Combining big data and lean startup methods for business model evolution

Article *in* Academy of Marketing Science Review · December 2017

DOI: 10.1007/s13162-017-0104-9

3 authors, including:

Steven Seggie
ESSEC - Ecole Supérieur des Sciences Economiques et Commerciales

Koen Pauwels
Northeastern University

AMS Review
https://doi.org/10.1007/s13162-017-0104-9

THEORY/CONCEPTUAL

Combining big data and lean startup methods for business model evolution

Steven H. Seggie[1] · Emre Soyer[2] · Koen H. Pauwels[3]

Received: 14 January 2017 / Accepted: 22 November 2017

Abstract
The continued survival of firms depends on successful innovation. Yet, legacy firms are struggling to adapt their business models to successfully innovate in the face of greater competition from both local and global startups. The authors propose that firms should build on the lean startup methodology to help adapt their business models while at the same time leveraging the resource advantages that they have as legacy corporations. This paper provides an integrated process for corporate innovation learning through combining the lean startup methodology with big data. By themselves, the volume, variety and velocity of big data may trigger confirmation bias, communication problems and illusions of control. However, the lean startup methodology has the potential to alleviate these complications. Specifically, firms should evolve their business models through fast verification of managerial hypotheses, innovation accounting and the build-measure-learn-loop cycle. Such advice is especially valid for environments with high levels of technological and demand uncertainty.

Keywords Business model · Innovation · Big data · Lean startup · Confirmation bias · Innovation accounting · Build-measure-learn-loop

> "Folks who were top growers had two capabilities that really stood out relative to their peers who were not growing as fast. One was, not surprisingly, around data and analytics. That might be a marker of what companies today who are really leaning forward—what are they investing capabilities in? We certainly see data and analytics as something that helps you squeeze out incremental growth on the margin. Then the second thing is the ability to collaborate cross-functionally to work in an agile method; this is another marker of top growers." (Mckinsey and Company 2017)

✉ Steven H. Seggie
seggie@essec.edu

Emre Soyer
emre.soyer@ozyegin.edu.tr

Koen H. Pauwels
k.pauwels@northeastern.edu

1 ESSEC Business School, Cergy-Pontoise, France

2 Faculty of Business, Ozyegin University, Istanbul, Turkey

3 D'Amore-McKim School of Business, Northeastern University, Boston, MA 02115, USA

Introduction

While there is no universally accepted definition of a business model, most researchers and managers agree that it includes a system of *processes* that serves as a business' *organizing logic for value creation and appropriation* (Zott and Amit 2010; Sorescu et al. 2011; Teece 2010; Blank 2010). One of the key sources of value for a firm is product innovation.[1] For a firm to ensure its long-term survival, it needs to be successful in its present markets through sufficient exploitation, while at the same time engaging in enough innovation to ensure that it is successful in the future (March 1991). In the 2016 Global Perspectives Barometer, about 800 "Leaders of Tomorrow" cited "innovation blindness" (the inability to recognize the need for a decision and staying passive in a quickly changing

[1] We use product innovation in the broadest meaning of the term to also include service innovation.

Published online: 12 December 2017

environment) as the most substantial risk for established companies in today's fast-paced markets (Neus et al. 2017, p. 35).

Product innovation separates top performing from bottom performing firms: A recent benchmarking survey revealed that the top performing quarter of firms saw 36.3% of sales coming from products launched in the last 3 years, versus only 10% for the bottom quarter performing firms (Edgett 2011). Yet successful product innovation is not an easy process for organizations. Castellion and Markham (2013) report that the average failure rate of product innovation projects is 40%, while Edgett (2011) reports that around half of all product innovation projects fail to reach the desired objective. While a risky undertaking, innovation is also essential for firms, as rapidly evolving new technologies, business models and competitors often end up threatening incumbents (Dyer et al. 2014).

Amongst these competitors are startups, which at first glance appear to have many advantages over the legacy corporations,[2] particularly with regard to the speed that they can operate at and bring business model innovations to market. For example, Airbnb's key partners are estate agencies and local hosts, neither of which are key partners for hotels. And while arguably the key resource for hotels is the real estate and the brand, the key resource for Airbnb is the technological platform that allows the matching of supply and demand (same is true also for transportation startups, such as Uber and Lyft). Startup success is driven both by the erosion of typical legacy firm advantages, such as capital and contacts, and by alternative business models that allow them to innovate at a substantially lower cost compared to only 10–15 years ago (Weiblen and Chesbrough 2015). As to the former, previously non-substitutable resources of legacy firms (Barney 1991) appear to have been substituted in a rich ecosystem of venture capital, contact networks and expert advice available to startups (Weiblen and Chesbrough 2015). As to the latter, new technologies have lowered the cost of reaching customers (e.g. online) and of designing and executing new solutions (e.g. cloud computing). In contrast, legacy companies suffer from innovation blindness caused by "holding onto outdated models and assumptions about how the world works" (Neus et al. 2017, p. 34) This raises the question of how legacy firms should evolve their business models to respond to the increasing competition from startups.

Some authors argue that incumbents should *imitate startups* in their organizations, in particular by adopting the lean startup methodology for product innovation (Furr and Dyer 2014; Blank 2013). In this way, the incumbent would be able to quickly adjust and adapt its business model to create and appropriate the most value (Blank 2013). We disagree, as a startup is by definition "an organization formed to search for a repeatable and scalable business model" (Blank 2010), while a legacy firm already has a business model that may require adjustments due to changing circumstances.

Moreover, we find the "death of legacy firms" to be exaggerated, as recent research shows that Christensen's (1997) famous theory of disruptive innovation fails to account even for many of the exemplary cases it highlighted (King and Baatartogtokh 2015). Unpredicted outcomes include disruptive innovations complementing incumbents by supporting their existing actions (e.g. credit scoring and business lending), and allowing incumbents to reach untapped markets (e.g. catalog sales and department stores). In addition, while startups have the ability to grow fast, the resulting growing pains also become a major challenge for them (Flamholtz and Aksehirli 2000). In contrast, legacy firms possess stable resources such as human capital, reputation, existing customers (customer equity) and known brands (brand equity) – which are extensively discussed in previous literature (e.g., Anand and Delios 2002; Carpenter et al. 2001; Kim and Ko 2012; Musteen et al. 2013), and can readily develop the skills of analyzing and acting on big data (Neus et al. 2017; Pauwels 2014), which we focus on in this paper.

While legacy companies should not imitate startups, they often need to *evolve* their business models to deal with the competition from startups, particularly with regard to the speed of innovation. Hence, we do not reject the notion that incumbents can learn from startups. We argue instead that incumbents can leverage their resources, specifically big data, while adapting rather than adopting the lean startup methodology.

Accordingly, we ask: *how can a legacy firm adapt its business model(s) to compete in innovation with the new generation of startups?* As visualized in Fig. 1, we begin by exploring the reasons many legacy incumbent firms fail to adapt their business models. We then discuss the lean startup methodology and how it can be used by legacy firms to adapt their business models. We then examine how firms can leverage the key resource of big data in the lean startup methodology and outline behavioral challenges to learning from big data. Our conceptual framework integrates these components by juxtaposing the 3 V's of big data (volume, variety and velocity) with the corresponding 3 behavioral challenges (confirmation, communication and control) and the recommendations from the lean startup methodology.

