations 70
5.1 Introduction 70
5.2 Service Operations Output Model 71
5.2.1 Preliminary definition of output 71
5.2.2 Cost objectives from queuing theory 72
5.2.3 Output value equivalence 73

5.3 Analytical Model for Incentive Mechanisms 75
5.3.1 Agent behavior and output 75
5.3.2 Utility and risk attitude 77
5.3.3 Analysis of traditional incentive schemes and metrics 78
5.4 Numerical Results and Insights from Agent Decision Theory Analysis 80
5.4.1 Impact of traditional incentive scheme on agent behavior 81
5.4.2 Impact of number of QA checks 82
5.4.3 Misalignment of QA with performance 85
5.4.4 Impact of other factors in basic model 86
5.4.5 Guarding against shirking and multi-task misalignment 86
5.5 Application of Insights from Theoretical Results 88

6 Creating a Culture of Experimentation and Organizational Learning 91
6.1 Organizational Learning Revisited 92
6.2 Relational Contracts: Theory and Literature 93
6.3 Application of Relational Contracts to Organizational Learning 96
6.4 Sustaining Relational Contracts with Issues of Credibility and Uncertainty . 101
6.4.1 Modeling relational contracts with Markov Decision Processes 102
6.4.2 Analytical results for Bayesian Nash Equilibrium 107
6.5 Conclusions 111

7 Case Study II: Aligning Incentives for Performance and Learning at Atlantic Energy 114
7.1 Analysis of the Current State 115
7.1.1 Formal metrics and incentives in the Atlantic Energy customer care center 115
7.1.2 Experimentation at Atlantic Energy 117
7.2 Defining Countermeasures for Performance Improvement 118
7.2.1 Aligning performance goals with optimal output and improvement . . 118
7.2.2 Increasing the value of experimentation and cooperation in organizational learning 120

8 Conclusion **122**

8.1 Retracing the Thesis Framework 122

8.2 Key Lessons 123

8.3 Generalization and Further Application 124

8.4 Closing 125

List of Figures

1-1 Framework for process-focused service quality improvement 15
1-2 Aligning process with quality drivers meets, and subsequently exceeds the quality-cost frontier 17
1-3 Shifting effort from output to innovation in the short-run leads to long-term improvements in performance 17

2-1 Step 1 of the thesis framework 20
2-2 2x2 matrix showing how to interpret different drivers based on the results of the statistical analysis 30

3-1 Step 2 of the thesis framework 32
3-2 The funnel, by which customers are potentially diverted off the efficient path at different points of the process 41
3-3 Traditional view of quality / cost tradeoff 43
3-4 Reaching and expanding efficient frontier through ideas from thesis framework 43
3-5 When First Contact Resolution (FCR) is not achieved, the customer re-enters the process, creating a multiplier on the input volume the process experiences 44
3-6 Efforts to increase FCR may also increase AHT at no additional cost to the customer care center's operations 46

4-1 Steps 1 and 2 of the thesis framework are covered during Case Study I . . . 48
4-2 High-level issue tree serving as road map for investigating CSAT drivers . . . 55
4-3 Pareto chart showing reasons for unfavorable CSAT scores related to account initiation process 57

4-4 Results of regression analysis on customer data and satisfaction scores reveals drivers of customer satisfaction in account initiation process 60
4-5 Steps along the account initiation process provide opportunities for customers to be "funneled" off the efficient path, and into higher likelihood of rework and errors 64
4-6 Priority queue for account initiation process filters on connection status to increase CSAT 69

5-1 Step 3 of the thesis framework 70
5-2 Erlang C queuing model predicts nearly linear relationship between staffing needs and handle time 74
5-3 When normalized by arrival rate, staffing curves in Figure 5-2 nearly collapse onto a single, nearly linear curve 74
5-4 Dynamics of rework creating a multiplier of $1/p$ on arrival rate of unique customers; $\lambda = \lambda_{unique}/p$ 75
5-5 Agent objective output shows unique maximum that firm would like to maximize with $\mathrm{argmax}_{a_1}(y) > m$ 77
5-6 A combination of time goal cut-off, uncertain performance measurement, and agent risk places agent's optimal action choice below the firm's, underserving the customer 81
5-7 Agent's optimal AHT very slightly decreases with an increase number of QA checks, but quickly approaches a limit 82
5-8 More forgiving QA structure shifts optimal agent target AHT lower, widening the gap from firms optimal 84
5-9 Less forgiving QA structure shifts optimal agent target AHT higher, potentially overshooting the firm's optimal AHT 84
5-10 With the lower AHT target, the agent's optimal action is to not slack 88
5-11 With the higher AHT target, the agent's optimal action is now to slack, filling in the extra time with a_2 rather than increasing a_1 88

6-1 Step 4 of thesis framework 91

6-2 Trading off short-term performance for experimentation leads to increasing firm capability, maximizing long-term output 94
6-3 The trust game (Kreps, 1990)[17], used to model relational contracts (Gibbons,2012)[9] 95
6-4 Relational contract as a two-person simultaneous repeated game modeled by a nondeterministic Markov Decision Process 104
6-5 The optimal strategy resulting in the highest-valued BNE changes as the system uncertainty and likelihood of poor results $1 - p^2$ increases 108
6-6 Relational contract modeled by nondeterministic Markov Decision Process with intermediate state of defection "M" 109
6-7 Optimal BNE strategies with intermediate, "belief-based" state added 110

7-1 Steps 3 and 4 of the thesis framework are covered during Case Study II . . . 114

List of Tables

2.1 Comparison of quality measurement in manufacturing and service operations 22

6.1 Actions of cooperation, defection, and punishment in manager-worker relational contracts in service operations experimentation 97

6.2 Potential payouts in a manager-worker relational contract sustaining organizational learning, as illustrated by trust game (Figure 6-3) 98

6.3 Actions of cooperation, defection, and punishment in manager-manager relational contracts in service operations experimentation 99

6.4 Actions of cooperation, defection, and punishment in manager-manager relational contracts in service operations experimentation 102

6.5 "Forgiveness" Strategy 108

6.6 Grim-Trigger Strategy 108

6.7 Mutual Defection Policy 108

6.8 3x3 "Forgiveness" Strategy 110

6.9 "Cooperate on Diagonal" Strategy 110

6.10 "Cooperate on Upper Diagonal" Strategy 111

6.11 3x3 Grim-Trigger Strategy 111

Chapter 1

Introduction

1.1 Motivation and Problem Statement

As companies continue to increasingly compete on customer service, service operations become more important from a quality standpoint. As service quality is increasingly seen as a competitive advantage, rather than an operating expense, companies face increased pressure to respond by determining the key drivers of customer satisfaction in their organization and acting on such drivers to retain customers, increase revenue, and improve the company's perception in an ever more transparent market. While manufacturing quality can be measured, controlled, and improved in objective, tangible ways delineated by classical management science, service quality is far more difficult to define, address, or link to the organization's overall success.

As a result, this thesis aims to address the following problem: How does an organization measure service quality, determine the parts of its operations that drive quality, and ultimately improve quality in a meaningful and sustainable way? To this end, in this thesis we provide both a theory for and a case study in addressing and improving quality in a service operations environment. Specifically, we define the problem scope to include the measurement, analysis, and improvement of service operations in an organization from a process-focused point of view, in which we define a service operations process a priori for improvement, and then prescribe operational changes that address the results of such an analysis. In doing so, we take an enterprise view of the organization, considering not only the mechanics of the operation or process in question, but the role of several stakeholders, including customers, the employees the conduct the operations, and the managers for the organization as well.

As an example of the motivating problem, we consider the company used in the case study: Atlantic Energy.[1] Atlantic energy is a strong-performing energy utility with significant US operations. While it has historically focused on providing safe, reliable, affordable energy to the millions of customers in its geographic footprint, it aims to increase its customer satisfaction in measurable, quantitative terms both for its commitment to serving customers and its desire to provide evidence of its service quality to regulatory bodies determining its revenue streams through rate cases. In the energy utility industry, a company scoring in the first quartile in customer satisfaction per JD Power[2] is given an allowed, regulated return on equity that is 0.5 percentage points higher than than that of companies in the fourth quartile.[12] In an industry with regulated assets that may be in the $10 billion to $100 billion range, this can lead to a $50 million to $500 million impact on shareholder value.

While the impact of exceptional- or poor- service on company value may be explicit in this industry, examples abound of its importance in any organization with a significant focus on serving the customer.

1.2 Prior Work in Service Quality Improvement

There have been many past works in service quality measurement and improvement that this thesis builds on. Parasuraman [21] introduced many of the most influential concepts in service quality measurement. Klaus and Maklan[16] expanded the traditional view of customer satisfaction beyond the company-focused view, creating a measurement system that focused on the customer's total end-to-end experience when considering customer satisfaction. While Green [10] highlights the application of conjoint analysis and statistical regression in measuring drivers of customer satisfaction, Jasrai [14] builds on conjoint analysis to consider the role of multiple regression when looking at customer satisfaction and its potential drivers.

[1] Company name has been obfuscated in this thesis.
[2] www.jdpower.com

The MIT LGO thesis of Avijit Sen[25] investigated the application of lean principles when designing new operational metrics for the service operations of a customer care center at Dell, Inc. This thesis demonstrated the ways in which designing call center metrics that focused on the ideas of value-added and non-value-added activities from Lean could better align operations with customer desires.

Past work has also considered the nature of service quality drivers, in addition to their measurement. Sivakumar [26] found that customer satisfaction can often be driven not only by events, but the frequency, timing, and sequence of them as well, which expanded the view of potential customer satisfaction drivers to include not only discrete events but measures of how they occur as well. Julien [15] investigated the relationship between service quality perception by customers and service quality perception by front-line employees, giving rise to the idea of using feedback from both sources in this thesis. Devine et. al. [6] explored the link between human psychology and service quality, an idea that is built on in this thesis.

This thesis also builds on many key past works in operations improvement, including Lean Manufacturing (Womack and Jones [29]), High-Velocity Organizations (Spear [27]), and Enterprise Architecting (Nightingale [20]). These works in particular have stressed high-level, systems thinking when considering the operations of an organization, which is paramount in the ideas introduced in this thesis. Finally, this thesis builds heavily on the concepts of relational contracts, agency alignment, and incentives as developed in the past by Gibbons and Henderson [8] and Holmstrom [13].

While past works provide a strong foundation for addressing the problem studied in this thesis, this thesis will build off of many of these individual concepts in the creation of the framework introduced in Figure 1-1. This framework summarizes the approach that will be taken in this thesis, for which the focus will be on the improvement of service operations processes by addressing the determination of service quality drivers, the improvement of operations to address such drivers, and the redesign of company-employee relationships and incentive mechanisms to ensure employee behavior- and organizational learning- consistent with high-performance service operations.

1.3 Hypothesis and Overview of Thesis Framework

To address this problem, we proceed by introducing the thesis framework, depicted in Figure 1-1, and introduce the following hypothesis: we propose that the application of this framework, as delineated in this thesis, provides a method for the improvement of service operations quality. In walking through the application of the thesis framework, we further propose to demonstrate the following:

1. Statistical analysis can be used to analyze and identify customer preferences and customer satisfaction drivers in a service operations process. (Chapter 2)

2. Results of such a statistical analysis can be addressed through the application of existing methods in operations management. (Chapter 3)

3. Game theory and decision theory analysis can be used to predict agent behavior in the consideration of principal-agent and incentives concerns in service quality improvement. (Chapter 5)

4. The creation of a learning organization[3] will rely on relational contracts, and in the face of uncertainty or issues of credibility and clarity, insight can be derived from autonomous agent theory in helping the organization to reach its goals. (Chapter 6)

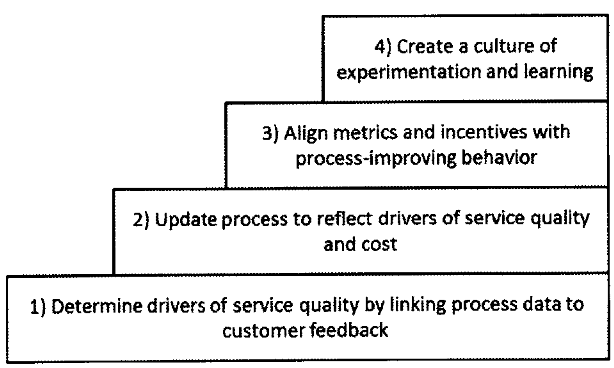

Figure 1-1: Framework for process-focused service quality improvement

[3]A learning organization will be defined in greater detail later, and draws from concepts in many operations management philosophies including Lean, High Velocity Organizations, and others

To this end, the thesis will address these goals through derivation of the thesis components, prediction of its components' results through theory and analysis, and ultimately, application of the framework to an actual organization in the form of a case study.

The thesis framework, in its full application, consists of four parts that necessarily build on each other to reach the organizational goal of service quality improvement. In the first step, we measure the current performance of the organization's service quality (typically through customer satisfaction feedback), collect data that explicitly links the service quality data to the process in question (the target process), and analyze the relationship with statistical analysis to determine the key drivers of service quality from the target process. In the second step, we use the analysis from the first step to address the specific areas in which the target process can be improved to act on the key drivers, and then redesign the process to address these opportunities. In the third step, we move beyond process improvements to consider the role of the organization's customer service agents on service quality, addressing issues of incentive design and principal-agent alignment within the organization. Finally, once the process has been analyzed, key quality drivers have been determined and acted on, and incentives have been aligned within the company, we aim to create and maintain a learning organization that iterates on these first three steps, in which employees and managers design and run experiments to put existing thoughts about the physics of company-customer interaction to the test, ultimately discovering the way- rather than designing the way- to an improved organization. Each of these four steps is described in detail in this thesis.

Our hypothesis predicts that service quality in the organization is a function of three internal variables: structural alignment of the process with quality drivers, alignment of employee effort with quality drivers, and the magnitude of employee effort; if the four steps in the framework are completed, we predict an increase in service quality as the steps address these three variables, in turn, as shown in Figures 1-2 and 1-3. The first two steps act to improve the process, which maximizes the effectiveness of employee effort, while the third step shifts employee effort to be more in-line with the customer's perception of quality; these three steps combined thus shift the state of the process to the pareto efficient frontier, as depicted in Figure 1-2. However, once the efficient frontier has been reached, the fourth step- the long-term improvement of the service operations through experimentation- continues to

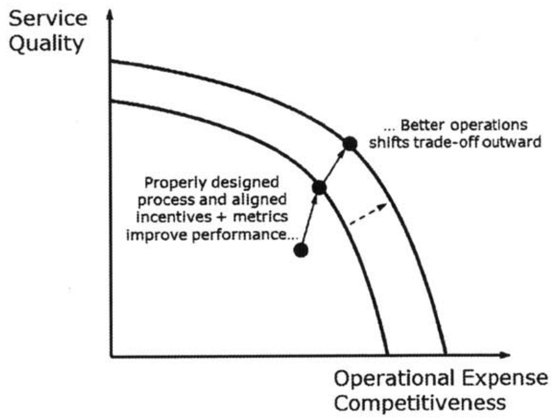

Figure 1-2: Aligning process with quality drivers meets, and subsequently exceeds the quality-cost frontier

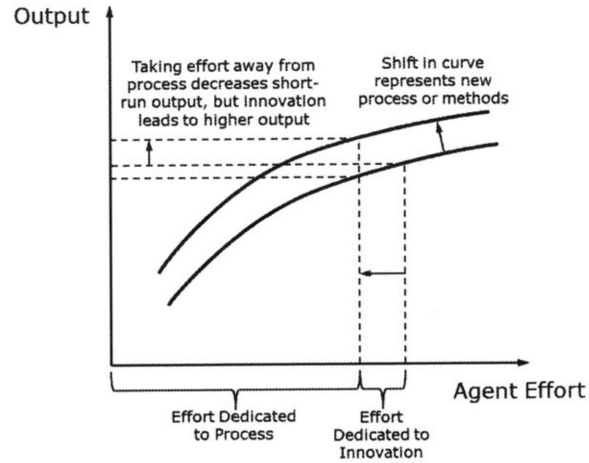

Figure 1-3: Shifting effort from output to innovation in the short-run leads to long-term improvements in performance

shift the frontier as discoveries lead to improved processes and customer service models. While this requires a short-term decrease in output as shown in Figure 1-3, the effort that is diverted to innovation and experimentation continues to shift the curve outward, as shown in both Figure 1-2 and 1-3, leading to a continued high level of output.

1.4 Contributions and Structure of Thesis

This thesis proceeds by describing each of the four steps of the thesis framework in succession, along with detailed examples in the form of case studies. In chapter two, the thesis begins by considering the methods by which an organization can measure and analyze the impact of service operations on customer satisfaction. In it, we define ways in which an organization may use qualitative analysis through customer and employee interviews, and quantitative analysis through linear regression or logistic regression to determine what impact various aspects, performance indices, or subprocesses of a target process can have on customer satisfaction for the organization. Here, we draw from concepts in marketing theory and operations research to develop simple methods for studying these links with an ultimate focus not on rigorous prediction, but rather managerial intuition.

From here, we proceed into chapter three, in which we progress to the second step of the thesis framework and the results from chapter two are used to drive the focus of operational improvement efforts. After opening with a brief discussion of current philosophies in operations management, the chapter addresses the application of such methods to address the gaps in quality performance identified by the analysis from the first step of the framework. This is followed by chapter four, in which the first two steps of the framework are applied to Atlantic Energy in the form of a case study. This chapter traces the improvement of service quality in the account initiation process at Atlantic through customer and employee interviews, statistical analysis of customer satisfaction surveys linked to process performance data, and several resulting operational improvement efforts to address the resulting insights gained from such analysis.

In chapter five, we proceed to the third step of the framework: the alignment of agent behavior with service quality through metrics and incentives. In particular, we consider the trade-off between cost and quality in operations at organizations like Atlantic and, after already addressing process-related improvements to quality in the second step, we consider ways to align agent behavior with actions that optimize the cost-quality balance. To do this, we build on concepts from game theory and decision theory analysis to develop an agent decision model, and then use the model to explore common issues with metric and incentive systems in service operations with the aim of gaining insight into ways to improve incentive mechanisms.

In chapter six, we progress to the final step of the framework, in which we consider organizational issues at play that either help or prevent an organization from successfully iterating on the first three steps via use of the scientific method to learn more effective ways of delivering high-quality service. Drawing again from economic game theory, we consider the role of relational contracts in sustaining a learning organization, and model the problem to consider if such a relationship may be sustained given the uncertainty and possibility of an experiment failing. We finally use the intuition from this analysis to consider the application of such ideas to an organization.

In chapter seven, we apply the ideas of the third and fourth steps of the framework to Atlantic Energy once again, considering the ways in which one might think about and address incentive misalignments at all levels of the organization in driving high-quality service, and improving on operations iteratively through experimentation. This is followed by the final chapter, the thesis conclusion.

This thesis offers a novel approach to service quality improvement that incorporates widespread data collection and statistical analysis, systems-level operational improvement, and elements of relational contracts and game theory in solving the agency and incentives problems that are often overlooked when measuring and improving customer satisfaction. Throughout the description of the framework's components, this thesis aims to provide insight into many novel concepts in the management of high-performance service operations.

Chapter 2

Measuring and Improving Service Quality

Qualitative and statistical analysis of process and customer satisfaction data to identify key service quality drivers

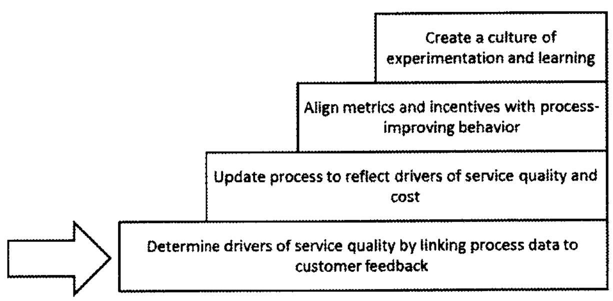

Figure 2-1: Step 1 of the thesis framework

This chapter outlines the first step of the thesis framework, in which the organization determines aspects of the targeted processes that drive service quality as perceived by customers. This process, in itself, is comprised of four steps, which we will describe for the remainder of this chapter; these steps are as follows:

1. Define an appropriate measurement of service quality.

2. Develop a hypothesis for drivers of service quality via qualitative methods.

3. Collect data that explicitly links customer satisfaction responses and process data to test hypothesis.

4. Use statistical analysis to determine the most important drivers of customer satisfaction in preparation of the framework's second step.

Within this process, service quality drivers are measured three separate ways: qualitatively through customer feedback, qualitatively through employee feedback, and quantitatively through statistical analysis. All three methods are delineated in this chapter and described with an example in the first case study.

2.1 Defining an Appropriate Measurement of Service Quality

Any discussion on the improvement of service quality- or any other performance factor- necessarily begins with an appropriate measurement from which performance and improvement can be based. While most service organizations will have a measure of customer satisfaction in place- which will often be difficult to change in the short run- we open with a brief discussion of the measurement of service quality. Once an explicit, quantitative measurement of service quality exists, the organization is prepared to use the measure for identifying and acting on the drivers of that measure.

One of the primary differences between service operations and manufacturing operations is that in manufacturing operations, customers judge quality based on the output product of the operation, whereas in service operations, customers witness the entire process and perceive quality based on both the process and the output. Such considerations are contrasted in Table 6.1.

These factors make quality measurement more difficult and subjective in service operations; however, one accepted practice is the use of customer satisfaction metrics as a measurement of service quality. Past research in this area stems from Parasuraman et. al.[21] In this paper, the authors explore the notion of service quality as defined by the difference between customer perception and expectation, based on their past research; in the ideal case, we

Characteristic	Manufacturing Operations	Service Operations
Customer perception of quality	Judged based on performance	Perceived based on experience
Typical measures	Hard measures (e.g. tolerances)	Soft measures (e.g. customer feedback)
Customer experience	Based on product output	Based on both output and process

Table 2.1: Comparison of quality measurement in manufacturing and service operations

would describe service quality the same way. Using this measurement, the authors created customer surveys on an ordinal, 7-point scale to measure independently both the customer's perception and expectation of an organization's services (e.g. "How well did the company do X?" and "How well should the company have done X?" with slightly different wordings). This approach avoids over-delivery on service items that customers view as secondary and identifies company underperformance on key issues as well. The authors found that the aggregate difference between perception and expectation amongst several service dimensions was an accurate predictor of the customer's overall satisfaction with the company, including the likelihood to be a repeat customer or to recommend the company to a friend.

Often, a customer satisfaction system is already in place prior to an improvement effort, even if it is less than ideal. Thus, the focus of the thesis framework is on the use of a given service quality metric to improve a targeted process rather than the design of a service quality measurement; that is, the focus here is on process improvement rather than measurement improvement. For this reason, this step assumes that a reasonable metric can be used based on existing customer satisfaction data, or quickly built to obtain both qualitative and quantitative customer feedback regarding an organization's operations; in addition, we further assume that such a metric can be tied to individual customer accounts to support the more advanced techniques developed in this chapter. In some cases, improving the metric itself will be meaningful due to its signal to customers, regulators, and other stakeholders; this includes JD Power scores and University of Michigan Consumer Sentiment scores in many industries. In these instances, it is most helpful to develop an internal measure that accurately tracks the dynamics of the external score in question and then use the internal measure in the steps outlined in the rest of this chapter. In other cases, the goal may be the improved business and customer attitude that customer satisfaction scores are meant to represent. While measurement design makes up another branch of research entirely, this thesis now progresses on the assumption that the organization has a measurement of customer satisfaction installed that will serve as a proxy to service quality, and that process improvements that increase such a metric are desirable.

2.2 Qualitative Determination of Customer Satisfaction Drivers and Development of a Quality/Process Data Set

Before completing any statistical analysis to identify service quality drivers, it is important to know which drivers to test with quantitative data. To do this, we first collect feedback from two important stakeholders in the organization's service operations: the customers and the company employees who interact with them. Hence, both groups are surveyed to gather insight into the main drivers of customer satisfaction, which is then used to develop an initial hypothesis of the drivers that the second step of the thesis framework (operational improvement, Chapter 3) should address to increase customer satisfaction.

Although there are many ways to accomplish this step, several methods are used in the case study presented in this thesis. First, customer service representatives within the actual organization are interviewed individually or in focus groups to determine their perception of the issues customers care the most about and the parts of the current processes which might be broken. At the same time, any customer comments available from customer surveys are read and sorted to develop an intuition of the dynamics at play in customer perception of quality. In the case study, customer comments were used to sort the reasons for dissatisfaction (defined as a score of 1-7 on 10 point scale). This grouping allows the data to be used to construct a Pareto chart, as shown in Figure 4-3, which helps to identify some of the most common causes for customer dissatisfaction. After quality drivers have been assessed through both customer and employee feedback, we can then develop a hypothesis regarding the main drivers of service quality from the results.

Once we have developed a hypothesis for the primary drivers of service quality, we then create a data set explicitly linking service quality measurements (e.g. customer satisfaction scores, or CSAT) to the quality drivers making up the hypothesis. One crucial concept presented in this thesis is that, while many organizations look at average customer satisfaction scores on a monthly or quarterly basis for comparison to average process performance metrics on the same time scale, such a method does not create a meaningful comparison between customer satisfaction and the processes in question. Instead of looking at data in an averaged sense, *we propose creating a data set that explicitly links customer satisfaction scores to process measures that each particular customer actually experienced on a customer-by-customer basis.* For example, suppose that the qualitative assessment from employee and customer feedback indicated twenty potential drivers of service quality including cycle times, numbers of interactions, employee errors, and different paths through the process. Then for the time period in question, process metrics for each of the twenty drivers (including the actual cycle times, binary variables indicating paths through the process, etc) should be collected for every customer account for which the company has customer satisfaction data. This creates a data set for which the variation in customer satisfaction scores can be seen based on the unique experience of each customer, which is much more meaningful when looking for relationships than monthly trends that combine hundreds or thousands of customers together.

$$\textit{Process Data} \longrightarrow \textit{Linked by Customer ID} \longleftarrow \textit{CSAT Scores}$$

2.3 Analytical Determination of Customer Satisfaction Drivers

Once we have created a data set explicitly linking service quality measurements (e.g. customer satisfaction scores) to process performance metrics quantifying customer experiences, the primary drivers of customer service are ready to be identified through statistical analysis. The goal is not to develop an accurate, quantitative model for prediction of customer satisfaction scores, but rather to develop a qualitative understanding of the primary drivers of service quality; this, in turn, enables managers to design processes and incentive mechanisms aligned with behavior consistent with customer satisfaction without requiring a rigorous, expensive set of experiments to develop a true predictive model.

