A System-Oriented Analysis of Team Decision Making in Data Rich Environments

By
Shirish K. Nilekar

M.S., Computer Science, Worcester Polytechnic Institute, Worcester, MA, 2006
B.E., Computer Science & Engineering, University Of Pune, India, 1998

Submitted to the System Design and Management Program in Partial Fulfillment of the Requirements for the Degree of
Master of Science in Engineering and Management
at the

Massachusetts Institute of Technology

May 2013
[June 2013]

Signature redacted

Signature of Author:___ ___________________________________
System Design and Management Program

Signature redacted

Certified By: _____ ___________________________________
Michael Davies, Thesis Supervisor
Senior Lecturer, Engineering Systems Division

Signature redacted

Accepted By: ___________________________________
Patrick Hale, Director
System Design and Management Program

A System-Oriented Analysis of Team Decision Making in Data Rich Environments

By

Shirish K. Nilekar

Submitted to the System Design and Management Program on January 30, 2013 in Partial Fulfillment of the Requirements for the Degree of Master of Science in Engineering and Management

Abstract

The information processing view of organizations [1] and subsequent works highlight the primary role of information processing in the effective functioning of markets and organizations. With the current wave of "big data" and related technologies, data-oriented decision making is being widely discussed [2] as a means of using this vast amount of available data for better decisions which can lead to improved business results. The focus of many of these studies is at the organization level. However, decisions are made by teams of individuals and this is a complex socio-technical process. The quality of a decision depends on many factors including technical capabilities for data analysis and human factors like team dynamics, cognitive capabilities of the individuals and the team. In this thesis, we developed a systems theory based framework for decision making and identified four socio technical factors viz., data analytics, data sensing, power distribution, and conflict level which affect the quality of decisions made by teams. We then conducted "thought experiments" to investigate the relative contribution of each of these factors to the quality of decisions. Our experiments and subsequent analyses show that while improved data analytics does result in better decisions, human factors have an out-sized contribution to the quality of decisions, even in data rich environments. Moreover, when the human factors in a team improve, the predictability of the positive impacts due to improvements in technical capabilities of the team also increases.

Thesis Supervisor: Michael Davies

Title: Senior Lecturer, Engineering Systems Division

This page has been intentionally left blank.

Acknowledgement

To Aai and Pappa - your faith in the necessity and value of good education made my decision to pursue and complete this degree so much easier.

To Dada, Vahini, Eshaan, and Aayush - your support and humor made the frequent frustration of homework in the last two years much more manageable.

To Prabha, this thesis is complete in large measure to your encouragement and persistence.

To my friends, Dmitriy, Neil, Karthik, Ilana, Sergey, Armando, JB, Jaya, Ritesh, Jon, and many more who made my experience at MIT so rich – this definitely has been one of the best phases of my life.

Knowledge indeed is power and to the professors at MIT, I owe you all a huge debt of gratitude for imparting your knowledge to us. To Pat Hale – thanks for your excellent leadership of the SDM program. The flexibility I enjoyed in coursework and tailoring the program to my needs is very much appreciated. To the wonderful and helpful staff at SDM, thank you for making all administrative things at MIT a non-hassle for students.

To the creator(s) of the epic Mahabharata, which I greatly enjoyed reading while writing the thesis report, - your immortal work will continue to inspire humanity for all eternity.

Last but not the least, a big thank you to Michael Davies for providing me this opportunity to work with you on this thesis and the SLaM Lab class. Your insights into management theories and models, ideas about visualizations, and the discussions during SLaM Lab have been instrumental in expanding my understanding of not just the technology behind businesses, but also their strategies and organizational process.

Table of Contents

List of Tables

List of Figures

1.0 Introduction

"Data is the new oil!"
Clive Humby, ANA Senior marketer's summit, Kellogg School.

While I was writing this thesis, I received a copy of the Mahabharata as a gift. In this ancient epic, the Kaurava and Pandava armies that stood poised for battle were very well matched in their valor and achievements. If the five Pandavas and their allies were mighty, most of whom had never seen defeat in battle, so were Duryodhana and his brothers, Bhishma, Drona, Karna, Bhurisravasa, Jayadhrata and their allies on the Kaurava side, equally valiant in their deeds and unmatched in valor. If the Pandavas had the wisdom of Krishna and the righteousness of Yudhisthira on their side, the Kauravas had the artful Shakuni and the experienced Bhishma and Drona on their side. The two armies were equally matched as they stood facing each other at Kurukshetra on the first day of battle and not one soul could predict which side would win, if any. Over the next eighteen days of high uncertainty and bloody battles, the Pandavas lost many of their near and dear ones including Abhimanyu, son of Arjuna and Subhadra, and Ghatotkach, son of Bhima and Hidimba. The Kaurava army, on the other hand, was completely routed.

Why did the war between two equally matched armies result in victory for one side and complete defeat of the other? While this epic is celebrated as the victory of good over evil all across India, an MBA class discussion about this situation today will very likely include the suggestion "The Pandavas made the right strategic choices all along". Moving one level deeper than "having the right strategy", they would notice that the Pandavas collectively made a series of decisions which balanced the outcome of the war in their favor – be it Yudhishthira's decision to keep his promise and accept exile when he lost his kingdom in a game of dice – a decision which won him important allies, or Arjuna's decision to use Shikhandin as a shield when battling Bhishma – a decision which gave him an upper hand in his duel against the invincible Bhishma. Moreover, the Pandava's objective was focused – to win back their dignity, their kingdom, and to punish Duryodhana and his brothers for the injustice they had done to Panchali. The

Kauravas on the other hand were mostly driven by jealousy, hatred and had internal conflict which prevented them from working together effectively. Karna, for example, refused to fight in an army led by Bhishma, and Duryodhana quarreled with his generals and questioned their resolve when battles were lost. This internal conflict and lack of aligned objectives resulted in the Kauravas making some impetuous and rash decisions in battles which cost them the war.

From what I know of other classics, I am quite certain that a similar dynamic of good and poor decision making with momentous consequences can be found in many of them - Homer's Illiad, the epic of Gilgamesh, Shakespeare's novels, Star Wars, Lord of the Rings, and many more. From the classics of yesterday to today's works of fiction and even the current corporate wars, some things seem to have stayed the same. This thesis provides quantitative insights into some human and technical capabilities which contribute to the quality of decisions made by teams of individuals.

1.1 Problem Definition

In the recent past, there has been an exponential growth in the amount of data being created and recorded [24]. The “big data” phenomenon as it is popularly know has resulted in a revolution in the entire data management ecosystem. Right from data storage hardware to the algorithms for extracting information out of this data to new applications which utilize and present this information in innovative ways, and the laws governing the storage and privacy of this data - people and organizations are re-thinking old ways of managing their data and coming up with innovative ideas for storing, processing, and using it. The sources of big data vary from textual data in user updates on social websites like Facebook and Twitter, log files on servers, click stream data on websites, sensor data like RFID tags, audio and visual recording by users using their smart phones, and many more sources. Enterprises today are swimming in this sea of data and are scrambling to collect it, store it, and mine it to get better insights into their business environment and operations. A few examples of such data mining applications include targeted advertisements for consumers of social websites based on their usage and browsing patterns, optimization of supply chains or other business processes, improved drug discovery, machine learning based image and scene recognition, predicting diseases, or even predicting purchase histories of consumers. Many of these applications, especially at the current scale, would have been highly impossible even a decade back.

Are businesses making better decisions with all this data at hand? Most professionals are aware that more data can result in more information but it may not always lead to better decisions. Any non-trivial decision in an organization usually involves one or more teams of people contributing their effort and inputs to the decision, even if the final accountability for the decision might rest with one individual. Arriving at a decision is usually a complex socio-technical process involving both technical expertise and human interactions. The central question we try to answer in this thesis is: *What is the relative contribution of technical factors like data analytics compared to human factors in determining the quality of decisions made by a team of individuals?* The hypothesis we want to test here is that even in data rich environments which allow opportunities for unprecedented and extremely intelligent data analytics, human factors have an out-sized contribution to the quality of decisions in team environments. If this primary hypothesis is validated, we also want to determine the relative contribution of human factors compared to technical factors to the quality of decisions made.

1.2 Research Methodology

The framing of the research question being addressed in our thesis was influenced by the author and his advisor's interest in data analytics and its effect on decision making. To investigate our hypothesis, we first developed a systems theory based framework for analyzing decision making in teams. This framework helped us identify some key technical and human factors that could affect decision quality. Our personal experience and observations helped us form four hypotheses about the relative importance of these factors and their contributions to the quality of decisions. We then designed a survey in the form of a "thought experiment" to measure the contribution of these factors. Industry professionals were asked to fill out a survey and the data collected from these surveys was used to test our hypotheses and gain deeper understanding into how decisions are made in teams. Of the four hypotheses, three were validated in our data and the results can be explained via existing laws while another one could not be validated based on the data.

1.3 Thesis Outline

This chapter introduced the problem we are addressing and provides an outline of what comes next. Chapter 2 discusses the concept of data oriented decision making and also explains our systems theory based framework for decision making in teams. We also briefly discuss the human side of decision making, identify the factors influencing decisions, and form hypotheses to test. Chapter 3 details our survey design and the analysis of survey data. In Chapter 4, we discuss related work in this field and conclude in Chapter 5 with our conclusions and possible directions for future work.

2.0 From Data to Decisions

"If you torture the data long enough, it will confess."
Ronald Coase

With the easy availability of large amounts of data in many industries, data analytics and especially *predictive* data analytics capabilities and skills are highly sought after by companies. Appendix A has a more detailed discussion of this "big data" phenomenon and some of its use cases. Managers in organizations are looking for new ways to mine the data they collect and own. In fact, Harvard Business Review's front page article for October 2012 titled "Getting Control of Big Data" mentions that "businesses are collecting more data than they know what to do with. To turn all of this information into competitive gold, they'll need new skills and a new management style". In this chapter we examine the process of decision making, especially in such data rich environments and then introduce a systems theory based framework which is useful for understand the relative importance of data analytics in decision making. Finally, we introduce the concept of decision quality, identify the factors which we believe contribute to decision quality and then form hypotheses about the relative importance of each of these factors in the quality of decisions.

2.1 Data and Decision Making

At a very basic level, a decision is a choice, either implicit or explicit, from amongst a set of alternatives in a given situation. These choices and alternatives could be explicitly explored and documented by a team of people or could be implicit in an individual's mind based on their previous experiences and their mental model of the situation. The DIKW (short for Data-Information-Knowledge-Wisdom) hierarchy, first popularized by Russell Ackoff, and shown in Figure 1 is a useful model for understanding how data gets converted into wisdom which guides human decisions. In this hierarchy, data represents the raw material which is converted into information by organizing and presenting it in a meaningful form. Knowledge and wisdom are the internalization of this information by individuals and are employed for reasoning about new data and information.

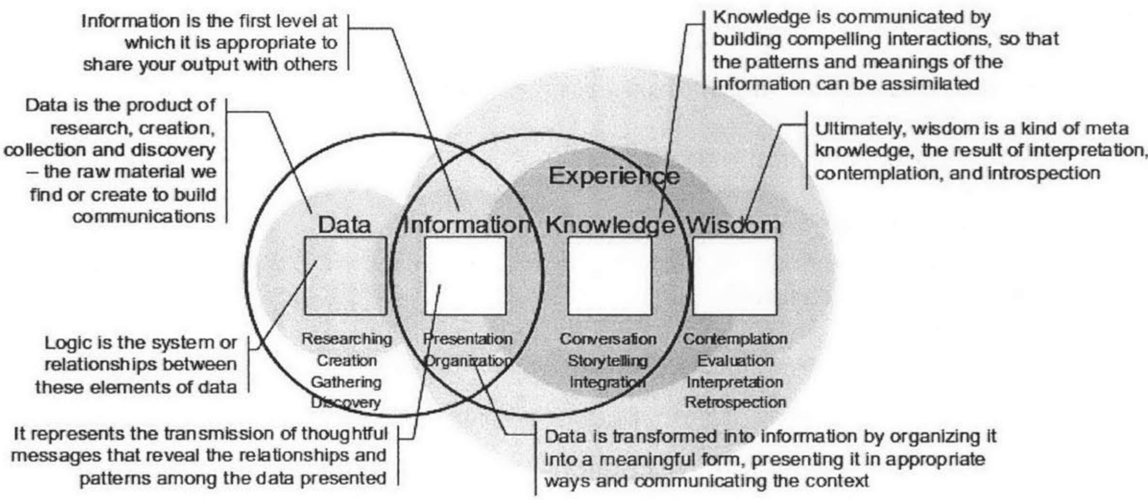

Figure 1: The DIKW Hierarchy.

(Source: SLaM Lab class by Prof. Michael Davies @ MIT)

At a more operational level, the process of decision making has been also been discussed in the literature. A detailed framework of the structure and process of corporate team decision making was proposed by Howard [12]. In this framework, there are three primary roles of people involved in a decision, viz, decision makers, decision staff, and implementors. The decision making process itself is a six step process consisting of:

1. Setting the appropriate frame.
2. Identifying creative and doable alternatives.
3. Gathering meaningful and reliable information.
4. Gaining clarity around values and trade-offs.
5. Using logically correct reasoning.
6. Committing to action.