We make three main contributions to the literature on business model evolution for new product innovation. First, we provide an integrated *process* for corporate product innovation that demonstrates how firms should combine the lean startup methodology and big data to adapt their business models. Second, we discuss *learning* from big data and three major learning issues that impede innovation; and we discuss how our use of the lean startup methodology combined with big data analytics can help overcome these. Finally, we provide propositions on the conditions

[2] We use the terms 'legacy', 'incumbent' and 'established' throughout the paper to denote the firm or company has been operating for over a decade (Henderson and Clark 1990), irrespective of its current size. As pointed out by an anonymous reviewer, our advice applies to both larger and smaller incumbents.

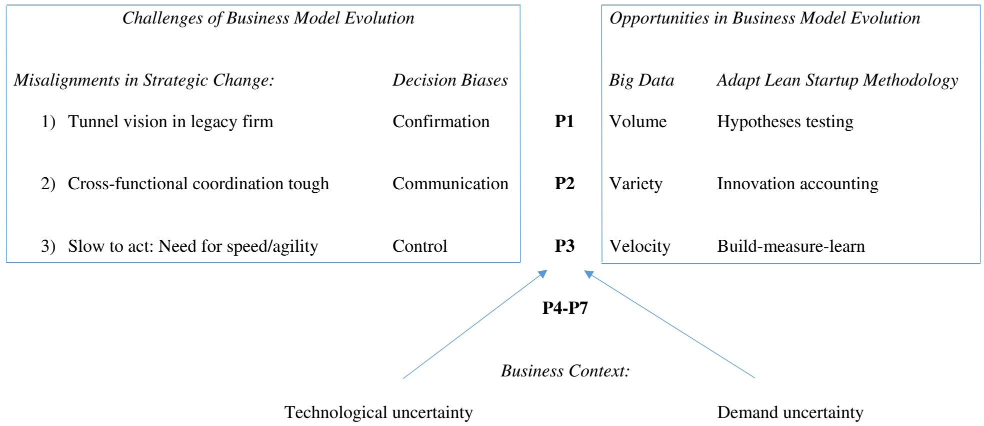

Fig. 1 Legacy firm business evolution challenges and opportunities in combining big data with lean startup methods (P = Proposition)

(of technological uncertainty and demand uncertainty) in which such a combination should especially benefit legacy firms in adapting their business models.

Why do legacy incumbent firms fail to adapt their business models?

The extant literature suggests that the execution of strategic change is complicated. In a study of nearly 8000 managers in 250 companies from 2010 to 2015, Sull and colleagues (2015) discovered several key problems with successful strategic change. First, managers believed successful execution was about alignment and used tools such as the balanced scorecard to ensure alignment. In fact, what was more important than alignment was an ability to *coordinate* between functional units. Second, managers believed that successful strategic change stemmed from having a plan and sticking to it, however, the reality was that no plan can truly predict events in the future and successful strategy change relies upon managers having the ability to *adapt*. Third, managers believed that communication and *understanding* were the same thing, however, in many instances even though managers considered themselves to be good communicators, strategic objectives were not clearly understood within the organization. Fourth, while a performance culture is important for successful strategic execution, the ability of employees and managers to be *agile, engage in teamwork and have ambition* is also crucial and needs to be rewarded. Finally, managers believed that strategic execution should be driven from the top to ensure success, while in reality *execution needs to be driven from the middle* and guided from the top.

Besides these problems, strategic change is hard to execute due to pressures from shareholders for short-term profits, which means that managers are frequently trying to redefine strategy, limiting the time to engage in successful strategic execution (MacDiarmid et al. 1998). This also leads to managers facing the difficult challenge of balancing strategic change with the demands of the organization at the present time.

Moreover, organizations are becoming more and more complex with greater activities across functional areas and geographies (Miles et al. 1997). Consequently, any structural change affects people, processes and structures across these boundaries, increasing the level of complexity and thus increasing the likelihood of failure (Abell 1999; Johnson-Cramer et al. 2007). Managers are also competing with other managers for scarce resources within the organization. As such, it is very difficult for managers to get resources to execute the change required. And, without these resources, the strategic change cannot be conducted successfully (Staw 1976). Finally, in legacy organizations it is often difficult to involve managers from all the functional areas at the early stages of strategy execution. This is essential to ensure commitment throughout the organization, however, it can be very cumbersome to organize and be perceived as slowing things down (Balogun 2006).

Examples where legacy incumbent firms were severely damaged by competition demonstrate the difficulties that firms have in adapting their business models. A 2007 Fortune magazine cover declared Nokia to be the undisputed king of the cell phone world, asking who, if anyone, could possibly catch Nokia in the near future. Ironically, the iPhone was launched in the same year, a few months prior to the release of the magazine. In a separate industry, hotels have struggled to compete against startups such as Airbnb, losing an estimated $450 million in yearly revenues (Mahmoud 2016). Similarly, Uber and other ride sharing apps continue to gain market share against taxicabs.

The reasons for these incumbent business model challenges are multifold: (1) the customer value proposition is no longer as attractive as it once was, (2) the key resources to deliver the value proposition (e.g., technology, people, channels etc.) no longer exist in the firm, (3) the firm's ability to make a profit has been dented, or (4) there are problems with the processes, such as training, budgeting, metrics, and norms (Johnson et al. 2008). Regardless of specifics, however, firms have struggled to adapt their business models in the face of new forms of competition.

An inability to adapt a business model, while of clear importance, has not been widely studied. Halecker et al. (2014) examine unsuccessful business' model innovation processes and demonstrate that an inability to adapt occurs for a plethora of reasons and due to a combination of many factors. These include negative customer reviews and hardships in adapting to a foreign market. Demil and Lecocq (2010) argue that a static view of what a business model *should be*, and a dynamic view of how it *is evolving* over time are complementary issues that need to be addressed simultaneously.

A business model goes through three stages of evolution, from the creation of the business model to the growing of the business unit, and finally to the increased desire for efficiencies (Christensen et al. 2016). This sort of linear process is appropriate for competing in a linear world where competition is known and takes place on a fairly level playing field. However, when faced with startup challengers, such incremental business model innovation may be more detrimental than beneficial to the health of the company. Therefore, we draw on the literature on *ambidextrous* organizations for inspiration as to how legacy companies could quickly and effectively compete with startups.

An ambidextrous organization competes both in the mature market where it performs at present, while at the same time developing new innovations for new markets and customers. A recent review of this literature by O'Reilly and Tushman (2013) shows that such organizations generally perform better, although this is contingent upon the context where the firm operates. In particular, ambidexterity is seen to have a greater impact on performance under uncertainty and when a firm has sufficient resources. Furthermore, ambidexterity in the organization aids strategic execution by addressing some of the challenges we noted previously. For example, Vinekar et al. (2006) note the ability of organizational ambidexterity to harness the benefits of agile development. Furthermore, the ambidextrous organization has integration mechanisms and teams due to its paradoxical strategies of improving execution of the present business and inventing new business (Smith et al. 2010), thus ensuring involvement from all throughout the organization.

For our purpose, i.e. how should firms adapt in response to the threats posed by startups, a key issue is *how* firms can actually become ambidextrous. Scholars have suggested two main ways. One is through structural adaptation, i.e. separating business units to engage in exploration and exploitation and making sure that each of these units has separate competencies and cultures. (O'Reilly and Tushman 2008). Another alternative is contextual ambidexterity where the problem is solved at the individual level rather than at the structural level. As such, the individual is provided with a context that allows him/her to make an appropriate decision as to how best to spend his/her time, i.e. on exploitation tasks or exploration tasks (Gibson and Birkinshaw 2004).