This section describes two methods for identifying service quality drivers. The first method assumes an ordinal service quality measurement, such as a customer satisfaction score on a scale of one to ten. The second assumes a categorical service quality measure, such as individually labeling a customer as a promoter or detractor. Both methods may be applied to the same data set, allowing for a more robust determination of service quality drivers.

Type of CSAT Score	**Regression Type**
Ordinal (e.g. scale of 1-10)	Linear regression (potentially with log transform to find elasticities)
Categorical (e.g. promoter vs detractor)	Logistic regression

2.3.1 Quantitative analysis using ordinal service quality measures

One method for deriving customer preferences from the data set collected above is taken from conjoint analysis as developed by the field of marketing analytics. This technique can be traced back to the seminal 1964 paper by the mathematical psychologists Luce and Tukey[19], and is further described and introduced to the marketing community by Green and Srivinasan's 1978 paper.[10] In its simplest form, this method consists of measuring consumer preference between alternatives, for example by survey or interview, and then performing regression of the variable measuring preference on the independent variables in question, such as features of a product or service. This regression indicates the impact of alternative choices on the customer's overall preference, which can be used to provide the customer with better offerings.

Here, we apply the same idea to service quality by treating the many customer experiences captured in our process/CSAT data set from before as data points revealing consumer preferences over many alternative customer experiences. Thus, by treating customer satisfaction as the measure of consumer preference, we can use the CSAT score from each customer as the independent variable in a regression on the process metrics linked to each customer account, which provide the alternatives that potentially make up the key quality drivers. The result is a much sharper insight into the true drivers of service quality compared to the more common approach of looking at monthly averages over all customers; this advantage is the primary justification for the effort needed to link service quality data to the customer experience on a customer-by-customer basis.

Given an ordinal measurement of service quality, such as numeric customer satisfaction scores, the most straightforward method for identifying drivers of customer satisfaction is linear regression. Although more complicated methods in data mining and machine learning exist, we proceed with linear regression for the following reasons:

- *Limited need for prediction:* Because qualitative insight into drivers is more important than predictive capability for the future steps in service quality improvement, correlation is acceptable in place of complex quantitative models.

- *Ease of execution:* Linear regression is simply computed using a number of common productivity packages.

- *Intuitive simplicity:* Linear regression is widely recognized and easily understood at an intuitive level, increasing the likelihood of managerial buy-in when using results to justify future improvement efforts to other stakeholders.

However, two important limitations of regression analysis in this case must be recognized to avoid misuse of the results. First, the analysis presented in this section involves the consideration of multiple predictive variables, creating the potential for issues with colinearity if multiple regression is used. Although the method presented here uses a single variable regression on each process metric, the interpretation of such results must still be tempered with intuition and judgment in the case of one or more process variables being correlated with each other. The second is that such an analysis makes no claim of identifying causation, and indeed is entirely unable to do so; however, given the goal of such an analysis- namely insight rather than experiment- this limitation is justified given the tradeoff in speed and efficiency of the analysis. As stated before, however, the insight from such an analysis is sufficient to progress to the future steps of the framework.

To begin, we assume a data set of N data points explicitly linking customer service scores to a set of M process measures on a customer-by-customer basis; as before, the process measures consist of both numerical and categorical values. Let y_i be the customer satisfaction scores, x_{ij} be the numerical measures of process performance, and w_{ik} be the categorical measures of process performance for $i = [1, N]$, $j = [1, M_{numeric}]$, $k = [1, M_{categoric}]$. A single-variable regression analysis is then carried out for each x_{ij} for $[j = 1, M_{numeric}]$ and w_{ik} for $k = [1, M_{categoric}]$. The analysis for each type of variable is described below.

Regression analysis with numerical dependent variables

For each $j = [1, M_{numeric}]$, a linear regression is carried out on a logarithmic transformation[1] of both the regressor x_{ij} and the regressand y_i. In this case, the coefficients from the regression analysis will represent elasticities between the dependent and independent variables, representing the percent change in the dependent variable corresponding to a percent change in the independent variable; this will allow for a more meaningful comparison of the regression coefficients when picking out the most important drivers later in the analysis.

Using classical regression theory, the Conditional Expectation Function (CEF) of the output (i.e. service quality score) given a single input is (for each j):

$$\ln\left(\mathbb{E}\left[y_i|x_{ij}\right]\right) = \alpha_j + \beta_j \ln\left(x_{ij}\right)$$

For a given set of points y_i, x_{ij}, the estimated parameters $\hat{\beta}_j$ and $\hat{\alpha}_j$ are:[28]

$$\hat{\beta}_j = \frac{\sum\limits_i \left(\ln(x_{ij}) - \ln(\bar{x}_{ij})\right)\left(\ln(y_i) - \ln(\bar{y}_i)\right)}{\sum\limits_i \left(\ln(x_{ij}) - \ln(\bar{x}_{ij})\right)^2}$$

and

$$\hat{\alpha}_j = ln(\bar{y}_i) - \hat{\beta}_j \ln(\bar{x}_{ij})$$

where $\bar{x}_{ij}$ and $\bar{y}_i$ are the arithmetic means of x_{ij} and y_i, respectively.

By itself, the coefficients β_j provide a measure of the influence of the factor x_j on the

[1]Here, we use a logarithmic transformation as the ordinal regressors may have different scales, and hence elasticities derived from the log transform make results easier to compare across factors

service quality score y by providing a measure of the elasticity of quality score with respect to an individual factor; the interpretation is that, if β_j were equal to 0.1 for example, a 1% increase in the factor measured by x_j would be associated with a 10% increase in quality score. Similarly, if β_j were instead -0.1, a 1% increase in x_j would be associated with a 10% decrease in quality score.

Although this coefficient shows correlation, two important points must be made. First, a large coefficient by itself does not show causation, nor does it rule out both factors being correlated with a third factor that has greater significance in the actual operational environment. Second, a large coefficient does not rule out the fact that many other factors will also impact quality score; indeed, it is expected that many factors will impact customer satisfaction, and many of them will show up in the regression analysis. Finally, a nonzero coefficient does not rule out the possibility that the factor being considered actually has little impact on customer satisfaction; to consider this, a second factor is considered along with the regression coefficients: the t-statistic for each factor's coefficient.

To consider if a nonzero coefficient of elasticity is actually meaningful, we test the null hypothesis in which the actual coefficient should be zero. To do this, we first calculate the t-statistic for factor x_j by:

$$t_j = \frac{b_j}{SE(b_j)}$$

where $SE(b_j)$ is the standard error of coefficient b_j. Using the t-distribution, we find the 2-sided p-value such that $p = \Pr(\beta_j^* = 0 | x_{ij}, y_j)$; that is, the probability that the true correlation for the entire population β_j^* is actually zero. The way the t-statistic is used in our assessment is described below, and shown in Figure 2-2.

Regression analysis with categorical dependent variables

Because many of the factors of interest will be binary (e.g. yes or no) rather than numerical, the logarithmic transformation is not always appropriate; instead, categorical factors can be considered using linear regression. In this case, consider the set of categorical factors $j = [1, M_{categoric}]$; to provide a statistical analysis of these factors through regression, the categorical process measurements w_{ik} are codified with dummy variables z_{ik} such that 0 and

1 are used in place of the binary categories. For example, to study the impact of contacting a customer, one could codify a customer being contacted as "1" and a customer not being contacted as "0"; the regression coefficient will then provide the expected increase in quality score when a customer is contacted rather than not contacted.

The CEF for a single input j is now:

$$(\mathbb{E}[y_i|x_{ij}]) = \alpha_j + \beta_j x_{ij}$$

For a given set of points y_i, x_{ij}, the estimated parameters $\hat{\beta}_j$ and $\hat{\alpha}_j$ are:[28]

$$\hat{\beta}_j = \frac{\sum_i (x_{ij} - \bar{x}_{ij})(y_i - \bar{y}_i)}{\sum_i (x_{ij} - \bar{x}_{ij})^2}$$

and

$$\hat{\alpha}_j = \bar{y}_i - \hat{\beta}_j \bar{x}_{ij}$$

where $\bar{x}_{ij}$ and $\bar{y}_i$ are the arithmetic means of x_{ij} and y_i, respectively.

2.3.2 Using the results of the regression analysis

Once the t-statistic for each factor has been calculated along with the regression coefficients or elasticity coefficients, the two are used together to identify the factors that potentially have the most meaningful impact on customer satisfaction and hence are worth the most consideration; this process is depicted by Figure 2-2. If a factor has both a high coefficient and a high t-statistic (accordingly, low p-value), this means it is highly likely the factor either has a meaningful impact on customer satisfaction, or is reflecting another factor that does, and hence should be considered first in the follow-up work described later in this thesis. If a factor has a high t-statistic, but also a low coefficient, it means that it is likely that the factor is correlated with quality score, but may have less of an impact compared to the factors with higher coefficients, and should be prioritized behind them when considering operational changes. If a factor has a high coefficient but also a low t-statistic (high p-value), this means it is possible that the factor has a large impact, but the data was so scattered when considering this factor that it is inconclusive; this factor should be prioritized third, and should be addressed by collecting additional data before considering any operational changes. Finally, factors that have both a small coefficient and a small t-statistic indicate factors that are unlikely to be correlated with or have an impact on customer satisfaction. Although these should be ignored when considering operational changes, these factors may also be long assumed by managers within the organization to impact customer satisfaction, requiring further attention through additional data to make the important case for being able to shift the focus away from such factors.

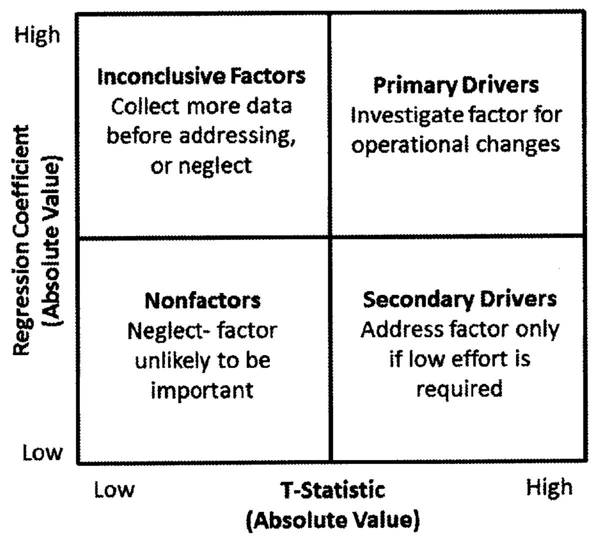

Figure 2-2: 2x2 matrix showing how to interpret different drivers based on the results of the statistical analysis

2.4 Conclusion

By the end of the first step of the thesis framework, we will have developed the metric that will be used to judge service quality[2], investigated the drivers of service quality in the target process through customer surveys, employee interviews, and statistical analysis, and identified the primary drivers of service quality for which we would like to update the target process to address. At this point, we are ready to proceed to the second step of the

[2]As will be demonstrated by concrete example in the case study in Chapter 4

framework: the actual process improvement to address the primary drivers of quality. While this first step is valuable in guiding process improvement, the next chapter will describe why improvement will be an iterative process requiring experimentation- and hence, why this first step will become a routine rather than a one-time exercise for organizations that want to drive service quality significantly higher.

Chapter 3

Updating Process to Reflect Drivers of Service Quality and Cost

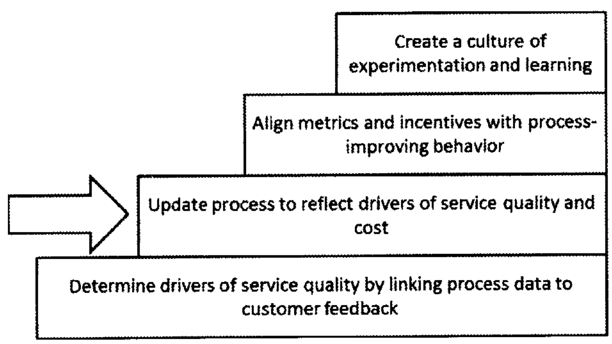

Figure 3-1: Step 2 of the thesis framework

Once service quality has been reliably measured and the drivers of service quality within the processes of question have been identified in the first step of the thesis framework, we are ready to progress to the second step: the process changes designed to address and act on the identified quality drivers. In this chapter, we first briefly cover some of the philosophies of operations management that provide a foundation for this step. Building on these past ideas, we introduce a few additional concepts to support this step, and finally close with a brief discussion on the role of operating cost within the largely quality-focused thesis framework. While the concepts in this chapter are general, they are best explained by example via case study in the following chapter (Chapter 4).

3.1 Existing Tools in Operations Excellence

The strategy for accomplishing the second step of the thesis framework takes root in several existing philosophies for driving operational excellence. Before tracing the methodology used in this thesis, we begin with a brief description of this past work.

3.1.1 Lean Thinking

Lean manufacturing is a socio-technical operations philosophy largely derived from the Toyota Production System in the early 1990's [29]; its primary tenets include the elimination of waste, or *muda*, from an organization through the removal of activities that do not add value, as well as a respect for all stakeholders involved in the organization to ensure the success of the improvement efforts in the organizational realm. In *Lean Thinking*, Womack and Jones lay out the following five lean principles for achieving operational excellence:

1. *Specify Value:* In lean manufacturing, value is defined as the products and services a customer ultimately wants; while there may be many customers to consider in a modern organization with several stakeholders, it is still the case that value is created by an organization when a customer wants it. Specifying value then allows one to distinguish between activities that add value and those that do not, allowing for an efficient use of resources leading to direct value creation for the customer.

   In service operations, value will be defined by the needs of the customer; thus, the first step of the thesis framework, the identification of service quality drivers, helps to define value, allowing for the process redesign to focus on activities that drive customer satisfaction. While many examples in service operations, such as that of Atlantic Energy in the case study, require necessary, non-value added activities required by regulation or true company needs, the same tools may be applied to determine the most efficient way to fulfill the needs of all stakeholders, including the company.

2. *Identify Value Stream:* The second principle involves tracing the entire process in which value is created and delivered to the customer to classify all of the activities into three categories: activities that add value, activities that do not add value, but are required

based on current technology or policy, and activities that do not add value and are unnecessary; Womack and Jones classify the last two activities as "Type 1 *muda*" (activities that do not add value, but are necessary) and "Type 2 *muda*" (activities that do not add value and are unnecessary). Thus, by tracing a process and looking for sources of waste, one can improve the operation's efficiency in delivering value to its customers by limiting Type 1 *muda* and eliminating Type 2 *muda*.

3. *Flow:* Once non-value added activities have been reduced or eliminated, the third lean principle involves linking the value-added activities together to achieve flow, or the continuous movement of material, information, or people through the process. Ideally, a process will achieve single piece flow, in which one item is worked on at a time in each part of the process at the rate needed to meet customer demand, oftentimes through the use of production cells that group all resources needed for a single piece together.

   In service operations, achieving flow means looking beyond the obvious boundary of the process to consider the customer's actions offline as well. When customers leave the process and require work offline, or customers are pulled into subprocesses in which their information waits to be processed within the organization, flow is disrupted in the process; therefore, there will be a strong relationship between achieving first contact resolution (FCR) and achieving flow in service operations.

4. *Pull:* The fourth lean principle, pull, indicates that production should be triggered by customer demand rather than a pre-set production schedule; when combined with flow, this creates a process with limited inventory, easy identification of errors, and less waste. In service operations, pull is typically achieved when customers trigger the process by requesting service, and thus inherent to the operation.

5. *Perfection:* The final lean principle is the continuous pursuit of perfection. As organizations employ the first four lean principles to improve the organization, they uncover further opportunities for improvement, creating a cycle of continuous improvement that persists over time, rather than consisting of an ephemeral event. This principle requires an organizational mindset that continues to use data and pursue improvement,

and very much coincides with the next paradigm summarized here, “High Velocity Operations”.

3.1.2 High Velocity Organization

In the book *High Velocity Edge*[27], Spear describes a “High Velocity Organization” as one that outpaces competition through both a structure that enables the highly complex systems of the organization to perform at a high level and a focus on discovery and experimentation that continuously enables it to improve; in this case, the role of a manager is not to merely make decisions, but to discover, learn, and facilitate the same behavior in others. Thus, according to Spear, an organization will "Discover- not think- its way to good design." This system, largely based on the author's work in studying the Toyota Production System, defines four capabilities that enable a High Velocity Organization to succeed:

1. *Specifying design to capture existing knowledge and building in tests to reveal problems:* High Velocity Organizations always begin with work that is both highly specified to capture the current best practices known by the organization and designed to reveal problems with the process. By designing work in this way, it becomes easier to root out errors and improve the organization so that it is always becoming more competitive. This is accomplished by first defining the output of the system clearly (similar to value definition in lean manufacturing), defining the pathways by which work moves through the system to become output (similar to identifying the value chain), defining the connections that trigger when and how work moves from one step to the next, and finally defining the work methods used at each step. This method is very much aligned with lean philosophy, with an emphasis on achieving flow to make issues in the system are clearly obvious to all involved so that they can be addressed immediately.

2. *Swarming and solving problems to build new knowledge:* While many organizations try to design a perfect process the first time, it is impossible to do so in reality, meaning that high-performing systems will require many iterations to achieve. As problems inevitably arise in the process, a High Velocity Organization responds by investigating the errors, diagnosing the underlying problems, designing a countermeasure to solve

the problem, testing the countermeasure in the actual process, and then observing the results to feed back into the next iteration of this process. Thus, improvement progresses by the scientific method, in which hypotheses are made before the change is introduced, an experiment is run to verify or refute the hypothesis, and the outcome of the experiment- whether it verifies or refutes the original hypothesis that led to the change- is incorporated into the organization's best knowledge of the system. While many organizations will work around problems or assume that they are not systematic, a High Velocity Organization attacks all problems on the assumption that they will occur again if not designed against.

3. *Sharing new knowledge throughout the organization:* As experiments are run and results are recorded, the next capability demands that the knowledge earned through experimentation is dispersed effectively throughout the organization so that all decisions and organizational designs are based off of the best information available to the organization. This capability, however, requires that not only are results shared, but the processes by which the results were obtained are as well to create a company that can problem solve, experiment, and learn at every level of the organization- not just amongst change managers.

4. *Leading by developing the first three capabilities in others:* Finally, the fourth capability requires teaching others in the organization to execute on the first three capabilities. By enabling the entire organization to design processes, problem solve through experimentation, and disseminate both the procedures and results to the rest of the organization, the entire company develops a strong capability in managing complex operations and avoids becoming capacity constrained on discovery when managers with these capabilities run out of available work hours to do more.

Many of these concepts are directly applicable to improving a service operations organization after the drivers of customer satisfaction have been identified in the first step of the framework. When addressing drivers of service quality such as errors or other behavior that frustrate customers, feedback loops will be built into the process as seamlessly as possible to provide feedback on where problems may be occurring. Experimentation is especially applicable after the first step; once both a clearly defined metric for customer satisfaction and a method for extracting information from it as described in the previous chapter are developed, it can be used iteratively as process changes are designed to address the quality drivers and experiments are run to verify the impact of the changes through a new set of customer satisfaction measurements. Therefore, the first two steps of the framework must occur continuously and iteratively in a truly outstanding service organization.

At the same time, the last two capabilities of a High Velocity Organization are equally important in developing an organization that successfully builds expertise over time and fosters an environment in which all levels of the organization can contribute to the improvement. Although this effort depends on the role of several stakeholders and encompasses cultural and political issues within an organization, addressing such issues is needed to ensure long-term success; these issues are addressed in the last two steps of the thesis framework.

3.1.3 Other related works in service quality

One final work of note for this thesis is the thesis of Avijit Sen of MIT LGO, 2009.[25] As part of a six-month research study at the customer contact center of Dell, Inc., Sen worked with the company to achieve a simultaneous improvement in service quality and cost competitiveness in customer service operations through metric redesign. By applying lean principles to his analysis of Dell's existing metrics, Sen developed a new set of metrics aimed at eliminating *muda* in the company's service operations and improving the tracking of service for each customer. Such concepts are both relevant and useful in this study, and have thus influenced the analysis of Atlantic Energy's service operations in the case study. One primary difference, however, is the focus of each study: while Sen focused on improving the call center metrics for the entire organization, this thesis focuses on the improvement of a specified process guided by customer satisfaction scores and existing contact center metrics, placing the emphasis on process improvement rather than measurement improvement. While the two approaches differ, both are equally important and even supporting of each other, and both should be considered in an organization.

3.2 Developing Operational Changes for Improved Service Quality

Drawing from past philosophies of operations management, we now create a few simple ideas to execute the second step of the thesis framework: the redesigning of processes in support of the drivers of high quality operations determined in the first step of the framework. It is important to note that every effort to align operations with customer perception of quality will be unique and necessarily tailored to the particular drivers uncovered; however, in this section, a few general concepts are given for accomplishing this step.

3.2.1 Defining and addressing value from drivers of service quality

Many of the existing philosophies for operations management begin with the idea of value or outputs, explicitly defining what the stakeholders involved in a process need to get out of the process in the first place. In this framework, process improvement begins much the same way, particularly with the first step of the framework in which drivers of quality according to customer perception are determined both qualitatively and quantitatively. At the same time, the organization must include the needed outputs from internal stakeholders as well; although these outputs may be classified as "non-value added activities" according to lean philosophy, it is useful to consider the minimum output that the process must produce.

As stated before, process redesign is necessarily a more ambiguous step than the specification of value, and hence will be highly dependent on iterative experimentation, as described by the High Velocity Organization philosophy; however, here we include a few concepts for beginning the redesign process based on a few non-comprehensive categories into which possible drivers of customer satisfaction fall: the elimination of undesirable steps or factors, the increased impact of desirable factors, and the decrease of critical cycle times.

1. *Elimination of undesirable steps or factors:* Many activities can potentially fall into this category, including frustrating subprocesses or requests of customers, error and rework, undesirable behavior from customer service representatives, unsatisfactory output, and many others as well. Those activities that can be eliminated should clearly be addressed

first. In other cases, such steps may be unavoidable, but potentially experienced by more customers than needed. In this situation, the process should, at a minimum, minimize the number of customers who experience undesirable parts of the process, as described by the funnel concept below.

2. *Increasing the impact of desirable factors:* When an analysis of quality drivers indicates that customers desire more of a certain factor, the obvious response is to determine the cost of providing more of that factor and then provide it if the cost-benefit balance is reasonable. In practice, however, it is important to change the process in small increments and collect more data after the process has been changed- oftentimes, the driver will offer diminishing returns, making it important to only expend resources on the factor while it is still impactful to do so.

3. *Reducing critical cycle times:* While most forms of waiting will decrease customer satisfaction, analysis reveals that some waiting times matter more than others. For those that matter most, two options exist depending on whether any customer segmentation is significant for the wait time in question. For cases in which a cycle time impacts one customer segment more than another, a priority queue may be appropriate if 1) the customers can be easily divided, and 2) no more than 25-30% of customers fall into the priority segment. In cases for which a reduced cycle time will increase customer satisfaction for all customers, one solution, if the process has already been trimmed of *muda*, is to add a capacity buffer at the step in question and eliminate batching in favor of single-customer flow. While adding a capacity buffer to reduce utilization may seem to contradict lean principles, it is a powerful tool in cases for which the reduction of customer wait time at a particular point is strategically superior to cost reduction; in many cases, the extra labor cost to increase slack capacity is small compared to the benefit of the cycle time reduction in question.

3.2.2 Reducing waste via the "funnel" concept

One common idea throughout this chapter has been the notion of reducing a process to the activities that actually create a desired output. In manufacturing operations, standard work is often defined and typically only one primary path exists to move material through the process. However, in service operations, standard work is nearly impossible as the item moving through the process is a person (or information about a person); thus, variation is not only expected, but likely to be high. As a result, service operations generally involve many possible paths or processes that a customer might experience to resolve the customer's issue or accomplish the desired output. In most cases, some of the possible paths will be more efficient, or perceived as higher quality (higher customer satisfaction) than others. To address this fact in process redesign, we introduce the concept of the funnel.

The funnel is presented in a generic form in Figure 3-2. The top of the funnel represents the point at which all customers enter the process in question for the first time. Customers who proceed all the way through the funnel are said to have completed the process on the "ideal path," or the path which most efficiently achieves the required output; this path will have the highest ratio of value-added activities to all activities. Customers along the ideal path are typically more satisfied and spend less time in the process. At some points in the process, however, customers are funneled off of the ideal path into other subprocesses where additional steps not required on the ideal path are taken; as a result, not all customers who enter the process complete it by the most efficient path possible. In addition to policies and defined subprocesses that funnel customers off of the ideal path, mistakes at any part of the process often funnel customers out as well. Once off the ideal path, customers are much more likely to experience rework, extra iterations in the process, extra steps, and other inefficiencies that the lower customer satisfaction and increase operating expense. An example of the funnel put to use is shown during the first case study in Chapter 4.

