# Data-driven Value-up Strategies

Ho-Young Ryou
Hanyang University Business School
222 Wangsimni-ro, Seongdong-gu, Seoul, Korea, 04763
hyryou94@handapartners.com

Moonkyung Choi
Ph.D. in Economics, Deputy director
Asset Management Team
The Ministry of Employment and Labor (MOEL)
cmk1280@korea.kr

Hyoung-Goo Kang
Founder of Handa Partners (www.handapartners.com)
Associate Professor, Department of Finance
Hanyang University Business School
222 Wangsimni-ro, Seongdong-gu, Seoul, Korea, 04763
Correspond to hyoungkang@hanyang.ac.kr

ABSTRACT

This paper proposes how to design and implement data-driven value-up strategies and addresses the following questions: How can a firm increase its value using data? How can one increase a firm's low valuation ratio (e.g., PB ratio, PE ratio, EV/EBITDA) based on data capabilities? Our suggestions are academically grounded and draw scholastic insights associated with architectural innovation, the behavioral theory of the firm and the knowledge-based view of the firm. Furthermore, our recommendations are logically derived from field observations such as how data science is abused in dealing with meso-level data, while it is underused in using macro-level and alternative data to accomplish machine-human teaming and risk management. Broadly, this paper addresses under-researched issues such as why some firms are better at drawing value from intangibles such as data, data-science capabilities and routines and how to value them. In conclusion, we provide an answer to how academia can guide practitioners (e.g., bankers and managers) in private equities and venture capitals and how practitioners could conduct 'science' rather than 'art' in increasing organizational value using data based on academic literature.

Keywords: data, value-up, data-driven value-up strategy, architectural innovation, a behavioral theory of the firm, the knowledge-based view of the firm.

## INTRODUCTION

The valuation gap between companies that are data-driven and those that are not is increasing (Brynjolfsson et al., 2011). When a firm's value is primarily dependent on its data and data-science capabilities, we call the firm's business model 'data-driven' (Schaefer et al., 2017; Sorescu, 2017). A data-driven value-up strategy aims to increase a target firm's value by maximizing the value of data and data-science capabilities while integrating them as a firm's strategic resources (Barney, 1986; Conner, 1991; Conner & Prahalad, 1996; Dierickx & Cool, 1989) to determine the competitive advantage of the firm.

The recent outbreak of COVID-19 has cemented the dominance of big technical companies that integrate data into their business model, i.e., those that are data-driven. For instance, the value of large technical firms increased throughout the crisis, and as of May 2020, technical firms such as Apple, Microsoft, Alphabet, Amazon, Facebook account for 20% of the S&P 500, becoming the true winners in the crisis. This current trend indicates that the success of a business depends on utilization of data, and this success would be further reinforced if COVID-19 is just the beginning of the series of uncertainties (e.g., climate change, global power struggle, artificial intelligence, and inequalities). Hence, data will continue to be an intangible asset or even strategic resource of a company, contributing to the value of a business.

Examples can be found in the finance sector. Although both banks and fintech companies compete in the same strategic group (Dess & Davis, 1984; Porter, 1985; Cool & Schendel, 1988), their valuation ratio differs dramatically. Hence, if traditional financial institutions could utilize data as well as fintech firms do, a considerable increase in value will occur. Data-driven value-up is particularly urgent for the banking sector, which has suffered from low valuations for a long time. Then, how should one design and implement data-driven value-up strategies? This paper addresses this important question, which the existing literature has ignored.

Private equities regard value-up as part of an important 'investment model' (Pomerance

& McCarthy, 2018). Venture capitalists often call value-up ‘company-making’ (Chesbrough, 2002; Hisrich & Jankowicz, 1990). Value-up is hence the value proposition of private equity firms and venture capitals. And it is no wonder that they recently become interested in value-up strategies based on data. According to our interviews with practitioners, several private equities regard data-driven value-up as their core investment model and advertise data-driven value-up schemes as the theme for fundraising. Similarly, a leading venture capitalist has started applying data science even for deal-sourcing and risk management. For instance, a banker in a large private equity firm said, “We are interested in firms whose data capabilities are underestimated. Putting data experts in the top management team of such firms would increase firm value.” Another said, “… Instilling data-driven and evidence-based routines in organizational processes is a value-up strategy.” Similarly, a venture capitalist said, “Data scientists can become good venture capitalists because we use data science for deal sourcing as well as for value-up.” Another VC said, “We are looking for startups with data whose value we can increase.” It is not a secret that sophisticated private equities and venture capitalists are aggressively hiring data scientists. Of course, such data-driven value-up would matter for other organizations as they care about the returns to stakeholders. These attempts for data-driven value-up strategies would be beneficial for investors and the national economy in line with the fourth industrial revolution.

Successful data-driven value-up strategies require appropriate guidelines. However, as far as we know, there are no studies or practical guidelines on this issue yet. This study aims to fill the gap in the literature. Furthermore, we critically evaluate some value-up attempts and suggest ideas what a company lacking data-science capabilities can do.

The remainder of this paper is as follows. In the next section, we discuss a common misconception about data-driven value-up strategies and evaluate where data science tends to be abused or misleading. The next section discusses where data science tends to be underused. The next proposes frameworks to develop and deploy data-science capabilities. The final section provides a conclusion.

