Implicit Costs of Data and Analytics

by

Bryan A. Kapicka

B.S. Industrial and Enterprise Systems Engineering,
University of Illinois at Urbana-Champaign, 2009

B.S. Mathematics,
University of Illinois at Urbana-Champaign, 2009

Submitted to the MIT Sloan School of Management
in Partial Fulfillment of the Requirements for the Degree of

Master of Business Administration

At the Massachusetts Institute of Technology

June 2014

Signature of Author Signature redacted
MIT Sloan School of Management
May 9, 2014

Certified by Signature redacted
Erik Brynjolfsson, Thesis Supervisor
Professor of Information Technology, MIT Sloan School of Management

Accepted by Signature redacted
Maura Herson, Director of MIT Sloan MBA Program
MIT Sloan School of Management

This page intentionally left blank.

Implicit Costs of Data and Analytics

by
Bryan A. Kapicka

Submitted to the MIT Sloan School of Management on May 9, 2014 in Partial Fulfillment of the Requirements for the Degree of Master of Business Administration

Abstract

Firms have been able to utilize data and analytics to achieve a variety of economic benefits. To realize this value, firms have to invest in the necessary information technology, process updates, and employee training. These costs are straightforward, but firms also incur implicit costs, the costs of mitigating potential risks and maximizing firm value from data and analytics. These costs are less well understood. This paper focuses on two of these costs, the mitigation of the adverse effects of metrics and the investment required to effectively complement people with information technology.

The first cost is the adverse effects of metrics and refers to the potential for metrics and analytics to compromise business objectives that they were originally intended to enhance. The analysis of this cost primarily utilizes Holmstrom and Milgrom's Multitask Principal-Agent Model to evaluate the impacts that incentives, metrics, measurability, and job design have on the firm's payoff. This model and ensuing analysis provide guidance for firms to avoid the pitfalls that accompany an increased reliance on data and analytics.

The second cost refers to the firm's investment to complement people with information technology to maximize their payoff from data and analytics. The evaluation of this cost discusses the conditions under which it is appropriate to complement, or substitute, humans with technology. In the scenarios where people are complemented by technology, this paper outlines additional practices and examples to highlight ways in which a complementary relationship between people and information technology can be cultivated. This discussion covers the efforts to shift people away from solely relying on intuition, while preventing them from blindly accepting data and empowering them to deal with the inherent complexity of new information afforded by data and analytics.

The analysis and discussion of each cost references existing research and case examples. This paper intends to further the understanding of these costs as well as identify future opportunities for research.

Thesis Supervisor: Erik Brynjolfsson
Title: Professor of Information Technology, MIT Sloan School of Management

This page intentionally left blank.

Acknowledgements

I would like to express my sincere appreciation to those whose guidance and support have aided me throughout this process and throughout my time at Sloan. Specifically, I would like to thank my thesis advisor, Erik Brynjolfsson, for his help and supervision. I owe a special thanks to Kristina McElheran and George Westerman at the Center for Digital Business for their invaluable feedback. I am truly grateful for their willingness to review my work. I would also like to thank Barbara Wixom at the Center for Information Systems Research for her guidance at the onset of this effort.

I would also like to acknowledge my classmates who were always happy to discuss ideas and willing to provide honest feedback.

Finally, I would like to thank my family. My parents and sister, Barbara, Robert, and Alexandra, are a constant source of encouragement and I cannot thank them enough for their support.

This page intentionally left blank.

Table of Contents

This page intentionally left blank.

1 Introduction

Data and analytics have created a variety of economic benefits and have proven advantageous for firms that can effectively harness the potential of such practices. Caesars Entertainment drives customer satisfaction by cultivating a superior customer experience and "turning customer data into customer loyalty" (Welch & Westerman, 2013). Caesars Entertainment's ability to provide this superior experience is largely attributed to its use of data and analytics, including the vast amounts of information they glean from their customer loyalty program, Total Rewards. Caesars Entertainment is one of many success stories in which firms have found unique and innovative ways to utilize data and analytics. Such successes have spurred significant interest among firms in exploiting the capabilities of data and analytics (Heudecker, 2013). The promised value of data and analytics can sometimes overshadow the fact that increased reliance on these tools also introduces costs for firms. These costs can hinder the firm's ability to maximize the benefits of data and analytics. To maximize the potential of data and analytics, firms have to address two of these costs, mitigating the negative effects that metrics can have on firm value and ensuring that information technology effectively complements people for the firm's benefit.

Data and analytics efforts require firms to invest in technology upgrades, process updates, and employee training, but the extent of these costs invites further exploration. Unavoidable expenses, or explicit costs, include the price of software and the fee for implementation. Implicit costs, on the other hand, are incurred by firms in their efforts to derive maximum value from and protect against additional risks that accompany an increased use of data and analytics. These costs are less straightforward, but existing research and published case examples provide a foundation upon which these costs can be explored.

In particular, this paper analyzes two implicit costs. These costs require firms to make investments to:

1. Mitigate Adverse Effects of Metrics – mitigation of the potentially negative impact on firm benefits that result from an increased reliance on the metrics that accompany data and analytics

2. Complement People with Technology (Data-Driven Decision-Making) – development of a balance between humans and technology that ensures people are complemented by data, analytics, and information technology for the maximum firm payoff

The analysis of these costs heavily relies on the model proposed by Holmstrom and Milgrom in their paper "Multitask Principal-Agent Analyses." This model focuses on the incentives and the relationship between principals and agents when the agent is balancing multiple tasks (Holmstrom & Milgrom, 1991). This provides a useful framework to explore the impacts of metrics and analytics related to these two costs. This paper's analysis also highlights relevant research and case examples.

These two implicit costs are not intended to be comprehensive, as there are likely additional costs that will reveal themselves as firms continue to increase their use of data and analytics. Instead they are meant to be an initial set of implicit costs that each requires further examination.

1.1 Motivation

The interest in data and analytics has resulted in a variety of publications, technologies, and services that only furthers the hype surrounding these topics (Heudecker, 2013). Individuals and organizations looking to capitalize on this interest are espousing what they believe to be best practices in support of their efforts to sell products or services that are well aligned with such suggestions. The interest in data and analytics, combined with these unproven best practices, puts firms at risk of blindly implementing data or analytical practices that may not have the promised positive return.

In this paper, I propose and analyze two implicit costs to increase the awareness and understanding of these costs by managers and firms. For each cost, I propose a set of practices that have the potential to address the costs that may attend an increased use of data and analytics. Again, these practices are not comprehensive. They are a summary of ways in which select firms have found success in addressing these implicit costs. When necessary, this discussion includes the context under which these costs exist and the conditions under which the associated practices apply. Outlining such limitations is an attempt to prevent managers and firms from overestimating the two implicit costs or misappropriating suggested practices. These limitations also serve as a boundary that should be further explored in future research.

2 Historical Context and Prior Research

2.1 History of Information Technology

Part of the promise of technology is its ability to enable firms to increase benefits or reduce costs by aiding, supplementing, and even replacing people (Heudecker, 2013). The use of data and information technology in firms has become common practice, but the extent and means by which they attempt to extract value from these assets is still evolving. This section examines the history of data and analytics, outlining firms' increased reliance on data across three distinct phases:

Phase I: Pre-Database

Before databases existed, executives primarily relied on instincts and experience, as limited data points were available. Data could be collected, but was generally recorded on paper and calculations were relatively simple. There were less complex inputs in this decision-making process and decision-makers could understand the source of these inputs.

Phase II: Information Technology

Throughout the 1950s and 1960s, firms began realizing the benefits of data thanks to continued improvements in data processing and storage such as the introduction of the database and subsequently the relational database, a database capable of representing relationships between data sets (Codd, 1970). These new assets provided firms with the ability to store and utilize data such as customer information and transactions. Since the introduction of information technology, technological advances have improved the firm's ability to utilize data and decreased the costs associated with data capture and storage. In the 1980s, the concepts of business intelligence and data warehousing were introduced to provide broader use of data across organizations and solutions leveraging these concepts have been utilized by firms ever since (Hayes, 2002). Firms primarily captured data essential to their sales and operations such as customer information, sales history, and inventory levels. Much of this data was captured to enable existing business processes, providing firms with data sets that could be further analyzed and incorporated into the decision-making process. It allowed decisions to be better informed by underlying information.

Phase III: Big Data and Analytics

In recent years, the concepts of analytics and big data have been introduced as another advancement in information technology. Recent evolutions in technology have enabled new ways to capture many different data sets from various sources (variety), increased the speed at which data is created and processed (velocity), and increased the quantity of data that can be stored (volume) (Laney, 2001). In particular, the Internet has allowed for the rapid creation of data, as well as new interactions for firms to track, such as browsing behavior. These characteristics collectively distinguish big data from traditional data. Doug Laney coined the 3-Vs approach to big data, which is further defined by Gartner as "high-volume, high-velocity and high-variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision making" (Gartner; Laney, 2001).[1] The supporting processes of analyzing, identifying patterns, and extracting insight from these large and varied data sets is referred to as analytics. Gartner asserts that "analytics leverage data in a particular functional process (or application) to enable context-specific insight that is actionable" (Kirk, 2006). The full potential of big data and analytics is unknown as these concepts are still evolving. Rigorous research is being done on the subject by individuals such as Adam Saunders and Prassana Tambe, but generally research on the subject is limited (Saunders & Tambe, 2013).

Benefits of Data

The increasing availability of and ability to capture and maintain data has been exploited by a variety of firms, many of which have achieved significant success in pursuing data-driven business opportunities. Most notably, Google's and Facebook's initial product offerings were almost exclusively dependent on data, providing people with new ways to search massive amounts of web information and to use data-driven interactions in a digital social network (Battelle, 2005; Locke, 2007). Amazon was able to disrupt traditional book sales and tailor

---

[1] Silicon Graphics chief engineer John Mashey arguably coined the term "Big Data" in his seminar entitled ""Big Data & the Next Wave of InfraStress." during the mid 1990s (Laney, 2013)

recommendations based on data collected on user browsing and purchase history, data that brick and mortar competitors were inherently unable to capture efficiently (Brynjolfsson & McAfee, 2012). Other companies have been built exclusively by offering data products and services. Palantir provides software for "integrating, visualizing and analyzing the world's information" (Palantir, 2014). Their software is purchased by, for example, government organizations to process intelligence information and corporations to identify fraud. Many startups have also achieved success in providing data services. Locu amasses menu data from restaurants and sources this information to a variety of websites, including YellowPages.com, Yelp, and TripAdvisor (Locu, 2014). While there has been an increase in firms built on data offerings, this is not an entirely new phenomenon. PASSUR was founded in 1967, offering predictive analytics on flight schedules to airlines (Brynjolfsson & McAfee, Big Data: The Management Revolution, 2012).