Therefore, one strategy for improving service operations is to identify the points in which customers are funneled off the ideal path and to then close off the leaks to reduce the number of customers being diverted, with more effort placed on closing off leaks to steps that customers have identified as particularly frustrating. This practice can take many forms; for example, policies that unnecessarily take customers off the ideal path can be changed, such as a policy that requires all customers to fill out a form with additional information when the information may only be required from a subset of customers. In other cases, the extra steps may be beneficial for the organization when imposed on all customers, but

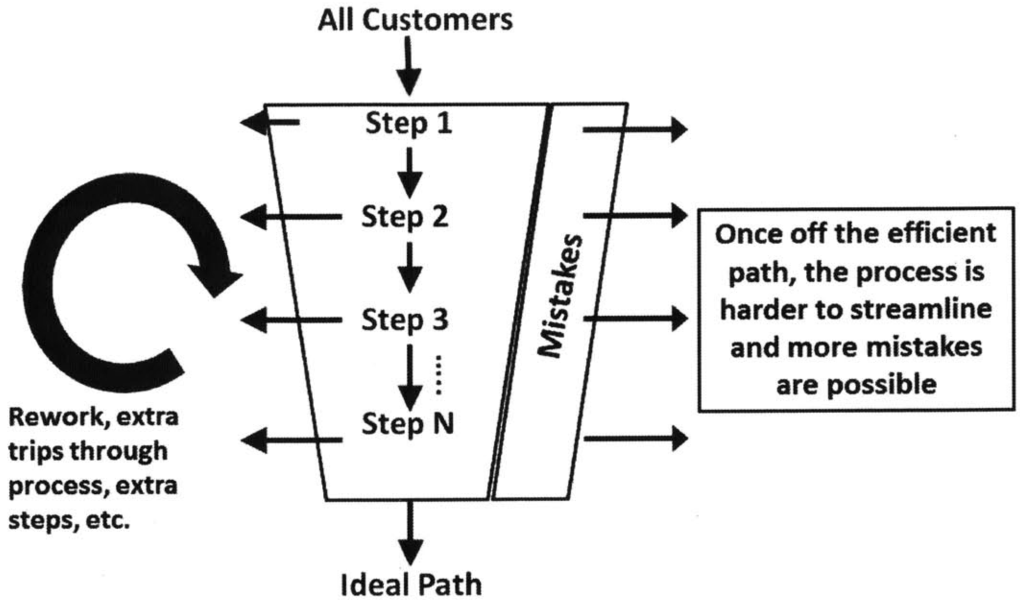

Figure 3-2: The funnel, by which customers are potentially diverted off the efficient path at different points of the process

perhaps more effective on certain segments compared to others, in which case the process can be more selective about the segments it takes off the ideal path. Many times, however, customers fall off the ideal path due to errors; by using the concepts from Spear, however, built-in inspection can help identify the points in the process that lose the most customers due to errors, and countermeasures can be designed, tested, and redesigned through iteration to help repair leaks in the funnel.

3.2.3 Improvement and learning through experimentation

Finally, the first two steps of the thesis framework- the identification of service quality drivers through measurement and analysis and the improvement of the process to act on these drivers- will almost certainly be most effective when iteration on the two steps is possible, as described by Spear's second capability of a High Velocity Organization. While the measurement and analysis of quality drivers gives a strong head-start into improving processes in service operations organizations, prescribed changes to the processes will require the same data collection used to initially diagnose the condition of the organization to both assess the impact of the changes and plan for the next iteration of changes as well. Experiments need to be carefully managed with stringent data collection and analysis procedures, especially compared to processes with less variation such as in manufacturing; however, such variation can increase the importance of experimentation as well. The topic of organizational experimentation and learning is covered further in the fourth step of the thesis framework.

3.3 Managing the Tradeoff between Quality and Cost

One of the greatest questions any manager faces in operations is the way to manage the tradeoff between cost and quality; indeed, this tradeoff is seen as one of the primary strategic decisions in operations.[4] In the context of this thesis, this tradeoff can play a major role in determining the extent to which customer satisfaction drivers can be pushed at the expense of cost. At first blush, it seems that pushing customer satisfaction upward can involve substantial increases in operating expenditures. While an explicit tradeoff between service quality and cost in quantitative monetary terms would be ideal- particularly in the case of the organization in the case study, which will be apparent in the next chapter- this tradeoff is typically made by managerial intuition in many organizations with effort being put into quality until it seems cost prohibitive to continue.

While the cost / quality tradeoff is an important topic deserving of much attention, the purely strategic part of this decision will be outside the scope of this thesis, as will any attempt to characterize the direct impact of service quality on revenue in a quantitative matter. Instead, this thesis will focus on the cases for which quality and cost are aligned, or for which it is reasonable to assume that acting on a service quality driver will not lead to a large increase in operational expense. These concepts are illustrated in Figures 3-3 and 3-4 below. In Figure 3-3, the problem of cost / quality tradeoff is shown for a typical organization in which the current state of the organization is below the operational efficient frontier. Rather than focus on this tradeoff, this thesis primarily addresses operational changes that are aligned with cost, or at the very least not significantly opposed to cost, resulting in an organization that reaches the efficient frontier as shown in Figure 3-4. In cases for which operational improvement is desired to improve service quality, as is the case

Figure 3-3: Traditional view of quality / cost tradeoff

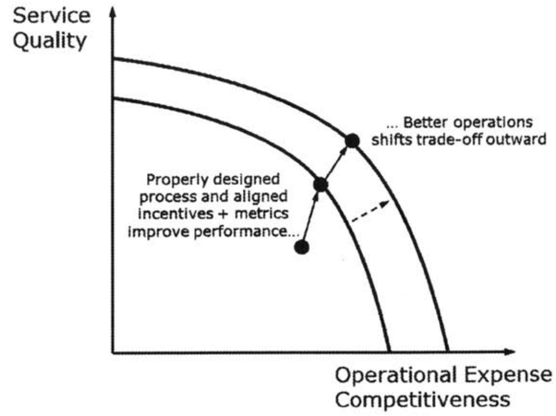

Figure 3-4: Reaching and expanding efficient frontier through ideas from thesis framework

in the case study with Atlantic Energy, we assume that, at a minimum, the focus of the improvement efforts is on quality, with a secondary goal focused on keeping costs nearly the same.

Thus, it is instructive to first understand where the sources of alignment between service quality and operational expense competitiveness arise from. Here, we focus on two primary reasons the two may be aligned in service operations: the impact of waste (or *muda*) on both quality and cost, and the impact of quality on cost. These two concepts are described below.

$$\begin{tikzcd} & Quality \arrow[dd, "Concept\#2"] \\ Waste \arrow[ur, "Concept\#1"] \arrow[dr, "Concept\#1"'] & \\ & Cost \end{tikzcd}$$

3.3.1 Quality / cost alignment from eliminating *muda*

As previously stated, one of the primary principles of lean operations involves the reduction of non-value added activities, or activities that do not lead to a successful output. Tracing the argument from before for the reduction of non-value-added activities to increase CSAT, we can see how quality and cost can be aligned in service operations due to their common relationship with waste. While in manufacturing, customers may not experience the presence of non-value-added activities, customers do experience some of the *muda* that may be present in service operations, leading to a new relationship between cost and quality via waste that may not be present in traditional operations thinking. For operational improvements involving increasing quality through the elimination of *muda*, therefore, the two goals are assumed to be aligned, or at least non-competing.

3.3.2 Quality / cost alignment from the impact of quality on cost

In service operations, because customers both flow through the process and ultimately judge quality, any process that does not successfully serve a customer will necessarily result in both low satisfaction and rework. Any rework resulting from low quality, in turn, will drive up the input volume into the system as customers must re-enter the process as seen in Figure 3-5. While rework has a multiplier effect on input volume in service operations as it does in manufacturing, the fact that customers are experiencing the process and judging the quality in service operations means that rework will cause both an increase in cost and a decrease in customer satisfaction or perceived quality. Therefore, it is possible to make a process both more efficient and more effective in serving the customer, allowing for opportunities in which an increase in cost when improving service quality may be offset by the decrease in cost resulting from fewer unsatisfied customers who need to re-enter the process as rework. While increasing customer satisfaction or service quality may require added resources, and hence added costs, in some cases a manager must recognize that poor service incurs a cost as well due to increasing volume from rework.

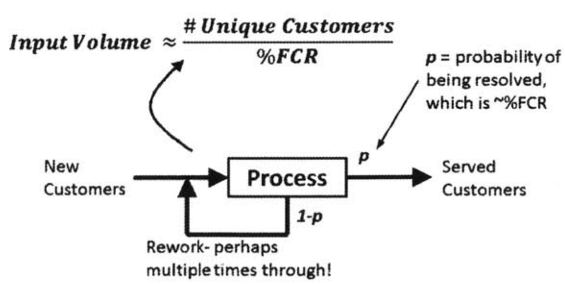

Figure 3-5: When First Contact Resolution (FCR) is not achieved, the customer re-enters the process, creating a multiplier on the input volume the process experiences

3.3.3 First contact resolution trade-off with average handle time

One particular manifestation of the quality / cost tradeoff, as described before, is the tradeoff between first contact resolution (FCR) and average handle time (AHT), or the amount of time spent on each customer. While FCR will clearly impact quality and AHT will impact cost, the relationship between the two resulting from the process dynamics shown in Figure 3-5 creates a quantifiable trade-off between the two. Using the service operations model explored in Chapter 5 derived from Erlang C queueing theory, we have developed the curve shown in Figure 3-6. This curve represents the rate of change in FCR with respect to increasing AHT that can be accomplished with no additional resources in a customer contact center at a constant service level per the Erlang C queueing model, normalized by the number of agents. That is:

$$\text{FCR Tradeoff} = \frac{1}{\#Agents} \frac{\partial AHT}{\partial \%FCR}\bigg|_{Constant\ SL}$$

Here, AHT is the Average Handle Time, %FCR is the percentage of calls answered correctly the first time, and "Constant SL" represents performance held at a constant service level, namely 80% of calls answered within 20 seconds as dictated by government regulators in the case of Atlantic Energy.

For example, for a tradeoff metric of .04 from the chart, a 200 agent customer center could increase the average handle time of the entire center by 8 seconds with the same number of agents and the same achieved service level if the increased time allows a 1 percentage point increase in the number of customers successfully served on the first contact. Armed with this knowledge, managers can use this information to drive cost-neutral or cost-saving changes that increase FCR- and hence quality- by breaking the common assumption that average handle time minimization is paramount in a customer care center. Further details, including a description of the model, can be found in Chapter 5.

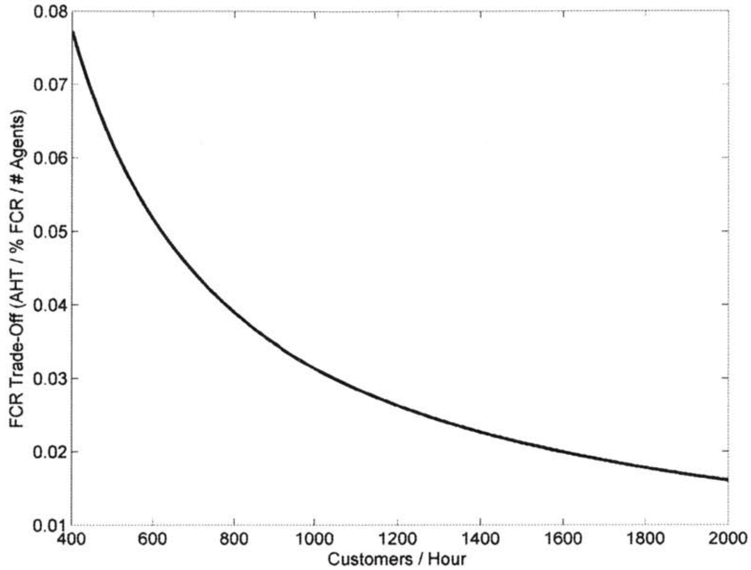

Figure 3-6: Efforts to increase FCR may also increase AHT at no additional cost to the customer care center's operations

3.4 Conclusions and Path Forward

At this point in the framework, service quality has been measured and its drivers have been identified, and the processes under consideration have been redesigned to address the findings of the analysis done during step one. Although there is no completely generalizable method for improving a process to reflect quality drivers, many concepts have been introduced that form a common theme for this step: customer and firm value, and required output, should be defined based on both the qualitative and quantitative analysis in the first step, and processes should be redesigned to avoid factors that negatively correlate with CSAT, increase factors that positively correlate with CSAT, and increase slack capacity by whole or by segment for critical cycle times. While the process should be designed so that as many customers as possible are kept on the most efficient path that achieves the required output, accomplishing all of these goals may require an iterative process of measuring CSAT, experimenting, learning, and repeating. To illustrate these first two steps of the framework, the next chapter summarizes a case study in which the framework was applied to a large energy utility, Atlantic Energy. Following the case study, the thesis progresses to the final two steps of the framework as it shifts its focus to the behavioral side of the framework using concepts from game and decision theory.

Chapter 4

Case Study I: Goal Definition and Operational Architecting at a Multinational Energy Utility

A case study to illustrate the first two steps of the framework

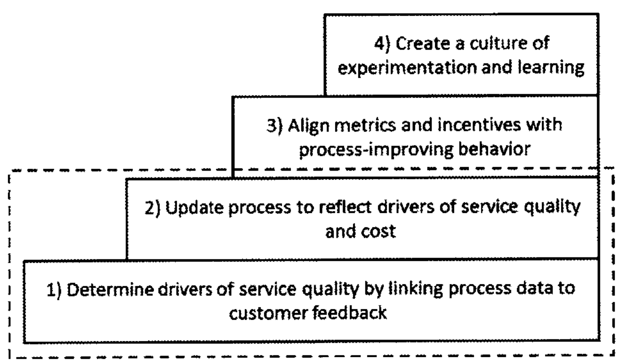

Figure 4-1: Steps 1 and 2 of the thesis framework are covered during Case Study I

4.1 Introduction

To test out the thesis framework, a project was conducted in partnership with a multinational energy utility, Atlantic Energy.[1] The partnership with Atlantic provided a business environment in which the methods of the framework could be designed, analyzed, and tested through multiple iterations, leading to a refined version of the framework built by observation and empirical evidence along with theory and intuition. A summary of the first two steps of the framework as applied at Atlantic is presented in this chapter as a case study to provide an example of the framework's application.

4.1.1 Company context

Atlantic Energy is a privately-owned, multinational utility serving over 3.4 million electric customers and 3.5 million gas customers in the U.S. throughout four states on the east coast. For both the gas and electric businesses, Atlantic will both generate and distribute the commodity to customers; while Atlantic acts as a distributor for all of its footprint as a regulated utility, customers have an option to choose Atlantic amongst a number of third-party suppliers for each commodity. Its U.S. business has over 15,000 employees and recorded a $2 Billion operating profit on $11 Billion in revenue (approximate) on regulated assets in FY 2012/2013.

Due to the nature of its business, Atlantic is an operations-intensive company with many processes and subprocesses that must be carried out to serve its customers. While many capital-intensive processes- such as the generation, maintenance, transmission, and distribution of gas and electricity commodities- are needed, these processes are not as visible to the customer as the supporting service operations. For strategic reasons described in the following section, the customer-facing service operations of the company were chosen for this project.

Because Atlantic is a regulated utility, it operates without competition for the distribution side of its business. Like all regulated utilities, it therefore has its pricing set through a process called ratemaking to avoid monopolistic pricing of energy. Ratemaking involves a

[1] As stated in the thesis preface, the company's name has been obfuscated throughout this thesis

rate case in which a regulatory agency sets the rate of return the utility is allowed to earn on its rate base, or net assets used to provide the public with energy. Once the allowable rate of return (R) has been set by the regulatory agency, the required revenue needed by the utility to earn the allowed rate of return will simply be the allowed return added to the necessary operating expenses, as given by the rate formula:

$$Required\ Revenue = (Operating\ Expense) + R * (Net\ Assets)$$

4.1.2 Objective and motivation

Because the return on equity of the company is largely set by the negotiated rate of return through rate cases, management will look to maximize the allowed rate of return. Although ratemaking is a complex legal process involving many variables, there is evidence that utilities providing higher levels of customer satisfaction are rewarded with a higher allowed rate of return[12], as a company in a non-regulated industry would also typically experience. As stated in the thesis introduction, utilities in the first quartile of JD Power's rankings earn a regulated return on equity that is 0.5 percentage points higher than than that of utilities in the fourth quartile. In an industry with regulated assets that may be in the \$10 billion to \$100 billion range, this can lead to a \$50 million to \$500 million impact on shareholder value.

In response, senior management at Atlantic Energy has defined four strategic goals for the company, two of which are cost competitiveness and customer responsiveness, or service quality. To align employee activities and initiatives with these goals, key performance indicators (KPIs), such as customer satisfaction metrics and operating expense targets, were defined for each goal and extensive improvement efforts were started to drive toward them. As a result, Atlantic partnered with the Massachusetts Institute of Technology to address these two strategic goals in the context of the account initiation (A.I.) process, the process by which over 3/4 million customers start electric and gas service in the US each year. This process entails all actions from the customers first call to service connection on-site.

Therefore, we will use this partnership as a laboratory within which to test the thesis framework by applying it to the improvement of Atlantic's account initiation process. Atlantic's management team has pre-chosen the account initiation process for study by MIT due to its wide exposure to customers (experienced by 750,000 customers annually) and the large number of customer satisfaction surveys generated as a result of the process; within this context, we set out to test the thesis framework on the process as described in the rest of this chapter.

4.2 Measuring and Improving Service Quality

In support of the strategic goal to improve service quality, the investigation begins with the first step of the framework: determination of the quality drivers through the use of customer and process data. In this section, we trace the process by which the primary drivers of customer satisfaction were identified at Atlantic Energy. This section begins by describing the organization's current measurement of customer satisfaction. After describing the service quality metric used in this investigation, the section then traces the three-pronged approach to determining service quality drivers at Atlantic, including two qualitative methods, employee focus groups and customer surveys, and one quantitative method, statistical analysis using customer satisfactions scores and customer process data.

4.2.1 Measuring service quality via customer satisfaction surveys

Like many organizations with widespread service operations, Atlantic Energy has an extensive system in place for measuring and tracking customer satisfaction which, as previously mentioned, forms a KPI on which 25% of managerial discretionary bonuses are based. Atlantic's customer satisfaction score (CSAT) is based on a monthly survey conducted by a third party company and consists of up to 46 multi-part questions gauging the customer's view of the company, services, interaction with employees, and general satisfaction with Atlantic; to limit the time spent on each survey, customers are generally asked a subset of the questions based on their experiences and answers to different questions. Some of the most relevant questions from the survey for this project include the following:[2]

- *On a scale of 1-10, what is your overall satisfaction with Atlantic Energy?*

[2]Note: questions listed are not necessarily verbatim from Atlantic's survey

- *What is your reason for giving Atlantic a (Insert Number)?*
- *On a scale of 1-10, how satisfied are you with the service provided by Atlantic Energy?*
- *Was the reason for your call resolved on the first call?*
- *On a scale of 1-10, please rate your experience with the representative you spoke with.* (If applicable)

Each month, approximately 2,000 customers are randomly selected amongst all customers who 1) had contacted Atlantic Energy by phone or web during the previous month, and 2) had indicated they would be willing to take a customer satisfaction survey in the future. A third party customer intelligence vendor then contacts all of the customers selected for a survey that month and administers the survey by phone, recording the numerical scores and verbatim comments from the customers during the process. Finally, the vendor compiles all responses into a spreadsheet and passes it to Atlantic at the end of the month for review by the customer data team.

From this data, the company key performance indicators for customer responsiveness defined by senior management are calculated and published throughout the company. Of these KPIs, the most important are the Net Satisfaction Score- given simply by the percent of customers giving a "detractor" score (1-4 on a 10 point scale) subtracted from the percent of customers giving a "promoter" score (8-10 on 10 point scale)- and the average overall satisfaction score on the ten-point scale.

Critique of current system

While the survey provided a wealth of much-needed data for the project, there are several shortcomings in its design that would ideally be fixed in a redesign of the survey. Some of the most important are as follows:

- *Ability to predict external measures of customer satisfaction:* While the customer surveys do offer a form of customer satisfaction measurement after customer interactions, no work had been done to see if Atlantic's customer satisfaction measurements correlated with those of sources such as JD Power, meaning that an increase in the customer satisfaction scores may not lead to an increase in the JD Power scores. While an internal satisfaction score would ideally be designed, tested, and redesigned according to this standard if there is value in signaling service quality during a rate case, the project proceeded with this internal measure due to both time constraints and the likelihood of the score providing useful information even without checking the correlation with external scores.

- *Convenience selection bias:* Customers are only eligible for the survey if they have contacted Atlantic during the month. While customers with recent contact with the company are a valuable source of information in these surveys, some customers who did not contact Atlantic during the month, but are willing to take the survey, would ideally also be polled.

- *Separation of customer perception and expectation:* The customer survey asks for numerical scores in each category; however, as described previously, the survey would ideally measure expectation and perception with different questions, as the difference between the two has been found to be more relevant than an absolute score of satisfaction.[21]

- *Lag between customer interaction and survey:* Because customers are contacted for the survey in the month following their original experience with the company, customers may be contacted 3-6 weeks after the initial event they are asked about during their survey. Ideally, customers would be contacted within a week of the interaction, after they have an opportunity to have their problem resolved but before the details of the interaction are forgotten.

Despite these shortcomings, the existing survey provided a source of customer satisfaction data that was helpful, meaningful, and available. Survey results tended to show consistency month to month and provided some insight into the customer's opinion about the organization. Because the survey was used in KPIs defined by senior management, changing the survey would require a lengthy, extensive effort that would not have been completed during the timeline of the project. For these reasons, the current CSAT survey was chosen to serve as the quantitative measure of service quality during the project.

4.2.2 Qualitative assessment through employee interviews

The investigation into service quality drivers begins with one of the more readily available sources of information available to the organization: customer-facing employees. In this case, some of the customer service representatives (CSR) who carry out the account initiation process were chosen to take place in focus groups with the intention of gathering insight into the drivers of customer satisfaction, including any potential problems with the current processes. To accomplish this, 43 customer service reps were interviewed both alone and in groups of up to 12 over a two-month period. Questions varied from group to group and were largely adaptive based on the flow of the conversation. As suggested previously when laying out the framework, small focus groups were preferred as more employee-hours of information could be gathered without requiring as much time from employees away from their work stations.

During the course of the employee interviews, a hypothesis for the drivers of service quality in the account initiation process was developed and refined; some of the high-level drivers are shown in an issue tree in Figure 4-2. A non-comprehensive list of insights obtained from the interviews and focus groups include the following:

- Barriers to communication between the customer-facing employees and the account processing employees in the organization frustrates customers and slows down the process as employees are unable to check on customer information and the status of their progress through the process when customers call in. Employees identified this lack of communication as a large driver of process rework and customer dissatisfaction.

- Employees indicated many errors stemming from improper handling of customers entering the process, resulting in unnecessary rework and multiple passes through the process. In particular, employees indicated that many other customer representatives would erroneously request unneeded information from a customer, resulting in more work for the customer and more work for the company in processing the additional customer contacts.

- Some employees indicated a pressure on meeting time-based performance metrics such

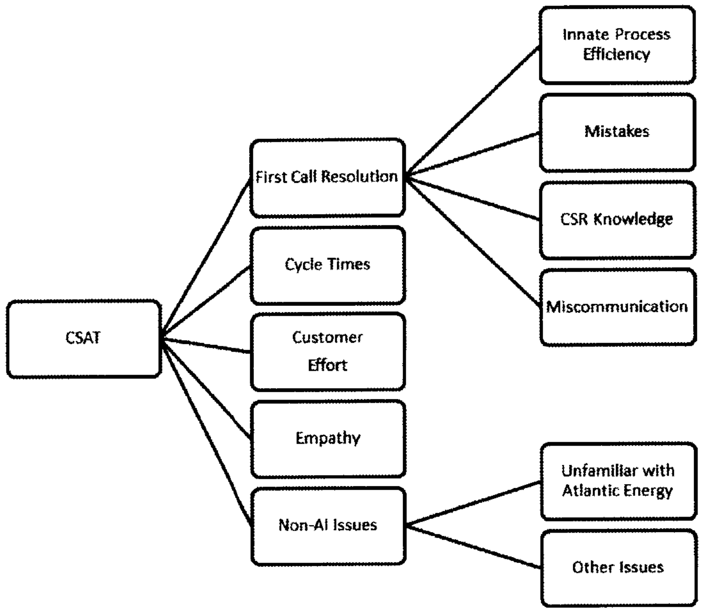

Figure 4-2: High-level issue tree serving as road map for investigating CSAT drivers

as average handle time (AHT) during work, perhaps leading to a trade off of emphasis on quality assurance scores as a result.

While the insights provided by the customer service representatives were incredibly helpful in pointing where to look, they were by nature largely anecdotal; thus, the investigation turned to data collection and analysis to further assess the process, using the insight from the employee interviews as a road map, as depicted by the logic tree shown in Figure 4-2.

4.2.3 Qualitative assessment through customer surveys

In addition to the employee focus groups, qualitative assessment of the account initiation process was accomplished through the gathering and classification of customer feedback via customer satisfaction surveys; this was carried out by identifying the primary reasons customers who had both recently completed the account initiation process and given Atlantic a non-promoter score (1-7 out of 10) gave for their dissatisfaction. To do this, six months of Atlantic's customer surveys were filtered by the customer's indicated reason for company contact, yielding 1,123 customer responses linked to the account initiation process. We then read through all customer comments given in response to the question "Why did you give

Atlantic Energy a score of X?" for all scores in Atlantic's neutral (5-7) and detractor (1-4) score ranges.

The result of this process was the Pareto chart shown in Figure 4-3. Some of the key insights from the analysis are provided below:

- Interestingly, nearly 1/3 of the customer responses that Atlantic categorizes as "non-promoter" did not have a complaint about the service. Many customers giving scores in the 4-7 range had neutral or even positive comments about the company, indicating that, for example, a score of 7 out of 10 represented what they were expecting from their utility, or that they simply didn't have many interactions with Atlantic and were satisfied with a middle-of-the road score to reflect how well they knew the company. This gives credence to the need for separating customer perception and expectation, but suggests an opportunity for increasing customer satisfaction by giving new customers a reason to see the company in a positive way, rather than a neutral way. This distinction between new and existing customers was further validated during quantitative analysis, in which existing customers were shown to give a statistically-significant higher score than new customers did.

- Unsurprisingly, customer service quality was the next most important reason customers gave for an unsatisfactory score. The subreasons for a poor score within "customer service" are also broken out in Figure 4-3 and appear to be equally divided between communication issues, lack of agent knowledge of the process, a mistake by the agent, an inefficiency with the process, and experiences with rude agents; the remaining responses were general complaints about customer service with no specific reason identified. The customer comments validated some of the insight provided by the change managers and customer service representatives at Atlantic, but other assumptions still remained unproven at this point prior to the quantitative analysis.