## ABUSING MESO-LEVEL DATA

GAP is a famous case regarding the 'innovative' use of data.[1] GAP is a global clothing and accessories retailer based in the United States. GAP is famous for brands such as Old Navy and Banana Republic. However, GAP has stagnated since the 2000s. During this period, fashion brands such as ZARA and Uniqlo overtook GAP. While GAP was in trouble, Mr. Art Peck became the CEO. Right after being appointed, Art Peck dismissed many creative directors, i.e., the gurus who predict the fashion trends. Design themes of their fashion houses are based on these predictions. Creative directors are considered to be the quintessential assets in fashion industries.

However, Art Peck insisted that GAP should rely less on the intuitions of creative directors or designers, and more on using data for decision making. In summary, Mr. Peck attempted to play a version of 'money ball' (Lewis, 2004) in the fashion industry. Obviously, the whole fashion industry was aghast and against Mr. Peck's view. There were worries that creativity might disappear, and the sector would be distorted while arguing 'Mr. Peck knows nothing about fashion.' Such reactions are in some sense natural because existing stakeholders with 'comfortable status' would resist any disruptive changes. For example, whilst incumbents always say that they need innovation, many of them are against it when an innovation indeed takes place. Innovators are criticized and dismissed for being naïve, not understanding the organization, industry, and practice.

Was Mr. Peck's attempt successful? It is hard to say that the attempt was successful because GAP's stock performance has been lukewarm compared to Inditex[2] and SPDR from

[1] The following article presents further stories about GAP's data strategy, https://hbr.org/podcast/2018/11/could-big-data-replace-the-creative-director-at-the-gap

[2] "Inditex is one of the world's largest fashion retailers, with eight brands (Zara, Pull&Bear, Massimo Dutti, Bershka, Stradivarius, Oysho, Zara Home and Uterqüe) selling in 202 markets through its online platform or its over 7,000 stores in 96 markets." https://www.inditex.com/en/about-us/who-we-are

2016/01/01 to 2020/06/08. Furthermore, Art Peck resigned in November 2019 taking responsibility for poor performance. Then, why was the data-driven strategy of Art Peck rather unsuccessful?

We argue that the mixed performance of Art Peck's data-driven strategy is closely related to the practice of misusing 'meso-level data'. Meso-level data arises at the group or organization level. A group or a firm's performance or behavioral data that are commonly found daily, weekly, and monthly are the examples. With few exceptions (e.g., stock prices), it is expensive to generate meso-level data in real time because one needs to aggregate micro-level data. Hence, meso-level data is usually used for establishing short- or mid-term strategies. Given the nature of meso-level data, it is hard to apply sophisticated techniques (e.g., deep learning) to such data. Even if one collects daily data for ten years, there is not enough data to develop (not 'estimate') any deep learning model. Imagine a problem of predicting a fashion trend of a customer segment for next autumn using the historical data about the segment's fashion trend for the last ten years. It would be very hard to develop an advanced machine-learning model. Similarly, imagine predicting the daily return of KOSPI 200 using 10-years of data such as hosts of financial, accounting and even unstructured big data. In such cases with short time-series data ($T$ = 252 x 10), one should be very selective and careful to include a large number of independent variables in estimating models. Then, one would need decent and academically grounded theories to complement the limitations of the stock-market data.

Nevertheless, data science is abused in many cases. The reason is simple. Meso-level data seems easy to analyze. Anyone who has learned basic statistics can create a simple algorithm. If one learns machine learning in a class, one can develop a simple trading strategy using daily stock returns or algorithms to predict 'trends' and to make 'simulated profits'. Indeed, such tasks are often given as a homework assignment in university classes. Meso-level data are sometimes easier to obtain than micro-level data (e.g., an individual's location information), which are frequently subject to privacy issues. Since it seems easy at first glance to obtain and to analyze meso-level data, it is no wonder that one can easily find firms and data consultants who advertise their 'data-driven' services in formulating strategies, trading algorithm, predictions, and intelligence that require producing meso-level insights for short- or mid-term forecasting horizons. However, these approaches have serious problems.

Let us assume we are faced with the problem of predicting the performance of a business strategy (e.g., marketing a particular fashion style). Let us indicate the size of the data as $N = T \times K$, where $T$ and $K$ represent the number of observations (length) and number of feature variables (feature size), respectively. Data is called 'big data' when $N$ is large. There are two ways to make big data (i.e., increase $N$) by increasing $T$ or increasing $K$. Academic researchers tend to highlight $T$, the length of a dataset. However, interestingly, industry practitioners tend to focus more on $K$. A possible reason for this is that it is clearly easier to increase $K$ than $T$ or that it is convenient to develop a marketing messages with $K$ such as "we have a new dataset to predict the stock market", "only our firm has access to the feature and will use it to develop new products", etc. For example, when predicting the performance of a company, one can easily increase $K$ by adding more variables such as corporate accounting data, crawled text data, macroeconomic data, and other unstructured data (Kang, et al., 2019).

Yet, increasing $K$ can complicate the problem rather than solve it because of the curse of dimensionality. As we collect more features ($K$) to predict the performance of a firm, the length of data, $T$, should increase in proportion to $K^2$. In addition, if we try millions of tossing coins simultaneously, one of them would fit the trend perfectly ($R^2$ = 100%). This is the curse of dimensionality. The problem is that $T$ is often difficult to increase (e.g., the length of time is limited). However, one can determine the distribution of a variable only when $T$ is large. For example, if one coin is tossed millions of times, the probabilities of getting heads and tails will converge to the actual probability. Then, how do we solve this problem that is inherent in meso-level data? Would a firm need to hire world-class experts in artificial intelligence or machine learning to solve this problem to predict the future or customer behavior?