Essentially, for firms to be ambidextrous and innovative requires them to be able to update their business model when necessary. Yet, engaging in such an evolutionary approach to business models is very challenging. For example, Hewlett-Packard was hailed in 1996 as one of three examples of successful ambidextrous organizations (Tushman and O'Reilly 1996). Only 6 years later, in 2002, Hewlett-Packard had a dysfunctional board and faced an abundance of criticism for its merger with Compaq (Bandler and Burke 2012).

In sum, while firms need to adapt their business models to compete in innovation against startups, there is no clear best practice for how firms should do this. Recently, corporations started to rethink their business model to "innovate like startups," with the promise that they "have everything to gain when they adopt the process, strategies, and mentality of high-growth startups" (Owens and Fernandez 2015, book summary). However, such learning may encounter many pitfalls, as legacy corporations fundamentally differ from startups in their organizing logic and in their learning processes. Indeed, March (2011) offers a wide variety of challenges legacy organizations face when acquiring actionable intelligence, processing past outcomes, and adapting to a changing business environment. The unpredictability of innovative processes, strictness of the corporate procedures and lack of organizational slack can limit the necessary experimentation and exploration for discovery and innovations (Nohria and Gulati 1996). Hence, to achieve ambidexterity, we argue that firms should build on the *lean startup methodology* to help *adapt their business models* while at the same time *leveraging the resource advantages* that they have as legacy corporations. In this way, we also deal with two outstanding reasons for failure of strategic change: a need for speed and agility, and difficulty to access resources.

A summary of these issues and the respective studies can be seen in Table 1.

The lean startup methodology and innovation

The lean startup methodology is a quick and iterative process that requires minimal resources compared to more traditional models of innovation (Blank 2013). It's a practitioner based idea yet to find its way into the academic literature.[3] It has been developed by entrepreneurs and scholars such as Steve Blank, Eric Ries, and Bob Dorf and is now widely used by startups as a means to help them to grow their businesses quickly and efficiently. Leading business accelerators such as 500 StartUps encourage the use of this methodology to help firms

[3] The one exception that we are aware of is an unpublished manuscript by Delvecchio et al. (2013) who compare the lean startup methodology with the stage-gate model.

Table 1 Selected literature synthesis of difficulties in executing strategic change and adapting business models*

Authors (Year)	Perception misalignment	Execution misalignment	Lack of ambidexterity	Evolving business model (BM)	Acting like a startup	Inability to adapt	Short-termism	Organization complexity	Limited resources
Abell (1999)								Leads to failure-prone complex systems	
Johnson-Cramer et al. (2007)								Leads to failure-prone complex systems	
MacDiarmid et al. (1998)		Due to short-term profits vs. long-term goals					Due to shareholder pressure		
Miles et al. (1997)								Due to the diversity of functions and geographies	
Owens and Fernandez (2015)					Due to different processes of legacy firms				
O'Reilly and Tushman (2013)			Leads to inability to manage uncertainty						
Staw (1976)									Due to competition among managers
Sull et al. (2015)	Due to communication problems	Due to coordination problems		Due to BM not driven from middle and guided by the top		Leads to problems due to uncertainty			
Vinekar et al. (2006)			Leads to frictions towards agility						

*Some studies focus on the reasons (denoted in the cells as 'due to'), others on the consequences ('leads to') of the difficulties with their growth. It is widely accepted as best practice for startups, and in fact, we observe the existence of a lean startup movement (Egusa 2013). It borrows many of its ideas from lean manufacturing and attempts to remove the fat from the product innovation process by eliminating wasteful practices.

While at first glance, the lean startup methodology possesses certain similarities to the stage-gate model (e.g., Cooper 1990; Cooper 2014), it is different. While both methodologies argue for an iterative approach to product innovation, the two most important differences between them are in how *decisions* are made and in what the *process* looks like. In the stage-gate model there is an assumption of product / market fit, therefore the decision at the end of each stage is to continue or terminate the project. However, with the lean start up methodology, the assumption is that the firm is still searching for a business model; it is still searching for product-market fit– therefore the decision is to continue or pivot.

Originally introduced in this context by entrepreneur Eric Ries (2009), pivoting means changing direction quickly, but staying grounded in what you learned (keep one foot in the past and place one foot in a new possible future) as you search for product-market fit. An example is the robotic lawn mower idea by Blue River Technology, who learned through 100 interviews that the initial target customer segment, golf course owners, simply did not see enough value in the product. Instead, farmers saw value in an automatic way to kill weeds without using chemicals, and the company pivoted to such product and customer segment, building and testing a prototype within 10 weeks (Blank 2013).

In the stage-gate methodology, the process is one of refinement, while in the lean startup the process is one of testing hypotheses (Delvecchio et al. 2013). Essentially, the lean startup methodology is based on two key notions: (i) business plans do not survive beyond contact with the first customers[4]; hence investing time in them is inefficient, and (ii) startups are not just small versions of legacy companies; instead they are constantly adapting, iterating, and learning from their customers (Blank 2013). As with any methodology, the lean startup has underlying principles. These are: (a) entrepreneurs do not know much about their market on day one and all they have are several good guesses and hypotheses, (b) to test these hypotheses, entrepreneurs need to get out of the building and talk to customers, and (c) products should be developed iteratively and incrementally rather than in long product development cycles (ibid.).

The principles of the lean startup methodology allow for the systematic reduction of risk and for the management of the ubiquitous uncertainty in the product innovation process. Proponents of the lean startup methodology (e.g., Ries 2011) argue that iterative development is managed through the Build-Measure-Learn loop. In this loop, a minimum viable product (MVP) is built and then that MVP is tested with customers and tweaks are made to change it in line with customer feedback. The MVP is a product with the minimum set of features that solves the problem of the customer.

A frequently cited example of the model in action is Zappos. Its founder, Nick Swinmurn, used customer development to test his assumptions about his business. In particular, Swinmurn had developed the hypothesis that customers would be willing to buy shoes online. In a more traditional type of innovation process, Swinmurn would have developed a fully functional website and database of footwear in stealth mode. Once it was completed (after investing a lot of money), he would have launched the website to an unsuspecting public and competition. Instead, he used lean startup methods and set about developing an MVP to test whether or not this particular hypothesis was supported. He took photographs of shoes in local shoe stores and posted these pictures to a web page. Then if a customer wanted to buy the shoes online, he went to the local store, purchased the shoes at full price from the store and sent them to the customers. Thus, he was able to validate his hypothesis that customers would be willing to purchase shoes online. Apart from Zappos, many other startups including the likes of Airbnb and Dropbox have used the lean startup methodology to grow relatively quickly into multi-billion dollar companies.

Recently there has been discussion (e.g. Furr and Dyer 2014) about how to bring the principles of the lean startup methodology into the product innovation of legacy companies. This appears a tough task as the fundamental dynamics, organizational structure, and incentive mechanisms of startups and legacy corporations are different. For example, startup companies are in search mode, looking for a scalable business model. This is not the case for established companies that already have a reasonably successful business model and are focused on the execution of that model.

Moreover, incumbent organizations do not have a similar degree of chaos and lack of knowledge or marketing schemes that startups have. While failure is a way to learn for startups, it could lead to employees losing their jobs in a legacy organization. These differences alone are enough to demonstrate that, much the same as a startup is not a small version of a legacy company, neither is a legacy company just a large version of a startup. Hence, despite 'startup envy' for their innovativeness and lower cost levels, legacy incumbent organizations cannot take these lean startup best practices without adapting them to their corporate environment.