In scientific research settings, Wagstaff [3] lists similar steps that should be ideally involved in machine learning research projects:

1. Frame the problem to be solved.

2. Phrase the problem as a machine learning task.
3. Collect data.
4. Choose or develop an algorithm for data processing.
5. Choose metrics, conduct experiments.
6. Interpret results and evaluate alternatives.
7. Publish and share results.

As the two frameworks above show, the process of decision making usually consists of broadly three phases - a preparation phase involving identification of the problem, a planning phase involving exploration of options, and an execution phase.

2.2 A Systems Theory Model for Business Decision Making

The two decision making processes from Howard [12] and Wagstaff [3] discussed in the previous section are a step-wise and logical description of the decision making process. However, they give an impression that decision making is a linear process, when in reality it is usually iterative and rarely follows these steps in the listed order. To account for this iterative nature of the process and to understand it better, we borrow the feedback control model from systems theory and adapt it to the process of team decision making. Figure 2 shows a commonly used representation of a feedback controlled process in engineering and control systems theory. It consists of a **controlled process** with a well-defined **boundary**, **inputs** and **outputs.** Even with a well defined boundary, every process is affected by **noise** from the environment and must be designed to handle it. **Sensors** read signals about the current state of the process and convert them into data, while automated **controllers**, which encapsulate a model of the controlled process direct the **actuators** to perform actions to keep the process in a desired state. In addition to automated controllers, there are also human controllers who monitor the process and can provide additional controls, if needed, to keep the process in a desired state. This interaction is shown via dotted lines in Figure 2. The controlled process can be as simple as maintaining of temperature of a room at a preset level or as complex as the flight avionics system of a modern aircraft, or a controlled nuclear reaction involving many thousands of sensors and hundreds of personnel to monitor the process.

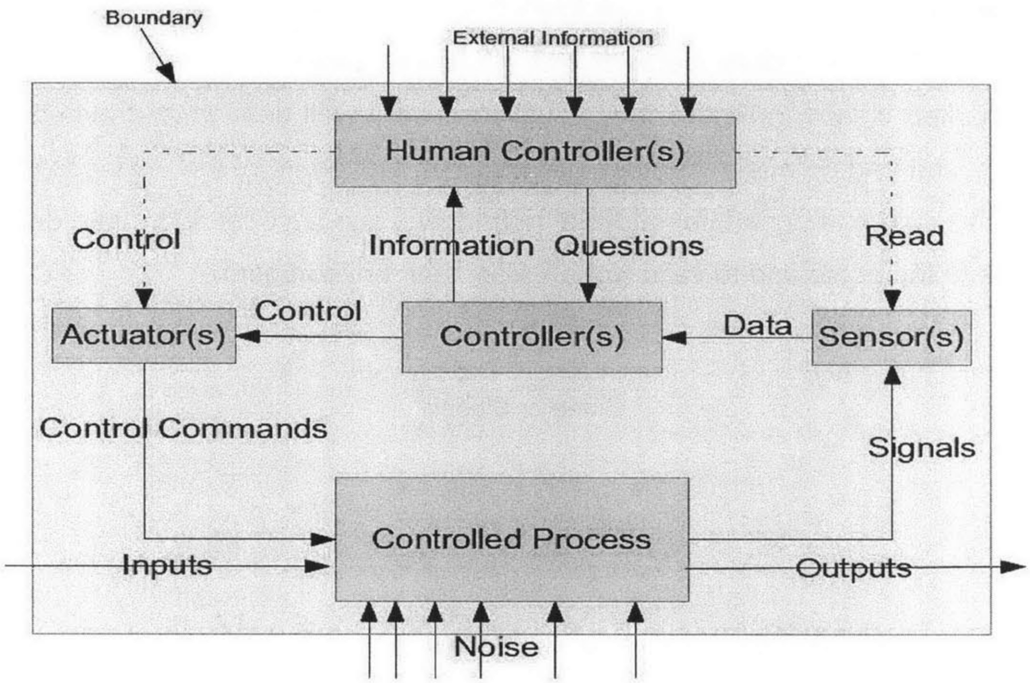

Figure 2: A Generic Feedback Controlled Process

In this thesis, we adapt this model of feedback control and use it as a framework to reason about decision making in teams. Figure 3 shows our adapted model for team decision making and is discussed in more details in the next two sections.

2.2.1 The Structure of Our Adapted Model

Our model has been influenced by Prof. Nancy Leveson's STAMP framework [4] for System Safety. In this adapted model, the controlled process is a **business process** such as a marketing campaign, a manufacturing process, or a product development process that is actively managed by a group of professionals. The **inputs** to the business process include the capital and human resources necessary to achieve the goal. The **output** is a desired business outcome – such as an increase in market share, greater adoption of product etc. **Sensors** convert signals from the process into data and pass this data on to controllers. The **controllers** are the data storage and analytics modules which have a model of the current state of the system and process data from sensors to make decisions. The objective of these decisions is to "guide" the business process towards desired outcomes. The **actuators** execute these decisions and perform the necessary actions in order to maintain or bring the business process to is desired state. The boundary between the business process and its environment, although marked clearly in Figure 3, is usually not very well defined and events in the environment can and do affect the business process. Managers of the business process must take this fluid and uncertain nature of boundaries into account and make decisions in response to the noise and to valid signals from their environment.

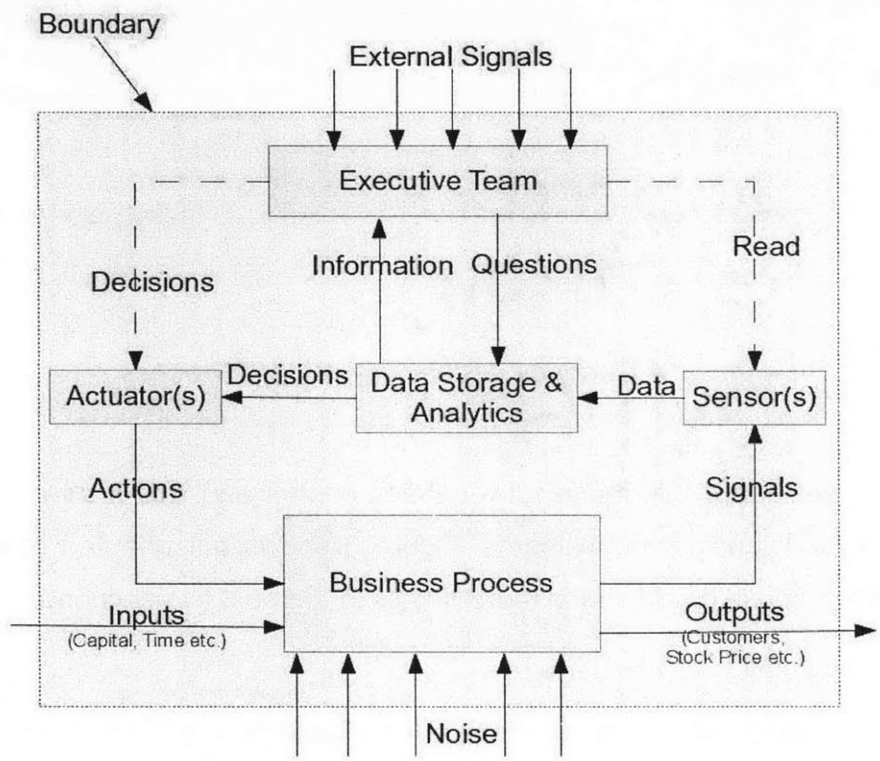

Figure 3: Adapted Feedback Control Process for Business Decision Analysis

2.2.2 The Behavior of Our Adapted Model

Let us consider a concrete example to explain how this adapted model works. The marketing executives of an on-line retailer have decided to launch a marketing campaign in which every customer who makes two or more purchases of more than $100 on the company's website within a one month time window gets a 10% discount on their second purchase. The objective of this campaign is to increase the amount of repeat business from existing customers in order to lower the total marketing budget. In this case, the input to the campaign is the money required to fund the campaign and the human effort needed to start and run the campaign. The expected output is an increase in repeat customer base. A sensor on the company's website distinguishes repeat customers from new online shoppers, typically based on IP Address information or some cookies stored in the user's web browser. If the customer is a repeat customer, an automated controller – in this case, some software script running on the web server - decides to notify the user of the available discount. When the user's shopping cart exceeds $100 in amount, a sensor captures that data and another automated controller makes the decision to provide the 10% discount to the customer. An actuator actually applies the discount of 10% and recalculates the new total amount to pay. The sensors, controllers and actuators in this example are all automated software programs and these decisions are made without human intervention. The human controllers – the marketing executives - regularly monitor the performance of the campaign. At the end of the campaign period, the human controllers must make a decision: Depending on the results of the campaign compared to the objectives, the new market conditions including reactions from competition, budgetary constraints, or other organizational or external factors, they must decide either to extend the campaign or discontinue it.

In our adapted model, the sensors, controllers, and actuators all have an automated component and a human component. When there is enough certainty about the process or some part of it, sensors, controllers, and actuators can usually be developed to automate that part of the business process. In many cases, the automated controllers may be able to process data but leave the actual decisions to humans – either due to legal reasons or due to the risk of a mistake being too high. As the uncertainty in the environment and in the process increases, or if some of the signals cannot be detected through physical sensors, the level of human involvement in the business process also increases. The human controllers typically combine information from sensors, controllers, along with other signals from the environment in making their decisions. So, humans perform all three roles – sensors, controllers and actuators in our adapted model.

Just as in engineering systems, the sensors, controllers, and actuators, in our model can work at different speeds and may contain state information which may be out of sync with the actual state of the process. This difference in processing speeds and lagging state information results in **information delays** in the system. These delays, when combined with the balancing or re-enforcing **feedback loops** in this process can result in business processes exhibiting patterns of either **oscillatory**, **goal seeking**, or **exponential behavior**, just as many engineering processes do.

Also, the models for business processes are usually not as well understood as those of physical processes. Consequently, controllers have a limited subset of variables to model the business process and there is usually a fair deal of uncertainty about the actual state and behavior of a business process at any given point in time. In such a situation, managers often rely on their **mental models** based on their prior experiences to guide their decisions. In order to gain deeper insights into their business processes, managers (and researchers) may also do exploratory data analyses, often using advanced data processing and information visualization techniques.

The model discussed in this section represents a single team in one organizational unit within a larger organization. However, this model can be repeated at each level of hierarchy within the larger organization. So, in some ways, this is a recursive model, not unlike the structure of the Viable System Model [30].

A variety of fundamental system level laws such as Ashby's Law of Requisite Variety, signaling theory, Arrow's paradox, and even the concept similar to entropy from the second law of thermodynamics can be potentially used in the discussion and analysis of the behavior of such socio-technical systems - but are outside the scope of this thesis.

2.3 The Human Side of Decision Making

The process of decision making discussed so far in this thesis is also studied in many scientific fields. Psychology students learn about sensing (data gathering), attending (limiting data), perceiving (interpreting data and information), cognition (the process of thinking), learning (making associations) and memory (retention of information) – the mental faculties which all contribute to the activity of decision making. Economic models think of humans as (mostly) rational entities who try to maximize their own utility in a given situation. Sociologists and philosophers often debate the ethical implications of decisions, while anthropologists analyze the cultural and historical influences on people's thinking and decisions. The total available literature on this subject is very vast and we are not trying to summarize it here, just provide an overview of some work which is relevant to this thesis.

2.3.1 Ladders, Loops, and Mental Models

Chris Argyris proposed the "Ladder of Inference" as a framework for understanding how people arrive at conclusions from observed data. The model is shown in Figure 4 and is self explanatory.

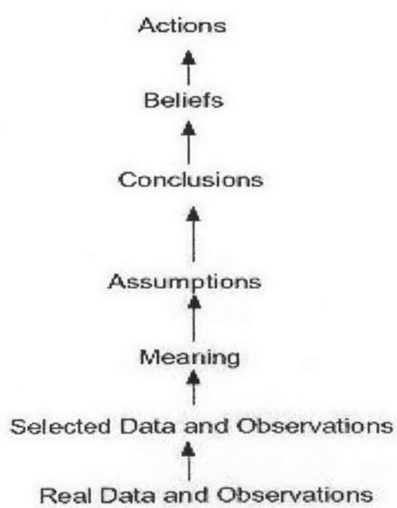

Figure 4: The Ladder of Inference

Even if this model is mostly linear, there are possible loops between the "Beliefs" step of the ladder and the "Selected Data and Observations" step. The loop from "beliefs" to "selected data" means that people tend to select data which confirms their belief systems. In psychology, this phenomenon is often referred to as confirmation bias. A loop from "selected data" to "beliefs" would signify that people update their beliefs based of new data and evidence – the process of learning. Argyris has also proposed a useful model for learning itself – the concept of double loop learning shown below in Figure 5.

Both these frameworks, the ladder of inference, and the learning loop model clearly indicate that the chance of a mismatch between the reality of a situation an individual's mental model of the situation is extremely high. In the early stage of any non-trivial decision, there is usually a fair amount of uncertainty and people's belief's and mental models have a huge impact on how the situation is perceived and the problem framed.