In this situation, academic theories or business acumen matter more than technology, i.e., "strategy, not technology, drives digital transformation" (Kane et al., 2015). Meso-level analysis does not require a world-class AI expert, but calls for the joint work of an experienced field expert and an econometrician who can deal with identification problems (Angrist & Pischke, 2008; Roberts & Whited, 2013). Thus, theories compensate for the lack of data. Intuitively speaking, if we can observe everything, why does one need a theory? Note that whether a strategy works or not is a problem of causality, which in turn calls for addressing identification problems. Simply collecting a large number of datapoints or using a highly complex model does not automatically guarantee the identification of causality because one needs specific techniques (e.g., "Mostly Harmless Econometrics", Angrist & Pischke, 2008) to solve identification problems.

For this reason, a firm needs to be careful when working with consulting firms and data scientists who do deep learning without considering theories. Often, their actions are no more than data mining and are mostly harmful. One requires special skills or experiences to build and test hypotheses in the framework of 'mostly harmless econometrics'. In this regard, we present a three-step strategy for a firm interested in hiring a data consultant to develop 'intelligence' on meso-level data analysis as follows.

> [Step one] Ask about the data structure (i.e., $N$, $T$, $K$ at $N = T \times K$) and the approach to address the curse of dimensionality.
>
> [Step two] Ask what theories or frameworks are used to design strategies or recommendations, which is equivalent to hypothesis formulation in academia (e.g., a firm's brand equity would increase in a customer segment if the firm conducted strategy *X*).
>
> [Step three] Ask what kind of identification strategies are used to test hypotheses.
>
> [Final step] Do not work with a data consultant who does not present plausible answers to the previous questions.

## MACRO-LEVEL DATA AND SCENARIO PLANNING

Scenario analysis is essential to designing mid- to long-term strategies. The inputs to scenario planning are views about macroeconomic or socio-political variables (Ramirez & Wilkinson, 2016). To formulate the views, one needs to collect and analyze macro-level data. However, logically speaking, if it is difficult to apply machine learning to meso-level data, it would be more difficult to forecast macro-level events. For example, suppose we are interested in oil prices, a popular variable for scenario planning for many firms. Thus, 20 years of data include only 5,000 observations (20 years x 252 business days). Thus, it is difficult to practice deep learning, which requires the estimates of thousands of parameters. While it is possible to apply deep learning to transaction data in seconds or real time data, they are not commonly used or practical for establishing strategies, especially mid- to long-term strategies.

Nevertheless, despite the above limitations, data-driven algorithms can be surprisingly helpful for formulating mid- to long-term strategies (Kang, et al., 2019). According to behavioral economics, people tend to exhibit behavioral biases in mid- to long-term decision-making. The longer the decision horizon, the larger the biases. Planning fallacies, hyperbolic discounting, and optimism bias are examples of biases. In particular, humans tend to make decisions using mental shortcuts on instant events, which is called availability heuristics. In conclusion, while it is difficult for machines to make mid- to long-term decisions, it is more difficult for humans. Hence, machine-human complementarity becomes important (Kleinberg et al., 2018; Lyons et al., 2019; Walliser et al., 2019). Data-driven algorithms in combination with human intuition will enhance decision-making and lead to data-driven value-up. The quotes below from the book Zero to One (Thiel & Masters, 2014) are helpful.

> But the most valuable companies in the future won't ask what problems can be solved with computers alone. Instead, they will ask how can computers help humans solve hard problems? … men and machines are good at fundamentally different things. People have intentionality – we form plans and make decisions in complicated situations. We're less

good at making sense of enormous amounts of data. Computers are exactly the opposite: they excel at efficient data processing, but they struggle to make basic judgements that would be simple for any human.

In summary, machine algorithms can assist and double-check the decisions of humans and organizations by incorporating behavioral theories (Cyert & March, 1963) and behavioral economics in data-based algorithms. Furthermore, an organization can continuously update the model with practical information from its industry and growing body of academic literature, which will in turn lead to enhanced absorptive capacity (Cohen & Levinthal, 1990) and learning in the organization (March, 1991). This iterative routine will be the basis of data-driven value up.

The area where data can create large value is enterprise risk management (ERM) (Kang, et al., 2019). Risk management is regarded as one of the most important elements in a decision-making process, especially for financial and human-resource decision making. However, we argue that data science is underused in ERM processes. In particular, firms tend to ignore Knightian uncertainty (Keynes, 1921; Knight, 1921) although the era of Knightian uncertainty (which includes pandemic, climate change, geopolitical competition, wealth gap, the future of capitalism, artificial intelligence, etc.) is close. According to Keynes (1921), Knightian uncertainty arises when the decision maker cannot quantify the uncertainty in decision making. For instance, Knightian uncertainty can arise when a state is too complex or ambiguous so that even its probability distribution cannot be specified. Keynes (1937) explains the Knightian uncertainty as follows.