The large volume of data has also allowed for the discovery of solutions to problems that were previously unsolvable. In particular, artificial intelligence algorithms and models were not particularly good at language recognition or vision. Translation algorithms were considered difficult. In these nuanced use cases, the use of large amounts of data provided enough context and information to develop new solutions to the challenges of language recognition, computer visual recognition, and translation (Harford, 2014; Ordonez, Deng, Choi, Berg, & Berg, 2013). These solutions were not possible when only applying large amounts of processing power. Google's self-driving cars are a recent example of the power of big data to solve new problems, as these cars take advantage of the ability to rapidly (velocity) process a wide array of signals (variety) in large quantities (volume) to match or exceed the driving abilities of humans (Rosen, 2012). These new solutions are the basis upon which firms can offer new products and services.

While data and analytics have recently played a major role in creating new business opportunities and solving previously unsolvable problems, firms have been taking advantage of data to improve their operations for decades. In the early 1980s, Seven-Eleven Japan was one of the early adopters of such practices, consolidating disparate point of sales information and collecting information on customer gender and approximate age (Nagayama & Weill, 2004). Caesars Entertainment is well known for the data collected on customers through its rewards program, using this data to improve its operations and offer targeted promotions. This data allowed Caesars Entertainment to provide high levels of customer service and run more effective marketing campaigns (Welch & Westerman, 2013). These examples are consistent with recent research showing that investments made in big data and analytics are associated with a higher market value (Saunders & Tambe, 2013).

Applicability and Limitations of Data

The success that firms experience as a result of applying information technology, big data, and analytics has led to much hype about these practices in the business community. As more advanced data and analytical technologies become available, firms have a greater opportunity to replace human intuition, analysis, and even decision-making with information technology. Some business processes have been fully automated such as certain accounting and payroll functions. Such processes no longer rely on interactions with people with the exception of initially setting up and continuously monitoring the process. While firms have greater potential to decrease the involvement of humans in any process, minimizing human involvement by substituting information technology wherever possible does not necessarily create more value for a firm.

There is evidence that fully automated tasks are not ideal in every situation given the complementary benefits of humans and information technology working in tandem. For example, while computer programs have been able to beat chess grandmasters since the late 1980s, combined teams of humans and computers were recently able to beat computers operating in isolation. Even more striking is the 2005 tournament success of human-computer team "ZackS." Without any grandmasters and with arguably weaker technology, the "ZackS" team defeated teams of grandmasters paired with powerful software (ChessBase, 2005). This example illustrates that there may be bounds on the extent to which technology can be substituted for humans. Playing chess is similar to making business decisions in that they both require interpretation of complex sets of information to achieve an optimal outcome. To maximize potential value, firms and managers should understand the extent to which information technology can effectively complement or replace people.

Understanding the extent to which technology can be applied to specific tasks is one example of the management challenges that firms face as they increase their use of data. Analytics and big data have introduced even more opportunity, but also have created a more diverse data landscape that poses new management challenges. Some of these challenges are similar to the challenges related to business intelligence and data warehousing, while others are unique to analytics and big data. With the use of larger volumes of data and more complicated analytics, resulting information becomes increasing complex. Individuals meant to consume this information are less likely to be able to understand such complexity given that limited information is available to them (Simon, 1955). This is not to suggest that this complexity is necessarily negative, but it has implications on the way in which individuals and firms utilize this information.

Firms also have to face challenges that arise from their increased use of metrics and incentives. Using data and analytics inherently introduces new metrics and incentives that require effective management. Measured inputs and associated processes are at risk of being influenced by the emphasis on these incentives and metrics. Managers and firms are responsible for ensuring that the intended business outcomes are not compromised by corroded or corrupted metrics and that the firm truly benefits from new data and analytical practices.

2.2 Research on the Value of Data

While data introduces new management challenges, recent research highlights the value created when firms effectively complement people with information technology, making firms willing to accept the introduction of such challenges. A 2012 study of data-driven decision-making, defined as "decision-making based on data and business analytics," showed that firms implementing this practice are 5% more productive and 6% more profitable (Brynjolfsson, Hitt, & Kim, 2011). A 2013 study analyzed the text within firm 10-K filings from 1996-2010. This study shows that in the years following an increase in data-related keywords, firms see significantly higher market values relative to their industry average (Saunders & Tambe, 2013). It concludes that firms that make investments in data are rewarded with higher market values, as well as eventually higher profitability.

This research also highlights the fact that firms have to do more than invest blindly in data to realize a return on their investment. Data-driven decision-making, by definition, requires more than investment in data. A study of intangible assets in 2002 evaluated the complements between information technology and supporting organizational practices (Brynjolfsson, Hitt, & Yang, 2002). When implemented together, the impact on market value was ten times higher than the impact of information technology only. This study also identified a clustering of firms that demanded a high premium. According to the study, the competitors of these firms were not able to quickly match the performance of this cluster, implying that "it must be difficult to imitate the specific technologies and practices they have implemented. Simply buying a lot of computers and implementing a stock set of practices" is not sufficient (Brynjolfsson, Hitt, & Yang, 2002).

There is additional evidence that no universally applicable approach to data and analytics exists, as a firm's complexity may inhibit its ability to implement information technology. The complexity of firms may also hinder an organization's ability to adopt information technology and may result in lower firm performance (McElheran, forthcoming). The study by Saunders and Tambe also highlighted that investments in data result in higher profitability, but generally three to five years after these investments are made. Saunders and Tambe offer an explanation consistent with this finding that "complementary organizational or managerial changes are required while the firm implements data-related initiatives" (Saunders & Tambe, 2013).

Current research also highlights that the value created is impacted not just by firm organizational practices, but also by the type of data used. In one evaluation of first-degree price discrimination, feasible profits were 12.2% higher when utilizing consumer web browsing behavior, but only 0.8% higher when only utilizing consumer demographics (Shiller, 2014). An earlier study by Martens and Provost (2011) shows that models based on past consumer transactions were more useful in targeting customers than models utilizing socio-demographic data. The most effective models combined both sets of data. After a critical mass of socio-demographic data is utilized, additional data does not substantially improve consumer-targeting efforts. On the other hand, transaction data continues to see marked improvement as more data is added (Martens & Provost, 2011). These two studies highlight the fundamental fact that not all data is equally valuable and adding more data will not necessarily increase value.

In addition to the emphasis on the types of data, the Martens and Provost study provides an example of the relative value of the sources of data sets. It highlights an example in which internal transaction data proves to be much more valuable than externally obtained demographic data. This relative value is even greater when considering the fact that internal transaction data provides a unique advantage to the owner of that data. Competitors cannot easily obtain this transaction data since it is not readily available from external sources. Thus, this information is more valuable and can be a source of competitive differentiation. Externally obtained demographic information offers limited opportunity for firms to differentiate since its external source makes it inherently available to everyone.

Collectively, this research shows that firms can derive value from data, but maximizing this value is influenced by a firm's ability to implement necessary organizational changes, as well as the type and source of underlying data sets.

2.3 Holmstrom and Milgrom

To analyze the attempts to maximize firm value from data and analytics, this paper utilizes Holmstrom and Milgrom's Multitask Principal-Agent Model. This model was created to explore issues related to incentives, asset ownership, and job design (Holmstrom & Milgrom, 1991). It was then used as the basis for Holmstrom and Milgrom's paper "The Firm as an Incentive System," which further explored these incentives (Holmstrom & Milgrom, 1994). I use the Multitask Principal-Agent Model and subsequent findings of Holmstrom and Milgrom since their research provides a solid analytical framework upon which the ways firms derive value from data and analytics can be explored.

Using the same basic model, Holmstrom and Milgrom's 1991 analysis of the Multitask Principal-Agent Model evaluated the impact of quantitative measures. One of their primary arguments was that when agents are balancing a variety of tasks with varying measurability, incentives result in agents favoring measured tasks and neglecting unmeasured ones. To prove this, the model uses an example in which agents can choose between two tasks where only the first task is observable. Since incentives for the second task are impossible, an increase in the incentive for the first task would result in a decrease in effort allocated to the second task. Thus, the outcomes of immeasurable tasks, such as those of the second task in the above example, are potentially compromised when competing tasks are incentivized. Thus, firms may benefit more by not incentivizing any tasks.

2.4 Campbell's Law

While it is less robust than Holmstrom and Milgrom's model, Campbell's Law is a commonly cited piece of research related to the adverse effects of metrics and incentives. In 1976, Donald Campbell published a paper entitled "Assessing the Impact of Planned Social Change" which outlined what would later become known as Campbell's Law. The central thesis is that the "The more any quantitative social indicator (or even some qualitative indicator) is used for social decision-making, the more subject it will be to corruption pressures and the more apt it will be to distort and corrupt the social processes it is intended to monitor" (Campbell, 1976).

Campbell cites a particularly compelling example of the corruption of metrics in United States police departments. These departments were commonly evaluated by their "clearance rates," the rate at which crimes are solved by the department. This pressure was shown to corrupt the "clearance rate" indicator as well as the process of criminal justice. During plea-bargaining, criminals were commonly rewarded with lighter sentences as they admitted to more crimes, which improved a department's "clearance rate." However, the number of crimes that were properly solved may not have been impacted, as it was common for criminals to admit to crimes they did not commit to receive a lighter sentence. Thus, changes in the quantitative social metric reflected a change in police activities instead of an actual change in criminal activities. The use of this quantitative metric compromised the process it was originally intended to measure. This further reinforces the importance of incentives and measurability as the agent, a law enforcement department, makes task choices. Campbell's Law is not discussed explicitly throughout this paper, as it does not provide a sufficient means of analysis relative to Holmstrom and Milgrom's Multitask Principal-Agent Model. Still, this model is consistent with Campbell's Law and therefore the forthcoming analysis indirectly incorporates Campbell's findings.

3 Methodology

Collectively, the history of information technology and existing research on the value of data provide the context upon which implicit costs of data can be further analyzed. The analysis of these costs primarily relies on the Multitask Principal-Agent Model and case studies that discuss the use of data and analytics in firms.