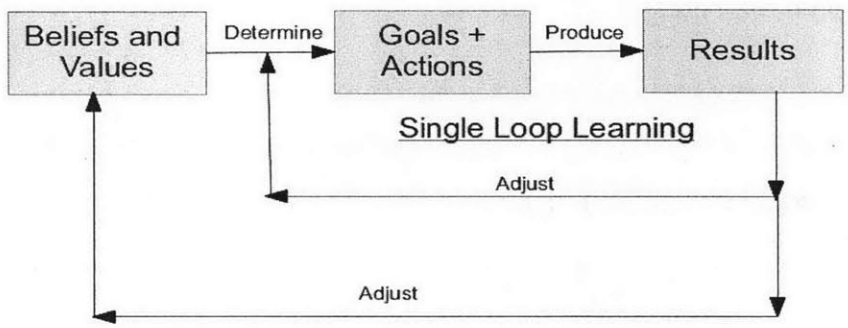

Figure 5: Double Loop Learning Model

2.3.2 Decision Making Styles

In addition to differing mental models for a given situation, individuals also have different decision making styles in identical situations. Busentz and Barney [6] compare two styles of decision making in most big organizations – entrepreneurial versus managerial and note that people with these two kinds of decision making styles rely on different heuristics and exhibit different biases in their decision making process. The entrepreneurial type employees are more comfortable with arriving at decisions based on limited data and show a bias towards over-representativeness while the managerial types are more cautious and guarded in their decision making. They also note that each style of decision making has its own advantages and disadvantages with each decision style more suited in some contexts than others. For example, in a resource constrained environment of a startup company, entrepreneurs often rely a lot on instinct as they do not have the luxury of a detailed data analysis before making a decision. Often, such data is not even available. On the contrary, if data and the analytical capabilities are easily available, it does not make sense not to use them in order to arrive at more informed decisions.

Individual personalities also play a big role in determining how a situation is perceived and how the problem is framed in a given situation. MBTI [10] is a commonly used framework to categorize people into sixteen preferred personality types while Belbin [9] provides a framework of nine preferred team roles for individuals working in a team. These two frameworks are a good starting point for understanding differing thinking styles and preferred roles of individuals in team settings.

2.3.3 Biases and Heuristics in Decision Making

Economic theory talks about utility functions and decisions as an optimization process that maximizes individual utility. This commonly held notion of individuals making rational decisions to maximize their own utility in a given situation was challenged in the seminal work of Herbert Simon [14] on "bounded rationality" in which he showed that decision making by humans is not an exercise in rational optimization. Instead, the rationality of individuals is restricted by the available information, their cognitive limitations and the amount of time they have to make a decision - most decision makers are "satisficers" instead of optimizers. Kahneman and Taversky [7] additionally showed how people use interesting heuristics to evaluate losses and gains differently when making choices between alternatives involving risks even when the probabilities of outcomes are known. A detailed discussion of these theories is beyond the scope of this thesis.

2.3.4 Distributed Cognition in Teams

Hutchins [11] observed on US Navy ships that "the outcomes that mattered to the ship were not determined by the cognitive properties of any single navigator, but were instead a product of the interactions of several navigators with each other and with a complex suite of tools". This observation led to the development of the theory of Distributed Cognition in which the traditional boundary and unit of cognitive science was expanded from an individual to a group of interacting individuals. The key insight of this theory is that cognitive processes do not happen only inside the brains of individuals but may also be distributed across members of a social group, may involve co-ordination between internal and external (material or environmental) structures, and may be distributed through time in such a way that the products of earlier events can transform the nature of later events [11].

Differing mental models, decision making styles, personalities, and motivations often cause conflicts when team members interact with each other. This conflict, if constructive, can lead to uncovering of new knowledge, alternatives, possibilities and a better shared understanding of reality. On the other hand, destructive conflicts between individuals often lead to poor decisions. Eisenhardt et. al [8] discuss the beneficial nature of constructive conflict in arriving at optimal decisions.

2.3.5 The Role Information Visualization in Decision Making

Albert Einstein said "My particular ability does not lie in mathematical calculation, but rather in visualizing effects, possibilities, and consequences." [18] .

Information visualization is an important technique in analyzing large data sets and any discussion on data oriented decision making would be incomplete without mentioning it. In the industry, information visualization is used extensively for analyzing large data sets and the following two quotes are an attestation to that fact:

"Texas Instruments manufactures microprocessors on silicon wafers that are routed through 400 steps in many weeks. This process is monitored gathering 140,000 pieces of information about each wafer. Somewhere in that heap of data can be warnings about things going wrong. Detect a bug early before bad chips are made." [20]

"Eli Lilly has 1500 scientists using an advanced information visualization tool (Spotfire) for decision making. With an ability to represent multiple sources of information and interactively change your view, its helpful for homing in on specific molecules and deciding whether we should be doing further testing on them." [20]

Similar examples of the utility of information visualization which provide useful insights into data can be found in many other industries and applications ranging from process improvements to fraud detection.

Card et al [23] provide an excellent explanation of how information visualization aids cognition. The table from [23] is reproduced below and clearly shows how valuable information visualization is to the process of data analysis and making decisions based on it.

How Information Visualization Amplifies Cognition	
Increased Resources	
• High Bandwidth Hierarchical Interaction	The human moving gaze systems partitions limited channel capacity so that it combines high spatial resolution and wide aperture sensing visual environments (Resnickoff, 1987)
• Parallel Perceptual Processing	Some attributes of visualization can be processed in parallel compared to text which is serial
• Offload work from cognitive to perceptual system	Some cognitive inferences done symbolically can be recoded into inferences done with simple perceptual operations (Larkin and Simon, 1987)
• Expanded working memory	Visualizations can expand the working memory available to solve a problem (Norman, 1983)
• Expanded storage of information	Visualizations can be used to store massive amounts of information in a quickly accessible form (example, maps)
Reduced Search	
• Locality of processing	Visualizations group information used together, reducing search (Larkin and Simon, 1987)
• High Density Data	Visualizations can often represent a large amount of data in a small space (Tufte, 1983)
• Spatially Indexed Addressing	By grouping data about an object, visualizations can avoid symbolic labels (Larkin and Simon, 1987)
Enhanced Recognition of Patterns	
• Recognition instead of recall	Recognizing information generated by a visualization is easier than recalling that information by the user.
• Abstraction and aggregation	Visualizations simplify and organize information, supplying higher centers with aggregated forms of information through abstraction and selective omission (Card, Robertson, and Mackinlay, 1991; Resnikoff, 1987)
• Visual schemata for organization	Visually organizing data by structural

	relationships (e.g. by time) enhances patterns.
• Value, relationship, trend	Visualizations can be constructed to enhance patterns at all three levels (Bertin, 1977/1981)
Perceptual Inference	
• Visual representations make some problems obvious	Visualizations can supports a large number of perceptual inferences that are extremely easy for humans (Larkin and Simon, 1987)
• Graphical Computations	Visualizations can enable complex specialized graphical computations (Hutchins, 1996)
Perceptual Monitoring	Visualizations can allow for the monitoring of a large number of potential events if the display is organized so that these stand out by appearance or motion.
Manipulable Medium	Unlike static diagrams, visualizations can allow exploration of a space of parameter values and can amplify user operations.

Table 1: How Information Visualization Amplifies Cognition

Source: Readings in Information visualization – using vision to think [23]

The human aspect of decision making, including information visualization is a vast area of study with many books devoted to each of them. This section was designed to provide a quick introduction to these topics in order to understand some parts of the remainder of this thesis work.

2.4 Decision Quality and Factors Leading to Errors

In studying the process of decision making, a natural question to ask is: what constitutes a good decision? How is the quality of a decision measured? From a classical economics viewpoint, a good decision is one which maximizes utility for all stakeholders in a given situation. An intuitive notion of a good decision for most people is one which achieves the desired outcomes. Howard [12] takes this idea further and defines the quality of a decision as a six dimensional quantity consisting of: 1. problem framing, 2. identifying alternatives, 3. gathering meaningful information, 4. gaining clarity on values and trade-offs, 5. using correct reasoning, and 6. committing to action. Thus, in this framework, a good decision is one which is well balanced on all these six aspects regardless of the eventual outcome of the decision.

A study of the characteristics of good decisions is useful, but a look at erroneous decisions is usually very instructive too. It is commonly accepted that there are two broad categories of errors in decisions – errors of commission and errors of omission. Errors of omission are usually more difficult to detect and are considered more expensive than errors of commission. Here is a quote from Jeff Bezos, CEO of Amazon during an an interview at a Wired magazine's "Disruptive by Design" event in New York city in 2009: "We've made many errors. People over-focus on errors of commission. Companies over-emphasize how expensive failure's going to be. Failure's not that expensive....The big cost that most companies incur are much harder to notice, and those are errors of omission."

By analyzing our adapted decision making model of Figure 3, and knowing elementary systems theory, it is easy to identify the sources of decision errors – both errors of commission and errors of omission. Shown below is a list of situations, adapted from Prof. Leveson's work in [4], which lead to erroneous or poor quality decisions

- If the sensors in our model operate inadequately, they can miss valid signals, or generate incorrect data which can result in poor decisions. Also, a failure to respond to unidentified hazards, or important signals from the environment can affect the business process is usually an expensive mistake.
- Having an incomplete, inconsistent, or incorrect process model in the controller often leads to erroneous decisions. This is true for both automated controllers and their process models as well as humans and their mental models of the process.
- If the actuator action is inadequate, execution the business process is difficult to maintain in a desired state and will definitely result in erroneous outputs. In other words, not knowing the actual state and capability of an actuator is a cause of poor decisions.
- Communication flaws in the decision making process often lead to incorrect decisions or no decisions being made even when one is required.

- Due to feedback loops and information delays in the system, time lags and measurement accuracies which are not accounted for, human and automated controllers can make erroneous decisions based on a their limited and out-of-sync knowledge of state of the business process.

2.5 Defining factors to test for their effects on Decision Quality

The discussions in the previous sections clearly show that the quality of decisions in our adapted model depends on the quality of its components – the sensors, controllers, and the actuators. All these components have to function adequately in the presence of noise as well as signals from the external environments. Additionally, they also have to be aware of the feedback loops in the system and account for information delays when making decisions that affect a business process.

In any team managing a business process, the automated sensors, controllers, and actuators constitute the technical resources of the team. Additionally, useful data analytics can help the team better understand the state and behavior of the process and thus manage it efficiently. We intend to test the contribution of data analytics capabilities to the quality of decisions.

People are also the sensors, controllers, and actuators in this framework. They sense noise and signals in the environment, formulate problems to be addressed, and make decisions to the best of their abilities to keep the business process in its desired state. This collective cognitive capability of teams, which we will call "data sensing" is essential for making better decisions. This "data sensing" capability is very similar to the concept of distributed cognition discussed in the previous sections. We intend to test the contribution of data sensing capabilities to the quality of decisions.

The decision making process also involves a lot of human interaction and discussion, sometimes involving conflicts within team members. The nature of conflict – whether constructive or destructive, and level of conflict within a team can have a big impact on the quality of decisions. A lack of conflict means that team members do not disagree with each other either because of fear or because they all think alike (the group-think phenomenon). This lack of conflict is also not a good situation for teams as it prevents them from exploring creative solutions to current or future problems. We intend to test the contribution of the conflict level within a team to the quality of its decisions.

The structure of the team, especially the power distribution of who makes decisions can also determine the nature and level of interactions between individual members of the team and the subsequent quality of decisions. We intend to test the effects of power distribution between team members to the quality of decisions made by the team.

To list it more explicitly, we intend to test the contribution of the following four factors to the quality

Data Analytics: This factor measures the technical skills, capabilities, and resources to collect, store, and process large amounts of data.

Conflict Level: This factor measures the amount of conflict in a team. This conflict could be either constructive or destructive in nature.

Power Distribution: This factor measures the power structure of the team. Teams with a more distributed power structure have a more democratic decision making process than those with a lesser one.

Data Sensing: This factor is a measure the contextual awareness of the team as a whole, their ability to sense the signals and noise in their environment, and to respond appropriately. Note that this capability can possibly be affected by the level of conflict and the power distribution within the team and we will be testing for these interactions between the capabilities as well.

One important human factor that we have not listed here is the motivation and incentives of team participants. A famous quote attributed to David Hume is "reason alone can never produce any action, or give rise to volition". We believe that motivation and incentives are "gating factor", in the sense that without the right motivation and incentive, no amount of decision making has any meaning. Hence we assume the requisite level of motivation amongst team members and do not explicitly include it in our factors to test. Additionally, since we are focused on the quality of decisions, and not the quality of execution, we will not test for the execution capabilities – even though they are extremely critical in determining the outcome of decisions.

2.6 Our Hypotheses about Factors

Based on our experiences, we believe that the contribution of human factors in decision making has increased, not decreased in today's environment – in part due to increasing complexity of today's businesses. The central argument of this thesis is that human factors play a very dominant role in determining the outcome of decisions. The specific hypotheses listed below intend to compare the contributions of human factors involved in the decision process with that of the technical factors.

We first start with stating the null hypothesis.

Null Hypothesis: Data Analytics, Data Sensing, Conflict Level, and Power Distribution have no effect on the quality of decisions made by a team.