> By "uncertain" knowledge, let me explain, I do not mean merely to distinguish what is known for certain from what is only probable. The game of roulette is not subject, in this sense, to uncertainty... Or, again, the expectation of life is only slightly uncertain. Even the weather is only moderately uncertain. The sense in which I am using the term is that

in which the prospect of a European war is uncertain, or the price of copper and the rate of interest twenty years hence, or the obsolescence of a new invention, or the position of private wealth owners in the social system in 1970. About these matters there is no scientific basis on which to form any calculable probability whatever. We simply do not know. Nevertheless, the necessity for action and for decision compels us as practical men to do our best to overlook this awkward fact and to behave exactly as we should if we had behind us a good Benthamite calculation of a series of prospective advantages and disadvantages, each multiplied by its appropriate probability, waiting to be summed. (Keynes, 1937; 213-214)

Intuitively, risk managers should collect information from media, industry sources and communities, to forecast Knightian uncertainty. However, in reality, managers do not have enough time to even properly read the Financial Times. This is where a machine can help humans. A machine can collect unstructured data and formulate an early warning system. Furthermore, it would constitute a routine to develop a data-driven culture (Waller, 2020) to address Knightian uncertainty by attempting to quantify it. The Ministry of Employment and Labor (MOEL) of Korea is a good example of adopting an early warning system (EWS) based on unstructured data. MOEL's asset management team has been using the EWS system since early 2019, and they successfully predicted the serious economic harm associated the COVID-19 risk as early as January 2020, two months ahead of the global economic shocks.

## FRAMEWORKS FOR DATA-DRIVEN VALUE-UP

This section presents a general framework about how to increase firm values and find investment opportunities grounded on data, the most important intangible asset for innovation. To address the issues, we suggest a framework summarized in Table 1.

-----------------------------------------
Insert Table 1 about here
-----------------------------------------

To use data to accomplish value-up and innovations, a firm needs to consider at least [1] what innovations to pursue (data-driven innovation strategy), [2] how to change its organization to implement the strategy (data-driven organizational strategy), and [3] how to keep the changes sustainable to be competitive in the market (data-driven organizational learning for sustainable competitive advantage). In the context of a value up strategy, we specifically highlight architectural innovation (Henderson & Clark, 1990) as a data-driven innovation strategy, a behavioral theory of the firm (BTF) (Cyert & March, 1963) as a data-driven organizational strategy, and the knowledge-based view of the firm (KBV) (Cohen & Levinthal, 1990; Kogut & Zander, 1996) to build data-driven organizational learning processes suitable for a sustainable competitive advantage. The following subsections respectively discuss the three approaches -- architectural innovation, BTF and KBV.

## Architectural innovation

Despite their field expertise, managers in low-valuation industries (e.g., banks) tend to lack a data science background (e.g., coding) and so have difficulty in understanding and applying technological innovations such as unstructured data and deep learning. A problem is that the more successful they have been, the less likely they are to adopt innovations, which in turn prompts their downfall (The innovator's dilemma, Christensen, 2013). Furthermore, the body of the literature grows too fast and is too complex for a practical manager to study and apply it in a systematic way.

On the other hand, data scientists without field expertise often mindlessly undertake data mining. They repeat trials and errors without understanding the value proposition of the fields and the contexts of the sectors where they apply their data capabilities. This naturally leads to spurious results. Not surprisingly, managers are already becoming skeptical about artificial intelligence (Conkle, 2020). In particular, we believe the problem is more serious in areas such as competitive strategies and investment decision-making, where it is important to combine data science with academic theories and practical experiences from economics and business management to resolve identification problems (e.g., causality vs. correlation).

Neither data science nor field experts alone can solve this problem. Only data scientists with industry experience, theory, and practice can solve the problem. However, only large technology companies or hedge funds can afford such experts. Alternatively, if any company somehow provides the service that solves the problem of combining data science expertise and field expertise regarding the fourth Industrial Revolution with easy UI/UX (like iPhones), it will be a tremendous success.[3] However, such a service is not offered now or is too expensive if it exists. For a while, the problem of combining field expertise with data science remains as an important challenge for both small startups, and large institutions, as explained by the innovator's dilemma. Given these challenges, what kind of data-driven innovations should a firm aim to develop to accomplish data-driven value-up?

Henderson & Clark (1990) classify innovation into four categories as shown in Panel A in **Table 2**. The strategy for pursuing innovation has two axes: the concept and the relationship between concepts. The concept axis is about either reinforcing or overturning concepts. The relationship axis is about either maintaining or changing the relationship between concepts. In the end, four types of innovations are derived: incremental innovation, modular innovation, architectural innovation, and radical innovation.

[3] This is the vision of a fintech startup, Handa Partners (http://www.handapartners.com).

-----------------------------------------

Insert **Table 2** about here

-----------------------------------------

Among these innovations, which one is the appropriate model to solve the problems of traditional firms with low valuation, and so are urgently interested in data-driven value-up? Panel B in **Table 2** describes our suggestions. First, radical innovation (i.e., overturning concepts and changing their relationship) is too expensive and time consuming for most organizations except for technology giants (e.g., big techs) or large hedge funds. With respect to data, radical innovation requires developing innovative artificial-intelligence techniques while simultaneously connecting new techniques in a creative manner, which not many firms are capable of or need at their current stage.

Second, incremental innovation, reinforcing concepts and maintaining their relationships, is what most of the companies are already attempting to do, but this strategy is vulnerable to the innovator's dilemma (Christensen, 2013). The theory of the innovator's dilemma warns of a situation in which a firm eagerly pursues the innovation targets that the current market desires, but eventually fails because of the very attempt. Hence, incremental innovation is possibly appropriate for firms that lack significant data capabilities compared to industry peers. They can start from small incremental strategies to learn from peers and possibly catch up to them if opportunities are open. However, incremental innovation itself would not achieve significant data-driven value-up and, again, is vulnerable to the innovator's dilemma.