3.1 Multitask Principal-Agent Model

The Model

Holmstrom and Milgrom's Multitask Principal-Agent Model explores the interactions between principals and agents. I selected this model as an appropriate analytical framework due to the model's robust ability to represent the tradeoffs in incentive schemes, the relationship between principals and agents, and the design variables that firms can utilize to achieve a desired outcome. I apply this model to the context of data and analytics where firms are represented as principals that utilize data and analytics with hopes of a larger firm payoff. In their efforts to maximize the benefits of data and analytics, firms have to account for tradeoffs in incentive schemes, handle their relationship with decision-makers (represented as agents), and manage design variables

In this model, the agent chooses to allocate his effort among a series of activities n = 1, . . . , N. These efforts are reflected in the vector $\mathbf{t} = (t_1, . . . , t_N)$. Since the principal is unable to directly observe the agent's choice of **t**, the agent's choice is indirectly observed by a series of measures $\mathbf{X} = (X_1, . . . , X_I)$. Each measure X_i is a function of the agent's task choices **t**, where $F_i(\mathbf{t})$ represents the agent's measured input. Examples of $F_i(\mathbf{t})$ include estimates of inputs, such as time worked, or agent's contributions to performance indicators. To account for measurement errors, the term ε_i is used:

$$X_i = F_i(t) + \varepsilon_i$$

Agents are rewarded with a base salary, denoted as β, and commission rates $\boldsymbol{\alpha} = (\alpha_1, . . . , \alpha_I)$. The resulting incentive scheme $s(\boldsymbol{X})$:

$$s(\boldsymbol{X}) = \sum_i \alpha_i X_i + \beta$$

The agent also incurs a private cost, C(t), dependent on his choice of t. The agent can incur a private benefit, but for simplicity's sake that benefit is recorded as a negative cost. The principal accrues a private benefit B(**t**). It is also possible for the agent's choice of **t** to impose costs on the principal, recorded as negative benefits.

The model accounts for tasks for which the agent enjoys a private return, but for which the principal does not benefit. These tasks are reflected in a series of measures $Z = (Z_1, . . . , Z_K)$. A similar measurement error ε exists and the same variable is used, but with a different index. These measurement errors are recorded as $(\varepsilon_{I+1}, . . . , \varepsilon_{I+K})$ to differentiate them from the error terms used in the previous measurement function. The agent's returns from these activities are reflected below:

$$Z_k = H_k(t) + \varepsilon_{I+k}$$

There are means, such as contracts, by which the principal can prevent the agent from receiving such benefits. This exclusion is reflected by the variable δ_k, where when $\delta_k = 1$, the agent can realize benefit Z_k and when $\delta_k = 0$, the agent cannot realize Z_k. These returns are excluded based on $\boldsymbol{\delta} = (\delta_1, . . . , \delta_K)$ and total return for the agent is the following function:

$$\sum_k \delta_k Z_k$$

This model also includes a measure for the principal's overall cost of monitoring, $K(\boldsymbol{\Sigma})$. This is a function of the complete set of error terms $\boldsymbol{\varepsilon} = (\varepsilon_1, . . . , \varepsilon_{I+K})$. Each error term is a normally distributed measurement of errors with a mean of zero. The variable $\boldsymbol{\Sigma}$ is introduced as the variance-covariance matrix of these random terms.

Reviewing the entire model, Holmstrom and Milgrom highlight the variables that principals and agents have control over. They refer to these variables as organizational design variables:

1. Base Salaries β
2. Commission Rates $\boldsymbol{\alpha} = (\alpha_1, . . . , \alpha_I), \alpha_i \geq 0$
3. Exclusion of Private Returns $\boldsymbol{\delta} = (\delta_1, . . . , \delta_K),\ \delta_k \in \{0,1\}$
4. Monitoring Intensity $\boldsymbol{\Sigma}$

The resulting financial payoffs for principals and agents are outlined below, denoted by P and A, respectively:

$$P = B(\boldsymbol{t}) - \sum_i \alpha_i X_i - \beta - K(\boldsymbol{\Sigma})$$

$$A = \sum_i \alpha_i X_i + \sum_k \delta_k Z_k + \beta - C(\boldsymbol{t})$$

The model assumes that the principal is risk neutral and the agent has a constant risk aversion with coefficient r. A is normally distributed, so the agent's utility measure is reflected in the certainty-equivalent form:

$$ACE(\boldsymbol{t}, \boldsymbol{\alpha}, \boldsymbol{\delta}, \boldsymbol{\Sigma}) = \sum_i \alpha_i F_i(\boldsymbol{t}) + \sum_k \delta_k H_k(\boldsymbol{t}) + \beta - C(\boldsymbol{t}) - \frac{1}{2} r V(\boldsymbol{\alpha}, \boldsymbol{\delta}, \boldsymbol{\Sigma})$$

Here $V(\boldsymbol{\alpha}, \boldsymbol{\delta}, \boldsymbol{\Sigma})$ reflects the variance of income. An efficient set of variables $(\boldsymbol{t}, \boldsymbol{\alpha}, \boldsymbol{\delta}, \boldsymbol{\Sigma})$ has to maximize the total certainty-equivalent, TCE:

$$\max\,[TCE(\boldsymbol{t}, \boldsymbol{\alpha}, \boldsymbol{\delta}, \boldsymbol{\Sigma})] \equiv B(t) + \sum_k \delta_k H_k(\boldsymbol{t}) - C(\boldsymbol{t}) - K(\boldsymbol{\Sigma}) - \frac{1}{2} r V(\boldsymbol{\alpha}, \boldsymbol{\delta}, \boldsymbol{\Sigma})$$

such that the agent maximizes his payoff:

$$\mathbf{t} = \underset{t'}{\operatorname{argmax}} \left\{ \sum_i \alpha_i F_i(\boldsymbol{t}') + \sum_k \delta_k H_k(\boldsymbol{t}') - C(\boldsymbol{t}') \right\}$$

The base salary and risk terms are dropped since they are not affected by the agent's actions. Assuming that F_i and H_k are concave and C is strictly convex, the agent's choice of **t** is in response to $\boldsymbol{\alpha}$ and $\boldsymbol{\delta}$, and can be denoted by $\mathbf{t}(\boldsymbol{\alpha}, \boldsymbol{\delta})$. This simplifies the set of independent variables in the objective TCE function to $\boldsymbol{\alpha}, \boldsymbol{\delta}, \boldsymbol{\Sigma}$:

$$\max_{\alpha,\delta,\Sigma}[T(\boldsymbol{\alpha}, \boldsymbol{\delta}, \boldsymbol{\Sigma})] \equiv TCE(\,\boldsymbol{t}(\boldsymbol{\alpha}, \boldsymbol{\delta}), \boldsymbol{\alpha}, \boldsymbol{\delta}, \boldsymbol{\Sigma})$$

This was the maximization problem at the center of much of Holmstrom and Milgrom's work and will be further analyzed throughout this paper. Holmstrom and Milgrom also included the impact of asset ownership in their model, but this is omitted, as it is not relevant for the ensuing discussion of data and analytics.

Suggested Practices derived from the Model

Holmstrom and Milgrom presented suggestions regarding the use of metrics that enable both principals and agents to realize the maximum payoffs:

1. Optimize with Organizational Design Variable – As previously mentioned, payoffs are optimized over the commission rates $\boldsymbol{\alpha}$, exclusion of private returns $\boldsymbol{\delta}$, and monitoring intensity $\boldsymbol{\Sigma}$.
2. Separate Tasks Based on Measurability – This separation can preserve the use of incentives, but only for agents exclusively charged with highly measurable tasks.
3. Without Separation, Lower Incentives – If an agent has a mix of measurable and non-measurable tasks, it is better not to incentivize the measurable tasks because it comes at the cost of non-measurable tasks.

Holmstrom and Milgrom's analysis also discusses the implications of the relationships that tasks have with one another, specifically the impact of tasks as substitutes and complements in the agent's private cost function. It shows that when two activities are complementary in the agent's private cost function, introducing incentives is beneficial. When two activities are substitutes, on the other hand, incentives lead to a preference for the measured task. When such substitutes exist, rewarding a specific activity or reducing the opportunity costs of other activities can provide incentives for that specific activity. To reduce the opportunity costs of other activities, principals can lower the incentives on these activities. In situations like the two-task example where one task cannot be measured and therefore cannot be rewarded, the model suggests that it is optimal to reduce or remove rewards for the measured tasks. These suggested practices are further discussed in the analysis of implicit costs.

3.2 Case Studies

This analysis also relies on a series of case examples in which firms utilize data and analytics. These examples include firms such as Google, Target, Wal-Mart, Caesars Entertainment, Proctor & Gamble, Continental, and Whirlpool. Given that there is limited research on the implicit costs analyzed in this paper, case examples provide a necessary set of reference materials that collectively sample the means by which firms have addressed the implicit costs attended by data and analytics.

4 Analysis of Implicit Costs

Firms seeking to gain value from data and analytics take on a variety of associated costs that go beyond software and implementation fees. Such fees, referred to as explicit costs, are obvious and unavoidable in any effort to increase the use of information technology. Other costs are less understood and many firms struggle to understand their implications. Firms still incur these implicit costs as they seek to maximize the value from and protect against additional risks of increased use of data and analytics. This analysis explores two of these costs: the mitigation of the negative effects that metrics can have on the total firm payoff and the investment required to ensure people are complemented by data and analytics to maximize firm benefits.

4.1 Implicit Cost 1: Adverse Effects of Metrics

The phrase "You can't manage what you don't measure" is commonly mentioned when discussing data and analytical capabilities (Brynjolfsson & McAfee, 2012). It implies that to manage something, it needs to be measured. However, the research of Holmstrom and Milgrom highlights the negative effects that such measurement can have in the context of other tasks and larger objectives. This adverse effect, combined with the fact that not everything is easily measurable, gives firms reason to be cautious before attempting to manage by measurement. Misaligned incentives can lead to behaviors that reduce overall firm value. Firms increasing their reliance data and analytics inherently increase their use of metrics, which may also increase the incentives associated with these metrics. These incentives include, but are not limited to, compensation, praise, and promotions. The Multitask Principal-Agent Model provides a means by which these incentives can be explored as firms increase their use of data and analytics.

As firms rely more on data and analytical methods, they are automatically increasing their reliance on metrics. The introduction of these metrics may introduce incentives that compromise the inputs that the metric was originally intended to measure. Consistent with Campbell's Law and the work of Holmstrom and Milgrom, these adverse effects have the potential to collectively hinder the firm's ability to achieve the goals that motivated a greater reliance on metrics in the first place. Social examples of this include an overemphasis by police officers on arrest metrics instead of fighting crime, by teachers on test performance instead of education, and by bankers on short-term profit instead of financial stability (Salmon, 2014). The 2012 Obama campaign website was optimized for donations, but as a result visitors found it difficult to discover the candidate's actual positions (Salmon, 2014).

Since the use of metrics can compromise inputs and intended outcomes, measurement design can be utilized to mitigate the aforementioned risks. To consider the impacts of measurement design choices, this analysis continues with examples of internal and external metric corrosion using the Multitask Principal-Agent Model.