If this null hypothesis is false, it means that at least one of the factors contributes significantly to the quality of decisions and we can then test for the relative contribution of these four factors via the following hypotheses:

1. ***Sensor Quality Hypothesis: The data sensing capabilities of a team contribute more to the quality of decisions than the data analytics capabilities of the team. Moreover, the variability in quality of decisions decreases as the data sensing capabilities of a team increase.***
2. ***Conflict Level Hypothesis: The conflict level within a team contributes more to the quality of decisions than the data analytics capabilities of the team. Moreover, the variability in quality of decisions decreases as the conflict level of a team becomes increasingly constructive.***
3. ***Power Distribution Hypothesis: The power distribution within a team contributes more to the quality of decisions than the data analytics capabilities of the team. Moreover, the variability in quality of decisions decreases as the power distribution within a team increases.***
4. ***Decision Speed Hypothesis: The speed of decision making within a team contributes more to the quality of decisions than the data analytics capabilities of the team.***

Of these four hypotheses, we will look for quantitative validation of the first three and a qualitative validation of the last one. The next chapter discusses the details of testing the validity of these hypotheses.

3.0 Survey and Results

"An ounce of information is worth a pound of data; an ounce of knowledge is worth a pound of information, an ounce of understanding is worth a pound of knowledge"

A popular aphorism

In this chapter, we discuss the design of our survey and the data analysis for testing the validity of the hypotheses presented in the previous chapter.

3.1 Survey Design - An Experimental Design Approach

To test our hypotheses, we formulated an experimental design setup in which the four factors identified in the previous chapter constitute independent variables and the "quality of decisions" is the dependent variable. The experiments we conducted were "thought experiments" (aka. "gedanken experiments") in the form of a survey. In these experiments, we set two levels to test for each of the four factors as follows:

1. Power Distribution within the team: This was measured on a scale of [1,10] where "1" denoted decision making power concentrated in the hands of one individual while "10" denoted a collaborative decision making setup. In our experiments, we tested for two levels of this variable: "2" and "6".
2. Level of Conflict within the team: This was measured on a scale of [-5,+5] where "-5" denoted a high level of destructive conflict within the team and "+5" denoted a high level of constructive conflict within the team. In our experiments, we tested for two levels of this variable: "-3" and "+1".
3. Data Sensing capabilities of the team: This was measured on a scale of [1,10] where "1" denoted poor data sensing and 10 denoted high data sensing. In our experiments, we tested for two levels of this variable: "4" and "8".
4. Data Analytics capabilities of the team: This was measured on a scale of [1,10] where "1" denoted poor data analytics and 10 denoted high data analytics. In our experiments, we tested for two levels of this variable: "5" and "9".

We preferred using an experimental design setup instead of collecting a large number of data points and performing an OLS regression analysis on this collected data because of two reasons:

1. An experimental design setup allows us a more detailed analysis of the effects and interactions of each of the factors to test in our experiments, including determination of the contribution of each factor to the response variable.
2. In addition to checking for the contribution of factors and their interactions to the quality of decisions, we can also calculate the signal-to-noise ratio (explained later) and analyze it which provides us insights into how the factors or their interactions contribute to the variability in quality of decisions.

The survey had a total of sixteen questions with each question corresponding to one "experiment" or "treatment". The level of factors in each question was chosen according to an L16 orthogonal array design. The questions on the survey were designed as a paired comparison test. In each question, two teams, Team "A" and Team "B", competed in an investment strategies game and the survey taker had to make a decision about which team s/he thought would win the game, and by what margin on a scale of 1-5 where "1" denoted very little chance of winning and "5" denoted a very high chance of winning. In all questions, Team A always had the same setting for the four variables while we varied the settings for Team B according to the L16 orthogonal array design. A sample survey question is shown below while the entire survey is shown in Appendix B.

On a scale of 1-5(1 = Slightly higher possibility of winning; 5 = Very high possibility of winning), which of the following two teams is likely to win?

	Team A	**Team B**
Power Distribution within the Team (Team Organization: 1=Concentrated Power; 10 = Distributed Power)	4	6
Level of Conflict within the Team (Team Interaction: -5=Destructive Conflict;+5=Constructive Conflict)	-1	+1
Data Sensing Capability of the Team (Non Technical Skills: 1 = Poor; 10 = Excellent)	6	8
Data Analytics Capability of the Team (Hard Technical Skills: 1 = Poor; 10 = Excellent)	7	9
Which Team Will Win? (1-5)	Winning Team: A B Scale: 1 2 3 4 5 1=Slightly higher possibility of winning. 5=Very high possibility of winning	

Table 2: A Sample Question from the Survey

An interesting aspect of this survey setup is the use of paired comparison tests in each question. Since this survey deals with factors which can have varying subjective interpretations, we believe that a paired comparison design, as shown above, in which the respondents had to make a clear choice between two teams, one of which (Team A) is kept invariant while the other is varied would provide more reliable answers than an open ended (or non-paired comparison) "Rate the Chances of Winning" kind of question.

3.1.1 Survey Responses

The survey was answered by 30 respondents, all from the author's professional network. Most of these survey takers are also enrolled in the SDM program at MIT. All survey takers had a minimum of five years of work experience and come from diverse industries such as banking, engineering, military, consulting, and education and most of them have been in positions of team leadership. The survey takers age range was from the low 30s to 60s and four out of the thirty survey takers were female. The sequence of questions on the survey handed out to participants was randomized to prevent the chance of the memory of a previous question affecting answers to the next question. Survey participants were also instructed to answer each question independently of the previous questions. The response to a question consisted of two parts: a decision about which team will win (A or B), and a decision on the scale of winning (1 = Slightly better chance of winning; 5 = very high chance of winning). The answers were then converted into a single number between 1 to 10 which denoted the chance of team B's winning. Using this conversion scheme, an answer of (A,5) which signifies a very high chance of Team "A" winning the game was converted to (B,0) - a very low chance of team B winning the game. Similarly, (B,5) was converted (B,10). Table 3 below shows the correspondence between survey answers and the values used for our analysis.

Survey Answer	Value Used For Analysis (= Team B's chance of winning on a 1-10 scale)
(A,5)	0
(A,4)	1
(A,3)	2
(A,2)	3
(A,1)	4
(B,1)	6
(B,2)	7
(B,3)	8
(B,4)	9
(B,5)	10

Table 3: Correspondence of survey answers to values used for analysis

In these experiments, we have made a simplifying assumption and equated the quality of decisions made by a team with its chance of winning the game. With this assumption, the responses from survey takers are a proxy measure of team B's quality of decisions.

3.1.2 Data Preparation for Analysis

After data was collected from all survey participants, we calculated the mean and the signal-to-noise ratio (defined as mean/stdev of responses and abbreviated to SNR) of all 30 responses for each of the sixteen settings. The structure of how the data was arranged for analysis is shown in Table 4 below. In this table, Factor A is "Power Distribution", Factor B is "Conflict Level", Factor C is "Data Sensing", and Factor D is "Data Analytics". Each row in this table corresponds to one "treatment" or "experiment" for Team B. Each column contains the response from a single survey taker and represents one "run" for that corresponding treatment. The "mean" and "SNR" values across all 30 runs for each row are the two response variables which we used for our analysis. The mean value is a measure of centrality of user responses while SNR provided us an idea of how consistent user responses were for each treatment.

#	Factor A Level	Factor B Level	Factor C Level	Factor D Level	Run #1	Run #2	...	Run #30	Mean (Avg. of 30 responses)	SNR (Mean / STDEV)
1	2	-3	4	9						
2	6	-3	4	5						
3	6	-3	4	9						
4	2	-3	4	5						
5	2	1	4	5						
6	6	1	4	5						
7	2	1	4	9						
8	6	1	4	9						
9	2	-3	8	5						
10	6	-3	8	5						
11	2	-3	8	9						
12	6	-3	8	9						
13	6	1	8	9						
14	2	1	8	5						
15	6	1	8	5						
16	2	1	8	9						

Table 4: Tabular Representation of survey data

3.1.3 Repeatability and Reproducibility in our experiments

It is useful to discuss the R&R – Repeatability and Reproducibility - of our experiments before we delve into a detailed analysis of the results. Reproducibility is a measure of an experiment to produce similar or identical results when performed independently by someone else. This survey was conducted amongst 30 individuals, and the signal-to-noise ratio (SNR) – a measure of the fidelity (i.e. lack of variability) of the response - varies between 1.3 and 6.8. The SNR would have been higher if the numeric response variable was a continuous one rather than a step one as on our survey. A SNR of more than "1" tells us that there is more agreement than disagreement of responses between individual survey takers and we believe this is an acceptable measure of reproducibility for the "thought experiments" on the survey.

Repeatability in experiments is a measure of agreement between experiments done by the same person under the same settings. Given the hypothetical nature of questions on our survey, some survey takers were indeed curious about the repeatability of their answers to the survey. A couple of survey participants even mentioned that their answers would be “wildly different” if they took the survey another time. To check for the validity of such a claim, we randomly chose one such skeptical participant, and without any advance notice, asked them to retake the survey two days after the original survey was taken. The answers from the two surveys are shown below.

	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	10	11	12	13	14	15	16
#1	3	1	3	1	2	4	6	7	2	3	6	3	9	4	6	7
#2	4	3	3	0	3	4	4	8	3	4	7	4	10	4	4	7

Table 5: Survey results for checking repeatability of answers

Of the sixteen answers by this respondents, four answers were identical, nine varied by one point – a relatively minor variation given the step nature of the responses, and three varied by two points. None of the answers varied by more than 2 points. When shown these results, the survey taker was surprised at the similarity of answers and responded “This is so weird”. Like this skeptical survey taker, we believe that most survey takers had well defined mental models of what factors contributed to better decisions. We believe that the potential variations in their answers, if they were to retake the survey, would have been relatively minor – thus giving us confidence that there is an acceptable amount of repeatability in our “thought experiments”.

3.2 Survey Data Analysis

From the 30 responses to each question we calculated the mean and SNR of these responses and performed a standard OLS regression and ANOVA analysis which is discussed below. We first discuss the analysis of the mean value followed by that of the SNR value.

3.3 Response Mean Analysis

3.3.1 OLS Regression Model for the Data

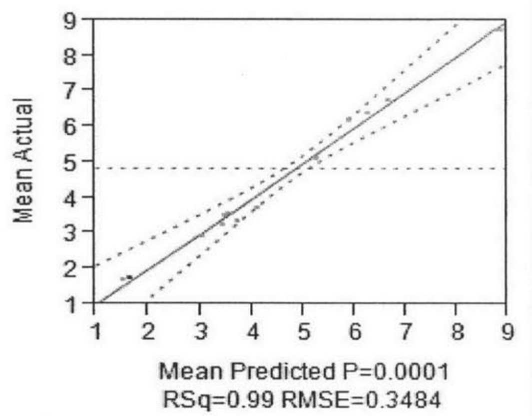

Figure 6: OLS Regression - Predicted Vs. Actual

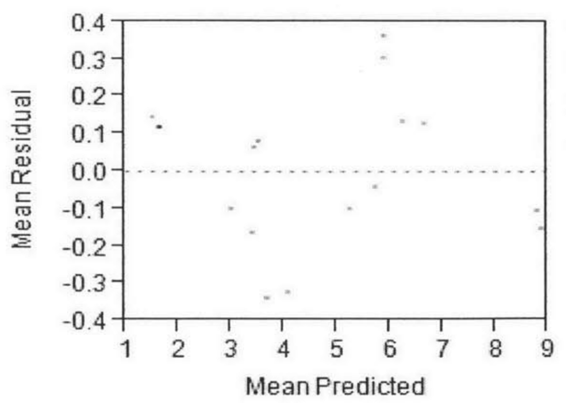

Figure 7: OLS Regression - Residual Vs. Predicted

RSquare	0.991771
Rsquare Adj	0.975313
Root Mean Square Error	0.348421
Mean of Response	4.844758
Observations (30 samples per observation)	16

Table 6: OLS Regression - Summary of Fit

The figures and table above show the Ordinary Least Squares (OLS) Regression fit for the mean of the data collected from the survey. A few points to note in the OLS fit:

1. A high RSquare value of 0.99 indicates a good fit of OLS to the data.
2. The low difference between RSquare and adjusted RSquare indicates that all explanatory variables contributed to the output.
3. The RMSE is low suggesting a low level of error in responses.
4. The random nature of the residual plot indicates a lack of heteroscedasticity, thus allowing us to analyze the data using ANOVA.

3.3.2 ANOVA Summary

Source	DF	Sum of Squares	Mean Square	F Ratio	Prob > F
Model	10	73.154136	7.31541	60.2603	0.0001*
Error	5	0.606985	0.12140		
C. Total	15	73.761121			

Table 7: ANOVA Summary for mean value

The table above shows the summary results of ANOVA for the survey data. The high F-Ratio and correspondingly low "Prob > F" value (of much less than 0.05, the threshold value for acceptance) indicates that the null hypothesis is not true and that at least one of the factors contributes significantly to the response. In the next section, we will explore these results in more detail and determine the relative contribution of each factor to the response.