Third, modular innovation, which is overturning concepts while maintaining their relationships, focuses on collecting new data that are not normally used or creating new data science methods given existing frameworks or business relationships. However, collecting innovative data is appropriate only for those that already integrate data collection with their business model. For instance, Facebook collects precious individual-level data as part of their business model. Users 'pay' their data to enjoy Facebook. Purchasing innovative data is infeasible either because they are usually the core resources of leading big technology firms, and therefore those firms have no reason to trade the data. Even if a firm purchases precious data, its value would decrease significantly once the data is transferred to other organizations. Suppose that even if a firm obtained Facebook's data, it would be very difficult to use the data as effectively as Facebook because the data is integral part of Facebook's business model and a core strategic resource of the firm. A core strategic resource is by definition socially complex, ambiguous, and not significantly transferrable (Barney, 1986) with "the characteristics of the asset accumulation process: time compression diseconomies, asset mass efficiencies, inter-connectedness, asset erosion and causal ambiguity" (Dierickx & Cool, 1989).

Furthermore, it is unnecessary to develop new artificial-intelligence technologies for most companies because researchers in academia are consistently developing new machine learning models, and they often post their findings in Github or Gitlab for free. Just following them would be enough for most firms. In fact, since the speed of knowledge accumulation is so fast, it would be hard for most firms to follow the speed, let alone overtake it with new technologies. Therefore, we recommend pursuing modular innovations for startups grounded in universities or research institutes. If a startup succeeds, a larger firm can easily import its product as a "new module" in its existing business architecture. This can be done conveniently because modular innovation by definition does not require changing existing relationships between concepts.

This leaves us with only one choice, architectural innovation, for most traditional firms suffering from low valuation. Note that it is relatively easy to customize and apply the existing models rather than developing a new machine learning model. Even undergraduate students can download and experiment with recent working papers and their codes posted at GitHub. In the end, it is important to grasp the specific business questions of the industry, and then to combine existing resources to solve the questions. This constitutes an architectural innovation, reinforcing concepts while changing their relationship. Hiring a world-class artificial-intelligence expert would not help to accomplish architectural innovation because she would be "pigeonholed within a company" (Waller, 2020).

However, even architectural innovation would not be easy if an organization lacks planning. The biggest challenge here is not technology. Rather, the hard problem is organizational challenges and strategy. For instance, some practical data scientists revealed during our interviews that most traditional financial firms are not properly utilizing "even the data they already own, let alone drawing the knowledge from academic or open sources". Again, "challenges reside in organizational issues" as several data scientists in banks put it. This issue will be described in the next subsection.

In conclusion, to accomplish architectural innovation, low-valued organizations do not need to hire expensive artificial intelligence experts. Not every firm needs world-class experts in artificial intelligence and data science. Instead, they need to combine field experts who correctly identify business questions with the researchers who can draw a body of knowledge from available resources to test the experts' intuition. To test the experts' intuition, the researchers should be able to address identification problems (Angrist & Pischke, 2008) because developing recommendations inevitably leads to testing hypotheses. If they work together, even minimal levels of technology and data could have a great effect. Therefore, architectural innovations demand that a firm's field expert and data scientists should creatively connect existing knowledge around specific business problems. This is again in line with the previously mentioned intuition of "strategy, not technology drives digital transformation" (Kane et al., 2015).

We argued that combining business problems and data strategies is important to undertake architectural innovations. Then, how do we characterize problems and identify data strategies? This question is answered in the next section.

## Behavioral Theory of the Firm (BTF)

Practical aspects of data strategies are very diverse and depend on the context of the firm. However, we lack academic literature to analyze and generalize such diverse organizational heterogeneity to develop data strategies. Without any framework for understanding the sources and patterns of practical diversity, how would it be possible to generalize the diversity, and to logically propose or develop a prescriptive data-driven value-up strategy?

This study attempts to present a perspective framework based on a classical theory, the Behavioral Theory of the Firm (BTF) (Cyert & March, 1963). BTF is one of the main classics of the Carnegie School[4] and has become one of the most popular frameworks in the field of behavioral science, strategic management, organization theory and information management. BTF includes large valuable implications that cannot be fully covered in this paper. Instead, this paper focuses on two well-known concepts in BTF: Knightian uncertainty (Keynes, 1921; Knight, 1921) and stakeholder conflict.

The existing literature argues that entrepreneurship is important in the case of high Knightian uncertainty (Knight, 1921; Mazzucato, 2011) and that the creation of shared value is important in the case of high stakeholder conflict (Porter & Kramer, 2011). Then, depending on high or low Knightian uncertainty and stakeholder conflicts, we can combine high and low entrepreneurship and shared-value creation. **Table 3** summarizes this intuition from the existing literature (Kang et al., 2018).

-----------------------------------------

Insert **Table 3** about here

-----------------------------------------

From **Table 3**, four approaches to data-driven value-up strategy can be derived regarding cases with high or low Knightian uncertainty and high or low stakeholder conflict. Those four approaches are what organizations can rely on to design specific data-driven value-up strategies. To summarize our recommendations:

First, a firm needs to identify its business problems based on field experts' intuition.

[4] "The Carnegie School was a so-called "Freshwater" economics intellectual movement in the 1950s and 1960s based at Carnegie Mellon University and led by Herbert A. Simon, James March, and Richard Cyert." https://en.wikipedia.org/wiki/Carnegie_School. See also Gavetti et al. (2007) about neo-Carnegie school.