4.1.1 Example 1: Internal Metric Design

Within a firm, metrics can have adverse effects that can be mitigated through thoughtful metric design. An example of these adverse effects is the tradeoff between volume and quality of output. Since quality is more subjective, measuring output volume is easier than measuring the quality of output. Using a modified example of Holmstrom and Milgrom's Multitask Principal-Agent Model, suppose that agents can select between the following tasks:

1. t_1 - High volume without compromised quality
2. t_2 – High volume at the cost of quality
3. t_3 – Low volume to ensure quality

The firm strongly prefers that agent's allocate their effort to task t_1, as it leads to the greatest benefit to the firm. Principal benefits are related accordingly $B(t_1) >$

$B(t_2)$ and $B(t_1) > B(t_3)$. In some cases the principal may actually experience a negative benefit from t_2 and t_3 depending on the extent of the compromises to quality or volume. For the agent, the private cost of task t_1 is also higher than those of the other tasks as task t_1 requires providing high output on two dimensions instead of just one. So $C(t_1) > C(t_2)$ and $C(t_1) > C(t_3)$. The principal may, in an effort to increase output, increase the monitoring intensity $\mathbf{\Sigma}$ and incentives related to volume. This volume incentive is reflected in a choice of incentives $\boldsymbol{\alpha}$ where $\alpha_1 = \alpha_2 > 0, \alpha_3 = 0$. The agent's choice of $\mathbf{t}$ is based on the following optimization:

$$\mathbf{t} = \underset{t'}{\operatorname{argmax}} \left\{ \sum_i \alpha_i F_i(\boldsymbol{t}') - C(\boldsymbol{t}') \right\}$$

Since only α_1, α_2 are non-zero and no activities are excluded, this is the equivalent of:

$$\mathbf{t} = \underset{t'}{\operatorname{argmax}} \{\alpha_1 F_1(\boldsymbol{t}') + \alpha_2 F_2(\boldsymbol{t}') - C(\boldsymbol{t}')\}$$

Each commission rate is associated with the measurement of its associated task:

$$\mathbf{t} = \underset{t_1, t_3, t_3}{\operatorname{argmax}} \{\alpha_1 F_1(t_1) + \alpha_2 F_2(t_2) - C(t_1) - C(t_2) - C(t_3)\}$$

The agent's attempt to maximize his payoff results in a full devotion to task t_2. Since task t_3 has no incentive, but carries a cost, it would result in a negative payoff for the agent and therefore is avoided by the agent. More simply expressed:

$$\alpha_2 F_2(t_2) - C(t_2) \quad > \quad \alpha_1 F_1(t_1) - C(t_1) \quad > \quad -C(t_3)$$

This incentive scheme effectively avoids the lower benefit to the principal of task t_3 $B(t_1) > B(t_3)$, but results in a preference for task t_2, which has a lower benefit to the principal than task t_1: $B(t_2) < B(t_1)$. This not only reduces the principal's payoff, it may actually make it negative as only $B(t_1)$ is guaranteed to be positive. To address this, if quality can be effectively monitored, t_2 can be excluded using quality audits. Such audits increase the overall monitoring costs $K(\boldsymbol{\Sigma})$. This can be formulated by introducing a $\delta_i \in \{0,1\}$ similar to that used in the exclusion of external returns. When introducing the quality control on task t_2, the values are $\delta_1, \delta_3 = 1; \delta_2 = 0$

$$\mathbf{t} = \underset{t'}{\operatorname{argmax}} \left\{ \sum_i \alpha_i \delta_i F_i(\boldsymbol{t'}) - C(\boldsymbol{t'}) \right\}$$

which removes the incentive for task t_2:

$$\mathbf{t} = \underset{t_1, t_2, t_3}{\operatorname{argmax}} \{ \alpha_1 F_1(t_1) - C(t_1) - C(t_2) - C(t_3) \}$$

This properly incentivizes each task for the agent:

$$\alpha_1 F_1(t_1) - C(t_1) \quad > \quad \alpha_2 F_2(t_2) - C(t_2) \quad > \quad -C(t_3)$$

If quality is difficult to measure and quality checks are imperfect, in an effort to maximize his payout, the agent will likely do the bare minimum to pass this quality check even if it accrues a negative benefit to the principal. This again highlights the importance of metric design.

An example of the previous discussion is the use of metrics in call centers that want to optimize their throughput and began rewarding agents for the number of calls completed. In this scenario, the model predicts that this reward will unfortunately result in agents picking up the phone to minimally interact with customers – in the extreme, to do so only to quickly hang it up (McKeon, 2012). This practice would bolster the agent's numbers, but compromise the call center's core purpose of providing customer support and subsequently hinder the ability for the firm to realize its larger goals.

This example highlights the importance of metric design. The intent of the incentive was to increase customer satisfaction, which in this case is based on the combination of wait time (volume of output) and the resolution of their call (quality of output). These two factors are at odds with one another, as reducing wait times may reduce the number of calls that are successfully resolved. If call center staff members are focused on increasing the volume of calls answered, they may not take the time necessary to successfully resolve each call they answer. Therefore, incentives to improve customer satisfaction by decreasing wait times does not account for subsequent compromises to quality. The model predicts that call center staff over-optimize for task t_2 even though it has a negative benefit to the principal.

In an attempt to maintain well-aligned incentives, the principal could monitor the agent using quality checks. The principal may be tempted to ensure call center staff remain on the line for a certain number of seconds or minutes, but again this would be a poor incentive structure as the model's optimization suggests that in the same instant that the minimum time requirement expires, the agent will hang up the phone regardless of whether or not the customer's issue has been resolved. Other options such as customer satisfaction surveys are more direct ways to ensure quality. Firms can incentivize agents to properly emphasize quality responses by compensating or promoting the agents with high customer satisfaction survey results. Using this approach, the firm is more likely to ensure that their incentive structure leads to a positive principal benefit $B(\mathbf{t})$. For the best outcome, such quality incentives need to also be monitored to ensure that they do not reduce the principal's benefit if the agent allocates more effort to task t_3 than task t_1 since $B(t_1) > B(t_3)$.

4.1.2 Example 2: External Metric Design

Firms that rely on external factors and measures as inputs, in particular capturing external data in their efforts, risk future corrosion of inputs. If external parties recognize the firm's choice of measurement and are in some way impacted by the firm's use of such measures, these parties will attempt to influence these inputs in their favor. This can be further examined using the model in the case in which firms are principals and external parties are agents.

Consider the case of the Multitask Principal-Agent Model in which the principal wants to encourage a particular task t_1, but the measurement for a less desirable task t_2 cannot be distinguished from that for task t_1. Thus, incentives for these tasks are the same. The principal receives benefit $B(t_1)$ from task t_1, but receives a much smaller or potentially negative benefit $B(t_2)$ from task t_2.

$$B(t_1) > 0$$

$$B(t_1) > B(t_2) \quad or \quad B(t_2) < 0$$

To encourage agents to increase their emphasis on task t_1, the principal introduces an incentive α_1 on measure F_1. Given the nature of the two tasks, F_1 is unable to distinguish between agent tasks t_1 and t_2. If **t** is limited to tasks t_1 and t_2, the agent's optimization is as follows:

$$\mathbf{t} = \underset{t_1,t_3}{\operatorname{argmax}}\{\alpha_1 F_1(t_1) + \alpha_1 F_1(t_2) - C(t_1) - C(t_2)\}$$

$$where \;\; \alpha_1 F_1(t_1) = \alpha_1 F_1(t_2) \quad when \;\; t_1 = t_2$$

Since the incentives for t_1 and t_2 are equal, the optimization is dependent on the relative costs of the two tasks. The task with the lower cost is the one that the agent will favor to maximize his payoff. If $C(t_1) > C(t_2)$, then the agent will prefer task t_2 and the principal will be rewarding a task that he never meant to incentivize, commonly at the expense of the task he initially hoped to encourage, task t_1. This scenario is a reasonable possibility given that task t_1 has a much higher benefit to the principal.

Similar to the approach in the discussion of internal metric design, the principal can disallow task t_2 via a contract or term of use, which introduces exclusion variable δ_i in this formulation of the model.

$$\mathbf{t} = \underset{t_1,t_3}{\operatorname{argmax}}\{\alpha_1 \delta_1 F_1(t_1) + \alpha_1 \delta_2 F_1(t_2) - C(t_1) - C(t_2)\}$$

If the restriction on task t_2 requires enforcement, the firm may need to invest in the means to monitor or audit agents who are wrongly taking advantage of task t_2. If this investment allows the firm to be completely able to distinguish between tasks t_1 and t_2, measurement F_1 would no longer depend on task t_2. This would eliminate the incentive on task t_2, as well as the need for the firm to explicitly exclude task t_2.

$$\mathbf{t} = \underset{t_1,t_3}{\operatorname{argmax}}\{\alpha_1 F_1(t_1) - C(t_1) - C(t_2)\}$$

Alternatively, in the case that the firm is aware of its inability to distinguish between tasks 1 and 2, it may only disclose the incentive on task t_1 in hopes that agents don't discover the incentive on task t_2. This effectively obscures $F_1(\mathbf{t})$ for agents, clouding their ability to optimize their payoff by emphasizing task t_2. As agents continue to maximize their payoff, it is unlikely that the incentive on task t_2 will go unnoticed. This is consistent with Campbell's law.

A good example of this scenario is the incentives that stem from Google's search engine, which ranks pages based on a wide variety of factors. External parties had a vested interest in rising to the top of search results and sought to influence the metrics by which Google measured site popularity and relevance. This is commonly referred to as Search Engine Optimization, but some attempts to optimize pages compromised that which Google intended to measure. It started with a simple algorithm weighing inbound links, and external parties responded by creating fake websites with inbound links that inaccurately raised their profile (Gyöngyi & Garcia-

Molina, 2005). This dynamic can be formulated in terms of the model with the following agent tasks:

1. Task t_1 – Good practice, obtaining legitimate inbound links
2. Task t_2 – Bad practice, falsifying inbound links

Where Google benefited from agents with legitimate inbound links

$$B(t_1) > 0$$

while falsified links limited the value Google was able to provide, compromising its credibility since its algorithm could be easily gamed.

$$B(t_2) < 0$$

For the agent, creating fake websites with inbound links requires minimal effort, while working to receive credible, organic links is significantly more costly.

$$C(t_1) \gg C(t_2)$$

Google has worked to prevent the influences of task t_2 by adding more complex factors. This includes giving higher weights to inbound links from more reputable sites. The complete set of factors and bearing that these factors have on Google search results is not fully disclosed to prevent agents from participating in similar bad practices that inaccurately manipulate the rankings to the agents' benefit (Hansell, 2007). Google is attempting to introduce causal ambiguity, making it difficult for agents to understand the relation between inputs and outputs since such an understanding may lead to an agent's preference for task t_2 (Lippman & Rumelt, 1982).