3.3.3 Effects Test

Source	DF	Sum Of Squares	Mean Square	F Ratio	Prob > F
Power Distribution	1	0.000065	0.000065	0.0005	0.9824
Conflict Level	1	29.984456	29.984456	246.9951	< 0.0001*
Data Sensing	1	25.567703	25.567703	210.6123	< 0.0001*
Data Analytics	1	16.194134	16.194134	133.3982	< 0.0001*
Power Distribution* Conflict Level	1	0.306638	0.306638	2.5506	0.1711
Power Distribution* Data Sensing	1	0.063500	0.063500	0.5148	0.5052
Power Distribution* Data Analytics	1	0.226392	0.226392	1.8649	0.2303
Conflict Level * Data Sensing	1	0.492261	0.492261	4.0550	0.1002
Conflict Level * Data Analytics	1	0.196735	0.196735	1.6206	0.2590
Data Sensing * Data Analytics	1	0.120252	0.120252	0.9906	0.3653

Table 8: Effects Test for mean values

Table 8 above shows the results of F-test to determine the contribution level of each of the four factors and their interactions. From this table data, we pool the statistically insignificant factors and interactions (those with "Prob > F" value of more than 0.05 – i.e.

the factors outside a 95% confidence interval) into an "error pool" and determine the relative contribution of each factor as shown in Table 9 below.

	DF	SS	F-Ratio	Prob > F	CONTRIB%
Conflict Level	1	29.98	246.99	< 0.0001	40.65%
Data Sensing	1	25.56	210.61	< 0.0001	34.65%
Data Analytics	1	16.19	133.39	< 0.0001	21.95%
Pooled Error	12	2.03			2.75%
Total	15	73.36			100%

Table 9: Factor Contribution Calculation

As seen from the above table, "Conflict Level" factor contributes about 41%, "Data Sensing" contributes about 35%, "Data Analytics" contributes about 22%, while error terms contribute about 3% to the mean response which is an average measure of the chance of team B winning the game. Since we consider a higher chance of winning as a proxy for better decisions, these relative values in Table 9 show the contributions of these factors to the quality of decisions.

3.4 Response SNR Analysis

In the previous section, we analyzed of the mean value user responses for each question on the survey and determined the relative contribution of each factor in determining the mean or average value of the user responses. In this section, we perform a similar OLS and ANOVA analysis for the signal-to-noise ratio (defined as mean/stdev, and abbreviated to SNR) of the responses. The SNR is a measure of dispersion (or lack of it) in the user responses and the analysis provides us a deeper insight into which factors and interactions are more reliable in determining the quality of decisions.

3.4.1 OLS Regression Model of the Data

Similar to the analysis for mean values, Table 10 and Figure 8 show an OLS regression fit for the SNR values of survey results. The Rsquare and Adjusted Rsquare are high, although the bigger difference between Rsq and Adj. Rsq denotes that not all explanatory variables may have contributed significantly to the response.

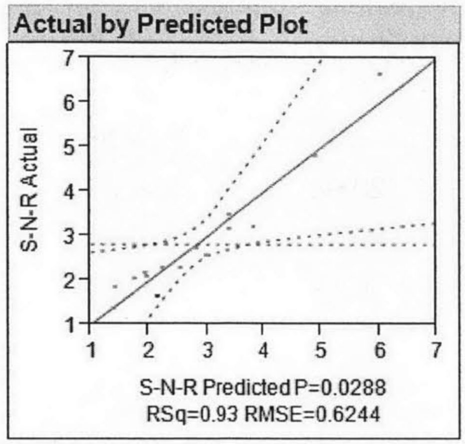

Figure 8: OLS Regression - Actual Vs. Predicted

RSquare	0.925331
Rsquare Adj	0.775993
Root Mean Square Error	0.624395
Mean of Response	2.8104
Observations (30 samples per observations)	16

Table 10: OLS Regression - Summary of Fit

3.4.2 ANOVA Summary

Source	DF	Sum of Squares	Mean Square	F Ratio	Prob > F
Model	10	24.157143	2.41571	6.1962	0.0288*
Error	5	1.949345	0.38987		
C. Total	15	26.106489			

Table 11: ANOVA Summary

Table 11 above shows a summary of ANOVA for the SNR values. The high "F Ratio" and low "Prof > F" ratio in ANOVA indicates that the "null hypothesis" for this case is false . In other words, at least one factor has a significant contribution in determining the consistency of the quality of decisions. In the next section we explore this in more detail and determine the relative contribution of each factor to the consistency of outcomes.

3.4.3 Effects Test

Source	DF	Sum Of Squares	F Ratio	Prob > F
Power Distribution	1	0.1212799	0.3111	0.6011
Conflict Level	1	5.8179743	14.9229	0.0118*
Data Sensing	1	6.9218811	17.7544	0.0084*
Data Analytics	1	5.5591309	14.2590	0.0129*
Power Distribution* Conflict Level	1	0.0539089	0.1383	0.7252
Power Distribution* Data Sensing	1	0.8753514	2.2452	0.1943
Power Distribution* Data Analytics	1	0.4691653	1.2034	0.3226
Conflict Level * Data Sensing	1	3.4413002	8.8268	0.0311*
Conflict Level * Data Analytics	1	0.4904713	1.2580	0.3130
Data Sensing * Data Analytics	1	0.4066827	1.0431	0.3539

Table 12: Effects Tests

The F-test for effect of factors shown in Table 12 above indicate that the three factors "Conflict Level", "Data Sensing", and "Data Analytics", as well as the interaction between factors "Conflict Level" and "Data Sensing" have a significant contribution in determining the SNR values in these experiments while "Power Distribution" did not have a meaningful contribution to the outcome. As we did for the mean analysis in the previous section, we did a more detailed analysis by pooling together non-significant factors into an "error pool" as shown in the table below.

	DF	Sum of Squares	F-Ratio	Prob > F	CONTRIB%
Conflict Level	1	5.81	14.92	0.0118	22.26%
Data Sensing	1	6.92	17.75	0.0084	26.51%
Data Analytics	1	5.55	14.25	0.0129	21.26%
Conflict Level * Data Sensing	1	3.44	8.82	0.0311	13.18%
Pooled Error	11				16.78%
Total	15	26.10			100%

Table 13: Factor Contribution Calculation

From Table 13 above, it is obvious that while all three factors – Conflict Level, Data Sensing and Data Analysis have almost similar contributions in determining the SNR value, "Data Sensing" has a slightly more contribution than the other two factors. Moreover, the interaction between "Data Sensing" and "Conflict Level" factors also has a statistically significant contribution to the outcome. This contribution by the interaction between "Data Sensing" and "Conflict Level" is better understood through a visualization shown in Figure 9. In this figure, if any two lines are parallel, or nearly parallel, it means that the interaction between the corresponding two factors is negligible. However, in the case of "Data Sensing" and "Conflict Level", when the conflict level is held at "-3" level, an increase in data sensing capabilities from "4" to "8" does not affect the SNR values much. But in the presence of constructive conflict, when the conflict is positive and held at "+1" while data sensing is changed from "4" to "8", the SNR value jumps significantly indicating that the interaction between these two factors has a significant effect on the SNR value.

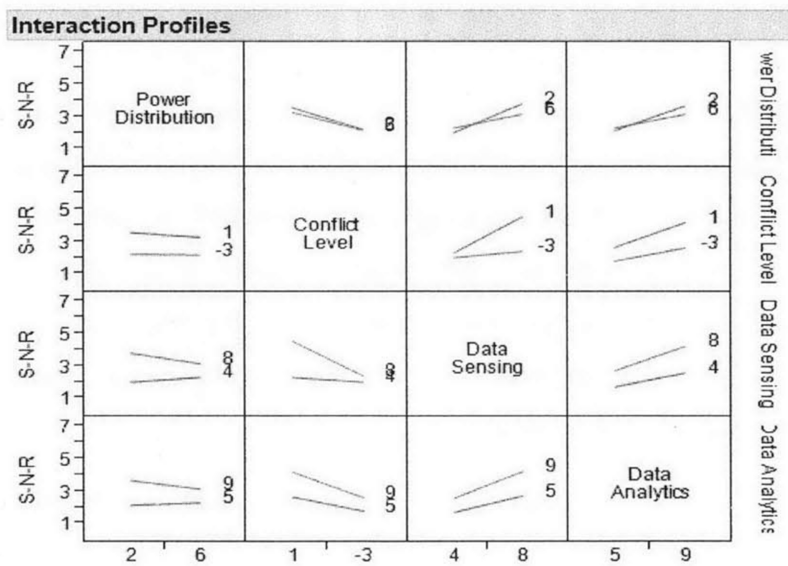

Figure 9: Factor Interaction Visualization

3.5 Hypotheses Validation and Discussion

The ANOVA analysis and the F-Tests show that the null hypothesis is not valid. We now discuss the validity of each of the four subsequent hypotheses to test the relative contribution of each factor.

Sensor Quality Hypothesis: The data sensing capabilities of a team contribute more to the quality of decisions than the data analytics capabilities of the team. Moreover, the variability in quality of decisions decreases as the data sensing capabilities of a team increase.

The effects tests for the mean response show that the F-Ratio for data sensing factor is 211 while that of the data analytics is about 133. After pooling the error, we see that the contribution of data sensing to the mean response is about 35% while that of data analytics is about 22%. A similar analysis of SNR shows a 27% contribution of data sensing factor and a 21% contribution of data analytics in determining the SNR. Both these results quantitatively validate the sensor quality hypothesis.

In addition to quantitative validation, it is also important to understand why this hypothesis is valid. Ashby's Law of Requisite Variety, stated informally as "variety absorbs variety", or "variety can destroy variety" implies that for a model to effectively control a process, it must have at least as many states as the process itself. In a dynamic process such as a business process in which the number of variables which affect it, and the subsequent large state-space, changes continually, the automated controllers can model only a part of the business process. These analytical models must be supplemented with human understanding and knowledge of the business process. A better awareness of reality through better sensing capabilities results in more accurate models – both mental models and analytical models – and better decisions.

Conflict Level Hypothesis: The conflict level within a team contributes more to the quality of decisions than the data analytics capabilities of the team. Moreover, the variability in quality of decisions decreases as the conflict level of a team becomes increasingly constructive.

The effects tests for the mean response shows that the F-Ratio for conflict level factor is 246 while that of the data analytics is about 133. After pooling the error, we see that the contribution of conflict level to the mean response is about 41% while that of data analytics is about 22%. A similar analysis of SNR shows a 22% contribution of conflict level factor and a 21% contribution of data analytics in determining the SNR. Both these results quantitatively validate the conflict level hypothesis.

This result is something that most professionals know intuitively and the survey results numerically highlight the enormous effect that conflict level has on the quality of team decisions. In fact, of the four factors, the conflict level factor has the greatest contribution in determining the mean response. Our reasoning of this is that destructive conflict in a team is likely to lead to errors of omission which are usually much more expensive than errors of commission. Failures lead people to re-evaluate their assumptions and uncovering of new facts and learnings. Conversely, a lack of failures just validates what people already know and no new learning happens. Thus, errors of omission mean that individual and cumulative learning of the team is reduced which can result in further deterioration of decision quality. On the other hand, in a constructive conflict environment, even if there are errors in decision due to uncertainty in the environment, these errors are likely to be noticed and corrected by the team. The dynamics of constructive conflict also result in better exploration of more alternatives and reduce the possibility of potentially costly errors of omission.

Power Distribution Hypothesis: The power distribution within a team contributes more to the quality of decisions than the data analytics capabilities of the team. Moreover, the variability in quality of decisions decreases as the power distribution within a team increases.

While it is commonly believed that empowering employees is a good business practice and results in better decisions within enterprises, our survey data indicates that there is no evidence that power distribution has any meaningful effect at the team level. There are possibly several explanations for this anomaly:

1. The way the survey question was framed, it is quite possible that survey takers did not take into account the possible implications of power structures on

individual incentives, flow of information, and the subsequent rewards or gains for efforts. These dynamics do come into picture in real life and may not have been considered by survey takers in formulating their answers.

2. It is possible that power distribution is a significant factor only when the level and frequency of personal interaction between the parties involved is low. In a small team environment, the personal interactions between team members results in a natural and mutually acceptable distribution of power and is not a major hindrance to decision making.
3. Many survey takers suggested in their comments that more concentrated power distribution implies a strong team leader who can mitigate the possible downsides of destructive conflict. So power distribution in team environments seem to matter only in the limited cases of negative conflict and hence it did not figure significantly in the cumulative results.

In any case, this anomaly in our understanding is an area of further investigation.

Decision Speed Hypothesis: The speed of decision making within a team contributes more to the quality of decisions than the data analytics capabilities of the team.

While this hypothesis could not be tested in the survey data, some comments by survey takers seemed to indicate that low power distribution led to faster decision making which equated to better decisions. The possible reasoning is based on the experience of many survey takers who think that some decision, even if it an incorrect one is better than no decision at all. This is equivalent to the previous observation of errors of omission being more expensive than errors of commission.

4.0 Related Research

"If I have seen further it is by standing on the shoulders of giants."

Sir Issac Newton

Restricting the scope of this thesis has been an interesting challenge. Just providing a complete overview of even all the research related to this thesis might take several chapters. We will present a small subset of some of the important work we have encountered during our work. This thesis borrows work from three broad fields of study:

1. Decision Theory
2. Systems Theory and its applications, notably the STAMP model [4] which has influenced our approach.
3. Data Oriented Decision Making in Organizations

Decision theory provides much of the theoretical foundations for the work in this thesis. The literature in this domain is extremely vast. Borrowing from Tang [13], Table 14 presents an overview of the three broad areas of Decision Theory work.