Therefore, we propose that incumbent firms adapt their business model using some of the best practices of the lean startup methodology to supplement the resources that legacy

[4] Prior to the use of the lean startup methodology, startups traditionally developed a business plan for their product/service while making various assumptions about customers before they launched the product or service. However, more often than not these assumptions were found to be wrong when the startup had its first contact with customers, i.e. when it tried to sell its product/service. This then rendered the business plan redundant.

organizations already have. In particular, we *consider the improvements the lean startup methodology can bring to innovation through its combination with big data*. We choose to focus on the resource of big data for three main reasons.

First, legacy corporations have access to big data and furthermore they have the ability to analyze and act on them (Pauwels 2014). This is a clear advantage that incumbent firms have over many startups that are focusing on core issues of the business. Second, big data analytics allows legacy companies to test hypotheses with less risk, including that of employees losing their jobs. Third, we are seeing more and more industries becoming competitive due to increased technological uncertainty, and the research suggests that as competition in an industry increases the advantages that incumbent firms have vis-à-vis big data become even more important. Germann et al. (2013) show that deploying marketing analytics increases return on assets by 8% on average, but by 21% in highly competitive industries. Therefore, it seems proper for us to consider how big data can complement the lean startup methodology in helping incumbent firms to adapt their business models to compete in innovation.

Learning from big data

Big data, and learning from it are hot topics, with assertions that we live in "The Age of Big Data" (New York Times 2012) and that it will revolutionize everything that we do, including product innovation processes (Mayer-Schönberger and Cukier 2013). According to a McKinsey Global Institute report (Manyika et al. 2011), big data will become a key basis for innovation in five major ways.

First, big data can make information *transparent and usable* at much higher frequency. Second, big data can help organizations *conduct controlled experiments* to make better management decisions and to timely adjust their business levers. Third, big data allows ever-narrower segmentation of customers and therefore much *more precisely tailored products or services*. Fourth, sophisticated analytics can substantially improve decision-making. Finally, big data can be used directly to improve the *development of the next generation* of products and services. For instance, manufacturers are using data obtained from sensors embedded in products to create innovative after-sales service offerings such as proactive maintenance (preventive measures that take place before a failure occurs or is even noticed).

Of course, the main advantage does not come from "the size of the data, but what you do with it" (Pauwels 2014). According to Gartner's (2015) hype cycle, big data is nearing the end of the hype stage and entering the phase of (potential) disillusionment if not used correctly. Indeed, the Boston Consultancy Group warns that "with the rise of big data, companies risk magnifying the impact of underlying inaccuracies and errors and falling into a big bad data trap" (Saleh et al. 2015, p.1). Hence, the question arises as to how firms can make use of this big data in their innovation processes to help them better manage the technological uncertainty and compete better against startups new to the industry. For instance, does richer data always benefit innovation? Are there any mechanisms through which data-rich environments can in fact harm innovation processes, rather than enhancing them? How could decision makers identify and mitigate potential traps? These are all questions that it is incumbent upon us to answer when we consider the role of big data in the innovation processes of legacy companies.

Accordingly, we discuss below three key characteristics of big data (3 Vs: volume, variety and velocity) and introduce the corresponding learning traps (3 Cs: confirmation, communication and control) these may deepen, which would distort innovative processes. We then discuss how the lean startup methodology can alleviate these. Table 2 summarizes our conceptual framework.

The 3 vs of big data

The big data sets that companies have access to today possess three main characteristics that set them apart from the analytics of the past: volume, variety and velocity (Laney 2001; McAfee and Brynjolfsson 2012). First, the sheer volume of data can be a challenge for managers (Manyika et al. 2011). Second, the sources of big data are much more *varied* than sources have been in the past. Whether this be in the form of tweets, GPS signals, Amazon reviews, or whatever the next big thing is, the variety is far greater than it has ever been raising challenges as to how to use this data in the innovation and new product development processes. A recent survey revealed that 67% of managers considered handling variety as their main big data challenge, before volume and velocity (New Vantage Partners 2016). Third is the velocity at which the data is available in real time or almost in real time. How to capture this data and incorporate it into the corporate innovation process is a real challenge. On the one hand, this will provide the firm with greater agility to compete better with startups, but on the other, the firm will likely lack processes and methodologies to make use of the data and will need to develop these (Du et al. 2015; Tirunillai and Tellis 2014).

3 corresponding key learning challenges

While big data offers much potential, it also needs to overcome several learning traps. From a decision-making perspective, any data-based learning involves two settings. In the first (learning setting), decision makers evaluate the information relevant to their objectives. In the second (target setting), they make predictions and/or choices based on these evaluations. A mismatch between learning and target settings often leads to

Table 2 Conceptual framework of big data, behavioral challenges and lean startup advice

Big data characteristics	Learning challenges	Lean startup methodology advice
Volume	Confirmation	Identify and test hypotheses
Increasing amount of data	*More opportunities confirm prior beliefs, while discarding disconfirming evidence*	*Use big data to come up with hypotheses to test through experimentation, instead of confirming them through the data alone*
Variety	Communication	Innovation accounting
Different types of data	*Increased complexity of data and analyses make it difficult to communicate insights for decision making*	*Incremental innovation helps break down the complexity into simpler processes and metrics, which can be communicated more easily*
Velocity	Control	Loop in build-measure-learn
Real time data	*Increased sensation of predictability due to immediate and rich information about past performance*	*Agility and ambidexterity, which help challenge data driven illusions and mitigate unprecedented crises*

Items in italics are explanations of the terms in standard text

systematic inferential errors (Hogarth et al. 2015). Recognizing these potential mismatches is crucial to mitigating them.

Consider for example a company aiming to improve its business model. To this end, managers decide to gather extensive and detailed evidence on the successful models and best practices, and try to apply these to their own firm. A problem arises, however, if this learning setting hides crucial information from them. One such discrepancy for instance can be due to survivorship bias: what if a wide variety of companies with similar practices have failed? It could even be that in the target setting where managers operate, such practices are actually detrimental to success and the minority that survived was merely lucky and achieved their objectives *despite* said practices (Einhorn and Hogarth 1978; Denrell 2003; Denrell et al. 2014; Soyer and Hogarth 2015). Crucially, more research on the subject could further reinforce such inferential errors and the resulting misconceptions would be harder to reverse, now that they are data-approved.

Hence, there are multiple ways in which data-fueled learning can impede innovation. We summarize them under three categories. In particular, volume may lead to deceptive confirmations, variety to miscommunications and velocity to illusions of control.

Confirmation

When data is plenty (volume), it becomes relatively easier for decision makers to use it to validate the answers they are looking for. However, if they selectively and systematically undervalue or disregard information that disconfirms their prior beliefs during this validation procedure, more information would lead to misconceptions (Nickerson 1998).

Successful innovation requires accurate validation of hypotheses. Baron (2000), however, reveals two ways in which big data may ease the confirmation of a false hypothesis. First, data on which the analyses are based may be constructed through a biased search. This would cause a mismatch between learning and target settings, as the data collection procedure filters out disconfirming evidence. Consequently, any conclusion that stems from such data would lead to predestined results.

Second, even when data incorporates representative disconfirming evidence, managers may still be able to confirm their prior beliefs through their choice of analysis. In this case, the mismatch between the two settings would be due to the underlying assumptions and statistical procedures employed. As a result, correlations can be confused with causations, or similar data may lead to two competing insights. A recent replication analysis uncovered that even rigorously conducted and peer reviewed scientific research may not be immune to data driven confirmations (Ioannidis 2005).