As Google has evolved, its algorithm has become increasingly complex and secretive, in attempts to discourage such compromising practices. Agents may be delayed in gaining a greater understanding of the incentives and measurement mechanisms, but they are still able to resolve these factors over time. To further prevent such bad practice, Google is constantly iterating on its algorithm while keeping it a secret, making it a moving target for agents looking to exploit Google's algorithmic approach.

Google has also been able to make the necessary investments to distinguish between good and bad practices, reducing the incentives for those participating in bad practices. To do this still required Google to increase its monitoring intensity $\boldsymbol{\Sigma}$, which if all else remained constant would decrease Google's overall payoff. Agent participation in the bad practice posed enough of a threat to Google's payoff that Google made an additional investment in monitoring intensity. In this case, agents participating in the bad practice were identified as those who duplicated content across multiple sites in an attempt to rapidly increase inbound links (Singhal & Cutts, 2011). These agents were still rewarded for participating in the bad practice of task t_2, but Google was able to lower the associated incentive, encouraging agents to place more emphasis on participating in the good practice of task t_1. If left unchecked, this bad practice may have compromised the benefits Google received from external parties. Google's Search Engine Optimization is one example, but other firms may require investments in monitoring intensity as they increase their reliance on data and analytics.

This example also illustrates another means by which the organizational design can be influenced by principals omitted by Holmstrom and Milgrom, the principal's ability to obfuscate the exact incentives and measurements by limiting information passed to agents. The research of Holmstrom and Milgrom explicitly discusses four organizational design variables: salary β, commission rates $\boldsymbol{\alpha}$, exclusion of private returns $\boldsymbol{\delta}$, and monitoring intensity $\boldsymbol{\Sigma}$. Limiting the design variables to these four misses another factor under the principal's control, the ability to influence the agent's understanding of incentives $\boldsymbol{\alpha}$ and measurement capabilities F_i. Holmstrom and Milgrom do not explicitly mention that principals can obscure the incentives and measurement mechanisms to encourage agent behavior that benefits the firm.

4.1.3 Mitigate Adverse Effects of Metrics

The increased use of metrics can compromise non-measured or non-measurable tasks and outcomes as outlined in the previous examples. While existing research outlines the potentially adverse impact that the use of data and analytics presents, the associated solutions proposed by Holmstrom and Milgrom are useful, but not sufficient in addressing the challenges posed by data and analytics. As such, firms require a more robust set of practices to mitigate the adverse effects of metrics.

Adverse Effects of Metrics Practice 1: Protect Unmeasured Tasks and Outcomes

Holmstrom and Milgrom propose that it is better to not introduce incentivized metrics when less measurable outcomes are at risk. For firms increasing their use of data and analytics, they should consider the potential impact that new metrics and incentives may have on tasks and outcomes. A negative impact on unmeasured tasks and outcomes may be reason to avoid implementing new metrics and incentives. However, firms may find that the risks are outweighed by the perceived value they will gain from data and analytics. The value of new data and analytical practices may be worth the potential risk of compromising certain tasks and outcomes. As such, firms may find it impractical to avoid data and analytics based on the tradeoffs reflected in Holmstrom and Milgrom's research.

Adverse Effects of Metrics Practice 2: Isolate Unmeasurable Tasks

Holmstrom and Milgrom suggested that another way to protect unmeasured tasks and outcomes is to separate highly measurable tasks from unmeasurable ones (Holmstrom & Milgrom, 1991). Their model showed that agents working on highly measurable tasks can be incentivized, while those who are working on less measurable tasks should be paid fixed salaries. Applying such a practice to the 2012 Obama campaign example would mean separating the responsibility of ensuring the candidate's position was easily discovered on the site. In this proposed scenario, at least one employee would be responsible for the maximizing donations through the website, while others would be responsible for making the candidate's positions easily discoverable. These responsibilities likely belonged to the same individuals, which led to less attention being paid to the unmeasurable outcome. The proposed solution of separating measured and unmeasured tasks assumes that these tasks can be split among agents, which may not be possible.

Adverse Effects of Metrics Practice 3: Optimize Job and Measurement Design

In the case that the first two practices are not possible, Holmstrom and Milgrom also suggest that firms evaluate the organizational design variables by which firms can better manage metrics and incentives. In both the call center and Google examples, the design of metrics resulted in an eventual corrosion of the outcome the metric intended to improve and required a redesign of measurement mechanisms. These are guiding examples that enable principals to better design and manage incentive and measurement systems. Such examples are particularly useful when unmeasured tasks cannot be separated or disincentivized. In the case of the call center, the metric was intended to improve customer satisfaction by decreasing commonly complained about wait time. The quick hang-ups decreased overall customer satisfaction. The measurements could be redesigned to ensure that overall customer satisfaction improved or to control for rapid hang-ups.

The suggestions of Holmstrom and Milgrom's research did not include another means by which organizations can manage metrics and incentives. The Google example highlighted the fact that firms can obfuscate their incentives and chosen means of measurement to achieve more optimal outcomes. Google has constantly evolved its measurement of inputs to discourage corruption of the same inputs. This example proves a firm may also benefit from making the design of their incentives and their measurements more complex and secretive, as well as from keeping this design in a constant state of evolution.

Adverse Effects of Metrics Practice 4: If Necessary, Accept Risks

Firms should account for the risk of increased reliance on data and analytics, but firms may willingly accept these risks if the value of technology is significantly compelling. To avoid data and analytics completely would be to ignore other research and case examples that highlight obvious benefits for firms using data and analytics. Holmstrom and Milgrom's research highlighted important risks associated with incentivizing metrics, but these must be weighed against the outcome of such metrics. In many cases, the risks are outweighed by the value that firms gain from increasing their use of data and analytics. Firms can utilize the outlined practices to mitigate these risks, but sometimes they may have to accept the risks given the strong potential for data and analytics to create value.

In summary, firms can utilize the following practices to avoid the adverse effects of metrics:

1. Protect Unmeasured Tasks and Outcomes – avoid or lower incentives on measurable tasks
2. Isolate Unmeasurable Tasks – only incentivize agents exclusively charged with highly measurable tasks
3. Optimize Job and Measurement Design

- Holmstrom and Milgrom Organizational Design Variables: commission rates $\boldsymbol{\alpha}$, exclusion of private returns $\boldsymbol{\delta}$, and monitoring intensity $\boldsymbol{\Sigma}$.
- Internal Metrics: balance quality and quantity
- External Metrics: obfuscate incentives and metrics by keeping them complex, secret, and in a constant state of evolution.

4. If Necessary, Accept Risks – given the potential value, the risks may be worthwhile

Many of these practices come directly from Holmstrom and Milgrom's work. These practices are prioritized such that if a firm, in any given setting, is unable to apply a particular practice, that firm can attempt the subsequent practice.

4.2 Implicit Cost 2: Complement People with Technology

There are a variety of factors that impact a firm's decision-makers' ability to properly consume data and analytics so that the firm realizes the maximum value from such practices. These factors make up the cost to effectively complement people with technology. At the core of this cost is the ability for a firm to properly supplement human decision-making with technology. To do so requires an understanding of the extent to which information technology can be applied for particular tasks and decisions. While some tasks are exclusively reliant either on humans or on computers, the majority of tasks require a balance between the two. For the tasks that require a balance, the firm has to ensure that its people are properly interacting with implemented information technology. With the mass of data presented by growing data sets and the opportunities of big data, firms are also faced with the challenge of how to effectively absorb complex and voluminous data.

To analyze the cost of complementing people with technology, this section is broken into the following topics:

1. Context: Task Automation and Strategic Decisions – The extent to which information technology can be applied as a complement to or a substitute for people.
2. Balance
   a. Shift from Intuition – Decision-makers moving away from a sole reliance on personal intuition and incorporating more data into their decision-making process.
   b. Prevent Overreliance – Decision-makers incorporating their intuition and experience instead of acting exclusively based on data and analytics.
3. Data Mass – The complexity and volume that accompanies increased data and analytics, in particular big data.

These topics will be discussed in terms of the high-level variables and incentives of the Multitask Principal-Agent Model. Throughout each of these, the firm is the principal, seeking to maximize its payoff P subject to the previously defined constraints. The agents are the decision-makers and include the firm's employees and contractors. These agents are intended to consume new data and analytical practices for the greater benefit of the firm $B(\mathbf{t})$. Agents will still optimize their payout based on the provided incentives $\boldsymbol{\alpha}$, exclusion of private returns $\boldsymbol{\delta}$, and private costs C(**t**). This model is used as the context in which the topics related to the cost of complementing people with technology can be discussed.

4.2.1 Context: Automation and Strategic Decisions

When discussing the use of data in organizations, a key consideration is the extent of technology's capabilities to complement or substitute for humans. Prior to information technology, people executed actions and made decisions based on limited data and analysis. Technological advancements created increased potential for people to utilize supporting data and analytics. In some cases, information technology has been able to completely automate tasks and decisions previously made by humans. While it may be tempting to increase the reliance on information technology, the degree to which people are substituted, or complemented, by this technology for any given process is highly context specific. In some cases, technology is a substitute for humans for a given task t_i, so that the firm can maintain the same benefit $B(t_i)$ without having to pay an agent. More commonly, information technology is complementary to human agents, allowing them to increase the overall benefit to the firm, but requiring the continued involvement of a human agent. There are also limitations to information technology's ability to aid people in every possible decision. This section discusses the bounds of each of these scenarios:

Context Scenario 1: Automation

In some instances, tasks have been fully automated and, with the exception of the setting up these processes, require no additional interaction with people. This level of automation is dependent on the level of repetition and degree of complexity of the task. For highly repetitive tasks with low levels of complexity, technology is a well-suited substitute since humans were previously acting robotically. This implies that the benefit $B(t_i)$ is the same regardless if the task is automated or executed by humans. So if the business process can be simply codified, the benefit when the task is automated is theoretically equivalent to the benefit provided by a human executing the same task. Firms may prefer to automate that task, substituting technology for humans, under the condition that the cost of the necessary technology does not exceed the cost of the human agent.