- Although many other drivers of customer satisfaction exist, such as pricing complaints, many of these fell lower on the Pareto chart due to the focus of the investigation on CSAT drivers *within the account initiation process*. Again, this is intentional due to Atlantic's desire to focus on the improvement of a preselected process (account

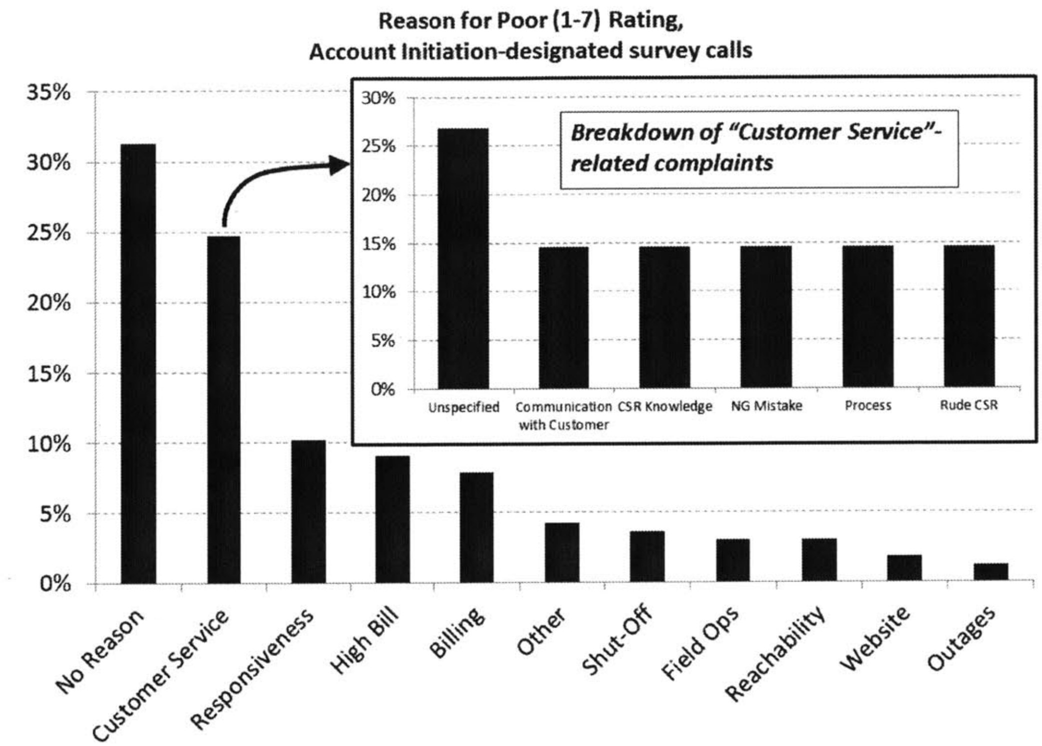

Figure 4-3: Pareto chart showing reasons for unfavorable CSAT scores related to account initiation process

initiation) rather than a more holistic approach to customer satisfaction for this project.

4.2.4 Quantitative assessment through customer satisfaction scores and process metrics

While qualitative assessment of service quality drivers through both employee and customer feedback is highly valuable, it does not deliver a complete picture by itself. Employee interviews offer insight from experience, but are driven by anecdotal evidence and are susceptible to biases from memorable events or personal opinions. Assessment of customer comments is more effective, as it clearly indicates the issue that actually pulled a customer into the dissatisfied range. However, even this can be incomplete, as comments generally point to one issue, and are often detached from the specifics of the process the customer experienced. While focus groups with the customer would be ideal, these can often be impractical, particularly on a large scale.

Therefore, to complete the investigation of customer satisfaction drivers in Atlantic's account initiation process, we conducted a statistical analysis on a data set linking process metrics to customer satisfaction survey scores. To do this, we first compiled the results of every customer satisfaction survey over a six-month period for which the respondent identified "account initiation" as the reason for contact; as before, this yielded 1,123 customer accounts for which account initiation-related CSAT scores were recorded. Next, customer service agents pulled up each of these customer account numbers in Atlantic's customer data system (CSS) to record 40 pieces of information relevant for the account initiation process, such as cycle times for each process step, the number and timing of customer interactions, the types of subprocesses the customer experienced, the amount of process rework and errors, customer behavioral scores, and demographic information. The exact fields for which customer data was collected were determined based on the results of the qualitative analysis; this helped guide the data collection, and provided a means to prove or disprove beliefs and hypotheses regarding service quality drivers determined during the employee focus groups and customer surveys.

The result was a data set that explicitly linked process-focused customer data to customer satisfaction on a customer-by-customer basis, rather than through aggregate results on a monthly basis during which customer satisfaction scores and process tendencies (e.g. average cycle times, percent of customers experiencing a certain subprocess) show variation. While the variation of customer data and process metrics in an averaged sense on a monthly basis has been studied previously, we argue that such analysis is inferior to analysis that studies the variation in each on a customer-by-customer basis. This data set provides the means to study the links between customer experiences and customer perception on an individual basis, providing meaningful insights into potential drivers of customer satisfaction. Although causation is not testable, as explained in the theory of Chapter 2, valuable evidence to support operational changes can be obtained.

Following customer data collection, a conjoint-type regression analysis was carried out using the method described in Chapter 2. As suggested by these methods, the ordinal customer satisfaction score on a 1-10 scale was held constant as the response variable while all of the fields of customer data, such as cycle times, amount of rework, first contact resolution, and progression through different subprocesses, were chosen one at a time to be the regressor. While log transformations were used on numerical regressors, such as cycle times, to provide comparable elasticities, categorical regressors, such as the progression through a given subprocess, did not use a transform, but rather used a binary dummy variable. These results are presented in part in Figure 4-4, in which the coefficient of regression, measuring the magnitude of the regressor's impact on customer satisfaction, is plotted against the t-statistic for the regression, measuring the statistical significance of the result. In this case, results appearing to the right side of the graph (high t-statistic) represent a low probability of zero influence of the regressor on the customer satisfaction response variable, thus indicating a statistically significant relationship between the two; results appearing at the top of the plot, in contrast, represent a larger magnitude of impact of the regressor on the response. Therefore, variables in the upper-right quadrant of the plot are likely to impact customer satisfaction, and the impact is large. Results in the lower-right quadrant, in contrast, are likely to impact customer satisfaction in a statistically significant way, but the impact is smaller. Results in the lower-left quadrant are most likely nonfactors, or have a very small impact on customer satisfaction, and thus can be neglected.

From the results shown in Figure 4-4, we draw the following conclusions about the account initiation process:

- Although this analysis serves as a test of correlation rather than causation, three factors stick out in the upper-right quadrant of the plot: 1) the cycle time on account processing for customers without service, 2) the rate of first contact resolution for customers, and 3) the percent of customers who experience errors in the process, and rework as a result. These three factors are described below; because of the significant correlation and potentially large impact each has on customer satisfaction, these three factors will be prioritized when investigating operational changes in the last three steps of the thesis framework.

    1. *First contact resolution:* First contact resolution (FCR) is achieved when a customer's issue or reason for calling is solved on the first call, email, web-based interaction, or office visit. Although FCR was tracked by Atlantic prior to the

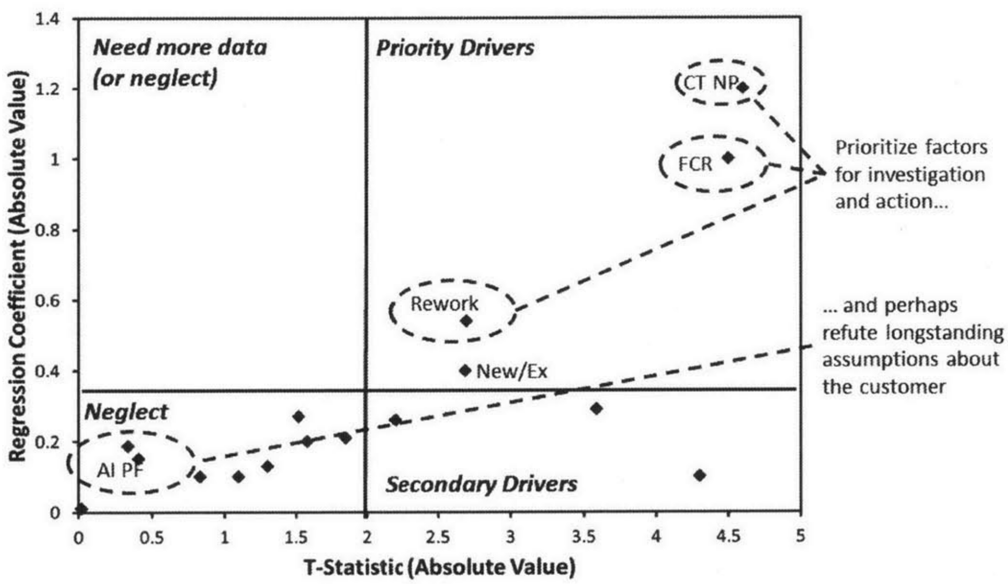

Figure 4-4: Results of regression analysis on customer data and satisfaction scores reveals drivers of customer satisfaction in account initiation process

case study due to its impact on cost and assumed impact on CSAT, its emergence as a primary driver of CSAT- as well as the evidence disproving other assumed drivers- heightened its importance in the follow-on operational changes. As shown previously in Chapter 2, FCR is worth increasing even if an increase in the average service time for the center is required.

2. *Process rework due to agent errors:* Intuitively, customers become more dissatisfied when errors occur that cause them to have more work, or more interactions with Atlantic, even in cases for which cycle time is not usually a concern. Combined with the evidence presented by the employee interviews and customer comments, clear sources of common errors- such as customer service agents erroneously sending customers through unnecessary steps- provide areas of opportunity for the follow-on work.

3. *Account processing cycle time for customers without power:* While the population of customers as a whole showed no statistically significant correlation between cycle time and CSAT, the data showed account processing cycle time to have a large impact on CSAT for customers without power, which is clearly an intuitive result. The regression analysis shows a significant drop in CSAT on even an *hourly* basis when customers are without power, which will be addressed among the case study's operational changes.

- While the distinction between new and existing customers appears to also fall in the upper-right quadrant of the plot, this is a driver that Atlantic cannot address through operational changes to the account initiation process, and is therefore not included in the next step. However, due to its potential impact, the implications of this relationship are considered outside the scope of this case study.

- Some drivers, such as the total cycle time and number of calls from the customer to Atlantic during the process, have a t-statistic > 2 (corresponding to a p-value < 0.05), indicating a significantly significant relationship between these drivers and CSAT for $p < 0.05$. However, their low regression coefficients place them in the lower-right quadrant of the plot, indicating that while there may be an impact, it is smaller than the impact of the three drivers listed above, and is therefore not prioritized for the next step of the framework.

- Many factors fall in the lower-left quadrant of the plot, indicating no significant relationship between these factors and CSAT; surprisingly, however, many of these factors were assumed to impact CSAT prior to the case study, and past operational excellence projects had been formed on the assumption of the link between these drivers and CSAT. For example, while customers abhor rework and mistakes, they seem to accept the extra work that subprocesses such as sending in proof of identification (requested of some but not all customers) impose on them without any significant change in CSAT-as long as they do not perceive errors or rework. In contrast, Atlantic previously assumed that sending customers through such sub-processes by itself decreased CSAT and assigned resources to addressing the issue based on this assumption. While other reasons for decreasing the number of customers sent through such subprocesses exist (such as the propensity for errors and rework whenever extra process steps are added), this distinction can enable a more efficient allocation of resources in addressing service quality drivers in the next part of the case study.

4.3 Updating Process for Improved Operations

Based on the results from the analysis of customer service drivers, three "levers" of service quality with the potential to create high impact have been identified for operational improvement of the account initiation process; as described previously, the three levers are 1) first contact resolution, 2) rework and customer service representative errors, and 3) account processing cycle time for customers without power. This section traces out the work done in conjunction with step two of the thesis framework, in which the process is updated to address the primary drivers of service quality revealed by analysis.

4.3.1 CSAT Lever #1: Increasing First Contact Resolution (FCR)

First Contact Resolution was found to have a large impact on customer satisfaction, yielding a two-point increase on the 1-10 scale when achieved according to the regression; with the large number of factors impacting customer satisfaction, this represented one of the largest impacts from any single driver. In particular, failure to achieve FCR tends to place customers on less efficient paths in which rework and errors are more likely, falling into the realm of the second CSAT Lever and further reducing satisfaction. For these reasons, FCR will be of primary concern for the rest of this case study, in the incentive concerns of Chapter 5, and in the consideration of service operations in general.

In general, obstacles to FCR fell into two categories at Atlantic, and potentially elsewhere: those obstacles arising from process issues, such as inefficient or error-prone processes, and incentive issues, in which employees feel a larger burden to meet cost-based metrics such as Average Handle Time (AHT) than they do quality-based metrics such as FCR or quality assurance checks. One major assertion of this thesis is that FCR and AHT need not be opposed to each other, but rather are actually aligned due to the fact that customers who are not served correctly the first time multiply the input volume into the system, hurting service level and cost more than time will. This tradeoff was considered quantitatively in Chapter 3 and will be addressed in detail in Chapter 5, while operational changes will be addressed in this section; the important consideration for now is the understanding of FCR's significant impact on both quality and cost as the discussion of operational change continues here.

Armed with the knowledge of the process gained from employee focus groups and customer comments before, we began the investigation into FCR by tracing out the current account initiation process. What was found was a complex process in which customers could follow one of many paths based on a number of factors. This concept is illustrated using the funnel concept from Chapter 3 in Figure 4-5 below. When customers enter the process, they will ideally take the most efficient path through the process, and therefore complete the process with the lowest amount of frustration for themselves, and the lowest amount of effort and cost for Atlantic. For example, it is more efficient for customers to use the website

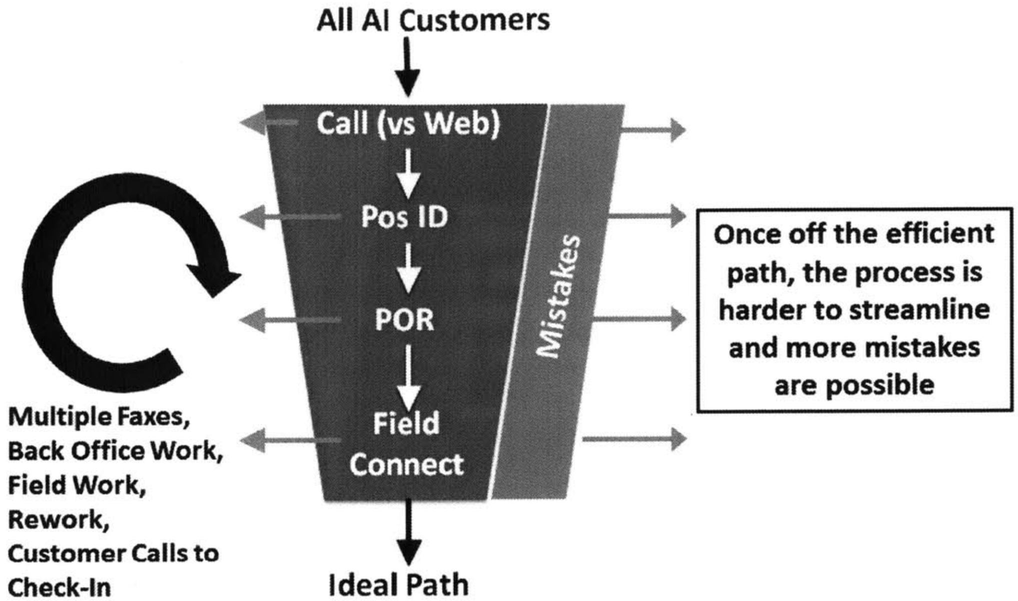

Figure 4-5: Steps along the account initiation process provide opportunities for customers to be "funneled" off the efficient path, and into higher likelihood of rework and errors

than to call Atlantic, it is more efficient to not require identification or proof of residency from customers, it is more efficient to have the meter unlocked between customers, and it is always more efficient to have fewer mistakes and unneeded calls from customers, especially when they are just calling to check in on account progress. Whenever the most efficient path is not followed, it usually results in extra work from either the customer or the company; at the same time, mistakes by customer service agents during the process can also siphon customers off the efficient path as well.

The idea of the process's efficient path and the funnel provides a road map for a systematic approach to increasing FCR in this case. By addressing each point in the process where customers are funneled off the ideal path to increase the total number of customers who pass through as efficiently as possible, FCR for the entire process can be increased. Below are some examples of ways in which the investigators addressed the funnel at Atlantic:

1. *Push toward web connect:* When entering the account initiation process, the first choice a customer makes is whether to start the process by calling Atlantic, which directs the customer to a call center for a customer service representative to complete the process,

or to complete an online from from the website. When a customer tries to complete the account initiation process by web first, he or she typically ends up calling much less (.1 calls / customer for web compared to 1.2 calls / customer for phone) and is much less likely to experience errors or rework during the process as the information is recorded directly into the system by the customer rather than taken verbally over the phone. While only 3% of customers currently use the web connect process, increasing this rate to 23% would result in a $1 million annual savings, according to the current process costing, and more importantly would lead to an increase in perceived service quality measured by CSAT.

As a result, several actions have been prescribed to increase the number of customer electing to use web connect, which will both increase the first contact resolution and decrease operational expenses. These actions include pre-recorded messages when customers are first placed on hold in the account initiation process that redirect callers to the website, redesign of the company website so that more customers find the link to start the process by web before calling in, and a push toward web connects through marketing efforts including move-in materials.

2. *Require less account processing:* A customer calling in to initiate an account may be asked to send in up to three additional pieces of information before the process is complete; these include the following:

   - *Positive identification check:* All new customers are asked to provide basic identifying information such as name, birthday, and social security number, from which a third party credit check service is used to verify the customer's identity in real time. If customers do not show up in the credit check system, which occurs about 15-20% of the time, the Atlantic policy requires customers to send in a copy of an acceptable piece of identification (such as a driver license or passport) by fax, email, or postal mail.
   - *Proof of Residency:* In some states, Atlantic has the option to request high-risk customers to send in a proof of residency, such as a residential lease, to prevent potential outlets of fraud. In areas of high risk for commodity loss (when new

customers move into a residence and consume power before attaching their name to the account), Atlantic may elect to turn a meter off when a customer moves out to force the incoming customer to start service in their name upon moving in, requiring additional steps such as proof of residency and a field connect for the next customer.

- *Collections:* When customers owe past arrears from a prior account, or usage has been registered on the meter for the residence the customer is moving into, additional steps may be required in the account initiation process.

While all of these subprocesses exist to ultimately protect the company from losses, they are equally capable of driving up costs and reducing customer satisfaction when not executed correctly, as triggering any of these three requests for additional information necessarily funnels a customer off of the efficient path and greatly increases the probability of additional contacts and rework to resolve the account.

As a result, these items provide important opportunities to keep more customers on the efficient path by reducing unnecessary redirection of customers into these subprocesses. Therefore, actions to reduce the number of customers pulled into these processes were prioritized; some of these actions include adding the use of additional sources of identification, such as credit card and bank account numbers, that can be used with the third party credit check system the first time a customer calls and a planned re-evaluation of the criteria for requesting leases for a policy that achieves a more optimal balance between the cost of not having proof of residence with the cost of the process inefficiency for requesting it of too many customers. Even more important than process changes, however, is the reduction of customer service agent errors in executing the current process which, based on the data collection, is the greatest cause of unneeded customer work; errors will be addressed by itself below.

3. *Error reduction through the tracking and measurement of sources of error:* As Figure 4-5 depicts, errors throughout the account initiation process can prematurely push customers off of the efficient path, resulting in extra steps and rework when first contact resolution might have been otherwise achievable. While error reduction is addressed

individually in the next step, it is worth noting that FCR and employee errors are clearly linked together with errors being one of the primary reasons for not achieving first contact resolution. Therefore, decreasing errors will positively impact the organization's first contact resolution.

Finally, while first contact resolution can be increased through the process changes described above, its improvement is equally as dependent on incentives and metrics that properly align agents with FCR goals; that is, to effectively address service operations from a quality standpoint, management must not only change the process, but change the behavior of customer-facing employees as well, much as the machines in a manufacturing environment might be fine-tuned to achieve the desired specifications. One example in this case is the redesign of metrics that potentially incentivize low cost and low average handling time at the expense of quality. Due to its importance, this topic will be considered on its own in step three of the framework (Chapter 5).

4.3.2 CSAT Lever #2: Reducing errors through feedback loops

After the investigation revealed that 30 - 40% of customers asked to submit documentation in the positive identification subprocess could have avoided the process completely and continued on the most efficient path, we decided to introduce mechanisms for measuring, tracking, and reducing errors. To do this, we recognized that, because all requested documentation will be reviewed by the account processing team, account processing has the potential to act as an inspection group while completing their current work. As a result, employees in the account processing group began to collect data on the most common errors by recording error type and employee ID whenever the review of a customer account indicated an error was made. By compiling this information, management is able to address the most common sources of errors within internal operations, or provide a measure of error to third party vendors completing operations on a contracted basis.

4.3.3 CSAT Lever #3: Implementing a priority queue for customers without power

The final CSAT driver identified by the regression analysis on CSAT scores is the sharply adverse effect that customer waiting time in the account initiation process has on the ultimate customer satisfaction scores when the customer is without power. While this analysis indicated a sharp reduction in CSAT score *by the hour* in this case, the process previously had no way to distinguish between customers currently with power- who are largely insensitive to process cycle time- and those without power when entering the account initiation process. Hence, all customers progressed through a single first-in-first-out (FIFO) queue, leading to the same cycle time for all customers. Clearly, this cycle time was unnecessarily (perhaps unacceptably) long for customers without power, and unnecessarily short for customers with power.

In response, we created a new subprocess within the accounts processing group to provide different cycle times for customers with and without service; this new process is shown in Figure 4-6. In it, all faxes and emails from customers into the accounts processing group first flow through a worker who looks up the customer account number on CSS to determine if the customer is currently with or without service, and then accordingly sorts the customer into the priority queue or the normal queue. This worker will intentionally have slack capacity (50% utilization) to move all customers without power into the priority queue quickly. The customer service representatives running the process then draw first from the priority queue until it is empty, which will be the case most of the time. When the priority queue is empty, they will then take customers from the normal queue. The entire process has enough slack capacity (70-80% utilization) to ensure cycle times are kept low for all types of customers, especially priority customers. Because the majority (80%) of customers will currently have power, the priority queue is predicted to reduce priority customer wait time from 4+ hours to under one hour while still having enough capacity to turn over the normal queue within a modest increase in cycle time.

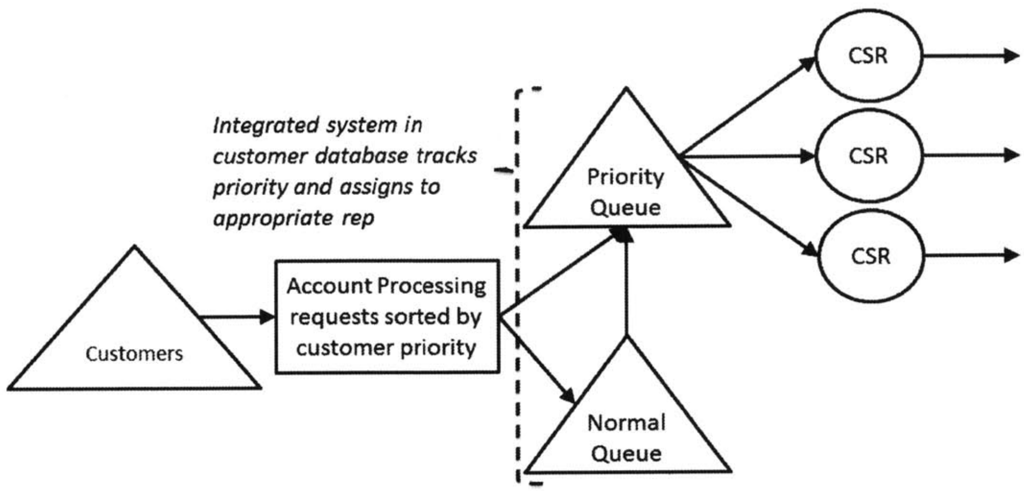

Figure 4-6: Priority queue for account initiation process filters on connection status to increase CSAT

4.4 Conclusion

Throughout this chapter, we have provided an example of the first two steps of the thesis framework as applied in practice to the account initiation process at Atlantic Energy. One final key insight from the case study is the importance of involving the personnel that actually run the process rather than only involving the change management teams; by involving the groups that actually own the processes in question, much better results were obtained during the case study. While the first two steps will help create an organization with an operational structure supporting higher quality services, two more steps in the framework remain to address non-structural issues, such as the alignment of employee behavior with high quality service and the creation of an organization in which experimentation and discovery is encouraged. These steps will be addressed in the next two chapters.

Chapter 5

Aligning Agent Behavior with High-Performance Operations

Principal-agent alignment through proper mechanism design

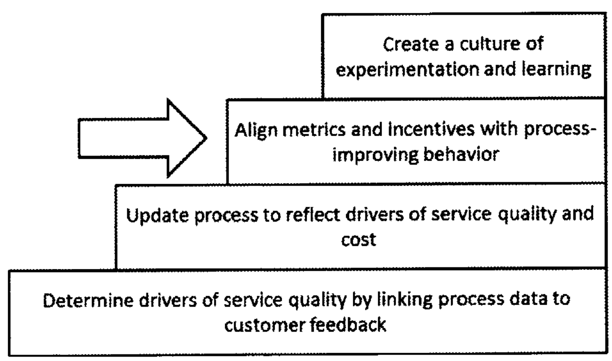

Figure 5-1: Step 3 of the thesis framework

5.1 Introduction

So far in this thesis, operational improvements have largely fit into two categories: process changes that address quality drivers and improve service quality, and behavior changes that pull the actions of customer service agents in line with high quality service. Going into step three of the framework, we assume that the organization has addressed all reasonable operational changes relating to process improvement to put agents in the best position possible to serve the customer; we now turn our attention to fine-tuning the system to get the desired output by changing the behavior of CSRs through incentive schemes and metrics.