Second, it measures Knightian uncertainty and stakeholder conflict in the problem.

Third, a firm designs a strategy combining entrepreneurship and shared value.

Fourth, data researchers draw a body of knowledge and resources and combine them around the formulated strategy to generate architectural innovation.

To accomplish architectural innovation and its associated data strategies, a firm's organization itself should be built for it. This is possibly the hardest problem as our interviewees noted. The next subsection presents our solutions.

## The Knowledge-Based View (KBV)

The Knowledge-Based View (KBV) (Cohen & Levinthal, 1990; Kogut & Zander, 1996) proposes that the combination of tacit knowledge and the ability to utilize this knowledge (which rival companies cannot easily follow) tends to determine the competitive advantage of a firm. Collecting data or building data-science capabilities are not the final, but an intermediate goal of a firm. Generating a sustainable competitive advantage is usually the goal. Hence, data should become a valuable knowledge resource on which a firm can base its strategies to build a competitive advantage over one's rivals.

The Knowledge-Based View (KBV) emphasizes the importance of organizational learning for transforming data to organizational knowledge. Importantly, absorptive capacity (Cohen & Levinthal, 1990) determines the extent of organizational learning such as the ability to collect and to apply information to create profit. Existing KBV literature highlights four determinants of absorptive capacity: **prior knowledge, incentive structure, organizational routine, and social**

**network**. Specifically, the performance of organizational learning is decided by [1] what one is studying (prior knowledge) (Eisenhardt & Santos, 2002), [2] whether one is incentivized to study hard (incentive structure) (Kapoor & Lim, 2007), [3] how is one studying (routine) (Grant, 1996; Grant & Baden-Fuller, 1995), [4] from and with whom one is learning (social network) (Yli-Renko et al., 2001).

Let us analyze each determinant in more detail. The first is prior knowledge. The fact that prior knowledge determines the absorptive capacity means that what data a company already owns would affect data-collecting and data-processing ability in the future. Many organizations claim that they do not have enough data, but in fact they often do have data that is useful. Firms just may not know what they have and what they can do with it. Sometimes, their managers intentionally ignore their data possibly because "their lives are already good enough."[5] Many firms do not realize that their prior data could determine the future trajectory of its data capabilities. This is related to another determinant, incentive structure. In addition, the importance of prior knowledge is also related to combinative capabilities (Kogut & Zander, 1996), which is the capability to combine existing knowledge sourced from inside and outside the organization to acquire new skills. This is again in line with the importance of existing data and architectural innovation. We recommend firms realize that they are already generating data in real time by their people, things they own, and organizational activities (e.g., Internet of Things). While firms complain about the lack of data, aren't they simply intimidated by the size of the data that they already have? Firms do not use 97% of data they own (Sebastian-Coleman, 2018), and 87% of organizations lack data-science capabilities.[6] How to use such data about their clients and organizations will determine the success of data-driven value-up.

The second determinant is incentive structure. A particular problem with incentive structure is that data managers are often uncooperative and hostile in sharing data. This is because

[5] Interview quote.
[6] https://www.gartner.com/en/newsroom/press-releases/2018-12-06-gartner-data-shows-87-percent-of-organizations-have-low-bi-and-analytics-maturity

data managers are concerned that others may invade their work or discover errors and issues in their practices if the data that they are in charge of is shared. In addition, it is natural that they feel a great threat (or at least burden) of disclosing the data that they have been managing to the experts with advanced degrees or publication capabilities. However, this situation should be resolved for any firm to increase the value of its data and capabilities. Therefore, an appropriate incentive structure should be built and shared with data managers so that data managers will openly share their data with experts and cooperate. Since incentives are inseparable from organizational culture, firms need to design organization culture about data, e.g., data-driven culture (Subrahmanyan & Jalona, 2020; Waller, 2020).

The third determinant is organizational routines (Dosi et al., 2001). One of the most common obstacles to the application of data science is poor organizational routines around data. For example, there are many cases where the table format and data structure are not systematically managed. If data is not well organized, data value fades and data science is not applicable. In many firms, IT experts decide data tables and formats, but the problem is that they are not the ultimate users of the data. Ultimate users should decide how to tabulate and structure data. This will make a firm's data analysis more efficient. If users are not capable of such data structuring, a firm should consult academic researchers who have published similar types of problems or data. For instance, given our experiences, firms do not know how to structure panel data efficiently. With the possible help of academic researchers or academic-minded consultants, the front office, not the back office, should design data dictionary, standard table format and folder structure in an organization. If data is organized and managed according to the format defined by the front office, cooperation becomes easier. The data can be shared to other departments with APIs (application programming interfaces) in standard format so that many data scientists can cooperate, and data may even be traded on data exchanges to generate extra revenues. We recommend starting from constructing a 'research dataset' which is free from security, privacy, and other regulatory issues, so that a firm shares it with internal and external researchers.

Eventually, the research dataset should be instrumental in overcoming data silos[7] and needs to evolve into the master data (metadata) of a firm, so that internal and external analysts broadly request access, and then analyze and crosscheck them to form a clear organizational consensus or a point of debate. This will prevent wasting time and energy.