Communication

A relatively less scrutinized factor that may cause a mismatch between learning and target settings is the communication of the big data generated insights. Even if data is collected and analyzed in an unbiased fashion, the way the results are summarized may lead to misperceptions about the reality in which the company finds itself. In particular, while big data scientists are statistically sophisticated, the decision makers and managers that act on the generated insights may be less so. As a result, crucial parts of the messages that stem from data analysis can be open to confusion or misinterpretation. This is especially prominent because big data sources show a lot of variety, from anecdotes and quotes available in social media text, to clickstream data on (prospective) customer online behavior, to social network analysis. To be useful for innovation, it is essential that data driven insights are not lost in translation.

Simulations can be used to translate big data analysis into decision-friendly insights (Hanssens and Pauwels 2016; Hogarth and Soyer 2015). Our technological capacity to process information also increases along with the 3Vs of big data. Such improvements in computing power allow for the construction of data driven simulations with which decision makers can interact. These tools let decision makers enter their inputs and observe outcomes produced by the statistical analyses conducted on big data. Experimental studies on the subject show that people trust such personally simulated experiences and make accurate judgments in situations involving uncertainties and complexities (Hogarth and Soyer 2011; Sterman 2011; Bradbury et al. 2014). Hence, simulations can be employed when decision makers need to communicate the uncertainties inherent in innovative processes.

Control

A major motivation behind any big data approach is to be able to more accurately predict the future based on available information about the past. Here, past is the learning setting, whereas future is the target, and evidence based control would depend on the approximate match between these two. In principle, the high velocity of big data should help, with managers and stakeholders obtaining very fast, sometimes real-time updates on key performance indicators (Tirunillai and Tellis 2014). Unfortunately, these online metrics often don't match with the company's goals (Peters et al. 2013) and predict brand performance poorly in the long run (Pauwels and van Ewijk 2013). A crucial question then becomes: can we measure all the factors that contribute to a future event we would like to predict? In fact, there are at least two elements we might not be able to accurately account for, which would cause a mismatch between what a company learns and what it wants to achieve.

The first is the set of past outcomes that are either not measured or no longer observable. This is similar to the situation we described above, which involved survivorship bias. Such a mismatch between data and reality leads to an illusion of control through common strategies of only those who succeeded.

Acknowledging this survivorship filter would allow managers to design mechanisms to correct for its adverse effects on data-generated insights. For instance, one important issue to consider when collecting and analyzing data would be to accurately reveal the base rate of success of a given innovative process, rather than focusing on its specific determinants. If such base rate were extremely low, then investing too much on a particular strategy would be wasteful. Instead, a lean business model that allows for a large number of trials and errors would be more appropriate (see below).

The second element, of which the measurement is problematic, is part of the uncertainties inherent in the outcomes (Makridakis et al. 2009). In particular, consider two types of uncertainty. We can estimate one of them and forecast its effects with accuracy. An example would be the daily fluctuations of a stock around its trend. The second type, however, we cannot accurately estimate and predict. An example would be the probability of an unprecedented fall in the same stock's value due to a global financial crisis. This latter phenomenon is often referred to as a Black Swan event (Taleb 2010). In particular, such level of unpredictability of the environment causes a mismatch between the past and the future, which cannot be mitigated directly through the 3Vs of big data.

In short, we summarize challenges to learning from big data in the product innovation processes under three key categories. First, there is a potential problem with confirmation bias whereby firms confirm their own assumptions and beliefs, for example, about the type of innovation customers would want. Second, there may be a communication problem whereby results of analysis of the big data are not communicated accurately to decision-makers, leading to misconceptions. And finally, there may be a problem in predicting the actual level of uncertainty in the innovation processes.

Learning from big data through the lean startup methodology

In this section, we explain how the lean startup methodology can be combined with big data in the corporate innovation process to address the learning challenges we outlined above. We formalize these explanations in three general propositions that are related to business model innovation.

We proceed by introducing the business model canvas (Osterwalder and Pigneur 2010). This canvas is used as part of the lean startup methodology to evolve the business model. The business model canvas consists of nine integrated components: customer segments, value propositions, channels, customer relationships, revenue streams, key resources, key activities, key partnerships, and cost structure. These are the building blocks that demonstrate the logic of how a firm creates and appropriates value, i.e. the business model. Underlying this canvas is the notion that the business model evolves in all or some of these components and the lean startup methodology drives this evolution. As such, when firms are launching a new product or service, they fill in the canvas with a series of guesses or hypotheses about each of the components, which they then proceed to test. Through iterative testing of these hypotheses, the firm will evolve the business model until it reaches a model that can be used to launch the product or service.

Lean startup methodology and confirmation

Experimentation, which is a key component of the lean startup methodology, can be effectively incorporated into legacy companies' innovation processes to counter a big data compounded confirmation bias. Authors including Pfeffer and Sutton (2006) and Ariely (2010) have noted that many big firms fail to conduct experiments. The lean startup methodology follows the scientific method in that it advocates the creation of hypotheses and the testing of these through experimentation. We propose that companies use big data analysis not directly to reach conclusions (e.g. we should reach our younger customers through an online channel) but instead to derive a set of hypotheses (e.g. H1: Customers under 25 years of age want to be reached about our product/service through an online channel, H2: The target customer is willing to pay five dollars for one day delivery, etc.). Thus, firms can analyze historical data to hypothesize about specific parts of the business model (in the example above H1 is a hypothesis about the channel to reach customers while H2 is a hypothesis about a revenue stream. Both are key components of the firm's business model). Testing the hypotheses derived from big data will allow the firm to evolve its business model efficiently.

This approach mitigates reaching false conclusions from big data. Even if managers were predisposed to find evidence supporting the answers they already had in their minds, this would only translate into a set of hypotheses about the possible business model for the innovation. If confirmation bias leads to false hypotheses, they would next be disconfirmed by the subsequent tests.

In this treatment of hypothesis testing, the lean startup methodology has a lot in common with scientific research that specifically looks for disconfirming evidence to a hypothesis (Popper 1959). For instance, the Dutch furniture company Inofec spent 80% of its marketing budget on expensive direct mail to lists of previous and potential customers, including hospitals, businesses and government offices. This notion was based on the assumption that such offline marketing was the appropriate channel to reach customers, and managers used the high share of offline-to-online sales to back up this claim (Pauwels 2014). However, merging data across online and offline platforms and analyzing all marketing effects showed that other actions, such as online paid search, were responsible for much of the offline sales (Wiesel et al. 2011). Talking to prospective customers revealed that the direct mail hit them at the wrong time, and that many disregarded it as a result. For every euro spent on direct mail, the company only got back about 55 euro cents in profits. In contrast, paid search was very effective because it targeted the right customer at the right time, even when most of those customers subsequently moved to the offline channel to negotiate a deal with the company's sales force. Such time-based targeting is especially important for products high in situational importance (Pauwels et al. 2016), i.e., people don't typically pay much attention to products like furniture or refrigerators until they truly need to buy them.

Use of the lean startup methodology would have prevented direct mail effectiveness to be taken for granted. Even if the managers had argued for direct mail as the best channel to generate offline sales, with a lean startup approach this would not have been taken as fact, but as a hypothesis to be tested and it would have been found to be faulty within a short period of time. This finding would have then allowed managers to develop further sets of hypotheses until they discovered that paid search was the most effective.

Proposition 1: Legacy companies that use big data to derive a set of testable hypotheses in their business models will outperform legacy companies that use big data to directly reach conclusions about their business models.