The area of data oriented decision making in organizations has been receiving a lot of attention. Eric Brynjolfsson's work in [2] details the effect of data oriented decision making in organizations, and also provides an overview of some of the major work in this area. Galbriath did pioneering work in this field and proposed the information processing view of organizations [1].

Information visualization goes hand in hand with data processing and decision making. This is another huge and multi-disciplinary area of study and has been receiving tremendous amount of interest due to the easy availability of data as well as improvements in display technologies in the recent past. Stuart Card [23], Robert Spence [20] and Edward Tufte [28] are the three primary sources of our study on this topic.

	Normative	Descriptive	Prescriptive
Focus	How people should decide with logical consistency	How an why people decide the way they do	Help people make better decisions Prepare people to decide
Criterion	Theoretical Adequacy	Experimental Validity	Efficacy and Usefulness
Scope	All Decisions	Classes of Decisions Tested	Specific Decisions for Specific Problems
Theoretical Foundations	Utility Theory Axioms	Cognitive Sciences Psychology about beliefs and preferences	Normative and Descriptive Theories Decision Analysis Axioms
Operational Focus	Analysis of Alternatives Determining preferences	Preventions of systematic human errors in decision making and inference	Processes and Procedures End-to-End decision lifecycle
Judges	Theoreticians	Experimental Researchers	Applied Analysts

Table 14: Overview of Decision Theory Work

Source: Tang, Victor. Corporate decision analysis: an engineering approach. Diss. Massachusetts Institute of Technology, 2006.

On the systems side, the early pioneers of what is now Systems Theory included Ludwig von Bertalanffy, Jay Forrester, Norbert Wiener, William Ross Ashby, amongst others. It is an inter-disciplinary field with roots in control systems, electrical and network theory but today has affected almost all scientific fields. Seminal work on the architecture of complexity of systems was done by Herbert Simon and is presented in [25]. The work which has influenced this thesis is the STAMP framework [4] for system safety by Nancy Leveson. While our thesis is not related to systems safety, Prof.

Leveson's work has influenced the choice of a systems theory based framework for understanding and reasoning decision-making in teams.

In this thesis, we have equated a favorable outcome with a good decision. However, the quality of decisions is more a more complex property than that. Howard [12] in their paper describe decision quality as a six dimensional quantity consisting of problem framing, exploration of alternatives, and commitment to action, amongst others. They also deals with the process of decision making in teams. They define the typical roles of members in a team – decision makers, decision staff, and implementors.

At a more individual level, the decision making styles of individuals in organizations is studied by [6]. In their interesting study on entrepreneurs versus managers, they notice how these two different sets of people user different heuristics and have different biases in decision making. Herbert Simon has also made important contributions to this area. [27] provides some keen insights into rational decision making in business organizations.

Finally, the idea of distributed cognition proposed by James Hollan, Edwin Hutchins, and David Kirsh [11] has been very useful for our in work on understanding the cognitive capabilities of teams.

5.0 Summary and Future Work

"If Martin Luther King had said 'I have some data' instead of 'I have a dream' do you think the speech would have been just as effective?"

Prof. Ralph Katz, in a lecture to SDM students at MIT

In this thesis, we adapted a systems theory influenced framework to analyze the relative contribution of technical and human factors on the quality of team decision making. Our central argument about the over-sized effect of human factors on the outcomes was validated, and we provided more insights into the relative contributions of four socio-technical factors viz, data analytics, data sensing, conflict level, and power distribution to the quality of decisions made by teams.

We have only scratched the surface of this field. The use of a control-systems approach to further investigate the variables and dynamics in decision making is an interesting area of further study in this area. Since many human and environmental factors contribute to decision, it will be an interesting study to investigate the contribution of these factors to the quality of decisions.

Our "thought experiments" in the survey were a useful way of collecting data analyzing data. However, the data collected in our survey showed a fair amount of variation. The sources of variation could be either in the way people interpreted the questions, or in their mental models about decision making. These variations from differing sources were not captured in our work and can be an exciting area of work which could lead to better insights in this field.

Another assumption in our study has been about the relative equivalence in changing or varying all four factors. In other words, there is an implicit assumption that the cost of improving the data analytics capability of a team by "one unit" is the same as the cost of increasing all other factors by their corresponding one unit. This is not true and more research for better proxies of these quantities, as well as approximate costs to change them can to be done. If that is possible, managers across enterprises can have more quantitative models and actionable information on improving the quality of decisions in their teams and organizations.

One surprising result we found in our work is that the power distribution within a team has no meaningful impact on its quality of decisions. This result is contrary to our expectation and can definitely be investigated further.

The question of how representative our selected survey takers are of a team in real world is an important one and has not been addressed. Real life case studies which either validate or invalidate any of these findings, possibly in different organizational cultures, will be a great learning opportunity and provide insights into specific problems to be addressed in future works.

Appendix A

"Big Data" is a term loosely used to denote large data sets and the technologies used to process the data. Since this is an emerging area, there is no uniformly accepted definition for the term. Shown below is a collection of samples of definitions of Big Data from Wikipedia [21] :

"Big data usually includes data sets with sizes beyond the ability of commonly-used software tools to capture, manage, and process the data within a tolerable elapsed time. Big data sizes are a constantly moving target, as of 2012 ranging from a few dozen terabytes to many petabytes of data in a single data set. With this difficulty, a new platform of "big data" tools has arisen to handle sense making over large quantities of data, as in the Apache Hadoop Big Data Platform."

"MIKE 2.0 an open approach to Information Management, defines big data in terms of useful permutations, complexity, and difficulty to delete individual records."

"In a 2001 research report and related lectures, META Group (now Gartner) analyst Doug Laney defined data growth challenges and opportunities as being three-dimensional, i.e. increasing volume (amount of data), velocity (speed of data in and out), and variety (range of data types and sources). Gartner, and now much of the industry, continue to use this "3Vs" model for describing big data. In 2012, Gartner updated its definition as follows: "Big Data are high-volume, high-velocity, and/or high-variety information assets that require new forms of processing to enable enhanced decision making, insight discovery and process optimization."In 2012 Courtney Lambert extended the Gartner definition to "4Vs" by adding the term 'virtual', thus scoping the discussion to only include online assets."

Big Data is used for any large data set and its analyses, including web logs, social media updates, machine generated sensor data, data from geo-exploration, data generated during drug discovery, genome data, amongst others. To get a better understanding how these large data sets are typically used by businesses, we looked at several use cases and found that the following three broad categories adequately describe most uses of Big Data:

1. **Exploratory Analyses** for scientific and strategic fact finding
2. **Operational Efficiency Improvements** by businesses
3. **End User Applications** which are now possible with Big Data

Exploratory Analyses

Many scientific efforts across various disciplines like life sciences, astronomy, physical sciences, geological sciences etc. today generate a large amount of data. A fair amount of effort is being put into investigating the data, its structure, and the behaviors evident in it with the aim of making new scientific discoveries. In the life sciences, genome data exploration or new drug discovery usually involve processing of large data sets. In the physical sciences, the Large Hadron Collider at Geneva and SETI@home are good examples of scientific experiments which resulted in large data sets amenable to exploratory analyses. Many businesses also possess large data sets which they would like to mine to extract better insights. A business problem like finding optimal transportation hubs in a city for a bike sharing service, based on the bike ride usage history of customers is a good example of an exploratory application [22].

The data sets in these exploratory analyses are marked by a high level of complexity in processing, high level of human involvement in data analysis and validation (low automation), and the cost of incorrect decisions based on these exploratory analyses can be pretty high.

Operational Efficiency Improvements

Decisions and actions which were previously labor intensive and required human intervention are now regularly being made by machines, whenever sufficient data is available. With data available for cheap, many businesses invest money in applications which process large data sets, provide fairly reliable answers to business questions and improve operational efficiencies. Examples of these include better recommendations for movies on Netflix, better recommendations on Amazon.com, fraud detection at credit card companies, automated surgeries, better risk management at insurance companies, and even automated airplanes, and so on. The relative complexity of the decision in these cases is moderate and the cost of incorrect decisions can be pretty high.

End User Applications

This abundance of data has resulted in completely new applications for individuals, both as consumers and as business users. Pandora – a music recommendation service, email spam filtering, "social" based recommendations, better calendaring, new applications in the travel space, better time scheduling, are just a few examples of these new applications. This is currently very high interest in this area and new innovations keep coming up every day. The complexity of the decision is not very high, the level of automation is very high, and the cost of a single incorrect decision is typically very low.

The Big Data Ecosystem

Matt Truck [29] shows a pictorial representation of some of the current players in the Big Data Ecosystem. Companies and enterprises operating in this space span the entire technology stack – from hardware infrastructure to systems software to data analytics and visualization tools to industry specific applications - and this set of companies is continuously evolving. Listed below is a broad categorization of companies involved with big data:

1. Hardware Infrastructure Manufacturers: This includes enterprises which manufacture the servers, storage arrays and devices, and the network equipment manufacturers which transmit this data.
2. Software Infrastructure Developers: These companies provide the software programs which can store and search large amounts of data. The Free and Open Source Software (FOSS) Community has made a very large contribution to innovation and new products in this area.
3. Analytics Developers: These companies provide the analytic and algorithmic expertise needed for data processing. Even in this area, the FOSS community has a significant contribution via valuable software libraries and products.
4. End Users: The end-users or clients of Big Data are the companies and people who utilize this data to make meaningful decisions for their enterprises.
5. Data Providers: These companies exist to collect, clean and provide data to others who might need it.

Appendix B

Shown below is the survey which was used to collect data for this thesis.

Introduction

The objective of this survey is to determine the relative importance of four socio-technical factors (Data Analysis Skills, Level of Team Conflict, Data Sensing Skills, and Amount of Power Distribution) on the quality of decision making within a team.

To answer questions in this survey, assume the following scenario:

"Team A" and "Team B" are cross functional teams (consisting of an industry expert, two quantitative analysts, a financial analyst, and a manager/team leader) of industry professionals competing against each other in an "Investment Strategy" game. The objective of the game is to make better investment decisions and make more money in order to win.

Game Overview

1) At the beginning of the game, each team is given 1M USD to invest in stocks of 25 companies from one specific industry sector.
2) Teams are allowed to change their investment portfolio as often as they wish.
3) Teams are provided extensive information on the performance of their investments and information and on market conditions all throughout the game.
4) The game is played over one entire year. The team with more money at the end of the year, wins the game.

Team Members: Both teams are of identical size (5) and are cross functional consisting of an industry expert, two quantitative analysts, a financial analyst, and a manager/team leader.

Investments: Teams may invest only in stocks of a pre-selected list of 25 companies provided to them at the beginning of the game.

Information Availability:

- Both teams have access to identical information during the game.
- Historical information on the performance of individual stocks is available.
- Both teams get extensive information on the performance of the market as well as performance of the 25 companies they can invest in. This new information is both quantitative (sales figures, trading volumes, balance sheets, market share etc.) and qualitative info about market sentiment, new opportunities, new product launches, expert opinions etc.
- Teams are free to use existing investment models to analyze data, modify them, or build new ones if they consider it necessary in order to win.

Winners and Losers: The winner is the team with more money at the end of the year. The winning team keeps all money they have at the end of the game. The losing team gets nothing.

Cross Team Communication: During the game, each team can view the other team's investment performance (but **not** each other's investment strategies). No other cross team communication is permitted during the entire game.

Each question in this survey has the following four "variables":

1. Data Analysis Skills on the Team (Hard Technical Skills): This variable measures the technical sophistication of the team in terms of knowledge of statistics, data analysis algorithms and knowledge of analytical software tools to process possibly large quantities of data and extract useful information from raw data. This skill is rated on a scale of 1-10, where:

- "1" is poor data analytics skills, models, and infrastructure and
- "10" is outstanding data analytics skills, models, and data processing capabilities.

2. Data Sensing Skill (Non-Technical/Cognitive Skill): This variable measures the cognitive ability of individual(s) on the team to "connect the dots", "separate signal from noise", form hypotheses to understand situations and to simplify complex situations into manageable problems. This skill is rated on a scale 1-10, where:

- "1" is poor understanding of the industry dynamics and an inability to comprehend the relative importance of new data and use it effectively for analysis.
- "10" is an excellent (but not fail-proof) ability to understand trends from raw data (either numeric data or qualitative information) before anyone else and to decide on an appropriate plan of action.

3. Level of Conflict within the Team (Team Interaction): This variable measures the amount of conflict in the team and is rated on a scale of (-5, 5);

- "-5" indicates a destructive conflict team environment often involving personality clashes within team members, and the possibility of a dysfunctional team.
- "+5" indicates a constructive conflict team environment where differing opinions are encouraged and explored without the conflict getting personal between team members.
- "0" implies an absence of conflict within the team either because everyone thinks alike or because nobody wants to disagree with other team members.

4. Power Distribution within the Team (Team Organization): This variable measures the "power hierarchy" between members of the team and is measured on a scale of 1-10;

- "1" is a team in which one person has absolute/most power and makes all important decisions for the entire team. In this situation, the person in power can make decisions without consulting other team members.
- "10" is a team in which power is equally distributed amongst all team members and no single team member can make a unilateral decision without getting consent of the majority of team members. Decision making requires consultation and inputs from all.