Technological innovation may increase the technology's ability to handle greater complexity and raise the potential benefit that technology offers relative to humans. This increase will create new opportunities and additional scenarios for automation as the firm benefit afforded by information technology $B_{Tech}(t_i)$ matches the benefit provided by humans $B_{Human}(t_i)$ for any task. Assuming the cost to the firm is lower for technology than humans, then

$$B_{\text{Tech}}(t_i) > B_{\text{Human}}(t_i) \quad \Rightarrow \quad \text{automate } t_i$$

Advancements in data science have enabled new tasks to be automated that were previously unable to be automated. This automation was made possible by information technology's increased ability to source a variety of inputs quickly and reference large volumes of supporting data sets in machine learning. An example of this is the development of the self-driving car, which automates a highly repetitive process that requires the simultaneous processing of a vast array of inputs previously impossible. Other examples include language recognition and translation.

The promise of technology and recent advancements may lead to over-automation by firms that perceive technology as a substitute for humans for tasks for which technology is more aptly a complement. With an increase in automation, there is a loss of control and oversight that was previously provided by human agents. While additional checks and controls can be embedded into the technology, they currently cannot match the human ability to deal with complexity. Thus, highly repetitive tasks, low in complexity are likely the only good candidates for automation.

Context Scenario 2: Strategic Decisions & Predictive Capabilities

While some tasks can be fully automated, there are others for which current information technology provides minimal benefit for firms. While data-driven insights are useful, this technology does little to complement humans in the decision-making process as companies are attempting to predict the future or make high-level strategic decisions. In fact, an overreliance on data caused a strategic mistake by one of the biggest proponents of big data and analytics, Caesars Entertainment's Gary Loveman. The numbers on the decision to obtain a gaming license in Macau suggested that it was a bad investment and Caesars Entertainment backed out, but the Chinese city ended up being a major success for competitors that did obtain licenses (Greenfeld, 2010). For many of these competitors the Macau-based operations are bringing in more revenue than their casinos in Las Vegas.

Upon reflecting on his decision, Gary admitted it was a bad one. "Big mistake. I was wrong, I was really wrong." He went further to say that "You had to have a kind of intuitive courage and I am not well suited to those kinds of decisions".

Big data predictions extrapolate upon existing data and are unable to account for the expected unknowns of the future. This is a consistent explanation for the fact that in the set of case studies sampled for this paper, few companies touted the use of data and analytics in defining their corporate strategy.

While its use in making such decisions is limited, there is still much discussion on the relation that this data has with the firm's corporate strategy. Data and analytics are still commonly a component of a firm's strategy – many firms have explicitly expressed that developing analytical capabilities is a high-level strategic initiative. In some cases, data is highlighted as a key strategic asset. The data or supporting models were also referenced as a means of supporting the organizations' understanding of their progress against strategic goals. In the case of Whirlpool, one of the highlighted uses of the data warehouse was to provide "support for the accomplishment of strategic business objectives" (Haley, Watson, & Goodhue, 1998).

Context Scenario 3: People and Information Technology as Complements

Given that most tasks are complex and decisions require interpreting a wide range of inputs, many tasks are not candidates for complete automation, but still benefit from information technology. For these tasks, people and information technology are complements, working together to achieve a greater benefit for the firm. This is the broadest scenario and invites further discussion, as it requires striking a balance between people, data, and analytics. A selection of case examples discuss the successful use of information to achieve a variety of outcomes:

1. Sales & Marketing - Capturing transactional history, website browsing history, and social demographics has provided companies with a wealth of information that they can use to understand consumer behavior to provide recommendations at Amazon (Brynjolfsson & McAfee, 2012) and tailor promotional materials at Target (Duhigg, 2012).
2. Operational Efficiency - Firms also seek operational advantages in the ways they manage their operations. Examples include improvements to pricing at Continental (Watson, Wixom, Hoffer, Anderson-Lehman, & Reynolds, 2006), manufacturing at Whirlpool (Haley, Watson, & Goodhue, 1998), and inventory at Wal-Mart (Hays, 2004).
3. Product Testing - In firms that consistently develop new products or refine existing ones, the process of product testing has become operationalized. Firms such as Amazon, Capital One, and Seven-Eleven Japan have found success in applying data and analytics throughout this process (Davenport, 2006; Nagayama & Weill, 2004).
4. Risk & Fraud Detection - Data and analytics are being deployed to organizations to aid in their efforts to prevent risk and fraud. Financial institutions and credit card companies are able to utilize data and analytics to detect identity fraud. Analytics in this instance, allows greater identification of such exceptions and can quickly prompt intervention, but is still aided by human interaction.

5. Strategy - In the series of analyzed case studies, there is some evidence of firms using analytics in their strategy. Continental Airlines uses data and analytics to support "strategic queries" (Watson, Wixom, Hoffer, Anderson-Lehman, & Reynolds, 2006). The CEO of Seven-Eleven Japan referred to information technology as "a tool to achieve business strategy" (Nagayama & Weill, 2004).

Among the referenced case studies, the two most common uses for data and analytics were sales & marketing and operational efficiency (Gartner, 2013). For firms, it can be complex to maximize the complementary benefits of combining humans and information technology. The next three topics in this section further explore the cost of complementing people with technology.

4.2.2 Balance – Part A: Shift from Intuition

Consuming data and analytics in decision-making requires striking a balance between intuition and data. When first introduced to data and analytics, intended consumers within the firm have to learn to incorporate this new input into their decision-making processes. Information consumers are expected to shift from an exclusive reliance on intuition and experience. This shift can be accomplished in a variety of ways.

Shift from Intuition Practice 1: Break Habit of Relying on Intuition

This can be equated to the Multitask Principal-Agent Model, where the firm is the principal and the agent is the information consumer. The information consumer is commonly an internal employee faced with a set of tasks. With the introduction of data and analytics, information consumers are expected to incorporate new data and analytics into their decision-making process for the greater benefit of the firm. This likely requires a new way of thinking and such change is not necessarily accepted willingly by all parts of the organization. This agent's choice of means by which they reach a decision can be modeled as a distinct set of tasks. The agent can make the decision the same way he did prior to the introduction of new data and analytics, denoted as task t_1. Alternatively, he could incorporate this new information into his decision-making as the firm intended, denoted at task t_2. Given that the information consumer is familiar with their prior way of doing things, it can be expected that the private cost of task t_1 is, initially, lower than that of task t_2.

$$B(t_1) < B(t_2), \qquad C(t_1) < C(t_2)$$

Similar to prior discussion of these costs, if all other variables are unchanged, the information consumer will prefer task t_1. Firms intend to encourage task t_2 and need to change the habits of these information consumers by compensating for the information consumer's private cost. Assuming the private cost of task t_2 is higher than that of task t_1, additional controls and incentive redesign may be required. These actions were discussed in previous sections in terms of the model. Examples of such actions include the enforced exclusion of task t_1 or reduced incentives for task t_1. Means by which firms can practically implement such actions include adding additional oversight, audits, or controls to ensure that such information is being included as an input into the decision-making process. In the examined case studies, organizations took on additional measures to ensure that information consumers properly utilized the information.

The information consumer may misperceive the relative costs of these tasks given his comfort with task t_1. If the cost of task t_2 is less than task t_1, the firm theoretically does not need to exclude or disincentivize task t_1, as information consumers will prefer that task given its lower cost. The consumer may require education to recognize the reduced private cost. This private cost $C(t_2)$ may also be initially higher but may reduce over time as the information consumer becomes more proficient in task t_2. In this case, the consumer will still initially prefer task t_1. If the firm wants immediate adoption, this requires an exclusion of or disincentive on task t_1. Alternatively the firm can provide training and support so that information consumers can realize a lower cost for task t_2. The firm still must account for the relative costs of implementing the necessary controls, training, and support in its pursuit of a higher payoff.

Training and support efforts are commonly more expensive for firms than implementing quick controls, but have proven effective in cases where firms intend to maintain the level of autonomy of these decision-makers. To ensure effective field operations, Seven-Eleven Japan requires all franchisees and their spouses go through a two-week centralized training before receiving on-the-job training. In addition they have the support of an Operation Field Counselor, who visits the store at least twice a week for at least two hours (Nagayama & Weill, 2004). These mechanisms are useful as all franchisees of Seven-Eleven Japan are expected to do analysis on their store's sales data to make better-informed product ordering decisions.

Shift from Intuition Practice 2: Present Consumable Information

Another way to reduce the cost of data-driven decision-making is to ensure that information consumers are presented with useful and relevant information. For information to be utilized, it must be consumable by decision-makers meaning that the information is both understandable and actionable. For instance, Target's pregnancy prediction algorithm evaluates the likelihood that customers are pregnant based on their purchasing history (Duhigg, 2012). For customers who are likely pregnant, they can also estimate the approximate due date. These two pieces of information are consumable and are easily utilized by sales and marketing functions to send out specialized promotional material. The interfaces and design through which the information consumers interact also impacts their ability to utilize the information. Seven-Eleven Japan uses multimedia information, making it easier to quickly identify goods and encouraging use by all store employees (Nagayama & Weill, 2004). Proctor & Gamble CIO Filippo Passerini also pushes to provide all information consumers with quality interfaces that include easily understood charts (Murphy, 2010).

Shift from Intuition Practice 3: Involve and Iterate with Users

Involving and iterating with information consumers is another means by which firms can increase the use of data and analytics by information consumers. A variety of the examined cases support this. Caesars Entertainment CEO Gary

Loveman encourages constant feedback among ground operators (Welch & Westerman, 2013). At Proctor & Gamble, Passerini doesn't wait until data is perfect before putting the information in the hands of users. In Passerini's words, Proctor & Gamble "intentionally put the cart before the horse, because it is a way to force change." He believes that as a result of this practice, information consumers are able to see what is possible and "use it as a catalyst to drive the right data convergence" (Murphy, 2010).

At Continental, field staff struggled to understand the need for business intelligence applications and in response, the data warehousing staff presented them an initial visual representation of the data. This convinced them of the value of, as well as the potential for, such tools. As a result, the field staff began submitting their own suggestions on how to better utilize data in managing Continental's hub operations (Watson, Wixom, Hoffer, Anderson-Lehman, & Reynolds, 2006).

Shift from Intuition Practice 4: Open the "Black Box"

Another practice to drive engagement and understanding among information consumers is to maximize their understanding of the analytical models and algorithms that are behind presented information, as well as underlying data source. This practice of opening the "black box" reduces the cost of utilizing data and analytics, as well as providing information consumers with additional insight into the way in which using these new models and algorithms impact pre-existing performance measures. At Whirlpool, the company purposefully created mechanisms that supported the exposure of metadata, such as data definitions and source systems, to information consumers (Haley, Watson, & Goodhue, 1998). At Seven-Eleven Japan, CEO Toshifumi Suzuki, highlighted the importance of being able to have an understanding. He stated, "We shouldn't use the technology unless we can understand what the information means on paper" (Nagayama & Weill, 2004).