In this chapter, we consider the tradeoff between first contact resolution (FCR) and average handle time (AHT) through typical incentive schemes used in call centers; however, while we have focused on call centers in this chapter's analysis, the results can be generalized to many other forms of service operations as well. Through the use of game theory and decision theory analysis, we consider the shortcomings of different common incentive schemes and the misaligned incentives they may provide, and offer guidelines for developing reward structures that help the agents to focus on quality while still keeping cost under control.

5.2 Service Operations Output Model

To begin the analysis, we start with a discussion on a suitable definition of output for service operations. While economic output can be defined by the profit contribution of the number of units produced by a process in manufacturing, such a definition is more difficult to obtain in service operations. Fortunately, for the purposes of this problem, an economic definition of system output need not correspond with an actual dollar amount, but rather only needs to correspond to the desired output in the system. Typically in service operations, the goals are to 1) provide high quality service, 2) to every customer that enters the system at a target service level, 3) at the lowest cost possible. To do this, we consider two candidate outputs: the number of completed calls in a given time, and the inverse of marginal operating cost.

5.2.1 Preliminary definition of output

One candidate definition to measure output by an individual would be the number of calls successfully completed per a unit of time, where successful completion is defined as a customer requiring no further interactions for the initial problem, with a high perceived level of satisfaction (such as would be the case for a "promoter"). Given an Average Handle Time (AHT) per customer for an agent executing the process, the number of calls completed per unit time is simply $1/AHT$. Using p as the fraction of customers the agent successfully serves, the number of successfully completed operations per unit time would be p/AHT. Given this expression, and the previous results in Chapter 2 showing first contact resolution as a dominant driver of service quality, a reasonable objective for the firm will be:

$$\max \frac{p}{AHT}$$

While we will expand on this definition of output before developing the models used in this chapter, it is first useful to next consider service operations from a cost standpoint using queuing theory; as we will see, p/AHT is actually a reasonable definition for output when considering output by either completed calls per time or marginal cost.

5.2.2 Cost objectives from queuing theory

To consider the the medium-term marginal costs of a service operation, we will approximate the dynamics of the operation with the Erlang-C queueing model. In this case, we assume marginal costs are proportional to the number of agents employed; thus, we assume that agents can be scaled up or down through hiring / attrition in the medium term[1], and that fixed costs are neglected from the analysis. Therefore, we will ultimately look to minimize the number of agents needed to sustain a given service level, which will typically be set by managerial decision or regulation. For illustrative purposes, we assume a service level in which the firm must answer 80% of the input calls within 20 seconds (in the case of the case study at Atlantic energy, this service level is set by regulation).

With an assumed service level, we now use Erlang C to calculate the number of agents needed to achieve the service level for a given input volume of calls. To begin, we assume an exponential distribution on interarrival times for customers calling in given by:[4]

$$\Pr(IA \leq t) = 1 - e^{-t/a}$$

[1]This is a reasonable assumption due to the high turnover in call centers; lower staffing needs can generally be met through attrition. This has other implications on service quality, however, and will be addressed elsewhere.

Under this assumption, the Erlang C model predicts the service level SL (percent of calls answered under a given time T_{SL}) for the operation given the input call volume (λ), the number of agents (or servers) available to handle the input volume (N_A), and the average time needed to serve a customer (AHT) with the following equations:

$$SL = 1 - E_C(N_A, u_\lambda) \cdot e^{-(N_A - u_\lambda)T_{SL}/AHT}$$

where

$$u_\lambda = \lambda \cdot AHT$$

$$\rho = u_\lambda / N_A$$

$$E_C(N_A, u_\lambda) = \frac{f_{poisson}(N_A|u_\lambda)}{f_{poisson}(N_A|u_\lambda) + (1-\rho)F_{poisson}(N_A - 1|u_\lambda)}$$

and $f_{poisson}(x|\lambda)$, $F_{poisson}(x|\lambda)$ represent the Poisson probability density function and Poisson cumulative distribution function, respectively.

Given this mathematical model, we can solve for the AHT corresponding to a given service level as a function of the number of agents and the incoming call volume. A numerical example of this is shown in Figure 5-2, assuming the 80/20 service level for various input call volumes. While the Erlang C model is nonlinear in truth, Figure 5-2 shows an approximately linear relationship between the number of agents needed to staff a customer care center and the average handle time for the center for various center arrival rates. Furthermore, when normalizing the results in Figure 5-2 by arrival rate, we find that the curves nearly collapse onto a single curve, as shown in Figure 5-3. While the relationship shown in Figure 5-3 is neither completely linear, nor completely independent of arrival rate, it does indicate that the required staffing level (number of agents) normalized by arrival rate can be approximated by a linear relationship with respect to average handle time, or $N_A/\lambda \approx \gamma \cdot AHT$ for some linear factor γ.

5.2.3 Output value equivalence

Given that staffing needs N_A, and hence the marginal costs of the customer service center, approximately go by $N_A/\lambda \approx \gamma \cdot AHT$, the center will ideally operate at a condition that

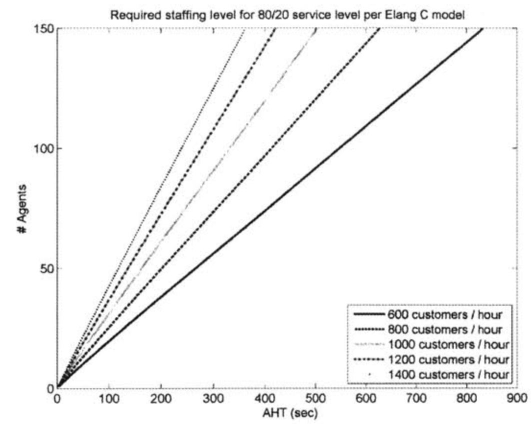

Figure 5-2: Erlang C queuing model predicts nearly linear relationship between staffing needs and handle time

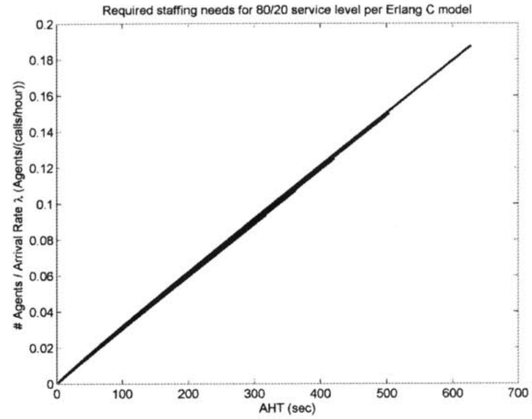

Figure 5-3: When normalized by arrival rate, staffing curves in Figure 5-2 nearly collapse onto a single, nearly linear curve

minimizes the staffing level with the objective:

$$\min\left(\gamma\lambda \cdot AHT\right)$$

For a given arrival rate of *unique* customers λ_{unique}, the arrival rate will be determined by the fraction of calls that are successfully answered p as defined before. Given the dynamics of rework depicted in 5-4, a customer entering the system has a probability p of being successfully served and $1-p$ of re-entering the system as rework; approximating p as constant each iteration[2], the number of resulting calls for a single unique customer is:

$$\sum [1 + (1-p) + (1-p)^2 + (1-p)^3 + ...] = 1/p$$

Therefore, the arrival rate into the system is approximated by $\lambda = \lambda_{unique}/p$ and the objective from before can now be given by:

$$\min\left(\gamma\lambda_{unique}\frac{AHT}{p}\right) \propto \min\frac{AHT}{p}$$

by noting that γ and λ_{unique} are nearly constant.

[2]This is not in general true, as p will tend to be conditional on previous attempts; however, for typical values of $p > .8$, this is a reasonable approximation.

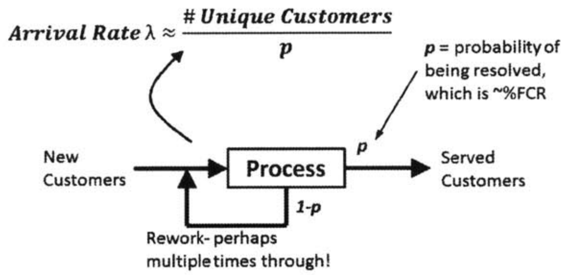

Figure 5-4: Dynamics of rework creating a multiplier of $1/p$ on arrival rate of unique customers; $\lambda = \lambda_{unique}/p$

Finally we compare our two separate definitions for the firm's objective and find that they are notionally equivalent:

$$\max \frac{p}{AHT} \sim \min \frac{AHT}{p}$$

Therefore, we now proceed with our analysis using p/AHT as our definition of output for two reasons: 1) output is more intuitive as a maximization objective and 2) we will find that p/AHT is numerically well-behaved compared to AHT/p.

5.3 Analytical Model for Incentive Mechanisms

5.3.1 Agent behavior and output

We begin with a simple one-action model for employee behavior; in this case, the employee's lone decision is to pick the amount of time spent serving each customer;[3] we pick the average time the agent spends on actions serving the customer customer to be a_1[4] which, in the single-action model, will be equivalent to AHT. Next, we assume that the probability p of an agent successfully serving a customer is monotonically increasing with a_1 as given by a cumulative distribution function. In this case, we further assume that that the probability of successfully serving a customer p from before is dependent on the time the agent spends on the customer such that $p = \Pr(\text{success}|a_1) = F_{cdf}(a_1)$. Brown (2005)[3] shows that this function follows a

[3]The ability to spend time on other activities, such as shirking work for personal time, is addressed later.
[4]We use a_1 to distinguish from other actions that will be introduced later.

lognormal distribution; hence, we use in our model:

$$p = F_{a_1}(a_1; \mu, \sigma) = \frac{1}{2}\left[1 + \operatorname{erf}\left(\frac{\ln a_1 - \mu}{\sqrt{2}\sigma)}\right)\right]$$

where erf(x) is the Gauss error function and, for a mean time length m and variance v:

$$\mu = \ln\left(\frac{m^2}{\sqrt{v + m^2}}\right), \ \sigma = \sqrt{\ln\left(1 + \frac{v}{m^2}\right)}$$

Finally, we combine this assumed distribution with the objective function of $\max p/AHT$ from before to define the output y the firm would like to incentivize:

$$y = \frac{F(a_{k \in \mathcal{A}})}{\sum_i a_i}$$

where $\mathbf{a}$ is the set of actions available to the agent, and $a_{k \in \mathcal{A}}$ represents the actions that add value, or are needed to successfully serve a customer. In this preliminary case for which there is one action a_1 that is assumed to be in the value-adding set; output is simplified as:

$$y = \frac{F_{a_1}(a_1; \mu, \sigma)}{a_1}$$

The nature of this output is shown in Figure 5-5; clearly, this function is maximized for:

$$\frac{\partial y}{\partial a_1} = \frac{1}{a_1}\frac{\partial F_{a_1}(a_1)}{\partial a_1} - \frac{F_{a_1}(a_1)}{a_1^2} = 0$$

We note that this output function is logarithmically concave with a unique optimal solution that we would like to incentivize the agent to take, representing the optimal trade-off between overservice (spending too much time on service with low return in quality) and underservice. Using $f_{a1}(a_1)$ as the lognormal probability distribution function, we find this optimal solution is a_1^* satisfying:

$$\frac{f_{a_1}(a_1)}{a_1^*} = \frac{F_{a_1}(a_1)}{a_1^{*2}}$$

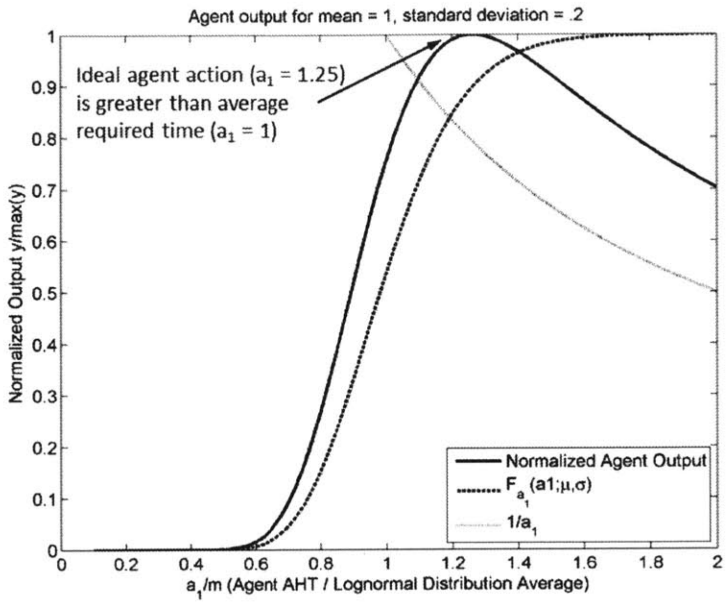

Figure 5-5: Agent objective output shows unique maximum that firm would like to maximize with $\text{argmax}_{a_1}(y) > m$

5.3.2 Utility and risk attitude

We assume the existence of a von Neumann-Morgenstern utility function $u : \mathbb{R} \mapsto \mathbb{R}$ for the agent in this model mapping the potential pay- and hence, the incentive metrics on which pay is based- to an equivalent utility for the agent;[5] given $u(a) > u(b)$, the agent will prefer a to b ($a \succ b$). We further assume that the agent is risk-averse with concave, monotonically increasing utility function $u(x)$ on wealth x and that the agent has a constant absolute risk aversion (CARA) utility function[22] such that:

$$-\frac{d^2u/dx^2}{du/dx} = \alpha = \text{constant}$$

where α is the Arrow-Pratt coefficient of absolute risk aversion. Finally, given these assumptions, we choose for our model an agent utility function of the form:

$$u(x) = 1 - e^{-\alpha x}$$

[5]Although each employee will have his or her own utility function, we will be developing representative results with a single utility function in this case.

Therefore, given a cumulative distribution function $F : X \mapsto [0,1]$ on the possible rewards for the agent, the expected utility on the distribution of possible outcomes F is:

$$\mathbb{E}_F[u(x)] = \int u(x)dF(x)$$

Because of the risk aversion, the agent will prefer a certain outcome to an uncertain outcome with the same expected value; that is $\mathbb{E}_F(u(x)) \leq u(\mathbb{E}_F(x))$. When considering alternative payouts and incentive schemes in this model, we will determine the agent's preferences by comparing expected utilities for the different distributions on potential outcomes; as before, we say that the agent prefers distribution F to distribution G if $\mathbb{E}_F(x) \equiv U(F) \geq U(G) \equiv \mathbb{E}_G(x)$.

5.3.3 Analysis of traditional incentive schemes and metrics

To begin, we would like to model an agent's response to a typical incentive scheme; in this case, we use an incentive scheme based on the one used by Atlantic Energy in the case study, from which a potential 10% salary bonus is available to agents who meet all objectives. In particular, this incentive scheme has two characteristics common amongst service operations organizations:

1. *Cut-off on average handle time:* The incentive scheme features a goal average handle time such that agents only gain credit when meeting the goal time. In the case for which agents have some uncertainty for their final AHT for the month, we hypothesize that agents will aim lower than is optimal on AHT due to the drop-off in utility-particularly when the agent is risk averse.

2. *Quality monitoring:* Clearly, time goals by themselves would incentivize an agent to keep calls very short, possibly hanging up early on customers often, as has been observed in other call centers in the past. Many customer care centers will randomly sample a handful of calls for each agent during the month for quality assurance grading. Although basing goals entirely on quality could possibly lead to unreasonable call times, an incentive mechanism that incorporates quality with AHT can lead to

desired agent behavior. However, we hypothesize that quality in its current form may be suboptimal if reward structures are nonlinear with success, as will be investigated later.

The representative mechanism we will study assigns a bonus cash payment to agents that is directly proportional to the total performance score achieved from AHT and quality; that is, a total score of 8 out of 8 earns a 10% salary bonus while a score of 0 earns no bonus, with all scores in between earning bonuses based on a linear scale between the two. In this case, an agent receives half of her score from the AHT rating, and half of her score from the quality rating as defined by the following criteria:

AHT Measurement	4 points	AHT < goal
	2 points	AHT > goal
Quality Measurement	4 points	4 out of 4 correct calls
	3 points	3 out of 4 correct calls
	2 points	2 out of 4 correct calls
	1 points	1 out of 4 correct calls
	0 points	0 out of 4 correct calls

To consider these effects, we model the agent's utility as a function of the average handle time that she aims for. To do this, we take the agent's utility and incentive structure as previously defined and add one last enrichment to the model: We assume that an agent choosing a target "effort" level $\bar{a}_1$ for the range of possible outputs shown in Figure 5-5 will experience an actual AHT a_1 where $a_1 = \bar{a}_1 + \epsilon$ with ϵ being a random variable with a normal distribution. This addition models the uncertainty an agent faces with respect to the actual impact of the effort level she chooses.

With this in mind, we define the following model for incentive mechanisms: We assume an agent receives a reward R based on the incentive scheme listed in the table above, and we assume that $R = R(a_1)$ based on the output relationships from before. Specifically, for this incentive scheme, we assume that:

$$R(a_1|q_i) = \frac{1}{2}q_i + \frac{1}{2}(2 + 2(a_1 < T_{goal}))$$

where q represents the number of the 4 calls that passes the quality assurance and $(a_1 < T_{goal})$ is a logical expression that is 1 if true and 0 if false. We further assume that the PDF distribution on a_1 given $\bar{a}_1$ is a normal distribution designated by:

$$pdf(a_1|\bar{a}_1) = h(a_1|\bar{a}_1) = \frac{exp(-\frac{(a_1-\bar{a}_1)^2}{2\sigma^2})}{\sigma\sqrt{2*\pi}}$$

Finally, we assume that each potential outcome to the quality assurance check (e.g. the number of correct calls observed) is dictated by a binomial distribution in which the probability of a call being correct given time a_1 spent by the agent on the call is given by the CDF distribution $F_{a_1}(a_1)$ as before. Hence, for q_i representing the quality score for $i = 1..N$ calls:

$$p(q_i|a_1) = \frac{N!}{q_i!(N-q_i)!}F_{a_1}(a_1)^{q_i}(1-F_{a_1}(a_1))^{N-q_i}$$

Therefore, the expected utility an agent receives from this system for each target $\bar{a}_1$ is given by:

$$\mathbb{E}_U(\bar{a}_1) = \sum_i \int_{-\infty}^{+\infty} u(R(a_1|q_i))h(a_1|\bar{a}_1)p(q_i|a_1)da_1$$

Given the uncertainty in actual a_1 compared to target $\bar{a}_1$, the expected output that the firm experiences is given by $\mathbb{E}_y(\bar{a}_1)$ below. We assume the firm is risk-neutral, and thus that the principal's utility is driven by $\mathbb{E}_y(\bar{a}_1)$, less the utility of the bonus wage structure, which we will address later.

$$\mathbb{E}_y(\bar{a}_1) = \int_{-\infty}^{+\infty} \frac{F_{a_1}(a_1)}{a_1}h(a_1|\bar{a}_1)da_1$$

5.4 Numerical Results and Insights from Agent Decision Theory Analysis

Now that the model is complete, we are prepared to explore the combined impact of the incentive scheme, service operation dynamics, and agent risk aversion on the agent's ultimate choice of a target $\bar{a}_1$. Here, we use the model to consider different issues that may arise in such an incentive scheme to develop an intuition that will drive a set of recommendations

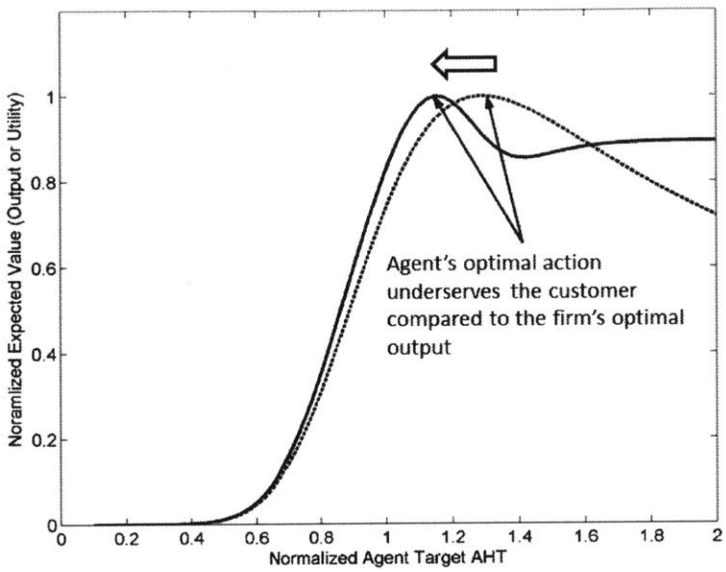

Figure 5-6: A combination of time goal cut-off, uncertain performance measurement, and agent risk places agent's optimal action choice below the firm's, underserving the customer

for completing step three of the thesis framework.

5.4.1 Impact of traditional incentive scheme on agent behavior

First, we plot numerical results for both the agent's expected utility $\mathbb{E}_U(\bar{a}_1)$ and the firm's expected economic output $\mathbb{E}_y(\bar{a}_1)$ for the traditional incentive scheme described above; these results are shown in Figure 5-6. In this case, the target cut-off in the incentive scheme is placed exactly at the firm's optimal output, meaning that the reward structure is trying to force the agent to target the firm's optimal AHT. However, as Figure 5-6 shows, the uncertainty in both the system and the performance measurement scheme, combined with a goal time cut-off, places the agent's maximum utility at an lower AHT than the firm's optimal AHT. Thus, a utility-maximizing agent will underserve the customer in response to such an incentive scheme.

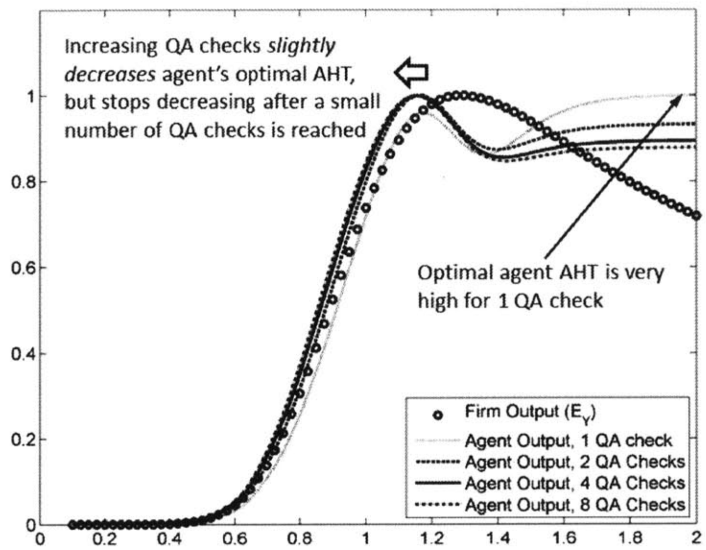

Figure 5-7: Agent's optimal AHT very slightly decreases with an increase number of QA checks, but quickly approaches a limit

5.4.2 Impact of number of QA checks

One characteristic of the incentive scheme we would like to study is the quality assurance process set up to prevent agents from keeping call length unreasonably short. We previously showed that the number of calls an agent "passes" out of a total N calls follows a binomial distribution. Interestingly, we find that for a small number of quality assurance checks, agents will actually increase AHT to avoid missing one of the calls, in the case that QA and the goal time cut-off are given equal weight in the incentive scheme. For a modest number of QA checks, however, the agent's utility quickly approaches a limit representing the utility the agent would derive from an incentive scheme in which every calls was sampled and checked; this is shown in Figure 5-7. Although it may seem counterintuitive that the number of checks driving the binomial distribution would have such a low impact on the agent's response, the expected utility does not vary greatly for modest levels of risk aversion, if the value assigned to each call scales linearly with the number of calls (e.g. an agent gets full credit for passing all calls, half credit for passing half of the calls, no credit for passing none of the calls, and so on).

The true cost of a small number of quality assurance call checks, however, is not in its impact on expected utility, but on the way that quality assurance tends to operate when fewer calls are used. Through the employee interviews at Atlantic Energy, we made two discoveries. First, customer service representatives do not like the idea of having only a few calls sampled to determine their quality assurance score, as the risk of having the occasional bad call sampled- and hence costing them part of their monthly bonus- in unsatisfactory, and creates the tension shown in Figure 5-7. Second, to address tension, managers tend to provide leniency on the quality monitoring checks, inflating QA scores beyond levels that would be expected given the actual first call resolution rates for the center. While this prevents employee dissatisfaction over potentially having a poor call sampled, it has a deleterious effect on quality incentives.

To demonstrate this, we consider several different scoring systems in our expected utility incentive model, with results shown in Figure 5-8. Here, the four scenarios considered, all with $N = 4$ sampled calls, are as follows in increasing order of leniency; they are identified by $R_q = (q_0, q_1, q_2, q_3, q_4)$ where q_i represents the "effective quality score" received for passing i of the four calls.

- *Baseline Case;* $R_q = (0, 1, 2, 3, 4)$*:* This is simply the baseline, linear QA incentive structure from before; this is also the best option of the alternatives considered here, as it pushes the agent's expected utility to be most in line with the firm's due to its close tracking of the expected quality of service as a function of AHT.

- *Minimum Reward;* $R_q = (2, 2, 2, 3, 4)$*:* This nonlinear structure more closely represents the structure that is *explicitly* in place at Atlantic energy, in which agents receive a minimum of 2/4 for QA if falling below the lowest threshold.

- *Leniency;* $R_q = (2, 3, 3, 4, 4)$*:* This nonlinear structure loosely represents the *implicit* structure that results when managers display lenience on top of the $R_q = (2, 2, 2, 3, 4)$ structure from before. In this case, agents who miss just one call might be given a pass from dropping down to the next level if they are "close," as has been suggested by the empirical evidence collected during employee interviews in the case study.