Lean startup methodology and communication

The lean startup methodology helps in mitigating problems arising from communication of data driven insights in two ways. First, as we already mentioned vis-à-vis confirmation, it advocates using big data analysis to develop hypotheses in the business model that will then be tested. The relative simplicity of these hypotheses alleviates much of the complexity that creates the miscommunications when results of analyses are presented. Furthermore, the experiments used to test the hypotheses are conducted relatively quickly and continuously in the organization. Such ongoing process increases the likelihood of identifying and correcting any misperceptions that may arise. In short, the changes in the communication process required by the lean startup methodology means more regular communication with less being communicated at each stage, thus reducing the cognitive burden and ensuring more accurate communication of the evolution of the business model.

Second, there is one other insurance mechanism built into the lean startup methodology that increases the likelihood of accurate communication: the concept of *innovation accounting* (Ries 2011). It consists of three steps that firms can use to measure the progress of innovation projects and communicate to managers as to how the innovation is progressing. The first step is to get a baseline of where a particular innovation project is. The second step is to make the changes necessary to get the MVP from stage one to an ideal product. The third step is the decision to pivot or persevere. In short, innovation accounting is used to help determine, measure and communicate progress of innovation.

For innovation projects at the MVP stage, research suggests that so-called love metrics are appropriate ones to measure, which include customer enthusiasm, net promoter score and likelihood of customers to pay for the product (Furr and Dyer 2014). These metrics would be measured to get a benchmark score for the innovation project and communicated to senior management. They would then be tracked as changes were made to the MVP in line with customer feedback and success would be measured with regard to improvement (or not) of these particular measurements. After multiple hypotheses are tested, the metrics would then be used to decide whether to persevere or pivot the business model. Given that in the previous section we argued for firms to use big data analysis to derive hypotheses rather than develop conclusions, the use of innovation accounting will complement this use of big data and help reduce the likelihood of miscommunication. By only tracking a limited number of metrics resulting from these hypotheses, innovation accounting allows for a simplicity, consistency and ease of visualization of the innovation process.

Proposition 2: Legacy firms that use the lean startup methodology and innovation accounting (building and pivoting a minimum viable product) will outperform legacy firms that use the traditional stage-gate innovation process.

Lean startup methodology and control

The lean startup methodology was developed partially in response to the increasing level of uncertainty that exists for startups. This uncertainty is partly technological, in the sense that technology in many industries is changing rapidly. Furthermore, startups operate under extreme demand uncertainty as they do not know whether or not and how much customers will buy from them. The iterative nature of the lean startup methodology whereby firms engage in a build-measure-learn loop (Ries 2011) allows for validated learning to take place and for firms to make incremental developments on the business model based upon feedback from customers. And as noted by Ries on p.46,

> It is a rigorous method for demonstrating progress when one is embedded in the soil of extreme uncertainty in which startups grow. Validated learning is the process of demonstrating empirically that a team has discovered valuable truths about a startup's present and future business prospects. It is more concrete, more accurate, and faster than market forecasting or classical business planning. It is the principal antidote to the lethal problem of achieving failure: successfully executing a plan that leads nowhere.

Since legacy firms also face a similar uncertainty, the same logic applies. Just as startups need to learn as much as possible about customers and what they want, legacy firms also need to obtain the same insights. A proper implementation of the lean startup methodology in the innovation process will help firms discover customer wants through the validated learning that takes place. This then complements the forecasting methods being used in the big data approach. As Neus et al. (2017, p. 35) advise: "Acknowledging the inherent risk in a truly new development implies making many small bets by seed funding parallel projects and accepting that most of them will not work. The key lies in then selecting and scaling those minimal viable products that have shown to work."

Hence, the lean startup methodology provides firms with the *agility* and ambidexterity necessary to navigate potential Black Swans and unprecedented crises they may face. The part of the organizational structure that is devoted to continuously innovate in an incremental fashion will provide two benefits in this context. First, these innovations will effectively serve as reality checks to any illusion of control on future prospects induced by big data. Second, in case of a crisis, ideas, products or services produced through the lean startup methodology can serve as a life jacket, helping the legacy firm avoid ruin. Thus, the learn startup methodology should enable legacy firms to connect agility with the right resources in an uncertain world.

Proposition 3: Legacy firms that use the lean startup methodology will manage uncertainty to generate innovation more effectively than legacy firms that do not.

Boundary conditions: demand and technological uncertainty

While our conceptual framework and advice is general, we acknowledge that these recommendations may be more important under certain conditions and less important under others, which leads us to formulate additional testable research propositions. We base our discussion of boundary conditions on the distinction between demand uncertainty and technological uncertainty. We examine the impact of these boundary conditions on the use of the lean startup methodology and big data analytics, given the components of the business model canvas that we introduced earlier in the manuscript (Osterwalder and Pigneur 2010).

The business model canvas contains nine components of the business model: customer segments, value propositions, channels, customer relationships, revenue streams, key resources, key activities, key partnerships, and cost structure. In our propositions, we argue how legacy firms would perform under different combinations of boundary conditions, while using the lean startup methodology and big data analytics for developing and testing hypotheses for some or all components of the business model canvas.

Demand uncertainty is the question of whether or not customers will buy a product. Higher demand uncertainty exists when there are a greater number of unknowns about the preferences and behaviors of customers (Furr and Dyer

2014). Technological uncertainty is the question of whether or not we can create a solution that customers will buy. High technological uncertainty exists when we are unsure about the technologies that may appear or that may be required to solve a particular customer problem (ibid.). For different innovations that firms engage in, there may be different combinations of technological uncertainty and demand uncertainty, thus necessitating conditions as to how the firm would integrate the lean startup methodology and big data analytics. We split both types of uncertainty into two levels: high and low, thus creating four combinations of technological and demand uncertainty (see Table 3). We discuss each of these and develop research propositions.

High technological uncertainty and high demand uncertainty

Industries such as software and medical equipment face both high technological and high demand uncertainty (Dyer et al. 2014). For example, few would have predicted that there would be a demand for robots performing surgery and developing these robots as solutions for such a critical task as performing operations requires spending large sums of money with no certain outcome (Furr and Dyer 2014). This high demand uncertainty in the medical equipment industry is reflected in relatively high firm turnover (13.1%) and revenue volatility over 10 years (90.7%), while the high technological uncertainty is reflected in relatively high industry R&D spend as a % of sales (8.2%) (ibid.).

A similar situation would be when a firm is entering into an unknown market with a completely new product (to the firm). In this condition, the firm is unsure whether the customers will buy the product or not (as it has little knowledge of the customers in this particular market and time) and also does not know if it can create a viable solution for the customers. In short, the firm does not know if the value proposition component of the business model canvas is able to deliver value to the customer and satisfy his/her needs. In addition, the firm is likely to have no real understanding of appropriate channels to reach customers, how to establish and maintain customer relationships and so on. In such a case, just making use of big data analytics to decide on the business model would be suboptimal. Instead, the firm should incorporate the lean startup methodology and big data analytics into the process of developing a business model. Big data analytics should be used to develop hypotheses in the business model canvas. Subsequently, the hypothesis testing leading to validated learning can be implemented with the use of innovation accounting to measure the progress of the innovation project. This will then lead to an evolution toward a business model that can be used to launch the product or service having mitigated much of the uncertainty. Hence:

Table 3 Boundary conditions: uncertainty and innovation

		Technological uncertainty	
		High	Low
Demand uncertainty	High	When a firm engages in innovation in a new market or industry with a completely new product, or in a market or industry where the potential for disruption appears high. *Integrate Lean Startup Methodology and Big Data analytics into developing and testing hypotheses for the full business model*	When a firm engages in innovation in a new market or industry with an incrementally improved product. *Only integrate Lean Startup Methodology and Big Data analytics in developing and testing hypotheses related to customer segments*
	Low	When a firm engages in innovation in a market or industry that it knows well with a completely new product. *Only integrate Lean Startup Methodology and Big Data analytics in the development and testing of hypotheses related to the value proposition.*	When a firm engages in innovation in a market or industry that it knows well with an incrementally improved product. *Focus only on Big Data analytics*

The recommendations (*in italics*) in the Table are meant to summarize our discussions and propositions. We recognize that firms' reliance on Big Data and Lean Startup Methods will be more nuanced and along a continuum – which, in turn, will increase or decrease the likelihood of certain performance outcomes

Proposition 4: In environments with high levels of technological uncertainty and demand uncertainty, legacy firms that fully incorporate the lean startup methodology and big data analytics into the process of developing the business model will outperform those legacy firms that do not.