## CONCLUSION

How can one increase firm value? How can one increase a firm's low valuation ratio (e.g., PB ratio, PE ratio, EV/EBITDA)? One idea is to increase the value of the intangibles of a firm. In particular, note that data is the most important intangible asset of high-valued firms. Therefore, to increase firm value, the firm's managers should find a way to enhance their data and data-science capabilities. In sum, a data-driven value-up strategy would be a promising and logical way to increase the value of traditional firms suffering from low valuation ratios. In this paper, we propose specific instructions about how to formulate and undertake a data-driven value-up strategy. Our instructions are academically grounded and draw the insight from the literature about architectural innovation, the behavioral theory of the firm and the knowledge-based view of the firm. Our specific instructions are logically derived from our field observations and interviews on how data science is abused in dealing with meso-level data, whereas it is underused for macro-level data and enterprise risk-management to accomplish machine-human teaming.

This paper is conceptual and draws insight from a literature review and qualitative research. We expect that future studies would enrich our proposal on data-driven value-up strategies grounded on deeper qualitative case research as well as survey and policy-oriented research. Furthermore, we argue that our proposal could be applied to government policies. Increasing

[7] https://techcrunch.com/2017/06/23/five-building-blocks-of-a-data-driven-culture/

value economy-wide is a significant mission of any government. We believe that governments can play instrumental roles in implementing data-driven value-up policies as well as supporting/leading private sectors so that they can thrive in the fourth industrial revolution based on data-driven value-up strategies.

## REFERENCES


Angrist, J. D., & Pischke, J.-S. (2008). *Mostly harmless econometrics: An empiricist's companion*. Princeton university press.

Barney, J. B. (1986). Strategic factor markets: Expectations, luck, and business strategy. *Management Science*, *32*(10), 1231–1241.

Brynjolfsson, E., Hitt, L. M., & Kim, H. H. (2011). *Strength in Numbers: How Does Data-Driven Decisionmaking Affect Firm Performance?* (SSRN Scholarly Paper ID 1819486). Social Science Research Network. https://doi.org/10.2139/ssrn.1819486

Chesbrough, H. (2002, March 1). Making Sense of Corporate Venture Capital. *Harvard Business Review*, *March 2002*. https://hbr.org/2002/03/making-sense-of-corporate-venture-capital

Christensen, C. M. (2013). *The innovator's dilemma: When new technologies cause great firms to fail*. Harvard Business Review Press.

Cohen, W. M., & Levinthal, D. A. (1990). Absorptive capacity: A new perspective on learning and innovation. *Administrative Science Quarterly*, 128–152.

Conkle, T. (2020). *Council Post: Tempering Your Company's Expectations For AI*. Forbes. https://www.forbes.com/sites/forbestechcouncil/2020/09/10/tempering-your-companys-expectations-for-ai/

Conner, K. R. (1991). A historical comparison of resource-based theory and five schools of thought within industrial organization economics: Do we have a new theory of the firm? *Journal of Management*, *17*(1), 121–154.

Conner, K. R., & Prahalad, C. K. (1996). A resource-based theory of the firm: Knowledge versus opportunism. *Organization Science*, *7*(5), 477–501.

Cool, K., & Schendel, D. (1988). Performance differences among strategic group members. *Strategic Management Journal*, *9*(3), 207–223.

Cyert, R. M., & March, J. G. (1963). A behavioral theory of the firm. *Englewood Cliffs, NJ*, *2*(4), 169–187.

Dess, G. G., & Davis, P. S. (1984). Porter's (1980) generic strategies as determinants of strategic group membership and organizational performance. *Academy of Management Journal*, *27*(3), 467–488.

Dierickx, I., & Cool, K. (1989). Asset Stock Accumulation and Sustainability of Competitive Advantage. *Management Science*, *35*(12), 1504–1511. JSTOR.

Dosi, G., Nelson, R., & Winter, S. (2001). *The Nature and Dynamics of Organizational Capabilities*. OUP Oxford.

Eisenhardt, K. M., & Santos, F. M. (2002). Knowledge-based view: A new theory of strategy. *Handbook of Strategy and Management*, *1*(1), 139–164.

Gavetti, G., Levinthal, D., & Ocasio, W. (2007). Neo-Carnegie: The Carnegie School's Past, Present, and Reconstructing for the Future. *Organization Science*, *18*(3), 523–536. JSTOR.

Grant, R. M. (1996). Toward the knowledge-based theory of the firm. *Strategic Management Journal*, *17*(S2), 109–122.

Grant, R. M., & Baden-Fuller, C. (1995). The knowledge-based theory of inter-firm collaboration. *Academy of Management Proceedings*, *1995*(1), 17–21.

Henderson, R. M., & Clark, K. B. (1990). Architectural innovation: The reconfiguration of existing product technologies and the failure of established firms. *Administrative Science Quarterly*, 9–30.

Hisrich, R. D., & Jankowicz, A. D. (1990). Intuition in venture capital decisions: An exploratory study using a new technique. *Journal of Business Venturing*, *5*(1), 49–62.

Kane, G. C., Palmer, D., Phillips, A. N., Kiron, D., & Buckley, N. (2015). Strategy, not technology, drives digital transformation. *MIT Sloan Management Review and Deloitte University Press*, *14*(1–25).

Kang, H.-G., Bae, K., & Ku, B. (2019). Artificial Intelligence and Financial Investment Strategy. 한국경제포럼, *12*(3), 93–110.

Kapoor, R., & Lim, K. (2007). The impact of acquisitions on the productivity of inventors at semi-conductor firms: A synthesis of knowledge-based and incentive-based perspectives. *Academy of Management Journal*, *50*(5), 1133–1155.