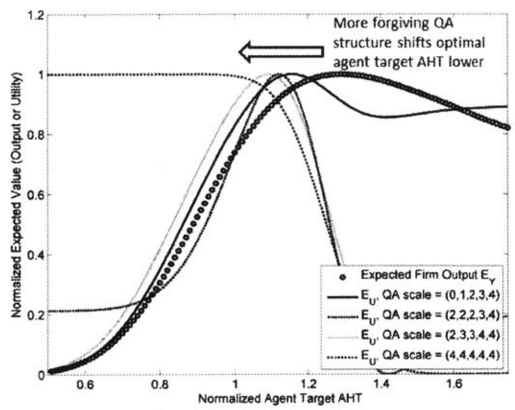

Figure 5-8: More forgiving QA structure shifts optimal agent target AHT lower, widening the gap from firms optimal

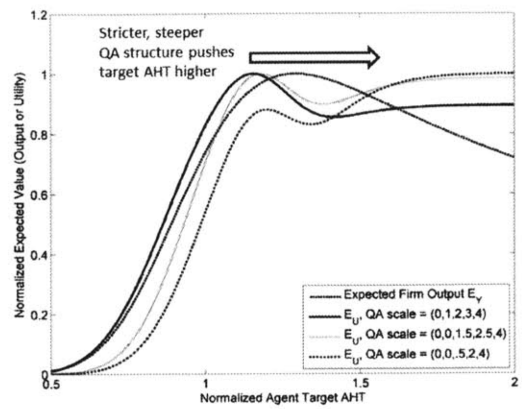

Figure 5-9: Less forgiving QA structure shifts optimal agent target AHT higher, potentially overshooting the firm's optimal AHT

- *No accountability;* $R_q = (4, 4, 4, 4, 4)$*:* This represents the extreme case in which quality monitoring is so lenient as to be irrelevant; obviously, an agent in this case will maximize his or her bonus incentive by staying far from the AHT goal cut-off, resulting in greatly underserved customers.

The results of this analysis using these alternative quality monitoring structures is shown in Figure 5-8. Clearly, as quality monitoring becomes more lenient, agents are incentivized to target a much lower AHT, widening the gap between the agent's optimal action and the firm's optimal position. A convex, or flat incentive structure reduces or eliminates the risk of losing points for low quality service, causing the agent to move farther away from the AHT cut-off. Here, the linear payoff, which most closely represents true quality, provides the highest target AHT, while the extreme lenient case results in the very behavior quality monitoring is supposed to avoid, that is, agents maximizing the number of calls they cycle through, perhaps by consistently hanging up early in the extreme case. Although agents will have some utility in providing quality service that this model does not include, the agent will still keep AHT artificially low to ensure the monthly salary bonus.

While nonlinear, lenient quality structures lead to underservice, steeper quality structures have the opposite effect, as shown in Figure 5-9. Here, a less forgiving quality monitoring structure causes agents to increase target AHT; however, as the structure becomes ever steeper, the optimal point overshoots the firm's optimal point entirely.

5.4.3 Misalignment of QA with performance

One last quality monitoring topic requires attention in this discussion; for this, we consider the role of any performance metric in measuring agent effort. In the service operations we are studying, as in many cases in which principal-agent problems arise, the firm is often able to only measure a metric or indicator of an agent's performance, rather than the actual output the firm would like to achieve. In this case, we would like to optimize the agent's success rate in providing high satisfaction-yielding customer service correctly on the first call. Because this cannot be measured directly, we use the sampling model in which a few calls are recorded and listened to by managers to check for adherence to a given set of qualities that the firm has deemed to make a high quality call.

While such a technique for quality monitoring is ubiquitous in such settings, and used at Atlantic Energy, it has one major shortcoming: the standard checklist by which the agent's call is graded is neither necessary nor sufficient to produce high-quality service as perceived by the customer. For example, standard quality monitoring may grade the employee on correctly using the customer's name, asking the customer certain questions, using certainty phrases, and many other cues that, while not unhelpful, may not contribute to quality service.

Such a checklist makes the process objective, but may also make it counterproductive. Homstrom and Milgrom[13] study a problem in which an agent is measured by a standard that differs from actual output. If one set of actions represents the checklist by which an agent is measured, and another set of actions represents the customer's perception of quality service, and each set is thought of as a mathematical vector in an abstract sense, then the weight that the firm should give to the metric when assigning an incentive structure is proportional to the alignment of the two sets (or mathematically, the cosine between the two vectors). While a firm can give heavy weight to aligned metrics, almost no weight should be given to misaligned metrics.

In such a search for objective quality measurements in service operations, in contrast to manufacturing operations, the quality monitoring checklists used are often at least moderately misaligned with customer-perceived quality. At Atlantic Energy, this resulted in two different outcomes: quality assurance scores that were much higher than those predicted by first contact resolution measurements, and the implicitly lenient grading scales studied above, both of which work to lower quality.

5.4.4 Impact of other factors in basic model

Before extending the model to include shirking, we quickly consider a few other factors:

- *Risk Aversion:* In the baseline case, the amount of risk aversion does not have a strong impact on the agent's optimal choice; however, this will not be the case when actions with no uncertainty in utility outcome, such as shirking, enter the picture.

- *Balance between quality and time goals:* In the previous case, the time goal and quality goal were given equal weight in the incentive structure. Clearly, if the time goal is de-emphasized in favor of the quality goal, agent target AHT will increase while the opposite will be true if quality is de-emphasized in favor of the time goal. While it may be tempting use this as a lever to move agent behavior around, doing so in practice has limited effectiveness when the firm's target AHT is not known for each agent.

5.4.5 Guarding against shirking and multi-task misalignment

While we have only considered a single value-adding action until this point, agents have the ability to choose between providing value-adding actions and non-value adding actions at work. When a worker slacks on his or her assignment, output and firm value decrease, but the agent's own utility may increase if he or she gains utility from slacking; furthermore, this effect may be exasperated under risk aversion, as slacking is often a sure reward, compared to a lottery of reward attached to nondeterministic output of value-added tasks.[5] In a call center, for example, agents may take extra time between calls or when customers are on hold for extra personal time, particularly when weak controls on AHT are in place.

To consider the impact of slacking, we enrichen our model from before to add a second action a_2 representing time spent not working. We assume that the agent's AHT is now $a_1 + a_2$ and that the agent's utility includes a term that is proportional to a_2; however, we assume that the probability of successfully serving a customer is still a function of a_1 alone. Thus we assume the following changes to our previous model:

- Output now takes the following form:

$$y = \frac{F_{a_1}(a_1; \mu, \sigma)}{a_1 + a_2}$$

- The agent's reward structure after slacking now takes the following form for some added slack time a_2 and some value of slacking v_{slack} to the agent (note that while a_1 is still nondeterministic based on some probability distribution, a_2 is deterministic):

$$R(a_1|q_i) = \frac{1}{2}q_i + \frac{1}{2}(2 + 2(a_1 + a_2 < T_{goal})) + v_{slack}a_2$$

- $h(a_1|\bar{a}_1)$ is the same as before as a_2 is set with certainty by the agent

- $p(q_i|a_1)$ is also the same as only value-adding actions (a_1) impact the probability of satisfying a given customer

- Expected firm output E_y and expected agent utility E_U are then as follows:

$$\mathbb{E}_y(\bar{a}_1, a_2) = \int_{-\infty}^{+\infty} \frac{F_{a_1}(a_1)}{a_1 + a_2} h(a_1|\bar{a}_1)da_1$$

$$\mathbb{E}_U(\bar{a}_1, a_2) = \sum_i \int_{-\infty}^{+\infty} u(R(a_1, a_2|q_i))h(a_1|\bar{a}_1)p(q_i|a_1)da_1$$

We now model two cases to consider the impact of an agent's ability to slack. In the first case, we model a lower AHT target (1.25 as before) corresponding to the a_1 target the firm would like the agent to take. Here, we see that the agent derives more utility from not slacking (a_2) than from slacking, and the agent's optimal action remains the same as before. However, in the second case, we increase the target AHT (to 1.75), as a manager might be tempted to from our previous results showing that the agent's behavior can be brought up closer to the firm's optimal target this way; this is shown graphically in the model results in

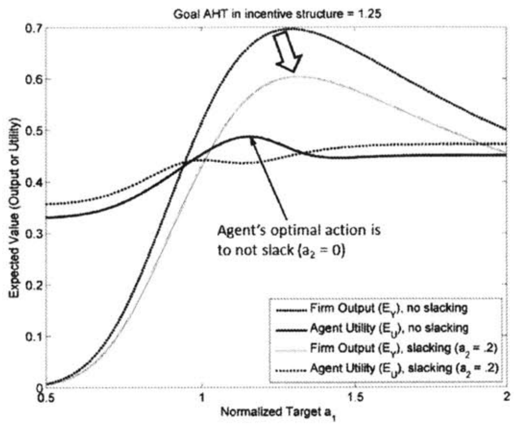

Figure 5-10: With the lower AHT target, the agent's optimal action is to not slack

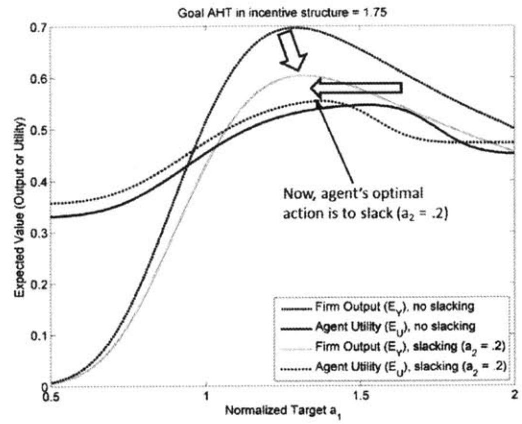

Figure 5-11: With the higher AHT target, the agent's optimal action is now to slack, filling in the extra time with a_2 rather than increasing a_1

Figure 5-10. However, when the agent has the ability to slack, he or she finds that it is more optimal to fill extra time with slacking rather than increasing a_1. In this case, the agent's choice of target a_1 remains low, and the firm experiences a reduced output curve as well, leading to lower firm output, rather than higher as intended. This is shown graphically in Figure 5-11 using numerical results from the model.

Therefore, while the current incentive scheme is suboptimal in the agent's average time spent on value-adding activities, increasing the goal AHT in the incentive structure alone may not shift the agent's optimal action to a higher value-added AHT, as slacking to fill in the added time may provide a higher expected utility.

5.5 Application of Insights from Theoretical Results

To conclude the discussion of incentive alignment in service operations, we draw together the results of our agency incentive model to develop a few conclusions:

- First, we recognize that firm value, as measured by either a reduction in the needed headcount for a given service level *or* the number of correctly serviced customer interactions per unit time, is largely a function of the agent's success rate and the average time spent handling each call. If the success rate is a monotonically increasing function of the amount of value-added time spent on each interaction, with a limit of unity, then there will be a tradeoff between losing value from being too slow and losing value from not delivering reasonable quality; there will furthermore be an optimal point at which an agent provides high quality service without overserving, in which there will be little marginal benefit to quality.

- Time-based goals introduce an incentive structure that will clearly act to keep agents below a given threshold; however, even when the target time is set exactly at the firm's optimal value, the dynamics of a rational agent's utility-maximizing process will tend to cause the agent to underserve the customer relative to the firm's optimal target.

- While time-based goals are imperfect when the optimal AHT is known, they can lead to even more suboptimal behavior when the optimal AHT is unknown, as is often the case in practice. Setting a target AHT too low will exasperate customer underserving; however, if the target AHT is increased to entice an agent to spend more value-added time on each customer, the temptation to slack grows larger as the target AHT is increased beyond the firm's optimal value.

- Sampling-based quality monitoring loses effectiveness as the checklist upon which a customer interaction is graded is less-aligned with the actual requirements of an interaction that is perceived to be high-quality by the customer. When it is difficult to objectively measure the value of a randomly-sampled call, it may be important to add a component of the quality monitoring incentive to either be based more on actual output, such as a measure of first contact resolution, or to make quality assessment more subjective in nature and dictated by management via concepts in relational contracts.[2]

- Quality monitoring is most effective in driving up quality when it is strict, and when the payoff is linear with respect to measured performance. Leniency in quality monitoring incentivizes underserving of customers, as does a scoring system that does not drive the score to zero as fewer customers are not properly served. Although an increase in the number of sampled calls may be necessary for the agent to feel the system is fair, the ability to grade strictly based on outcome with no safety net in the payoff scale will lead to more optimal behavior.

Chapter 6

Creating a Culture of Experimentation and Organizational Learning

Relational contracts to incentivize discovery at all levels of the enterprise

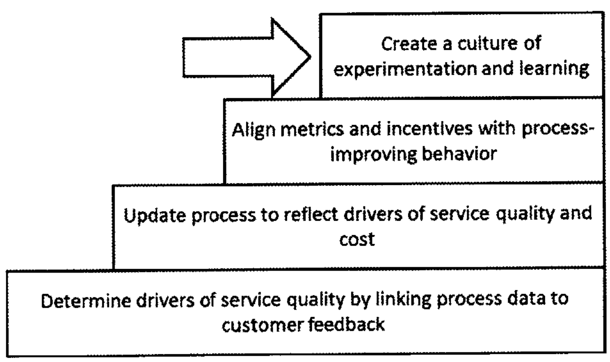

Figure 6-1: Step 4 of thesis framework

After the first three steps of the framework, the key drivers of service quality have been identified, processes have been redesigned to provide the best supporting structure for employees, and incentive mechanisms have been set to align employee behavior with outputs that are important for accomplishing these goals. In the fourth step, we consider the challenge of tying these steps together through iteration and experimentation using relational contracts and informal incentive mechanisms, rather than formal control as before.

6.1 Organizational Learning Revisited

A common theme throughout this thesis has been the philosophy taken from Spear[27] that exceptional service operations cannot be merely designed, but rather must be discovered through experimentation. Practically, this means studying existing problems in the processes in question, developing hypotheses for ways in the system can be improved, and then carrying out changes- perhaps in pilot groups- to test out new improvements to the system. This requires a number of relationships between stakeholders at different levels within the organization; an example of the many parties needed to sustain a learning organization can be illustrated by the following steps involved in testing a new idea:

1. Change management determines a potential improvement to the process.
2. Senior management gives permission to change managers to test changes to the process.
3. Front-line management asks workers to alter work methods from standard work to test new process.
4. Sufficient data is collected to determine effectiveness of new process.
5. Entire process is repeated.

However, the practical application of tying together these multiple levels to achieve experimentation in an organization presents several challenges. In practice, experimenting in an organization requires senior managers to allow front-line, operational managers to plan and test changes in the company's operations, and requires front-line employees to carry out work that differs from standard work correctly to ascertain the impact of the changes in question. While the process is straightforward, it necessarily requires multiple levels of trust to succeed. Managers must sacrifice some performance in the short-run in order to experience improved performance in the long-run, as illustrated in Figure 6-2.

The challenges arise as managers need to trade short-term performance for long-term performance in an environment where results are uncertain; specifically, the experiment may yield poor results, and due to the private information held by each party, it is difficult to know for sure if the failure was caused by poor decisions by managers, poor execution by front-line managers, insufficient effort or mistakes by front-line employees, or just bad luck in picking a reasonable change that didn't work as anticipated. Unfortunately, while formal incentive mechanisms worked for the problems of effort alignment seen in Chapter 5, they are necessarily unable to do so in this situation, in which the desired output of a learning effort is very much subjective in nature. In this case, formal, objective controls will not be able to guide the behavior needed at every level of the organization; instead, the desired organizational behavior requires actions driven by relationships and subjective incentive mechanisms instead. To approach this, we will take a different approach from that in Chapter 5, and will approach the problem with the idea of subjective relational contracts instead.

$$\begin{tikzcd} & \textit{Mechanism Design} \arrow[r] & \textit{Framework Step } 3 \\ \textit{Game Theory} \arrow[ur, "\textit{Objective Control}"] \arrow[dr, "\textit{Subjective Control}"'] & & \\ & \textit{Relational Contracts} \arrow[r] & \textit{Framework Step } 4 \end{tikzcd}$$

6.2 Relational Contracts: Theory and Literature

One way to consider the challenges present in operational experimentation is through relational contracts. [9] In contrast to the formal incentive and control structures discussed in Chapter 5, a relational contract is "a shared understanding of parties' roles in and rewards from collaboration, so rooted in the details of the parties' relationship that it cannot be shared with a court."[9] That is, rather than relying on formal measurements and reward structures, a relational contract allows a mutual relationship to proceed between two parties based on discretion and reputation instead. In another work,[8] Gibbons and Henderson ex-

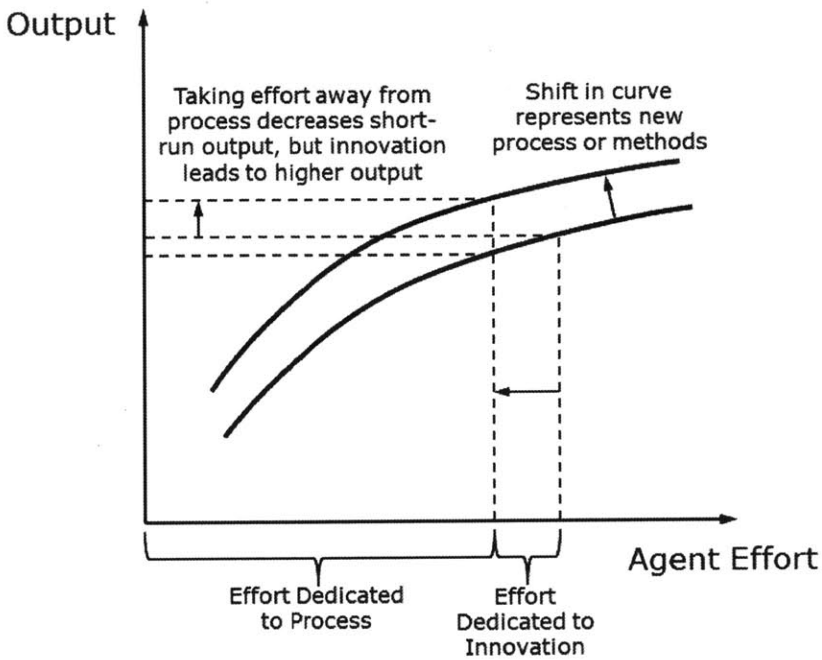

Figure 6-2: Trading off short-term performance for experimentation leads to increasing firm capability, maximizing long-term output

pand on this idea to develop a hypothesis for the reason that "Seemingly Similar Enterprises" (SSE) experience "Persistent Performance Differences" (PPD). They claim that one reason some organizations are observed to outperform others like them in practice is the ability of these organizations to build and maintain such relational contracts. These relationships, in turn, allow the organizations to maintain high performance work systems and solve problems with long-term results in mind rather than succumbing to short-term pressures; indeed, these are characteristics that would be nearly impossible to incentivize with only formal controls and objective, court-enforceable contracts, hence requiring trust-based relational contracts to exist.

While relational contracts allow for subjective metrics and agreements that could not be enforced by a court, this lack of formal control gives rise to a credibility problem; that is, such an agreement gains flexibility at the expense of an explicit, enforceable contract. Gibbons and Henderson model this using the "trust game" described by Kreps;[17] this game is diagrammed in Figure 6-3. In the simplest version of this game, the first player chooses either to trust the second player, and hence begin a relationship; in this case, the

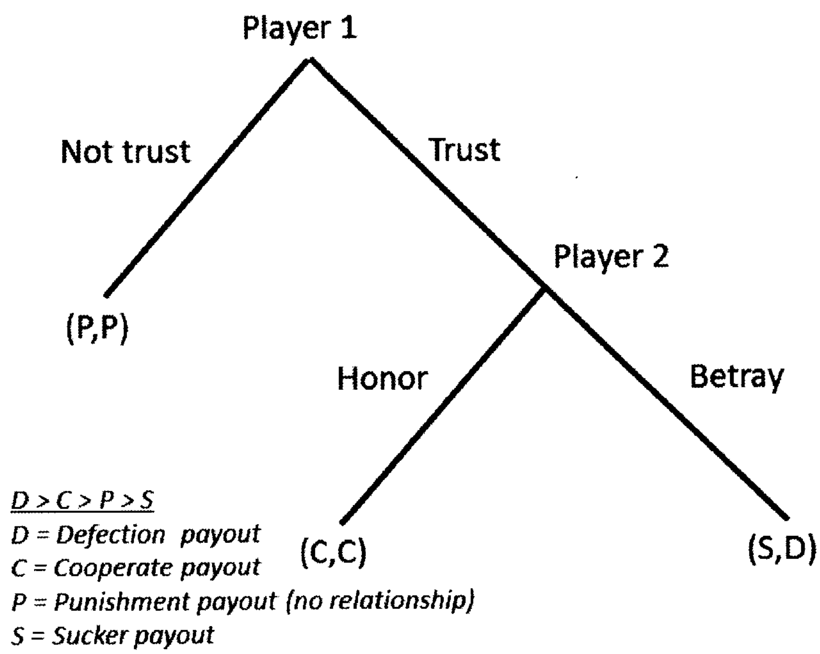

Figure 6-3: The trust game (Kreps, 1990)[17], used to model relational contracts (Gibbons,2012)[9]

relationship must be initiated by trust due to lack of formal control that would be possible by an objective, court-enforceable agreement. Player 2 then has the option to honor the agreement, in which both players receive the "cooperative" payout (C), or betray it, in which case Player 2 receives the higher "defection" payout (D) and Player 1 receives the much lower "sucker" payout (S). If the players choose not to trust each other, perhaps due to a prior defection, there is no relationship and both players receive the neutral "punishment" payout, which is assumed to be lower than the "cooperative" payout that would be achieved under a working relationship. However, the "punishment" payout (P) is still assumed to be higher than the "sucker" payout. When the game is played just once, backward induction will present the equilibrium solution. Player 2's highest payoff comes from selecting "betray" and Player 1, seeing this outcome, will hence choose not to trust, and therefore no relationship will be started when only one interaction is planned.

6.3 Application of Relational Contracts to Organizational Learning

To apply the trust game to an organization attempting to conduct experiments, we model Player 1 as the front-line manager and Player 2 as the agent conducting the experiments. Because actions needed to experiment in an organization require a shared understanding that cannot be written into an employee contract- such as the freedom to deviate from the specified process and reduce output to put time toward executing the changed process and collecting results- forming the needed relationship will require a degree of trust from both parties. The manager must trust that the front-line worker will take the effort to follow the new process with fidelity- perhaps at the expense of performance metrics- and will not take advantage of the newly introduced leniency. At the same time, the front-line worker must trust that the manager will act fairly, excusing any potential drops in performance or failed experiments and rewarding the employee when innovation is achieved, rather than using the results to make the employee worse off than if the experiment had failed. Thus, both parties in the relational contract will play the role of both Player 1 and Player 2 with opportunities to cooperate, defect, and punish defection with future non-cooperation; examples of such actions in sustaining a learning organization are shown in Table 6.1, and examples of such payouts from Figure 6-3 are shown in Table 6.2

While we have demonstrated the equilibrium of the single-shot game to result in non-cooperation, as is indeed observed in many cases in organizations, Gibbons has shown that, when the game is infinitely repeated (or repeated many times without a known ending), reputation concerns and the shadow of the future can sustain cooperation. For an interest rate r by which each player discounts future payouts, Gibbons shows that a cooperative relationship can be sustained if:

$$\left(1 + \frac{1}{r}\right) C > D + \frac{1}{r} P$$

The expression below states that the present value of cooperation in every game is worth more than a defection today plus the present value of no cooperation (mutual punishment) forever,

Front-Line Worker		
Cooperate:	**Defect:**	**Punish:** (in response to defection)
• Follow specified process changes accurately • Take appropriate time needed to carry out experiment and innovate	• Do not follow process • Take advantage of leniency to slack on work	• Do not contribute to experiments / innovation in proces • Only do work as measured by metrics
Manager		
Cooperate:	**Defect:**	**Punish:** (in response to defection)
• Use leniency in judging employee performance • Reward employees who innovate • Pass company gains on to employees	• Raise performance criteria after experimentation • Punish workers for failed experiments • Do not reward appropriate worker effort	• Refuse to experiment or test additional process changes • Take away employee empowerment through additional automation • Enforce metrics and measurements harshly

Table 6.1: Actions of cooperation, defection, and punishment in manager-worker relational contracts in service operations experimentation

if the inequality holds. While this clearly assumes both players employ a "grim-trigger" strategy in which both cooperate until one defects, after which mutual non-cooperation exists forever, it is instructive in showing the general conditions under which cooperation is possible. The expression can be reduced to $r < (C-P)/(D-C)$, demonstrating that cooperation is sustained when players are sufficiently patient (small r), and when the benefits from cooperation ($C-P$, or the value of cooperation above punishment or non-cooperation) is large compared to the temptations of defection ($D-C$, or the value of defection above cooperation). Therefore, while both parties may have something to gain from improving the organization through experiments, achieving sustained cooperation in the absence of formal contracts requires that the benefits are a substantial enough improvement on the status quo for both parties to avoid "defections" such as workers slacking off, or managers halting experiments and punishing workers for not meeting their typical metrics. This will be the difference between organizations that are able to sustain such a culture and those that don't even begin for fear of lacking credibility in its institution.