Shift from Intuition: Nature of Intuition Practices

While the previously discussed practices ensure that information is being properly consumed by decision-makers, a subset of these practices are necessary conditions for firms to utilize data and analytics. Whether it is an updated process or intensive training, at least some mechanism is required to introduce data and analytics as a worthwhile complement to human intuition. For the information to be effectively incorporated into the decision-making process, it needs to be presented in a readily consumable format. While these are necessary conditions, firms can still choose the extent to which they invest in these practices. Increased investment should logically increase the use of data by decision-makers, as illustrated by the Seven-Eleven Japan case example.

Some of the discussed practices are conceptually complementary and reinforce one another. By involving and iterating with users, firms will be able to more readily understand how consumable their information is and make adjustments if necessary. Through the process of involvement, the value of data and analytics can be more readily experienced by information consumers, which may naturally decrease the reliance on intuition. While breaking the practice of relying on intuition and presenting usable information are considered necessary conditions, firms can still allocate more resources beyond the minimum required investment to realize additional complementary benefits. For example, ensuring information is presented in a high quality and usable format will aid in the training process, making it easier to break the intuition habits of information consumers. The complementary nature of these practices is encouraging for firms that have implemented some of these practices but wish to enhance their efforts.

Shift from Intuition: Risks of Intuition Practices

Some of these practices have downsides that managers may take into consideration before pursing such a practice. There are potential risks associated with the involvement of end users. Many of these users may try to steer the project towards the old way of doing things or in a direction that does not align with the initiative's intent. The variety and quantity of suggestions may also be too much for the organization to handle. In Continental's case, they highlighted the challenge of finding "the time to support the ideas that users have" (Watson, Wixom, Hoffer, Anderson-Lehman, & Reynolds, 2006). It is also not enough to simply involve users in the discussion as users logically expect their input to be incorporated. If they perceive that such input is not being effectively incorporated, new data and analytical practices will be perceived as misaligned with their job, raising the private cost of utilizing new data and analytics.

Sometimes the practice of extensive user involvement is unnecessary. In cases where there are relatively few final decision-makers that need to absorb the information, little engagement is required across the organization for these decisions to have the intended impact. For example, when Wal-Mart discovered that they sold a surprisingly high amount of beer and strawberry pop tarts when natural disasters were expected, the analysis and decision-making process to stock more of these items involved few people (Hays, 2004). The cost to complement people with technology is much greater when a large number of decision-makers throughout an organization are expected to depend on data and analytics.

The practice of opening the "black box" can compromise the potential power and precision of underlying algorithms or models. It is preferred under a few conditions, namely when there is greater human involvement in the decision or the model is expected to evolve. With greater human involvement, firms may decide to forgo complexity in hopes that it will drive adoption among decision-makers. A complex algorithm, no matter how precise, is worthless if decision-makers disregard it.

4.2.3 Balance – Part B: Prevent Overreliance

Prevent Overreliance Practice 1: Avoid Blind Acceptance

While firms have to increase their information consumer's reliance on data and analytics, they must also be cautious about an over emphasis on these practices. Information consumers should not use these tools as substitutes for their intuition and experience. They should not blindly accept this information, and must clearly understand that correlation and causation are different and sampling issues can impact information. When increasing the reliance on data and analytics, firms are taking on the risk of information consumers making missteps in these practices.

This can be outlined in terms of the Multitask Principal-Agent Model where the agents are again information consumers who have choices between a series of tasks. Relevant here is the choice between incorporating data in their decision-making, denoted as task t_2, and blindly accepting data information, denoted as task t_3. Task t_3 has less benefit to the firm than task t_2, since acting on data alone removes the information consumer's valuable experience from the decision-making process. If task t_3 were more beneficial to the firm, it would be a candidate for automation. This is not the case since people are better at dealing with broader complexity than current algorithms and models. The CEO of Seven-Eleven Japan mentioned this explicitly when discussing their main data and analytics tool, the POS system. "Don't rely on the POS system. Information technology is a merely a tool to achieve business strategy" (Nagayama & Weill, 2004).

Since task t_2 requires more thought and effort on the part of the information consumer, the private cost $C(t_2)$ is higher than $C(t_3)$. Firms should be concerned about this as information consumers optimize their payoff by favoring task t_3 if both tasks are permitted and equally incentivized. Again the firm can exclude or disincentivize task t_3. In some instances, the presented data and analytics may not be sufficient for the information consumer to make a decision. This insufficient level of information eliminates task t_3 as an option and forces information consumers to provide some level of judgment.

Prevent Overreliance Practice 2: Beware of Limitations of Data and Analytics

Information consumers should also be aware of the limitations of data and analytics. Depending on the nature of information provided to information consumers, they may be tempted to essentially automate the decision-making processes. Given the current limitations of information technology, that is only advised in the highly repetitive scenarios. Otherwise, strictly relying on data opens up an organization to risks of over-automation. As was previously discussed, data has little use in larger corporate strategy decisions due to its limited predictive capabilities, mainly with its inability to account for unknowns. Information consumers need to recognize these limitations of data to ensure they are properly applying it.

Prevent Overreliance Practice 3: Understand Extent of Data Insights

Information consumers also need to understand the extent of insights provided by data. In particular, correlations can be quickly identified when analyzing data, but information consumers can wrongfully assume that these correlated events have a causal link. Information consumers also have to account for inputs that can adversely influence data inputs such as sampling error and sampling bias.

4.2.4 Data Mass

Data Mass Practice 1: More Data is not Necessarily Better

Part of the promise of big data is the increase in data volume. While this increase helped solve human translation and build self-driving cars, the "more is better" mantra is not necessarily proven to always create value for the firm, yet it undoubtedly adds additional costs. The study by Martens and Provost disproves the premise of this mantra in regards to data. For transaction data, more data significantly improved the studied model, but additional demographic data beyond a certain critical mass of data did not improve the model significantly (Martens & Provost, 2011). Given that obtaining additional demographic data commonly requires firms to investment in collection or procurement efforts, continued investment will eventually result in a negative return on investment. With more data, there is greater complexity for firms to manage, especially if additional data is from new sources or in varied formats.

Data Mass Practice 2: Not All Data has the Same Value

Another important consideration when guarding against excessive data is the relative value of that data. This value commonly varies by the type, quality, and source of data as evidenced by prior research. The previously mentioned study on first-degree price discrimination showed that consumer web browsing data was significantly higher than consumer demographic data (Shiller, 2014). The study by Martens and Provost showed that internal transaction data is more valuable than demographic information (Martens & Provost, 2011).

Data Mass Practice 3: Beware of Data Quality and Source

The variety of data can make it difficult to ensure quality, especially if data is sourced from external vendors. Without proper quality assurance mechanisms, using external data may have adverse impacts. Recently OfficeMax was in the news for the unfortunate results of procuring data from an external source. In January of 2014, Mike Seay received a letter address to him from OfficeMax that include the phrase "Daughter Killed In Car Crash." in the address (Schectman, 2014). The statement was true, as Mr. and Mrs. Seay lost their daughter Ashley in February of 2013, but the fact that it was included in the mailing was painful for the Seays. This led to negative press and a lot of questions for OfficeMax, who had obtained the mailing list from an external vendor and were not able to fully explain how this information was entered. Friends had sent digital photo frames of Ashley and when evaluating the source of data, it was likely that retailer who recorded the piece of information. This invasion of personal privacy caused the couple pain and put OfficeMax in a position where they had to answer for data that they had not intended to procure.

Firms hoping to protect themselves from excessive data should consider the source and uniqueness of each new data set. The Martens and Provost example also highlights the importance of internal data sets and their potential as unique competitive advantages (Martens & Provost, 2011). This internal data also avoids the aforementioned risks that accompany the use of external data.

Data Mass Practice 4: Data Procured should not Exceed Firm Consumption Capacity

In addition to volume and quality issues related to data, firms need to have the capacity to effectively integrate and consume this data within their organization to derive value from it. The NSA has collected massive amounts of data, but evidence suggests it has not made the investments necessary to consume data at the same rate as it is collecting it (Salmon, 2014). It is costly for firms to integrate extra data sets. If the additional data has no particular purpose at the time of integration, this data may not result in additional benefit. On the other hand, if a firm is in a position to uniquely capture data that is not currently highly valued, but assessed to be a potentially valuable asset in the future, it could capture and store that data, but should strongly consider deferring the costs of integration until its use becomes relevant.

5 Conclusion

5.1 Summary

To maximize the value derived from data and analytics, firms are faced with a complex set of implicit costs and management challenges that go beyond basic software and implementation fees. These implicit costs are less understood, but previous research and case examples are means by which they can be further analyzed. This analysis not only supports the collective understanding of these costs, but also proposes means by which firms can address two such costs as firms seek to maximize their total payoffs from data and analytical practices.

An increase in data and analytical practices inherently introduces additional metrics that can have unforeseen negative effects on firm payoffs, but firms can mitigate these effects. Firms can protect unmeasured tasks and outcomes by avoiding or lowering unnecessary incentives on measured tasks. If the firm requires that these tasks be incentivized, the tasks can be separated among agents. If incentives are unavoidable and the tasks cannot be separated, firms can optimize job and measurement design to mitigate the potential negative impact on firm benefits. Holmstrom and Milgrom present the design variables of commission rates, exclusion of private returns, and monitoring intensity. This paper's analysis of external metrics utilized by Google highlights additional means of optimization, specifically complexity, secrecy, and constant evolution in analytics. The analysis, also suggests that, contrary to the recommendations of Holmstrom and Milgrom, the potential firm value afforded by data and analytics are enough for firms to take on the risks associated with increased incentives.

Firms also can take action to ensure that people are effectively complemented by information technology so that the firm realizes the maximize benefit from the use of data and analytics within the organization. To do this first requires an understanding of the appropriate context under which people and information technology should operate as complements. This excludes tasks that are candidates for automation and high-level strategic decisions. For all remaining tasks, firms must establish a balance between people and technology, requiring people to break from the habit of relying solely on intuition while integrating enough intuition such that they are not blindly accepting data. Firms also have to avoid the "more is better" mantra and recognize data should only be procured at the rate at which the firm can consume it. When prioritizing data, firms should consider the relative value, quality, and source of each data set.

The full extent of these two costs, as well as the understanding of other, unaccounted-for costs, will resolve with time as more firms utilize data and analytics and as more research is completed. Understanding and addressing such costs will be essential as firms strive to realize the true potential afforded by data and analytics.