The nature of power distribution within the team can affect the speed of decision making, the flow of information, communication patterns, as well as level of commitment of individual team members within the team.

Sample Survey Question

A sample question from this survey (along with a sample answer) is shown below:

On a scale of 1-5(1 = Slightly higher possibility of winning; 5 = Very high possibility of winning), which of the following two teams is likely to win?

	Team A	Team B
Power Distribution within the Team (Team Organization: 1=Concentrated Power; 10 = Distributed Power)	4	6
Level of Conflict within the Team (Team Interaction: -5=Destructive Conflict;+5=Constructive Conflict)	-1	+1
Data Sensing Capability of the Team (Non Technical Skill: 1 = Poor; 10 = Excellent)	6	8
Data Analytics Capability of the Team (Hard Technical Skill: 1 = Poor; 10 = Excellent)	7	9
Which Team Will Win?	Winning Team: A B Scale: 1 2 3 4 5 1=Slightly higher possibility of winning. 5=Very high possibility of winning	

Reasons For Your Choice of Winning Team (Optional, but recommended):

1) The amount of power distribution of the winning team is more conducive for making better decisions.

2) The level of conflict of the winning team is more conducive for making better decisions.

3) The data sensing skills of the winning team is more conducive for making better decisions.

4) The data analytics skills of the winning team is more conducive for making better decisions.

Other Reasons: (Please Specify):

Team B is better than Team A in all aspects. They have better Data Sensing skills, better data analysis skills and the team environment is also productive. The greater power distribution seems to be leading to a positive team environment. With each employee being more empowered all team members are likely to work together as a team in order to process new information that comes along, consider more investment alternatives, and most likely make better investment decisions than the other team.

Shown below is a slightly more difficult sample question.

On a scale of 1-5(1 = Slightly higher possibility of winning; 5 = Very high possibility of winning), which of the following two teams is likely to win?

	Team A	Team B
Power Distribution within the Team (Team Organization: 1=Concentrated Power; 10 = Distributed Power)	4	2
Level of Conflict within the Team (Team Interaction: -5=Destructive Conflict;+5=Constructive Conflict)	-1	-3
Data Sensing Capability of the Team (Non Technical Skills: 1 = Poor; 10 = Excellent)	6	4
Data Analytics Capability of the Team (Hard Technical Skills: 1 = Poor; 10 = Excellent)	7	9
Which Team Will Win? (1-5)	Winning Team: A B Scale: 1 2 3 4 5 1=Slightly higher possibility of winning. 5=Very high possibility of winning	

Reasons For Your Choice of Winning Team (Optional, but recommended):

Reasons For Your Choice of Winning Team (Please mark one or more):

- The amount of power distribution of the winning team is more conducive for making better decisions.
- The level of conflict of the winning team is more conducive for making better decisions.
- The data sensing skills of the winning team is more conducive for making better decisions.
- The data analytics skills of the winning team is more conducive for making better decisions.

Other Reasons: (Please Specify):
Team B has superior data analysis skills and seems to have one or two people in charge of the team and making decisions. However, the team environment is not productive and the team does not seem to work well together. They may not be able to utilize new information constructively to make optimal decisions.

On the other hand, Team A is not significantly better either. They have better Data Sensing skills, but inferior data processing skills, and the team environment is somewhat less unproductive than team B but with no single person seemingly in charge of the team.

Hence I chose Team B is the likely winner with a very slight advantage of winning.

Survey Begins On Next Page

Survey Question 1/16

On a scale of 1-5(1 = Slightly higher possibility of winning; 5 = Very high possibility of winning), which of the following two teams is likely to win?

	Team A	Team B
Power Distribution within the Team (Team Organization: 1=Concentrated Power; 10 = Distributed Power)	4	6
Level of Conflict within the Team (Team Interaction: -5=Destructive Conflict;+5=Constructive Conflict)	-1	+1
Data Sensing Capability of the Team (Non Technical Skills: 1 = Poor; 10 = Excellent)	6	8
Data Analytics Capability of the Team (Hard Technical Skills: 1 = Poor; 10 = Excellent)	7	9
Which Team Will Win? (1-5)	Winning Team: A B Scale: 1 2 3 4 5 1=Slightly higher possibility of winning. 5=Very high possibility of winning	

Reasons For Your Choice of Winning Team (Please mark one or more):

The amount of power distribution of the winning team is more conducive for making better decisions.

The level of conflict of the winning team is more conducive for making better decisions.

The data sensing skills of the winning team is more conducive for making better decisions.

The data analytics skills of the winning team is more conducive for making better decisions.

Other Reasons: (Please Specify):

(Setting # 13. Ignore this information)

Survey Question 2/16

On a scale of 1-5(1 = Slightly higher possibility of winning; 5 = Very high possibility of winning), which of the following two teams is likely to win?

	Team A	Team B
Power Distribution within the Team (Team Organization: 1=Concentrated Power; 10 = Distributed Power)	4	6
Level of Conflict within the Team (Team Interaction: -5=Destructive Conflict;+5=Constructive Conflict)	-1	-3
Data Sensing Capability of the Team (Non Technical Skills: 1 = Poor; 10 = Excellent)	6	8
Data Analytics Capability of the Team (Hard Technical Skills: 1 = Poor; 10 = Excellent)	7	9
Which Team Will Win? (1-5)	Winning Team: A B Scale: 1 2 3 4 5 1=Slightly higher possibility of winning. 5=Very high possibility of winning	

Reasons For Your Choice of Winning Team (Please mark one or more):

The amount of power distribution of the winning team is more conducive for making better decisions.

The level of conflict of the winning team is more conducive for making better decisions.

The data sensing skills of the winning team is more conducive for making better decisions.

The data analytics skills of the winning team is more conducive for making better decisions.

Other Reasons: (Please Specify):

(Setting # 12. Ignore this information)

Survey Question 3/16

On a scale of 1-5(1 = Slightly higher possibility of winning; 5 = Very high possibility of winning), which of the following two teams is likely to win?

	Team A	Team B
Power Distribution within the Team (Team Organization: 1=Concentrated Power; 10 = Distributed Power)	4	6
Level of Conflict within the Team (Team Interaction: -5=Destructive Conflict;+5=Constructive Conflict)	-1	+1
Data Sensing Capability of the Team (Non Technical Skills: 1 = Poor; 10 = Excellent)	6	4
Data Analytics Capability of the Team (Hard Technical Skills: 1 = Poor; 10 = Excellent)	7	5
Which Team Will Win? (1-5)	Winning Team: A B Scale: 1 2 3 4 5 1=Slightly higher possibility of winning. 5=Very high possibility of winning	

Reasons For Your Choice of Winning Team (Please mark one or more):

The amount of power distribution of the winning team is more conducive for making better decisions.

The level of conflict of the winning team is more conducive for making better decisions.

The data sensing skills of the winning team is more conducive for making better decisions.

The data analytics skills of the winning team is more conducive for making better decisions.

Other Reasons: (Please Specify):

(Setting # 6. Ignore this information)

Survey Question 4/16

On a scale of 1-5(1 = Slightly higher possibility of winning; 5 = Very high possibility of winning), which of the following two teams is likely to win?

	Team A	Team B
Power Distribution within the Team (Team Organization: 1=Concentrated Power; 10 = Distributed Power)	4	2
Level of Conflict within the Team (Team Interaction: -5=Destructive Conflict;+5=Constructive Conflict)	-1	-3
Data Sensing Capability of the Team (Non Technical Skills: 1 = Poor; 10 = Excellent)	6	4
Data Analytics Capability of the Team (Hard Technical Skills: 1 = Poor; 10 = Excellent)	7	9
Which Team Will Win?	Winning Team: A B Scale: 1 2 3 4 5 1=Slightly higher possibility of winning. 5=Very high possibility of winning	

Reasons For Your Choice of Winning Team (Please mark one or more):

The amount of power distribution of the winning team is more conducive for making better decisions.

The level of conflict of the winning team is more conducive for making better decisions.

The data sensing skills of the winning team is more conducive for making better decisions.

The data analytics skills of the winning team is more conducive for making better decisions.

Other Reasons: (Please Specify):

(Setting # 1. Ignore this information)

Survey Question 5/16

On a scale of 1-5(1 = Slightly higher possibility of winning; 5 = Very high possibility of winning), which of the following two teams is likely to win?

	Team A	Team B
Power Distribution within the Team (Team Organization: 1=Concentrated Power; 10 = Distributed Power)	4	6
Level of Conflict within the Team (Team Interaction: -5=Destructive Conflict;+5=Constructive Conflict)	-1	+1
Data Sensing Capability of the Team (Non Technical Skills: 1 = Poor; 10 = Excellent)	6	4
Data Analytics Capability of the Team (Hard Technical Skills: 1 = Poor; 10 = Excellent)	7	9
Which Team Will Win? (1-5)	Winning Team: A B Scale: 1 2 3 4 5 1=Slightly higher possibility of winning. 5=Very high possibility of winning	

Reasons For Your Choice of Winning Team (Please mark one or more):

The amount of power distribution of the winning team is more conducive for making better decisions.

The level of conflict of the winning team is more conducive for making better decisions.

The data sensing skills of the winning team is more conducive for making better decisions.

The data analytics skills of the winning team is more conducive for making better decisions.

Other Reasons: (Please Specify):

(Setting # 8. Ignore this information)

Survey Question 6/16

On a scale of 1-5(1 = Slightly higher possibility of winning; 5 = Very high possibility of winning), which of the following two teams is likely to win?

	Team A	Team B
Power Distribution within the Team (Team Organization: 1=Concentrated Power; 10 = Distributed Power)	4	2
Level of Conflict within the Team (Team Interaction: -5=Destructive Conflict;+5=Constructive Conflict)	-1	+1
Data Sensing Capability of the Team (Non Technical Skills: 1 = Poor; 10 = Excellent)	6	8
Data Analytics Capability of the Team (Hard Technical Skills: 1 = Poor; 10 = Excellent)	7	9
Which Team Will Win? (1-5)	Winning Team: A B Scale: 1 2 3 4 5 1=Slightly higher possibility of winning. 5=Very high possibility of winning	

Reasons For Your Choice of Winning Team (Please mark one or more):

The amount of power distribution of the winning team is more conducive for making better decisions.

The level of conflict of the winning team is more conducive for making better decisions.

The data sensing skills of the winning team is more conducive for making better decisions.

The data analytics skills of the winning team is more conducive for making better decisions.

Other Reasons: (Please Specify):

(Setting # 16. Ignore this information)

Survey Question 7/16

On a scale of 1-5(1 = Slightly higher possibility of winning; 5 = Very high possibility of winning), which of the following two teams is likely to win?

	Team A	Team B
Power Distribution within the Team (Team Organization: 1=Concentrated Power; 10 = Distributed Power)	4	6
Level of Conflict within the Team (Team Interaction: -5=Destructive Conflict;+5=Constructive Conflict)	-1	-3
Data Sensing Capability of the Team (Non Technical Skills: 1 = Poor; 10 = Excellent)	6	8
Data Analytics Capability of the Team (Hard Technical Skills: 1 = Poor; 10 = Excellent)	7	5
Which Team Will Win? (1-5)	Winning Team: A B Scale: 1 2 3 4 5 1=Slightly higher possibility of winning. 5=Very high possibility of winning	

Reasons For Your Choice of Winning Team (Please mark one or more):

The amount of power distribution of the winning team is more conducive for making better decisions.

The level of conflict of the winning team is more conducive for making better decisions.

The data sensing skills of the winning team is more conducive for making better decisions.

The data analytics skills of the winning team is more conducive for making better decisions.

Other Reasons: (Please Specify):

(Setting # 10. Ignore this information)

Survey Question 8/16

On a scale of 1-5(1 = Slightly higher possibility of winning; 5 = Very high possibility of winning), which of the following two teams is likely to win?

	Team A	Team B
Power Distribution within the Team (Team Organization: 1=Concentrated Power; 10 = Distributed Power)	4	2
Level of Conflict within the Team (Team Interaction: -5=Destructive Conflict;+5=Constructive Conflict)	-1	+1
Data Sensing Capability of the Team (Non Technical Skills: 1 = Poor; 10 = Excellent)	6	8
Data Analytics Capability of the Team (Hard Technical Skills: 1 = Poor; 10 = Excellent)	7	5
Which Team Will Win? (1-5)	Winning Team: A B Scale: 1 2 3 4 5 1=Slightly higher possibility of winning. 5=Very high possibility of winning	

Reasons For Your Choice of Winning Team (Please mark one or more):

The amount of power distribution of the winning team is more conducive for making better decisions.

The level of conflict of the winning team is more conducive for making better decisions.

The data sensing skills of the winning team is more conducive for making better decisions.

The data analytics skills of the winning team is more conducive for making better decisions.

Other Reasons: (Please Specify):

(Setting # 14. Ignore this information)

Survey Question 9/16

On a scale of 1-5(1 = Slightly higher possibility of winning; 5 = Very high possibility of winning), which of the following two teams is likely to win?