Furthermore, the challenge does not stop at the interaction between change managers

<table>
<tr><th>Agent</th><th>Middle-Manager</th></tr>
<tr><td>
Payoff "C" (Cooperation):

Performance bonuses for contributing to innovation and increasing firm value
Benefits of higher firm performance passed down in form of improved salary or working condition
Increased job satisfaction from improved operations

</td><td>
Payoff "C" (Cooperation):

Higher operational performance
Improved understanding of process behavior
Higher employee satisfaction and retention

</td></tr>
<tr><td>
Payoff "D" (Defection):

Temporarily benefit from decreased work intensity
Avoid uncomfortable situation or reporting poor results
Decrease work difficulty

</td><td>
Payoff "D" (Defection):

Decrease costs of salary, performance, or overhead
Create reputation for toughness to influence future agent behavior
Use improved operations to increase employee's work intensity

</td></tr>
<tr><td>
Payoff "S" (Sucker):

Lose out on improved pay or working conditions after contributing to improvement efforts
Face tougher metrics or standards after helping to improve operations

</td><td>
Payoff "S" (Sucker):

Lower worker productivity and output as result of employee slacking
Receive unreliable results from experiments that hinder improvement efforts

</td></tr>
<tr><td>
Payoff "P" (Punishment):

Decreased opportunities for improved operations, job satisfaction, or empowerment
Stricter working environment and harsher enforcement of metrics
Formal punishment or job termination

</td><td>
Payoff "P" (Punishment):

Decreased opportunities for improved operations
Less leniency in defining standard work, employee appeals to union rules and formal employment standards
Human Resources complaints
Lower employee satisfaction, retention, and productivity

</td></tr>
</table>

Table 6.2: Potential payouts in a manager-worker relational contract sustaining organizational learning, as illustrated by trust game (Figure 6-3)

Middle-Manager		
Cooperate:	**Defect:**	**Punish:** (in response to defection)
• Run the best experiments that most efficiently use company resources • Do not conduct low-quality experiments • Faithfully report all results	• Select projects by personal preference rather than firm benefit • Use "experiments" as excuse for regularly poor performance	• Do not run experiments or devote effort to organizational improvement • Maximize performance solely based on KPIs- even if flawed • Quit position

Senior Manager		
Cooperate:	**Defect:**	**Punish:** (in response to defection)
• Use leniency in judging manager performance during experiments • Reward managers who innovate • Budget resources in support of organizational learning	• Punish managers for failed experiments • Do not reward appropriate effort • Cut funding	• Refuse to allow future experiments • Enforce KPIs and performance measures harshly • Fire manager / only hire "by the book" managers

Table 6.3: Actions of cooperation, defection, and punishment in manager-manager relational contracts in service operations experimentation

and front-line workers. In many organizations, including Atlantic Energy based on our observations, the relationship between senior managers and the front-line managers may be just as challenging, or even more so. In this case, middle managers face intense pressure from metrics and key performance indicators used by senior management to judge performance, and may be wary of conducting experiments that might fail. On the other hand, senior managers may worry that softening the hard edge of performance measurement may cause middle managers to conduct pet projects that do not use the company's resources in the most efficient manner, or that softer performance measurement may cause middle-managers to be risk-seeking as they face a convex payout from leniency on missed targets, but potential recognition and promotions for success. Examples of this game are laid out in Table 6.3. Indeed, this issue is similar to the agency issue that occurs between managers and shareholders as well, but such a game has received much attention in the past[1] and is not developed in detail here.

We have shown that cooperation may be sustainable if the interactions happen on a continual basis for an indeterminate amount of time (i.e. "infinitely" repeated game in which the discount rate includes players' beliefs on when the game will end); however, sustaining a relational contract to maintain an environment of organizational learning faces at least three major challenges:[1] credibility, clarity, and uncertainty.

1. *Credibility:* Credibility asks "Do I trust you?" and is modeled in this case by the dynamics of the trust game. Each party must trust that the other will faithfully uphold his or her end of the relationship for the relational contract to be sustained, otherwise defection and mutual punishment will ultimately take over the game.

2. *Clarity:* Clarity asks "Do I understand what you are promising?" and captures the difficulty in both parties understanding exactly what it means to cooperate or defect. Indeed, given the subjective nature of relational contracts in the first place, it is nearly impossible to explicitly state what it means to trust the other player or honor an agreement, and it is equally difficult to know what the other party is promising with such agreements as "I will be lenient with performance measurement if you try out this new process"- what does lenient mean in this case? For this reason, problems of credibility are compounded as players may be unsure what is being agreed on in the first place.

3. *Uncertainty:* Although not previously listed in the same vein as credibility and clarity by Gibbons, one final challenge we wish to address is uncertainty, or "How do I know whether the outcome reflected your actions?" Experiments are expected to occasionally fail or produce unexpected results; however, because the actions of workers and managers are rarely measurable or observable, it is often very difficult to know if the results of a project are due to the process changes not producing a desirable output- which is very valuable to know- or due to a mistake or sub-standard effort by the manager or worker involved (which we have deemed a "defection" thus far).

While we have shown how reputation and the expectation of future rewards can incentivize cooperation to overcome the credibility problem, the combination of all three problems

[1]The first two of which, credibility and clarity, were originally described by Gibbons

makes organizational learning a difficult relational contract to sustain. When poor results happen, and it is not possible to distinguish the root cause between poor process and poor effort, a policy that is too lenient will invite defection (slacking or irresponsible project selection), and any policy that is too strict will assure non-cooperation. Indeed, the classic grim-trigger strategy in relational contracts assures cooperation cannot be sustained when a "non-cooperative" result may happen inadvertently, and the lost future value ensures cooperation may not even be sustained from the beginning of the relationship. In the same fashion, confusion on what is promised- the clarity problem- may just as easily be mistaken for a credibility breakdown, much in the same way that uncertainty may be confused for a defection. Considering this more advanced problem, therefore, will require a more advanced technique.

6.4 Sustaining Relational Contracts with Issues of Credibility and Uncertainty

We will now enrich our model to consider relational contracts for which issues of credibility, clarity, and uncertainty must all be addressed to sustain a cooperative, learning environment. Because the ongoing relationship happens nearly continuously, and both parties make nearly simultaneous decisions to trust, honor, or betray each other, we simplify the relational contract from before and consider it as a repeated prisoner's dilemma game (rather than the sequential trust game) as shown below. In this case, we consider the two-person, simultaneous game in which players have the same actions described in Tables 6.1 and 6.3 previously; however, decisions to trust / honor and defect / punish happen at nearly the same time, so we collapse the game to the repeated prisoner's dilemma shown in Table 6.4.

In this repeated game, we are most interested in finding policies or strategies for each player that ensures long-term cooperation between parties to sustain the relational contract, even when faced with issues of credibility and clarity; here, we define a policy to be the action that an agent takes based on both her current state and her belief as to the action the other players will take given their current states. Mathematically, we define a policy π_i

		Principal	
		Trust / Honor	Defect / Punish
Agent	Trust / Honor	C , C	S , D
	Defect / Punish	D , S	P , P

Table 6.4: Actions of cooperation, defection, and punishment in manager-manager relational contracts in service operations experimentation

by the action a_i agent i takes out of the set of all available actions $\mathcal{A}$ given the agent's state s_i, the other players' states s_{-i}, and the agent's belief of the other player's strategic policy π_{-i}; thus, $a_i = \pi_i(s_i, s_{-i}, \pi_{-i})$. In this case, finding policies that sustain mutual cooperation leads to an organization in which the different levels of stakeholders (agents, managers, and senior managers) maintain working relationships that facilitate experimentation and trade off short-term gains for greater long-term gains. If all players are rational, utility-maximizing agents, such a relationship is sustained only in the case in which a Bayesian Nash Equilibrium (BNE) is sustained, for which each player has chosen a strategy that maximizes her payoff from the relationship give her belief of the other players' private information and actions. Again mathematically, the agent's optimal strategic policy π_i^* assuming reward function R_i is given by:

$$\pi_i^*(s_i, s_{-i}, \pi_{-i}) = \underset{a \in \mathcal{A}(s)}{\operatorname{argmax}} \sum_{s'} \Pr(s_i' | a, s_i, s_{-i}, \pi_{-i}) R_i(s', s_{-i})$$

Therefore, a Bayes Nash Equilibrium is achieved for:

$$\pi_i^*(s_i, s_{-i}, \pi_{-i}^*) = \pi_{-i}^*(s_{-i}, s_i, \pi_i^*)$$

6.4.1 Modeling relational contracts with Markov Decision Processes

We now proceed by attempting to determine the existence of BNEs for the game in Table 6.4 that will allow for sustained cooperation in the face of credibility issues and uncertainty, noting that while multiple BNEs may exist for a given repeated game, we will attempt to find the relational contract sustained with the mutual policies yielding the highest expected payoff for all stakeholders. For this, we now model each player as a rational, utility-maximizing autonomous agent, and we model the game as a 1st order nondeterministic Markov Decision Process (MDP) as shown in Figure 6-4. Here, we define four discrete mutual states for the game depending on the combined level of cooperation for each player (the four states are the four combinations possible when each player may "cooperate", given by a perceived act to trust or honor, or "defect", given by a perceived act to defect or punish; hence, <cooperate, cooperate>, <cooperate, defect>, <defect, cooperate>, <defect, defect>). We distinguish between the four combined states representing the perceived final state after each player acts, and the actions of "cooperate" or "defect" to make explicit the cases in which a player may choose to cooperate in the relational contract, but for which the stochastic nature of the process leads to an unfavorable result that would be perceived as indistinguishable from a defection by the other player, allowing the model to represent uncertainty in the game. Of course, we assume that the states are observable in this case and represent the signals each party receives, while the actions of each party are private, so that if an unfavorable result occurs, the other player cannot tell which action was taken.

Given these four mutual states and two actions, the game's dynamics are shown by the MDP in Figure 6-4. Here, Player 1, which we take to be the agent, chooses an action; if the agent chooses to cooperate, the principal (or manager) perceives a result looking like a cooperation with probability p and perceives a result looking like a defection with probability $1 - p$. The ultimate state at the end of this transition, however, depends not only on the outcome of the agent's action as perceived by the principal, but on the principal's action as perceived by the agent. If both actions are perceived to be cooperative, the mutual state will transition to state $< S, S >$ again; if the principal was perceived to defect, the new state would be $< S, D >$. As an example, we show in Figure 6-4 the state transition dynamics for the agent's cooperation succeeding with probability p, and the principal employing a grim-trigger with nondeterministic effects.

Finally, we employ one last simplifying assumption for the case of this study; here, we assume that the game is symmetric, such that the probabilities of inadvertently perceived defection or poor outcomes are similar for each player, and the ratios between payouts such as C, D, S, P are similar for each player as well. While this assumption is reasonable for the relational contract described above for a learning organization, the methods in this thesis

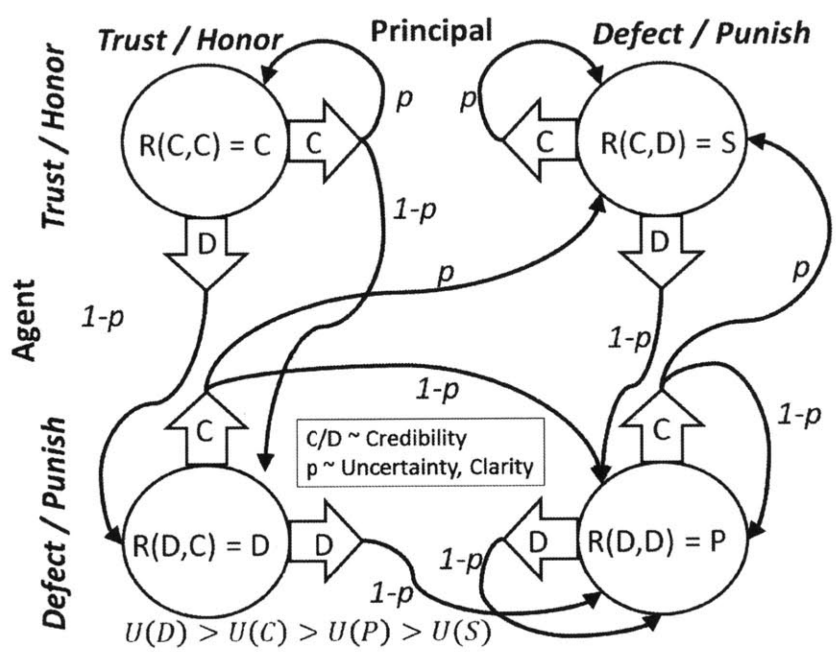

Figure 6-4: Relational contract as a two-person simultaneous repeated game modeled by a nondeterministic Markov Decision Process

can be extended to consider cases in which the game is moderately to severely asymmetric; because the purpose of this analysis is insight and discussion, we have not included it here.

We now proceed with the MDP which, by definition, is a finite-state process in which transitions between states are dictated by both a finite set of actions (in this case, cooperate and defect) and probabilistic transitions based on the actions. An MDP follows the Markov property, in which the following state only depends on the last state and the action taken:

$$\Pr(s_{t+1}|s_{1:t}, a_{1:t}) = \Pr(s_{t+1}|x_t, a_t)$$

The conditional distribution of the next state based on all past states and actions is only governed by the last state and action. In a similar fashion here, we assume that each players action, and the resulting shift in the state of cooperation during the game, depends only on the last state of cooperation. This is not a limitation for any candidate strategies depending on more than the last state, as any process depending on the last n states can be formed as an n^{th} order MDP, which may always be equivalently written as a first order MDP by, for example, defining a state as a sequence of two different states.

One key consideration of an MDP is that it only models a single player making moves at a time; however, this is not a limitation for our purposes of finding BNE. In this case, we assume a policy for the second player, that is, the action the second player will take for each state he may find himself in is set *ex ante*. Of course, we subject the second player to the same stochastic element on actions described before, such that a policy that dictates a cooperative move may end in defection unintentionally, but by assuming a policy for the second player, the MDP can be described completely in terms of conditional transitions and action of the first player alone. Given that the states for both players s_i and s_{-i} have become a joint state that is entirely defined by s_i, and that the other player's policy π_{-i} is uniquely defined *ex ante*, the policies from before simplify to:

$$a_i = \pi_i(s_i, s_{-i}, \pi_{-i}) = \pi_i(s_i)$$

and:

$$\pi^*(s) = \operatorname*{argmax}_{a \in \mathcal{A}(s)} \sum_{s'} \Pr(s'|a, s) R(s')$$

for which the BNE is given by:

$$\pi_i^*(s_i) = \pi_{-i}^*(s_{-i}) = \underline{\pi^*(s)}$$

As a result, the solution to the MDP (the optimal policy for the first player given the policy assumed for the second player) is only relevant if they are the same, which signals a BNE as desired. By checking each candidate policy (in this case, $2^4 = 16$ candidate strategies), the policies forming BNE can be tracked and sorted by the resulting lifetime expected value to determine the best mutual outcome.

To find the optimal policy $\pi^*(s)$ for the repeated game, we find the action at each state s that maximizes the total expected utility gained over all of time $t = 0 \ldots \infty$ given the other player's strategy $\pi_{-i}(s) = \pi(s)$ by symmetry. We define the total expected utility of executing a policy π starting in state s with discount γ:[2]

$$U^{\pi}(s) = \mathbb{E}\left[\sum_{t=0}^{\infty} \gamma^t R(s_t)\right]$$

If the utility of each state is found, then given that the probability distribution of transitioning to a new state s given the current state s and action a is $P(s'|a,s)$, the optimal policy is simply given by:

$$\pi^*(s) = \underset{\pi}{\operatorname{argmax}}\, U^{\pi}(s) = \underset{a\in\mathcal{A}(s)}{\operatorname{argmax}} \sum_{s'} \Pr(s'|a,s)R(s')$$

which we see is simply the definition of the optimal policy from before. This gives the action for each state that maximizes expected utility of the subsequent states. The optimal utility function can be found iteratively through dynamic programming using Bellman's Equation.[24] To do this, the initial utility $U_0(s)$ is given by the intrinsic reward at each state (e.g. $R(C,C) = C; R(C,D) = S$), and then values are iteratively updated by:

$$U_{k+1}(s) = R(s) + \gamma \max_{a\in\mathcal{A}(s)} \sum_{s'} \Pr(s'|a,s)U_k(s')$$

This continuously looks ahead one step and defines the utility of a state as the sum of its intrinsic reward and the discounted expected value of the state with the highest expected utility that it is able to transition to through one of the actions available. Contraction is proven to hold for this operator, meaning that it will eventually converge to the optimal utility for each state if the discount is less than unity. The optimal strategy (policy) is then simply the action in each state that gets to the state with the highest expected utility of the states that are reachable from the current state.

Intuitively, modeling the repeated game as a MDP means the following: If we map out all the states of interaction between the two players, and define the probability distribution for a player transitioning from one state of cooperation to the next for each action, holding the action that the opponent will take from that state fixed, then Bellman's Equation will

[2]The discount γ is the ratio of the value of a reward in game iteration $t+1$ to the value of the same reward in t

iteratively update the expected value of each state by looking one step ahead and adding the expected value of the next state over all possible non-deterministic transitions. Further, these values for each state will converge to the optimal expected value of each state, allowing for the optimal policy of one player to be determined given the policy that the other player is using. If the two policies match, then a BNE has been found; thus, the problem reduces to searching through all possible policies to find the ones that form equilibrium, and then sorting them by the policies with the highest expected optimal utility $U^{\pi^*}(s)$ for the expected starting state s; in this investigation, we started from the state of mutual cooperation.

6.4.2 Analytical results for Bayesian Nash Equilibrium

Using Bellman's Equation to solve the MDP representing our repeated games as described above, we have determined the best BNE for the players as a function of both the discount factor and the probability that mutual cooperation inadvertently breaks down; these results are shown in Figure 6-5. In this case, given that each player has a probability p of successfully cooperating when he intends to, the probability that at least one player inadvertently defects under intended cooperation, potentially ruining cooperation, is $1 - p^2$. The best BNE is defined as the equilibrium with the highest discounted expected value over all future time.

As seen in Figure 6-5, three different strategies may possibly form the optimal BNE depending on the amount of uncertainty in the game. For very high levels of uncertainty, when failure that is indistinguishable from betrayal or punishment is very likely, the optimal strategy for each player is immediate, mutual defection in which no cooperation is ever sustained. For intermediate levels of uncertainty, the optimal solution becomes a grim trigger strategy, in which players cooperate until the first breakdown in cooperation, after which cooperation ceases.[3] Finally, for low to moderate levels of uncertainty, a "forgiveness" strategy becomes optimal in which every deviation from cooperation- whether intended or not- is met with one period of mutual punishment, followed by an attempted return to cooperation (i.e. both players choose trust / honor). This strategy reduces the temptation of cheating by including a period of punishment that ensures cooperation will be in the best interest of

[3]Note that while mutual defection is still a BNE in this case, it is no longer the optimal BNE as grim trigger offers a higher expected value

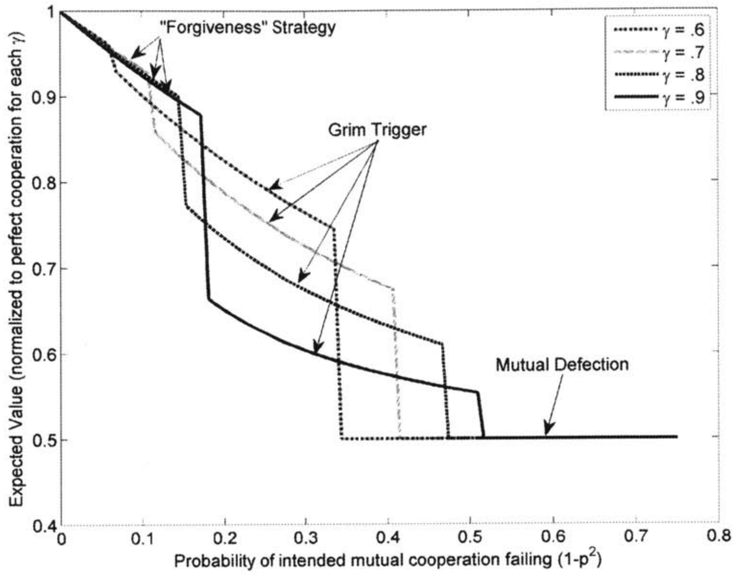

Figure 6-5: The optimal strategy resulting in the highest-valued BNE changes as the system uncertainty and likelihood of poor results $1 - p^2$ increases

	Perceived State	Principal: Trust / Honor	Principal: Defect / Punish
Agent	Trust / Honor	Cooperate	Defect
	Defect / Punish	Defect	Cooperate

Table 6.5: "Forgiveness" Strategy

each party, but allows for enough future value through the restoration of the relationship to enable long-term cooperation. At the same time, cooperation is always easier to sustain at greater levels of uncertainty as the discount rate γ increases; γ can be thought to increase as players' patience increases, or the number and frequency of iterations increases. Matrices describing these policies as a function of the four game states are shown in Tables 6.5, 6.6, and 6.7.

	Perceived State	Principal: Trust / Honor	Principal: Defect / Punish
Agent	Trust / Honor	Cooperate	Defect
	Defect / Punish	Defect	Defect

Table 6.6: Grim-Trigger Strategy

	Perceived State	Principal: Trust / Honor	Principal: Defect / Punish
Agent	Trust / Honor	Defect	Defect
	Defect / Punish	Defect	Defect

Table 6.7: Mutual Defection Policy

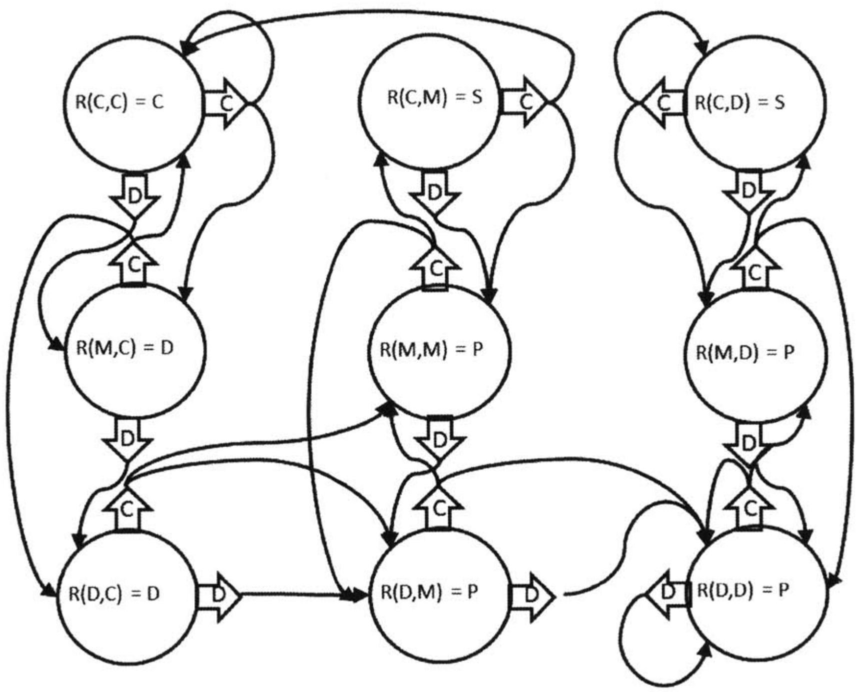

Figure 6-6: Relational contract modeled by nondeterministic Markov Decision Process with intermediate state of defection "M"

Extending the model

Finally, we consider the case in which the parties can remember more than one iteration of the game at a time to allow for potentially more optimal strategies when more uncertainty is introduced into the system. In this case, we consider the case in which players distinguish between one or two consecutive breakdowns in the mutual relationship supporting learning, with the thought that doing so may help sustain long-term cooperation even moreso than the single period punishment (or "forgiveness") strategy did. To model this, we consider a third, "Middle" state (M) between the original two states, representing the state players transition to when they defect (intentionally or inadvertently) the first time and the other player is uncertainty whether they intend to defect forever. Clearly, a player may transition to the pure "Cooperate" or "Defect" states based on his or her next action and the stochastically determined outcome. Payoffs for the intermediate state are the same as for the defect state, as the intermediate state represents a defection; however, additional strategies are now possible to deal with the uncertainty. The MDP that models this new case is shown in Figure 6-6.

Using dynamic programming as before, we find the optimal BNE for varying levels of

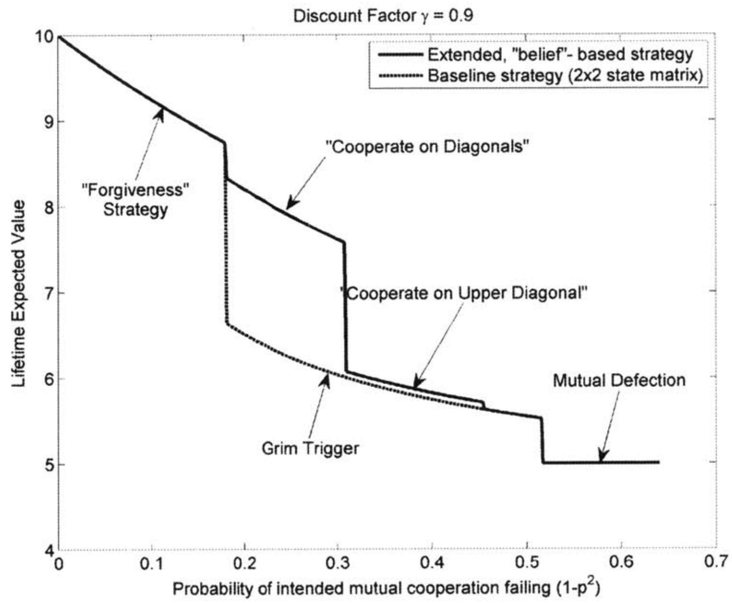

Figure 6-7: Optimal BNE strategies with intermediate, "belief-based" state added

Agent \ Principal	C	M	D
C	Cooperate	Defect	Defect
M	Defect	Cooperate	Cooperate
D	Defect	Cooperate	Cooperate

Table 6.8: 3x3 "Forgiveness" Strategy

Agent \ Principal	C	M	D
C	Cooperate	Defect	Defect
M	Defect	Cooperate	Defect
D	Defect	Defect	Cooperate

Table 6.9: "Cooperate on Diagonal" Strategy

uncertainty as before, and plot the results in Figure 6-7. We see that the new strategies allowing the use of the intermediate state always do as good or better than the original set of strategies, with the largest gains occurring for the range of $1 - p^2$ for which the "forgiveness" strategy from before breaks down; here, a "cooperate on diagonals" strategy or two-period punishment strategy emerges in which any perceived defection is met by two periods of mutual punishment before returning to a state of cooperation.[4] Intuitively, this means that more uncertainty raises the temptation for cooperation breakdown as cheating becomes more difficult to detect. Here, we allow for a longer period of mutual punishment before returning to trust as a result.

The new optimal policies called out in Figure 6-7 are listed in Tables 6.8, 6.9, 6.10, 6.11 based on the new 3x3 state space described in Figure 6-6.