5.2 Analysis Limitations and Future Research

There is likely a sampling bias in this analysis since case examples of unsuccessful efforts or the use of data for more menial tasks, such as automating operational actions without any use of analytics, are less likely to be published or widely discussed. It is still worth noting that among these cases there is significantly more focus on analytics for operational efficiency and marketing. Still, there are minimal examples of analytics being utilized by firms when making strategic decisions.

These cases highlight a few examples of the use of data within organizations and some of the context in which firms achieved success. The suggestions outlined in this paper should not be blindly accepted as sweeping truths. Instead, this analysis is meant to be an initial discussion of the costs associated with data and analytics, and case examples that highlight such costs. It is possible that other firms have faced similar conditions and attempted the outlined practices, but with different results. Further research is required to understand such cases.

Given this sampling bias and the limitations of this analysis, there is room for additional research related to the application of the Multitask Principal-Agent Model and the outlined practices. The analysis of this paper is at a relatively high level and mostly conceptual, leaving room for a more rigorous exploration of the model in the outlined contexts. Other potential areas for future research include an evaluation of the pervasiveness of the two outlined costs and the validity of suggested practices. Additional work could also include a deeper analysis of the bounds to which technology can be applied for a variety of tasks, in particular tasks that can be fully automated or are more strategic and utilize limited information technology.

Appendix

A1. Cost and Associated Practices Outline

Cost 1: Adverse Effects of Metrics – mitigation of the potentially negative impact on firm benefits that result from an increased reliance on the metrics that accompany data and analytics

Practices

5. Protect Unmeasured Tasks and Outcomes – avoid or lower incentives on measurable tasks
6. Isolate Unmeasurable Tasks – only incentivize agents exclusively charged with highly measurable tasks
7. Optimize Job and Measurement Design
    - Holmstrom and Milgrom Organizational Design Variables: commission rates $\boldsymbol{\alpha}$, exclusion of private returns $\boldsymbol{\delta}$, and monitoring intensity $\boldsymbol{\Sigma}$.
    - Internal Metrics: balance quality and quantity
    - External Metrics: obfuscate incentives and metrics by keeping them complex, secret, and in a constant state of evolution.
8. If Necessary, Accept Risks – given the potential value, the risks may be worthwhile

Cost 2: Complement People with Technology (Data-Driven Decision-Making) – development of a balance between humans and technology that ensures people are complemented by data, analytics, and information technology for the maximum firm payoff

Practices

1. Context: Task Automation and Strategic Decisions
    a. Highly repetitive tasks that can be codified are candidates for automation

    $$B_{Tech}(t_i) > B_{Human}(t_i) \quad \Rightarrow \quad \text{automate } t_i$$

    b. Recognize that predictive capabilities are limited to extrapolation, hindering the application of data and analytics in defining high-level strategies
    c. Except for automation and predictability issues, decision-making commonly benefits from combining humans and information technology

2. Balance
    a. Shift from Intuition
        i. Break habit of relying on intuition
        ii. Present consumable information
        iii. Involve and iterate with users
        iv. Open the "black box"
    b. Prevent Overreliance
        i. Avoid blind acceptance of data and analytics uninformed by intuition
        ii. Beware of limitations of data and analytics for automation and high level-strategy scenarios
        iii. Beware of the difference between correlation and causation
        iv. Account for sampling error and sampling bias
3. Data Mass
    a. More data is not necessarily better
    b. Not all data has the same value
    c. Beware of data quality and source
    d. Data procured should not exceed firm consumption capacity

References

Battelle, J. (2005). The birth of Google. *WIRED-SAN FRANCISCO-, 13*(8), 102.

Brynjolfsson, E., & McAfee, A. (2012). Big Data: The Management Revolution. *Harvard Business Review, 90*(10), 60-68.

Brynjolfsson, E., Hitt, L., & Kim, H. (2011). Strength in Numbers: How does Data-Driven Decisionmaking Affect Firm Performance?. *Available at SSRN 1819486.*

Brynjolfsson, E., Hitt, L. M., & Yang, S. (2002). Intangible Assets: Computers and Organizational Capital. *Brookings Papers on Economic Activity, 2002*(1), 137-198.

Campbell, D. T. (1976). Assessing the Impact of Planned Social Change. *Occasional Paper Series,#* 8.

ChessBase. (2005, June 19). Dark Horse ZackS Wins Freestyle Chess Tournament. Retrieved May 1, 2014, from *ChessBase*: http://en.chessbase.com/post/dark-horse-zacks-wins-freestyle-che-tournament

Codd, E. F. (1970). A Relational Model of Data for Large Shared Data Banks. *Communications of the ACM, 13*(6), 377-387.

Davenport, T. H. (2006). Competing on Analytics. *Harvard Business Review,84*(1), 98.

Duhigg, C. (2012). How Companies Learn Your Secrets. *The New York Times,16.*

Gartner. (n.d.). IT Glossary. Retrieved May 1, 2014, from *Gartner.com*: http://www.gartner.com/it-glossary/big-data

Gartner. (2013). Toolkit: Big Data Business Opportunities From Over 100 Use Cases. *Gartner, Inc.*

Greenfeld, K. T. (2010, August 6). Loveman Plays `Purely Empirical' Game as Harrah's CEO. Retrieved May 1, 2014, from *Bloomberg.com*: http://www.bloomberg.com/news/2010-08-06/loveman-plays-new-purely-empirical-game-as-harrah-s-ceo.html

Gyöngyi, Z., & Garcia-Molina, H. (2005). Link Spam Alliances . *31st international conference on Very Large Databases.* VLDB Endowment.

Haley, B. J., Watson, H. J., & Goodhue, D. L. (1998). The Benefits of Data Warehousing at Whirlpool.

Hansell, S. (2007). Google Keeps Tweaking its Search Engine. *New York Times.*

Harford, T. (2014, March 28). Big Data: Are We Making a Big Mistake? Retrieved May 1, 2014, from *The Financial Times*: http://www.ft.com/intl/cms/s/2/21a6e7d8-b479-11e3-a09a-00144feabdc0.html#axzz30chjc7es

Hayes, F. (2002, April 15). *The Story So Far.* Retrieved from Computerworld: http://www.computerworld.com/s/article/70102/The_Story_So_Far

Hays, C. L. (2004). What Wal-Mart Knows About Customers' Habits. *The New York Times, 14.*

Heudecker, N. (2013). Hype Cycle for Big Data, *2013. Gartner.*

Holmstrom, B., & Milgrom, P. (1991). Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design. *JL Econ. & Org., 7*, 24.

Holmstrom, B., & Milgrom, P. (1994). The Firm as an Incentive System. *The American Economic Review*, 972-991.

Kirk, J. (2006, February 7). 'Analytics' Buzzword Needs Careful Definition. Retrieved May 1, 2014, from *InfoWorld*: http://www.infoworld.com/t/data-management/analytics-buzzword-needs-careful-definition-567

Laney, D. (2001). 3D Data Management: Controlling Data Volume, Velocity and Variety. *META Group Research Note, 6.*

Laney, D. (2013, April 3). Who Owns (Really Owns) "Big Data". Retrieved May 1, 2014, from *Gartner*: http://blogs.gartner.com/doug-laney/who-owns-really-owns-big-data/

Lippman, S. A., & Rumelt, R. P. (1982). Uncertain Imitability: An Analysis of Interfirm Differences in Efficiency Under Competition. *The Bell Journal of Economics,* 418-438.

Locke, L. (2007). The Future of Facebook. *Time Magazine, 17.*

Locu. (2014). Locu Homepage. Retrieved May 1, 2014, from *Locu.com*: https://locu.com/

Martens, D., & Provost, F. (2011). Pseudo-Social Network Targeting from Consumer Transaction Data.

McElheran, K. (forthcoming). Do Market Leaders Lead in Business Process Innovation? The Case(s) of E-Business Adoption. *Management Strategy.*

McKeon, A. (2012, May 15). Call Center Agents' Behavior Should be Focus -- Not Clunky Metrics. Retrieved May 1, 2014, from *SearchCRM.com*: http://searchcrm.techtarget.com/news/2240150260/Call-center-agents-behavior-should-be-focus-not-clunky-metrics

Murphy, C. (2010, December 2). Procter & Gamble CIO Filippo Passerini: 2010 Chief Of The Year. Retrieved May 1, 2014, from *InformationWeek*: http://www.informationweek.com/it-leadership/procter-and-gamble-cio-filippo-passerini-2010-chief-of-the-year/d/d-id/1094572

Nagayama, K., & Weill, P. (2004). 7-Eleven Japan Co., Ltd.: Reinventing the Retail Business Model. *Center for Information Systems Research Working Paper, 338.*

Ordonez, V., Deng, J., Choi, Y., Berg, A. C., & Berg, T. L. From Large Scale Image Categorization to Entry-Level Categories.

Palantir. (2014). About Palantir. Retrieved May 1, 2014, from *Palantir.com*: http://www.palantir.com/about/

Rosen, R. J. (2012, August 9). *Google's Self-Driving Cars: 300,000 Miles Logged, Not a Single Accident Under Computer Control.* Retrieved May 1, 2014, from The Atlantic: http://www.theatlantic.com/technology/archive/2012/08/googles-self-driving-cars-300-000-miles-logged-not-a-single-accident-under-computer-control/260926/

Salmon, F. (2014, January). Why Quants Don't Know Everything. Retrieved May 1, 2014, from *WIRED.com*: http://www.wired.com/2014/01/quants-dont-know-everything/

Saunders, A., & Tambe, P. (2013). A Measure of Firms' Information Practices Based on Textual Analysis of 10-K Filings.

Schectman, J. (2014, January 24). How a Family Tragedy Landed on a Retailer's Mailing. Retrieved May 1, 2014, from *The Wall Street Journal*: http://online.wsj.com/news/articles/SB10001424052702303743604579351050709924372

Shiller, B. R. (2014). First-Degree Price Discrimination Using Big Data.

Simon, H. A. (1955). A Behavioral Model of Rational Choice. *The Quarterly Journal of Economics, 69*(1), 99-118.

Singhal, A., & Cutts, M. (2011, February 24). Finding More High-Quality Sites in Search. Retrieved May 1, 2014, from *Google Official Blog*: http://googleblog.blogspot.com/2011/02/finding-more-high-quality-sites-in.html

Watson, H. J., Wixom, B. H., Hoffer, J. A., Anderson-Lehman, R., & Reynolds, A. M. (2006). Real-time business intelligence: Best practices at Continental Airlines. *Information Systems Management, 23*(1), 7.

Welch, M., & Westerman, G. (2013). Caesars Entertainment: Digitally Personalizing the Customer Experience. *MIT Center for Digital Business*.