	Team A	Team B
Power Distribution within the Team (Team Organization: 1=Concentrated Power; 10 = Distributed Power)	4	2
Level of Conflict within the Team (Team Interaction: -5=Destructive Conflict;+5=Constructive Conflict)	-1	+1
Data Sensing Capability of the Team (Non Technical Skills: 1 = Poor; 10 = Excellent)	6	4
Data Analytics Capability of the Team (Hard Technical Skills: 1 = Poor; 10 = Excellent)	7	9
Which Team Will Win? (1-5)	Winning Team: A B Scale: 1 2 3 4 5 1=Slightly higher possibility of winning. 5=Very high possibility of winning	

Reasons For Your Choice of Winning Team (Please mark one or more):

The amount of power distribution of the winning team is more conducive for making better decisions.

The level of conflict of the winning team is more conducive for making better decisions.

The data sensing skills of the winning team is more conducive for making better decisions.

The data analytics skills of the winning team is more conducive for making better decisions.

Other Reasons: (Please Specify):

(Setting # 7. Ignore this information)

Survey Question 10/16

On a scale of 1-5(1 = Slightly higher possibility of winning; 5 = Very high possibility of winning), which of the following two teams is likely to win?

	Team A	Team B
Power Distribution within the Team (Team Organization: 1=Concentrated Power; 10 = Distributed Power)	4	6
Level of Conflict within the Team (Team Interaction: -5=Destructive Conflict;+5=Constructive Conflict)	-1	-3
Data Sensing Capability of the Team (Non Technical Skills: 1 = Poor; 10 = Excellent)	6	4
Data Analytics Capability of the Team (Hard Technical Skills: 1 = Poor; 10 = Excellent)	7	5
Which Team Will Win? (1-5)	Winning Team: A B Scale: 1 2 3 4 5 1=Slightly higher possibility of winning. 5=Very high possibility of winning	

Reasons For Your Choice of Winning Team (Please mark one or more):

The amount of power distribution of the winning team is more conducive for making better decisions.

The level of conflict of the winning team is more conducive for making better decisions.

The data sensing skills of the winning team is more conducive for making better decisions.

The data analytics skills of the winning team is more conducive for making better decisions.

Other Reasons: (Please Specify):

(Setting # 2. Ignore this information)

Survey Question 11/16

On a scale of 1-5(1 = Slightly higher possibility of winning; 5 = Very high possibility of winning), which of the following two teams is likely to win?

	Team A	Team B
Power Distribution within the Team (Team Organization: 1=Concentrated Power; 10 = Distributed Power)	4	2
Level of Conflict within the Team (Team Interaction: -5=Destructive Conflict;+5=Constructive Conflict)	-1	-3
Data Sensing Capability of the Team (Non Technical Skills: 1 = Poor; 10 = Excellent)	6	4
Data Analytics Capability of the Team (Hard Technical Skills: 1 = Poor; 10 = Excellent)	7	5
Which Team Will Win? (1-5)	Winning Team: A B Scale: 1 2 3 4 5 1=Slightly higher possibility of winning. 5=Very high possibility of winning	

Reasons For Your Choice of Winning Team (Please mark one or more):

The amount of power distribution of the winning team is more conducive for making better decisions.

The level of conflict of the winning team is more conducive for making better decisions.

The data sensing skills of the winning team is more conducive for making better decisions.

The data analytics skills of the winning team is more conducive for making better decisions.

Other Reasons: (Please Specify):

(Setting # 4. Ignore this information)

Survey Question 12/16

On a scale of 1-5(1 = Slightly higher possibility of winning; 5 = Very high possibility of winning), which of the following two teams is likely to win?

	Team A	Team B
Power Distribution within the Team (Team Organization: 1=Concentrated Power; 10 = Distributed Power)	4	6
Level of Conflict within the Team (Team Interaction: -5=Destructive Conflict;+5=Constructive Conflict)	-1	+1
Data Sensing Capability of the Team (Non Technical Skills: 1 = Poor; 10 = Excellent)	6	8
Data Analytics Capability of the Team (Hard Technical Skills: 1 = Poor; 10 = Excellent)	7	5
Which Team Will Win? (1-5)	Winning Team: A B Scale: 1 2 3 4 5 1=Slightly higher possibility of winning. 5=Very high possibility of winning	

Reasons For Your Choice of Winning Team (Please mark one or more):

The amount of power distribution of the winning team is more conducive for making better decisions.

The level of conflict of the winning team is more conducive for making better decisions.

The data sensing skills of the winning team is more conducive for making better decisions.

The data analytics skills of the winning team is more conducive for making better decisions.

Other Reasons: (Please Specify):

(Setting # 15. Ignore this information)

Survey Question 13/16

On a scale of 1-5(1 = Slightly higher possibility of winning; 5 = Very high possibility of winning), which of the following two teams is likely to win?

	Team A	Team B
Power Distribution within the Team (Team Organization: 1=Concentrated Power; 10 = Distributed Power)	4	2
Level of Conflict within the Team (Team Interaction: -5=Destructive Conflict;+5=Constructive Conflict)	-1	-3
Data Sensing Capability of the Team (Non Technical Skills: 1 = Poor; 10 = Excellent)	6	8
Data Analytics Capability of the Team (Hard Technical Skills: 1 = Poor; 10 = Excellent)	7	9
Which Team Will Win? (1-5)	Winning Team: A B Scale: 1 2 3 4 5 1=Slightly higher possibility of winning. 5=Very high possibility of winning	

Reasons For Your Choice of Winning Team (Please mark one or more):

The amount of power distribution of the winning team is more conducive for making better decisions.

The level of conflict of the winning team is more conducive for making better decisions.

The data sensing skills of the winning team is more conducive for making better decisions.

The data analytics skills of the winning team is more conducive for making better decisions.

Other Reasons: (Please Specify):

(Setting # 11. Ignore this information)

Survey Question 14/16

On a scale of 1-5(1 = Slightly higher possibility of winning; 5 = Very high possibility of winning), which of the following two teams is likely to win?

	Team A	Team B
Power Distribution within the Team (Team Organization: 1=Concentrated Power; 10 = Distributed Power)	4	2
Level of Conflict within the Team (Team Interaction: -5=Destructive Conflict;+5=Constructive Conflict)	-1	+1
Data Sensing Capability of the Team (Non Technical Skills: 1 = Poor; 10 = Excellent)	6	4
Data Analytics Capability of the Team (Hard Technical Skills: 1 = Poor; 10 = Excellent)	7	5
Which Team Will Win? (1-5)	Winning Team: A B Scale: 1 2 3 4 5 1=Slightly higher possibility of winning. 5=Very high possibility of winning	

Reasons For Your Choice of Winning Team (Please mark one or more):

The amount of power distribution of the winning team is more conducive for making better decisions.

The level of conflict of the winning team is more conducive for making better decisions.

The data sensing skills of the winning team is more conducive for making better decisions.

The data analytics skills of the winning team is more conducive for making better decisions.

Other Reasons: (Please Specify):

(Setting # 5. Ignore this information)

Survey Question 15/16

On a scale of 1-5(1 = Slightly higher possibility of winning; 5 = Very high possibility of winning), which of the following two teams is likely to win?

	Team A	Team B
Power Distribution within the Team (Team Organization: 1=Concentrated Power; 10 = Distributed Power)	4	6
Level of Conflict within the Team (Team Interaction: -5=Destructive Conflict;+5=Constructive Conflict)	-1	-3
Data Sensing Capability of the Team (Non Technical Skills: 1 = Poor; 10 = Excellent)	6	4
Data Analytics Capability of the Team (Hard Technical Skills: 1 = Poor; 10 = Excellent)	7	9
Which Team Will Win? (1-5)	Winning Team: A B Scale: 1 2 3 4 5 1=Slightly higher possibility of winning. 5=Very high possibility of winning	

Reasons For Your Choice of Winning Team (Please mark one or more):

The amount of power distribution of the winning team is more conducive for making better decisions.

The level of conflict of the winning team is more conducive for making better decisions.

The data sensing skills of the winning team is more conducive for making better decisions.

The data analytics skills of the winning team is more conducive for making better decisions.

Other Reasons: (Please Specify):

(Setting # 3. Ignore this information)

Survey Question 16/16

On a scale of 1-5(1 = Slightly higher possibility of winning; 5 = Very high possibility of winning), which of the following two teams is likely to win?

	Team A	Team B
Power Distribution within the Team (Team Organization: 1=Concentrated Power; 10 = Distributed Power)	4	2
Level of Conflict within the Team (Team Interaction: -5=Destructive Conflict;+5=Constructive Conflict)	-1	-3
Data Sensing Capability of the Team (Non Technical Skills: 1 = Poor; 10 = Excellent)	6	8
Data Analytics Capability of the Team (Hard Technical Skills: 1 = Poor; 10 = Excellent)	7	5
Which Team Will Win? (1-5)	Winning Team: A B Scale: 1 2 3 4 5 1=Slightly higher possibility of winning. 5=Very high possibility of winning	

Reasons For Your Choice of Winning Team (Please mark one or more):

The amount of power distribution of the winning team is more conducive for making better decisions.

The level of conflict of the winning team is more conducive for making better decisions.

The data sensing skills of the winning team is more conducive for making better decisions.

The data analytics skills of the winning team is more conducive for making better decisions.

Other Reasons: (Please Specify):

(Setting # 9. Ignore this information)

Bibliography

1. Galbraith, Jay R. "Organization design: An information processing view." *Interfaces* 4.3 (1974): 28-36.
2. Brynjolfsson, Erik, Hitt, Lorin M. and Kim, Heekyung Hellen, Strength in Numbers: How Does Data-Driven Decision making Affect Firm Performance? (April 22, 2011).
3. Wagstaff, Kiri. "Machine learning that matters." *arXiv preprint arXiv:1206.4656* (2012).
4. Nancy Leveson. A Systems Model of Accidents, International Conference of the System Safety Society, Denver, 2002.
5. Argyris, Chris. "Double loop learning in organizations." *Harvard business review* 55.5 (1977): 115-125.
6. Busenitz, Lowell W., and Jay B. Barney. "Differences between entrepreneurs and managers in large organizations: Biases and heuristics in strategic decision-making." *Journal of Business Venturing* 12.1 (1997): 9-30.
7. Kahneman, Daniel, Paul Slovic, and Amos Tversky, eds. *Judgment under uncertainty: Heuristics and biases*. Cambridge University Press, 1982.
8. Eisenhardt, Kathleen M., Jean L. Kahwajy, and L. J. Bourgeois III. "How management teams can have a good fight." *What Makes a Decisive Leadership Team* (2009): 33.
9. Belbin, Meredith. "Belbin Team Roles." *Belbin Home Page. Retrieved April* 16 (2004): 2004.
10. **MBTI** Manual: A Guide to the Development and Use of the Myers-Briggs Type Indicator Instrument
11. Hollan, James, Edwin Hutchins, and David Kirsh. "Distributed cognition: toward a new foundation for human-computer interaction research." *ACM Transactions on Computer-Human Interaction (TOCHI)* 7.2 (2000): 174-196.
12. Howard, Ronald A. "Decision analysis: Practice and promise." *Management science* 34.6 (1988): 679-695.

13. Tang, Victor. *Corporate decision analysis: an engineering approach*. Diss. Massachusetts Institute of Technology, 2006.
14. Simon, Herbert A. *Models of Bounded Rationality, Vol. 3: Emperically Grounded Economic Reason*. Vol. 3. MIT press, 1997.
15. Argyris, Chris. "The executive mind and double-loop learning." *Organizational dynamics* 11.2 (1982): 5-22.
16. Ross, Rick. "The ladder of inference." *The fifth discipline fieldbook: Strategies and tools for building a learning organization* (1994): 242-246.
17. Crapo, Andrew W., et al. "Visualization and the process of modeling: a cognitive-theoretic view." *Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2000.
18. Pinker, Steven, 1997. How The Mind Works, W. W. Norton and Company, New York.
19. Brady, Timothy F., et al. "Visual long-term memory has a massive storage capacity for object details." *Proceedings of the National Academy of Sciences* 105.38 (2008): 14325-14329.
20. Robert Spence, Information Visualization: Design for Interaction (2nd Edition), Prentice-Hall, Inc., Upper Saddle River, NJ, 2007.
21. http://en.wikipedia.org/wiki/Big_data; Accessed December 12, 2012
22. "Hubway Data Visualization Challenge." *Hubway Data Visualization Challenge*. Hubway, Inc., n.d. Web. 07 Jan. 2013.
23. Card, Stuart K., Jock D. Mackinlay, and Ben Shneiderman. *Readings in information visualization: using vision to think*. Morgan Kaufmann, 1999.
24. Cisco. "Visual Networking Index (VNI)." *White Paper*. Cisco, 14 Feb. 2012. Accessed 07 Jan. 2013.
25. Simon, Herbert A. "The architecture of complexity." *Proceedings of the american philosophical society* (1962): 467-482.
26. Ackoff, Russell L., Jason Magidson, and Herbert J. Addison. *Idealized design*. Wharton School Pub., 2006.
27. Simon, Herbert A. "Rational decision making in business organizations." *The American economic review* (1979): 493-513.
28. Tufte, Edward R. *Envisioning information*. Graphics Press, 1990.

29. Truck, Matt. "A Chart of the Big Data Ecosystem, Take 2." *Matt Turck*. N.p., 15 Oct. 2012. Web. 08 Jan. 2013.
30. Beer, Stafford. "The viable system model: its provenance, development, methodology and pathology." *Journal of the operational research society* (1984): 7-25.