[4]The one exception is the case in which cooperating players are perceived to defect at the same time, in which both will immediately return to cooperation

		Principal: C	M	D
Agent	C	Cooperate	Defect	Defect
	M	Defect	Cooperate	Cooperate
	D	Defect	Cooperate	Defect

Table 6.10: "Cooperate on Upper Diagonal" Strategy

		Principal: C	M	D
Agent	C	Cooperate	Defect	Defect
	M	Defect	Defect	Defect
	D	Defect	Defect	Defect

Table 6.11: 3x3 Grim-Trigger Strategy

6.5 Conclusions

In response to the subjective, relational nature of incentive mechanisms needed to sustain a learning organization, we have analyzed an organization's ability to sustain the relationships needed for a culture of experimentation and discovery through the notion of relational contracts. By modeling the relationships as an MDP, we have considered the ways in which an organization can maintain informal controls dictating the needed tacit understandings behind experimentation even in the face of problems of credibility, clarity, and uncertainty. A few of the main ideas from our investigation are as follows:

- *Understand what relational contracts are needed, and with whom they need to be built:* In this section, we applied the concept of relational contracts to at least two relationships within the organization, namely, the interaction between managers and front-line workers, and the interaction between senior managers and middle managers running the operations. We traced out what roles and responsibilities each party may play in the relationship via Krep's Trust Game in Tables 6.1, 6.2, 6.3. While the relational contract will be unique for each organization that attempts to incorporate regular opportunities for learning into its operations, we believe that successful implementation requires careful consideration of both relationships addressed here, and possibly others as well depending on the organization in question.

- *Although the agreement is informal and metrics are subjective, make the terms as explicit as possible:* While the issue of clarity negatively impacts the credibility of both sides in a relationship, making the roles and responsibilities as explicit as possible can combat clarity issues and support the needed relationships. Although agreements cannot be objectively enforced, making roles and responsibilities clear will help each side to have a basis for trusting the other.

- *Create opportunities for future value in the face of failed experiments with "forgiveness":* When high-quality experiments are run with a reasonable likelihood of success, the optimal Bayes Nash Equilibrium from our analysis demonstrated that a "forgiveness" or one-period punishment strategy can sustain cooperation without incentivizing employees to take advantage of the leniency. While this strategy will neither directly detect cheating by employees nor prohibit cheating from occurring, it will incentivize employees to seek the rewards of the relational contract such as improved performance pay, improved operations and working conditions, greater empowerment, and improved job satisfaction. Thus, the incentives of long-term cooperative payouts will outweigh not only the short-term temptations to betray, but the fear of losing the relational contract as well, as the BNE ensures each party will pick up the cooperation after the "cooling off" period following a failed experiment or perceived failure in cooperation.

- *Run the highest-quality experiments possible (maximize probability p of success):* Clearly, a firm benefits the most when the highest-quality experiments with the highest likelihood of successfully improving firm operations are run. However, our results have shown that, as p decreases, we not only lose firm value but get closer to a threshold beyond which the relational contract cannot be sustained long-term, and the cooperation holding the learning organization together breaks down altogether, resulting in a strict adherence to formal rules and the typical "us versus them" mentality seen too often in the principal-agent dynamics of service operations. While the occasional high risk, high reward experiment is acceptable, attempting too many will lead to too high of a temptation for defection (or perceived defection by the employees) if the poor success rate deflates the reward of cooperation compared to short-term defection.

- *Keep the process of experimentation frequent, with smaller, higher probability experiments if possible:* Running frequent experiments has many structural advantages over larger, higher-risk experiments when only considering the scientific method. Small experiments are easier to design, test, and analyze, and can allow the organization to create a bias for action when it faces a lower threshold needed to deem a change deserving of a pilot. However, frequent experiments also helps to sustain the cooperation needed in a learning organization as well, as it has the effect of increasing the discount factor γ from the analysis, increasing the value of the future by ensuring each party's belief that the relationship will both continue to exist and provide ample opportunity to demonstrate the nature of the relationship through frequent opportunities for operational improvement.

- *Use more creative policies when the experimental success rate is lower:* While a "forgiveness" or one-period punishment strategy is optimal when the probability of success is sufficiently high, we found that the BNE reverts to a grim-trigger strategy for lower p, making cooperation unsustainable in the future. However, we also showed that introducing a set of strategies for which intermediate belief states tracking historical results can help sustain cooperation and increase the value of the relational contract for intermediate probabilities of success. For example, we found that, below the p for which "forgiveness" strategy is no longer the optimal BNE, a "cooperate on diagonals" strategy keeping track of up to two successive unfavorable outcomes lost some value compared to that which would have been achieved by the "forgiveness" strategy, but created much more value in the relationship that was possible with the grim-trigger strategy. Indeed, as the number of intermediate states increases, we postulate that additional value can be achieved for lower probabilities of success p. Therefore, in cases for which high-quality experiments capable of sustaining cooperation via "forgiveness" strategy are not available, we recommend a strategy such as "cooperate on diagonals" or two-period punishment.

Chapter 7

Case Study II: Aligning Incentives for Performance and Learning at Atlantic Energy

A case study to illustrate the last two steps of the framework

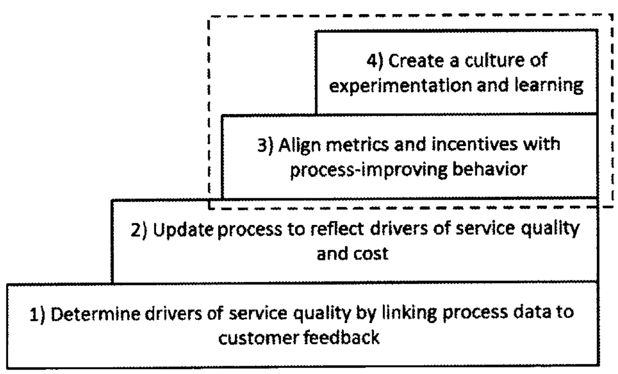

Figure 7-1: Steps 3 and 4 of the thesis framework are covered during Case Study II

In Chapter 4, we demonstrated the way in which the first two steps of the framework were applied to Atlantic Energy; here, we conclude by summarizing ways in which the concepts from the last two steps may be applied to the operations as well. We will walk through the implications of the analysis done in Chapters 5 and 6, and then consider potential solutions to address these concerns. Here, we start by analyzing the current state of the organization through the lens of the game and decision theory analysis in the previous two chapters.

7.1 Analysis of the Current State

7.1.1 Formal metrics and incentives in the Atlantic Energy customer care center

As described before, the account initiation process of interest in this case study is performed by the customer care center, a call center located at one of Atlantic's sites; although some calls are handled by a third party contractor, we consider the operations of the internal call center due to faster path to implementation it allows, with some comparisons added to the third party contractor as well.

Currently, the customer service agents at the call center are judged by two primary metrics: their average handle time (AHT) and their quality assurance score based off of the judging of four phone calls per month. Employees have a time target for which any AHT under the target will earn the employee a score of 4 out of 4 points possible on the quality monitoring evaluation; meeting a less stringent time goal earns the employee a score of 2 out of 4. Similarly, the quality assurance score is based on the scoring of four phone calls each month; scoring is based on a number of pre-defined characteristics, such as whether the employee used the customer's name, provided the correct information, avoided silence, and showed many other behaviors as well. The employee then receives a score of 2, 3, or 4 out of 4 based on their average score on all four phone calls (2 is the minimum score given in this case). The employee then receives a monthly bonus of up to 10% of salary based on the scores for these two metrics. In addition, there is a small additional bonus of approximately 2% of salary linked to a site-wide goal for first contact resolution as well.

On the surface, this scoring system attempts to cover the two main factors affecting output considered in Chapter 5: time and quality. The current system of metrics does provide control over these factors, and certainly helps to provide accountability for quality into the system and controls the time spent on each call to allow efficient operations and reasonable staffing forecasting. For these reasons, there are many qualities to like in Atlantic's current system. However, the results from our previous analysis reveal some potential issues that misalign employee behavior with firm output.

First, the metric uses an AHT cut-off goal, which has been shown to lead to grouping of times around the goal and a potential reduction in time spent on each call from the optimal. Indeed, this prediction has been confirmed by interviews and observations in the call center. Some employees expressed a frustration with the AHT metrics, claiming that they are very short for new employees; because of the high turnover rates common in call centers- and service operations in general- many employees fall into this category. Furthermore, many employees surveyed stated that they spend less time than is optimal on many calls as a result. During observations in the call center, we witnessed firsthand multiple calls in which the agent finished a call prematurely or abruptly passed it on to another party without warning when the problem was taking too long, leading employees to self-select calls that benefit their performance metrics and deflect or underserve those that don't.

Exasperating this problem is the potential lack of strict quality monitoring present. Because a large portion of the bonus is based on subjective quality monitoring, managers have been found to often show leniency in the judging of calls out of concern that judging an employee poorly on one bad call will hurt the employee's pay for the month, particularly when only four calls are judged. Combined with the QA scoring scale that stops at a minimum score of 2 out of 4, this has the effect of creating the nonlinear payoff on quality described in detail in Chapter 5, where our analysis predicted that leniency in scoring and deviation from a reward structure that is linear with respect to performance leads to suboptimal results, in which employees reduce AHT further and place more emphasis on time.

One overarching challenge is the expectations of management and employees regarding bonus structures. Employees, as a rule, do not like variance in take home pay, and as a result, the bonuses may even out over time near the maximum payment. This can lead to problems with quality assurance leniency; soon, the most important metric will informally become time, leading to the issues described above. While many employees at Atlantic Energy genuinely desire to serve the customer, and do so every day, the incentive structure still provides a pressure for low AHT, and as a result, quality assurance scores tend to be much higher than one would predict given first contact resolution rates or measurements of rework and errors as seen by the investigation.

7.1.2 Experimentation at Atlantic Energy

While the performance metrics in place at Atlantic provide controls over performance at both the front-line employee and manager level, they are not currently very forgiving to experimentation and the possibility of failure. Atlantic has a very talented process excellence team that is quite skilled in analyzing operations, identifying problems, and coming up with potential fixes. However, operational improvement could be even more effective in an environment that allows pilots and operational changes to be executed more frequently and more quickly.

At the manager-employee level, employees have the performance metrics described previously to contend with, such as AHT and quality monitoring. Although pilots may introduce new changes that can improve firm performance, there are currently no incentives tied to long term performance improvements, and there is the risk of losing performance and bonus pay in the short run if the experiments do not work as planned. Using the concepts from analyzing relational contracts, the reward of cooperation over the status quo $C - P$ is small, making sustained cooperation difficult to initiate. Rather than solving issues of credibility associated with subjective performance evaluation during rapid iterations, there appears to be no shared understanding of expectations associated with constantly trying out new ideas, described before as issues of clarity.

While the manager-employee relationship can benefit from being more explicit, even if subjective, experimentation ultimately can be held back by the relational contract at the management level. Senior managers have defined strict KPIs for customer service and cost competitiveness that impact discretionary bonuses for managers and tend to be unforgiving of failed experiments. While the potential reward for cooperation is improvements along all of the KPIs in the long run, such benefit needs to be sold to many stakeholders to agree on a pilot or experimental change to a process, which is difficult to achieve in practice. Because the KPIs tend to be based more on goals rather than absolute performance, the payoffs tend to be concave, falling off in times of failure but leveling off if future improvement is made. This will tend to make employees risk averse when facing decisions of experimentation, lowering the perceived value of experimentation ($C - P$); when combined with few or no implicit promises of second chances when experiments do occasionally fail, a bias for inaction can result.

7.2 Defining Countermeasures for Performance Improvement

7.2.1 Aligning performance goals with optimal output and improvement

To counteract some of the issues described for the current state, several countermeasures can be implemented. First, increased emphasis should be placed on quality by implementing a stricter set of guidelines for quality measurement. At the time of this thesis writing, Atlantic was considering doubling the monthly number of quality monitoring calls to eight; this will provide a large step toward creating more granularity in the quality measurement system. This granularity should be used to judge performance on a stricter scale, providing a linear bonus reward based on the agent's performance for the month, with more weight for each call placed on the outcome of the call rather than a predefined list of actions that may not represent the customer's definition of quality. Because the analysis in Chapter 2 showed that customers primarily care about having their call answered correctly, this should be the primary component of the quality score for each call.

The linear payoff (e.g. 2 good calls out of 8 yields 25% of the available points), larger emphasis on call result, and stricter approach to grading will more effectively align employee behavior with quality; however, it will necessarily provide more variance into monthly bonus pay, which risk averse employees will not prefer. While we feel this is a necessary change to make the current bonus structure more effective, we offer one potential tradeoff to deal with the current cultural expectation placed on bonus pay. While the quality monitoring will have the new characteristics described, one potential solution is to base the monthly bonus over a series of multiple months, using more of a Bayesian approach to determining performance rather than a pure frequentist, allowing management to account for past performance in judging between poor performance and bad luck for one month. A multi-month weighted average, potentially weighted to bias towards more recent results, can potentially smooth out the variation in pay month-to-month and reward employee performance improvement at the same time, allowing for the stricter, more meaningful approach to quality monitoring to happen.

Next, the two-bracket approach to time-based goals is a good start toward avoiding the negative effects of AHT cutoffs seen in the previous analysis. To continue this idea, one potential solution is to incorporate more time ranges, such as four or eight, that provide more granularity into the time-based measurements. This can have two potential effects: first, it avoids the drop-off effect in payout that forces employees to keep times lower than is optimal by softening the impact when missing a goal; if employees become used to falling within a range of scores that increment a fraction of a point at a time, then, coupled with the stricter quality monitoring, the tradeoff between losing a fraction of a point for a small increase in AHT to meet the quality measurements can happen, with the incentive to keep calls as short as possible still in effect. At the same time, having many goals rather than one goal can allow for scores to increase when exceeding the current AHT goals; that is, there is incentive to innovate and perform beyond one firm-wide goal cutoff. While this effect must certainly be capped to avoid extremely short calls (as predicted by the previous analysis), this cutoff need not be at the traditional single cutoff, lessening the importance of picking the correct institute-wide AHT.

The center-wide first contact resolution incentive is a good metric, and perhaps should be emphasized more in the future. Without explanation of the goal, agents may be frustrated by the metric as they may not feel they have individual control over it, due to its reliance on center-wide performance. However, if the link between employee effort and innovation-particularly knowledge sharing- is emphasized, this metric has the potential to incentivize quality-producing behavior. Increased visibility and emphasis on the metric, including framing each improvement effort as a way to improve the metric, can help on this front.

Finally, one overarching idea from this analysis is the benefit of a performance bonus when it can be thought of as a continuous scale than a single target. When goals are set as cutoffs, they incentivize behavior to hit the targets, limit incentives to improve on the targets, and cause employees to expect the target rewards at the potential expense of the effectiveness of the performance measurement system. Changing the understanding of the incentive bonus system is very difficult, as employees care deeply about the issue and any cap set on the bonus pay will provide an implicit target that the system will gravitate toward. However, woven through this discussion is one important recommendation that requires attention: increasing the granularity of the performance metrics and resulting rewards, and allowing them to extend to both sides of targets, can provide much more powerful incentive structures when influencing employee behavior.

7.2.2 Increasing the value of experimentation and cooperation in organizational learning

As stated before, creating a culture of experimentation and learning requires relational contracts and shared understandings that reward cooperation in maintaining such a culture. While the cutoff-based performance goals keep the reward of cooperation $(C - P)$ low and create the fear of more difficult goals as performance improves, explicitly adding incentives for improvement into the system can help.

At the manager-employee level, employees must simultaneously see that they will gain from innovation and understand that they will not be punished when experiments fail. Changing the performance metrics from cutoff goals to a continuous scale helps this objective by providing a benefit to employees when the experiments succeed and performance improves. At the same time, an understanding that bonuses will not be decreased if the experiments fail- and that participation in experiments will subjectively be taken into account, can help form the relational contract as well. However, as seen in Chapter 6, reneging- whether perceived or real- will almost certainly kill the culture of experimentation and result in a reversion to the status quo. Therefore, management must credibly commitment to these promises, but should implement the concept of the "forgiveness" strategy from before as well. Apparent breakdowns must result in at least a temporary hiatus from experimentation- and a delayed opportunity to gain from the payoff $C - P$- to prevent undetected defection in face of the agency problem and private information involved in such a relationship. However, opening up the possibility for future pilots and experiments can help the organization to implement the studies it needs over time to iterate on the changes suggested by the first three steps of the framework and achieve improvement.

Finally, these concepts must be applied at the management level as well. Performance goals that provide concave- or even flat- rewards rather than more linear ones can help provide the incentive to risk experiments in the short run to gain in the long run. Implicit agreements on pilots are even more important at this level as well; the concepts of one- or two-period punishment from Chapter 6 can be especially helpful in this case, as managers who implement poor or unsuccessful experiments should not immediately receive approval to continue without oversight, but continued experimentation will be necessary to eventually find the solutions that will lead to long term performance.

With both of these relational contracts, the ideas presented from theory still hold true: although the responsibilities and rewards associated with the relationships will necessarily be subjective, they still must be explicit, credible, and clear. Rewards must be tied to innovation and improvement to provide the incentive for cooperation, and breakdowns must be addressed to lower the temptation of the wrong effort, but forgiven to allow the relationship to continue. With a few steps in this direction, the organization may be able to implement experiments more easily and achieve the iterations needed for this framework to ultimately find the best solutions.

Chapter 8

Conclusion

8.1 Retracing the Thesis Framework

As organizations increasingly compete with service quality, the need to align an organization with the factors that drive customer satisfaction increases as well. Companies have long implemented customer feedback and satisfaction surveys into their service operations, and with the advent of increasingly powerful statistical tools and the advent of big data analytics, the pursuit of such feedback has only accelerated. However, such an effort that is only built on feedback averaged over many customers may not offer the whole picture. An improved system tracks the actual drivers of service quality on a customer-by-customer basis, uses proper statistical tools to interpret such drivers, applies best practices from managerial science in addressing such drivers, and considers incentives and relationships when designing organizational mechanisms to create the desired system of high-performing service operations.

This thesis has addressed this need through its four-piece framework. With the first step, we have addressed the proper way to obtain and interpret feedback on service operations from all relevant stakeholders, including employees and customers, and we have demonstrated some of the statistical tools that reveal the key drivers of service quality in a particular organization. In the second step, we have explored some of the best practices in management science to improve the organization's service operations by adjusting processes to put the customer service representatives and managers of the organization in the best position to successfully serve the customer; this was accomplished through concepts from lean manufacturing, systems thinking, and organizational learning. In doing so, we further addressed the true nature of many strategic trade-offs in service operations, such as that between quality of services and cost of services, and explored how proper insight into the relationship between the two can help a manager to choose a more optimal position between the two.

In the third step of the framework, we have built on the changed process by recognizing that process by itself is insufficient for superior service performance when people- rather than process- ultimately deliver quality service. Through the use of relational contracts and incentive mechanisms, we have considered the ways in which the agents of a service organization can be better aligned with behavior that most benefits the organization, rather than simply focusing on simple proxies for cost performance. Finally, we have considered the role of organizational learning in improving service operations, and have investigated the ways in which the stakeholders of an organization can be aligned with behavior that facilitates learning.

8.2 Key Lessons

Throughout all of this, we have seen that a simplistic view of service quality is insufficient; rather, a systems-level view that encompasses all stakeholders and all facets of service delivery is needed. An organization must not simply collect feedback on service operations, but analyze and interpret it correctly in the context of feedback from all stakeholders, and once conclusions have been drawn from such an analysis, a company must act on it by changing processes and systems to change the customer pain points identified by such an analysis. However, we have also shown that changing the process is not enough; although the process places employees in a position to succeed, the service is ultimately delivered by the people rather than the process and, without a proper alignment between the people in the organization and the goals of the operation, the process by itself will be insufficient.

Throughout this thesis, we have provided as an example the application of the thesis framework to Atlantic Energy. While the case study at Atlantic has aided in the description of the thesis framework by way of concrete example, it has revealed several key takeaways as well. The case study has shown the effectiveness of the feedback analysis techniques from Chapter 3, in which the analysis strongly pointed to key service quality drivers that agreed with employee feedback. It has also shown how concepts from operations management can be concretely applied to allow an organization to not just measure, but act on customer feedback. However, it has also shown the importance of the systems-level view this thesis advocates, and the importance of fine-tuning an organization's service operations through incentive mechanisms, relational contracts, and proper principal-agent alignment, and the difficulty in achieving the potential benefits of service quality improvement efforts without successfully addressing these issues. Ultimately, such properly designed mechanisms that incentivize behavior that actually benefits the organization through customer service and organizational learning, in addition to proper measurement and analysis of the factors that actually drive service performance, can enable the competitive advantage in service quality that many organizations seek.

8.3 Generalization and Further Application

Although this thesis has used service operations in the call center of a power utility as an example, the framework presented here can be applied to service operations in a broader sense as well; one primary example is in the airline industry. Although the competitive economics of the industry has long forced an intensive effort at cost reduction and rapid turnarounds in service, industry players have long been scrutinized from a service quality standpoint as well[23], creating the tension that has been addressed throughout this thesis. From reservations to check-in, baggage handling, boarding, and in-flight experience, customers interact with many facets of the organization and demand a high level of service while retaining the ability to easily shop around for tickets each flight as well. Aligning so many parts of an organization with all of the employees involved in delivering service cannot be accomplished through a simplistic view of service operations. Determining what is actually important to a customer and what trade-offs between customer desires and operational economics are actually at play requires the types of analysis described in this thesis; addressing all parts of the organization to design the proper incentive mechanisms in which the employees operate does as well. In the future, we would be very interested in applying these concepts to such a complex organization in testing its ability to address address this challenge as, building on this thesis, we would argue that the multi-faceted, system-level view incorporating correct feedback analysis and principal-agent alignment issues at the employee level is a superior way to address such an organization.

The areas for potential application extend further, however. We propose that almost any organization that involves the use of people to deliver quality service operations to customers under the tension of competing goals- such as time and cost- can benefit from the concepts of this thesis. While the customer support functions of other companies may be most closely related to Atlantic's operations, companies in hospitality, financial services, retail, education, and even professional services may benefit as well.

8.4 Closing

While this thesis does not offer the complete roadmap to solving every service operations problem, it aims to expand the view of operational managers to include principal-agent and relational contracts issues, and it aims to demonstrate ways in which to use data more effectively. With an expanding view of service operations management, organizations may move beyond the measurement of customer feedback and beginning steps of operational improvement to a sustainable shift in organizational behavior that accomplishes what all such efforts aim to do: improve service quality for the customers that ultimately give the organization a need for its operations.

Bibliography

[1] A. Agrawal and C. Knoeber. Firm performance and mechanisms to control agency problems between managers and shareholders. 1996.

[2] G. Baker, R. Gibbons, and K. Murphy. Subjective performance measures in optimal incentive contracts. 1994.

[3] L. Brown, N. Gans, and et. al. Statistical analysis of a telephone call center: A queueing-science perspective. 2005.

[4] G. Cachon and S. Terwiesch. *Matching Supply with Demand: An Introduction to Operations Management.* McGraw-Hill, New York, New York, 2013.

[5] Bernard Caillaud and Benjamin E. Hermalin. Hidden action and incentives. 2000.

[6] John Devin, Lal Shyam, and Michael Zea. The human factor in service design. (1):118–123, 2012.

[7] R. Gibbons. Incentives in organizations. 12(4):115–134, 1998.

[8] R. Gibbons and R. Henderson. *The Handbook of Organizational Economics*, chapter What Do Managers Do? Exploring Persistent Performance Differences among Seemingly Similar Enterprises. Princeton University Press, 2014.

[9] Robert Gibbons and Rebecca Henderson. Relational contracts and organizational capabilities. 23(5):1350–1364, 2012.

[10] P. Green and V. Srinivasan. Conjoint analysis in consumer research: Issues and outlook. *Stanford Graduate School of Business Research Paper No. 435*, 1978.

[11] Sanford J. Grossman and Oliver D. Hart. An analysis of the principal-agent problem. 51(1):7–45, 1983.

[12] Andrew Heath. How customer satisfaction drives return on equity for regulated electric utilities. *JD Power and Associates*, 2012.

[13] B. Holmstrom and P. Milgrom. Multitask principal-agent analyses: Incentive contracts, asset ownership, and job design. *Journal of Law, Economics, and Organization; Cambridge University Press*, 1991.

[14] Lokesh Jasrai. Predicting customer satisfaction towards mobile value-added services: An application of multiple regression. 13(1), 2014.

[15] Anne Julien and Christina Tsoni. Front-line employee versus customer perceptions of quality attributes. 20(2):48–59, 2013.

[16] Philipp Klaus and Stan Maklan. Towards a better measure of customer experience. 55(2):227, 2013.

[17] D. M. Kreps. *Corporate culture and economic theory*, chapter Perspectives on Positive Political Economy, pages 90–143. Cambridge University Press, 1990.

[18] Jonathan Levin. Relational incentive contracts. 93(3):835–857, 2003.

[19] R. Luce and J. Tukey. Simultaneous conjoint measurement: A new type of fundamental measurement. *Journal of Mathematical Psychology*, 1964.

[20] D. Nightingale and J. Srinivasan. *Beyond the Lean Revolution.* Amacom, New York, NY, 2011.

[21] A. Parasuraman, V. Zeithaml, and L. Berry. Servqual: A multiple-item scale for measuring consumer perceptions of service quality. *Journal of Retailing*, 64(1):12+, 1988.

[22] Erica L. Plambeck and Stefanos A. Zenios. Performance-based incentives in a dynamic principal-agent model. 2(3):240–263, 2000.

[23] Steven E. Prokesch. Competing on customer service: An interview with british airways' sir colin marshall. *Harvard Business Review, The Magazine*, 1995.

[24] S. Russel. *Artificial Intelligence: A Modern Approach.* Prentice Hall, Upper Saddle River, New Jersey, 2010.

[25] Avijit Sen. Identifying system-wide contact center cost reduction opportunities through lean, customer-focused it metrics. Leaders for global operations master's thesis, Massachusetts Institute of Technology, 2009.

[26] K. Sivakumar, Mei Li, and Beibei Dong. Service quality: The impact of frequency, timing, proximity, and sequence of failures and delights. 78(1):41–58, 2014.

[27] S. Spear. *The High Velocity Edge.* McGraw-Hill, 2009.

[28] G. Vining and S. Kowalski. *Statistical Methods for Engineers.* Brooks/Cole Cengage Learning, Boston, MA, 2011.

[29] J. Womack and D. Jones. *Lean Thinking.* Simon & Schuster, New York, New York, 2003.