Low technological uncertainty and low demand uncertainty

Industries such as precious metals, and personal services (e.g., hair-styling and manicure/pedicure providers) face both low technological uncertainty and low demand uncertainty (Dyer et al. 2014). For example, hair stylists are faced with a relatively well-known demand and are continuing to use similar technologies (Furr and Dyer 2014). This is reflected in relatively low firm turnover (4.4%) and revenue volatility (59.7%), and relatively low R&D spend as % of sales (0.3%) (ibid.).

This scenario could involve a situation where the firm is engaging in an incremental improvement of a product or service that it already provides and is going to launch this improvement in the market that it is already operating in. In such a situation, the firm knows both the customers and the technology well and illusions of control are less likely to emerge. Hence, less importance can be attached to lean startup methods. Due to the reduced levels of both technological and demand uncertainty, firms can focus their attention on big data analytics to drive decisions vis-à-vis the product innovation processes and the appropriate business model. In particular, taking advantage of the variety and velocity of their data would be enough to mitigate tunnel vision and to make key decisions in a faster and timely fashion.

Proposition 5: In environments with low technological and demand uncertainty, legacy firms that incorporate the lean startup methodology and big data analytics into the process of developing the business model will not perform better than firms that only focus on big data analytics.

High technological uncertainty and low demand uncertainty

Industries such as insurance and aircraft manufacturers face low demand uncertainty and high technological uncertainty (Dyer et al. 2014). The aircraft manufacturers are challenged by uncertainty on the technological side, leading them to spend huge amounts on developing new aircraft, but at the same time still being able to predict demand for the aircraft relatively well (Furr and Dyer 2014).

In such a condition, the firm is developing a completely new product for customers that it already knows well. Here, the issue becomes when to use the lean startup methodology in conjunction with big data analysis, and when big data analysis would suffice on its own. In such a scenario, the firm is dealing with a known group of customers, and through its data analytics efforts, the firm already knows how to build relationships with these customers and through what channels it should reach them. Therefore, for these components of the business model canvas, it is not necessary for the firm to make use of a combination of lean startup methods and big data analytics; big data analytics on its own will suffice. However, there are parts of the business model canvas that are unknown, as the firm is offering a new product for these customers. Here, big data analytics can be used to estimate the needs and wants of the customers (i.e. the response of these customers to the value proposition component of the business model canvas), which would help firms design hypotheses to be tested using the lean startup methodology.

Proposition 6: In environments with high technological uncertainty and low demand uncertainty, legacy firms that incorporate the lean startup methodology and big data analytics in validating the value proposition will outperform those legacy firms that do not.

Low technological uncertainty and high demand uncertainty

Industries such as coal and restaurants face low technological uncertainty and high demand uncertainty (Dyer et al. 2014). While the technologies for eating at restaurants has not changed much, restaurants often find it difficult to predict demand due to the many factors that play a role in deciding whether a customer is going to eat out or not in a particular restaurant (Furr and Dyer 2014). Similarly, a firm may be taking an established product with some incremental improvements into a new market with customers that it does not know. As with the high technological uncertainty and low demand uncertainty scenario, the firm would once again require a blend of lean startup methods with big data analytics. Here, the product itself is not the issue (e.g. we know coal), but the question is whether or not the firm can find a customer segment in the new market that will buy the product, among other alternatives. Hence, the firm will need to use big data analytics to develop hypotheses about the customer segments component of the business model canvas in the (new) market and then test these hypotheses using lean startup methods and innovation accounting. However, for product based decisions, the firm can mainly focus on big data analytics.

Proposition 7: In environments with low technological uncertainty and high demand uncertainty, legacy firms that incorporate the lean startup methodology and big data analytics in testing hypotheses related to customer segments will outperform those legacy firms that do not.

Conclusion

In this manuscript we asked the question *how can a legacy firm adapt its business model(s) to compete in innovation with the new generation of startups that are threatening its value creation and appropriation?* We recommend that firms use a combination of the lean startup methodology and big data analysis to improve the process component of developing the business model. The usual caveats apply, as we generalize across legacy firms in different industries and situations of resources and competitors. Instead of simply urging to "innovate like a startup," our advice recognizes both the uncertainty legacy firms face and the substantial resources they have over startups. The existing literature has shown that conditions of high uncertainty and sufficient resources increase the competitive advantage of ambidextrous organizations, i.e. organizations that both compete in mature markets and innovate for new markets (O'Reilly and Tushman 2013). However, a key unresolved issue is *how* legacy firms can actually become ambidextrous.

To achieve ambidexterity, we argue that firms should build on the lean startup methodology to help adapt their business models while at the same time leveraging the resource advantages that they have as legacy corporations. One of the resource advantages that legacy firms have over startups is the access to big data. It seems that everywhere we look in the business press these days there are discussions of big data and how firms can best make use of this resource to stimulate innovation and growth. At the same time, however, concerns mount over the ability of firms to handle and learn from all the data that they can access. In fact, big data could further perpetuate confirmation biases, miscommunications and illusions of control, instead of mitigating them. Furthermore, action needs to be taken to ensure that firms make good use of big data before disillusionments set in.

Our propositions list the potential advantages of combining big data and lean startup methods to develop testable hypotheses in an evolving business model. They also provide an interdisciplinary guideline through which legacy firms can evolve their business models. Proposition 1 mitigates learning challenges within a traditional business model by testing big data generated hypotheses before implementing them. Proposition 2 argues that incorporating innovation accounting within the business model will reduce and prevent failed innovations. Proposition 3 claims that, rather than being a victim of it, legacy firms can harness the power of uncertainty, which is often underestimated in a more traditional, big data approach. Not taking advantage of the recommendations offered in these propositions leaves legacy firms vulnerable to competition from startups that effectively operationalize these methodologies and to misconceptions that are born due to mismanagement of big data.

In our analysis, we also look at the boundary conditions of Propositions 1–3 in terms of technological uncertainty and demand uncertainty. Accordingly, we provide further hypotheses on when our claims would make the bigger difference in the innovation processes and performances of legacy firms. In particular, augmenting big data analytics with the lean startup methodology would bring all the advantages discussed in Propositions 1–3 when both types of uncertainty are high (see Proposition 4). Proposions 5–7 provide a framework for a more nuanced evolution of business models when the progress of technology and/or demand patterns are more predictable.

In sum, we provide a framework that firms can act upon to incorporate big data and the lean startup methodology into their innovation processes. Big data is not a panacea: it does not eliminate the main learning challenges induced by confirmation, communication and control. The lean startup methodology complements big data analytics, helping it realize its potential for corporate innovation under technological and demand uncertainty. In this way, firms will be better able to evolve their business models to ensure the greatest chance of successful innovation and the ability to not only survive but thrive in uncertain environments.