Keynes, J. M. (1921). *A treatise on probability*. Courier Corporation.

Keynes, J. M. (1937). The general theory of employment. *The Quarterly Journal of Economics*, *51*(2), 209–223.

Kleinberg, J., Lakkaraju, H., Leskovec, J., Ludwig, J., & Mullainathan, S. (2018). Human decisions and machine predictions. *The Quarterly Journal of Economics*, *133*(1), 237–293.

Knight, F. H. (1921). *Risk, uncertainty and profit*. Courier Corporation.

Kogut, B., & Zander, U. (1996). What firms do? Coordination, identity, and learning. *Organization Science*, *7*(5), 502–518.

Lewis, M. (2004). *Moneyball: The art of winning an unfair game*. WW Norton & Company.

Lyons, J. B., Wynne, K. T., Mahoney, S., & Roebke, M. A. (2019). Trust and human-machine teaming: A qualitative study. In *Artificial Intelligence for the Internet of Everything* (pp. 101–116). Elsevier.

March, J. G. (1991). Exploration and exploitation in organizational learning. *Organization Science*, *2*(1), 71–87.

Mazzucato, M. (2011). The entrepreneurial state. *Soundings*, *49*(49), 131–142.

Pomerance, J. M., & McCarthy, D. (2018). Private Equity Investing 2018: Rethinking the Private Equity Investment Model to Build Value. *The Journal of Private Equity*, *22*(1), 10–16. https://doi.org/10.3905/jpe.2018.22.1.010

Porter, M. E. (1985). *Competitive Advantage: Creating and Sustaining Superior Performance* (Illustrated Edition). Free Press.

Ramirez, R., & Wilkinson, A. (2016). *Strategic Reframing: The Oxford Scenario Planning Approach*. Oxford University Press.

Roberts, M. R., & Whited, T. M. (2013). Endogeneity in empirical corporate finance1. In *Handbook of the Economics of Finance* (Vol. 2, pp. 493–572). Elsevier.

Schaefer, D., Walker, J., & Flynn, J. (2017). A data-driven business model framework for value capture in Industry 4.0. *Advances in Transdisciplinary Engineering*, *6*(2017), 245–250.

Sebastian-Coleman, L. (2018). *Navigating the Labyrinth: An Executive Guide to Data Management*. Technics Publications.

Sorescu, A. (2017). Data-driven business model innovation. *Journal of Product Innovation Management*, *34*(5), 691–696.

Subrahmanyan, S. N., & Jalona, S. (2020, February 28). Building a Data-Driven Culture from the Ground Up. *Harvard Business Review*. https://hbr.org/2020/02/building-a-data-driven-culture-from-the-ground-up

Thiel, P. A., & Masters, B. (2014). *Zero to one: Notes on startups, or how to build the future*. Broadway Business.

Waller, D. (2020, February 6). 10 Steps to Creating a Data-Driven Culture. *Harvard Business Review*. https://hbr.org/2020/02/10-steps-to-creating-a-data-driven-culture

Walliser, J. C., de Visser, E. J., Wiese, E., & Shaw, T. H. (2019). Team structure and team building improve human–machine teaming with autonomous agents. *Journal of Cognitive Engineering and Decision Making*, *13*(4), 258–278.

Yli-Renko, H., Autio, E., & Sapienza, H. J. (2001). Social capital, knowledge acquisition, and knowledge exploitation in young technology-based firms. *Strategic Management Journal*, *22*(6–7), 587–613. https://doi.org/10.1002/smj.183

**Table 1: Data value-up framework**

| # | Considerations | Approaches | Key literature |
|---|---|---|---|
| 1 | Innovation strategy | Architectural innovation | Henderson & Clark (1990) |
| 2 | Organizational change | A behavioral theory of the firm (BTF) | Cyert & March (1963) |
| 3 | Sustainable competitive advantage | Knowledge-based view of the firm; absorptive capacity | Cohen & Levinthal (1990); Kogut & Zander (1996) |

**Table 2: Innovation frameworks**

Panel A: Henderson & Clark (1990)'s classification

| | | Core concepts | |
|---|---|---|---|
| | | Reinforced | Overturned |
| Linkage between core concepts and components | Unchanged | Incremental | Modular |
| | Changed | Architectural | Radical |

메모 포함[JB1]: Please consider reorganizing text to the left of the table. It is difficult to read as is.

Panel B: Innovation types and data application

| Innovation types | Description | Data application | Target Area |
|---|---|---|---|
| Incremental | Conventional meaning and connection | Catch up data capabilities | Catch-up projects |
| Modular | Relationship unchanged, but updated and enhanced information | Generate new data or different interpretation of existing data | Fintech startups |
| Architectural | Constant core meaning; different relationships between meanings | Reconfigure existing system | Large, but traditional financial institutions |
| Radical | New architecture; new concepts | Design new system of knowledge and technologies | Big techs |

**Table 3: Organizational strategy for value-up on data science**

| | Knightian uncertainty high | Knightian uncertainty low |
|---|---|---|
| Stakeholder conflict high | High entrepreneurship + high shared value = Social entrepreneurship and non-market strategies | Low entrepreneurship + high shared value = Shared economy or opportunities in social impact |
| Stakeholder conflict low | High entrepreneurship + low shared value = Experiments and explorations | Low entrepreneurship + low shared value = Process innovation, digital twins |