Both our general propositions and proposed boundary conditions set up a future research agenda to test our framework. For one, measurement tools can be designed to operationalize the extent of lean methodology in legacy firms. Next, empirical researchers can relate this process to firm outcome measures, such as survival, profits and abnormal stock returns. An especially interesting empirical question to us is the relative importance of each of the lean methodology components (Propositions 1–3) in contributing to legacy firm performance. Moreover, our framework does not specify in which time frame legacy firms should observe the benefits of acting on our recommendations. Finally, empirical research can validate our Propositions 4–7 by relating the performance-enhancing benefits to the moderating factors of demand uncertainty and technological uncertainty.

References

Abell, D. F. (1999). Competing today while preparing for tomorrow. *MIT Sloan Management Review, 40*(3), 73–81.

Anand, J., & Delios, A. (2002). Absolute and relative resources as determinants of international acquisitions. *Strategic Management Journal, 23*(2), 119–134. https://doi.org/10.1002/smj.215.

Ariely, D. (2010). Why businesses don't experiment. *Harvard Business Review, 88*(4).

Balogun, J. (2006). Managing change: Steering a course between intended strategies and unanticipated outcomes. *Long Range Planning, 39*(1), 29–49. https://doi.org/10.1016/j.lrp.2005.02.010.

Bandler, J., & Burke, D. (2012). How Hewlett-Packard lost its way. Retrieved December 24, 2016 from http://fortune.com/2012/05/08/how-hewlett-packard-lost-its-way/.

Barney, J. (1991). Firm resources and sustained competitive advantage. *Journal of Management, 17*(1), 99–120. https://doi.org/10.1177/014920639101700108.

Baron, J. (2000). *Thinking and deciding*. Cambridge: Cambridge University Press.

Blank, S. (2010). What's a startup first principles. Retrieved September 27, 2017 from https://steveblank.com/2010/01/25/whats-a-startup-first-principles/.

Blank, S. (2013). Why the lean startup changes everything. *Harvard Business Review* May, 4–9.

Bradbury, M. A., Hens, T., & Zeisberger, S. (2014). Improving investment decisions with simulated experience. *Review of Finance*, rfu021.

Carpenter, M., Sanders, G., & Gregersen, H. (2001). Bundling human capital with organizational context: The impact of international assignment experience on multinational firm performance and CEO pay. *Academy of Management Journal, 44*(3), 493–511. https://doi.org/10.2307/3069366.

Castellion, G., & Markham, S. K. (2013). Perspective: new product failure rates: influence of argumentum ad populum and self-interest. *Journal of Product Innovation Management, 30*(5), 976–979. https://doi.org/10.1111/j.1540-5885.2012.01009.x.

Christensen, C. M. (1997). *The innovator's dilemma*. Boston: Harvard Business School Press.

Christensen, C. M., Bartman, T., & Van Bever, D. (2016). The hard truth about business model innovation. *MIT Sloan Management Review, 58*(1), 31.

Cooper, R. G. (1990). *Stage-gate systems: A new tool for managing new products* (pp. 44–54). May–June: Business Horizons.

Cooper, R. G. (2014). What's next? After stage-gate. *Research – Technology Management* January-February, 20–31.

Delvecchio J., White, F., & Phelan, S.E. (2013). Tools for innovation management: A comparison of lean startup and the stage gate system. Working paper.

Demil, B., & Lecocq, X. (2010). Business model evolution: in search of dynamic consistency. *Long Range Planning, 43*(2), 227–246. https://doi.org/10.1016/j.lrp.2010.02.004.

Denrell, J. (2003). Vicarious learning, undersampling of failure, and the myths of management. *Organization Science, 14*(3), 227–243. https://doi.org/10.1287/orsc.14.2.227.15164.

Denrell, J., Fang, C., & Liu, C. (2014). Perspective—Chance explanations in the management sciences. *Organization Science, 26*(3), 923–940.

Du, R. Y., Hu, Y., & Damangir, S. (2015). Leveraging trends in online searches for product features in market response modeling. *Journal of Marketing, 79*(1), 29–43. https://doi.org/10.1509/jm.12.0459.

Dyer, J., Furr, N., & C. Lefrandt. (2014). The industries plagued by the most uncertainty. Retrieved December 24, 2015 from https://hbr.org/2014/09/the-industries-plagued-by-the-most-uncertainty.

Edgett, S. (2011). *New product development: Process benchmarks and performance metrics*. Houston: American Productivity and Quality Center, The Product Development Institute.

Egusa, C. (2013). Beyond lean startups: Eric Ries' movement heads to Fortune 500, government and beyond. Retrieved January 25, 2016 from http://venturebeat.com/2013/12/17/beyond-lean-startups-eric-ries-movement-heads-to-fortune-500-government-and-beyond.

Einhorn, H. J., & Hogarth, R. M. (1978). Confidence in judgment: Persistence of the illusion of validity. *Psychological Review, 85*(5), 395–416. https://doi.org/10.1037/0033-295X.85.5.395.

Flamholtz, E. G., & Aksehirli, Z. (2000). Organizational success and failure: an empirical test of a holistic model. *European Management Journal, 18*(5), 488–498. https://doi.org/10.1016/S0263-2373(00)00038-4.

Furr, N., & Dyer, J. (2014). *The innovator's method: bringing the lean startup into your organization*. Boston: Harvard Business Review Press.

Gartner. (2015). Hype cycle for emerging technologies.

Germann, F., Lilien, G. L., & Rangaswamy, A. (2013). Performance implications of deploying marketing analytics. *International Journal of Research in Marketing, 30*(2), 114–128. https://doi.org/10.1016/j.ijresmar.2012.10.001.

Gibson, C. B., & Birkinshaw, J. (2004). The antecedents, consequences, and mediating role of organizational ambidexterity. *Academy of Management Journal, 47*(2), 209–226. https://doi.org/10.2307/20159573.

Halecker, B., Bickmann, R., & Hölzle, K. (2014). Failed business model innovation – a theoretical and practical illumination on a feared phenomenon. Paper presented at the R&D Management Conference 2014 – Management of Applied R&D: Connecting high value solutions with future markets. Stuttgart, Germany on 03-06 June 2014.

Hanssens, D. M., & Pauwels, K. H. (2016). Demonstrating the value of marketing. *Journal of Marketing, 80*(6), 173–190. https://doi.org/10.1509/jm.15.0417.

Henderson, R. M., & Clark, K. B. (1990). Architectural innovation: The reconfiguration of existing product technologies and the failure of established firms. *Administrative Science Quarterly, 35*(1), 9–30. https://doi.org/10.2307/2393549.

Hogarth, R. M., & Soyer, E. (2011). Sequentially simulated outcomes: Kind experience versus nontransparent description. *Journal of Experimental Psychology, 140*(3), 434–463. https://doi.org/10.1037/a0023265.

Hogarth, R. M., & Soyer, E. (2015). Using simulated experience to make sense of big data. *MIT Sloan Management Review, 56*(2), 49.

Hogarth, R. M., Lejarraga, T., & Soyer, E. (2015). The two settings of kind and wicked learning environments. *Current Directions in Psychological Science, 24*(5), 379–385. https://doi.org/10.1177/0963721415591878.

Ioannidis, J. P. (2005). Why most published research findings are false. *Chance, 18*(4), 40–47. https://doi.org/10.1080/09332480.2005.10722754.

Johnson, M. W., Christensen, C., & Kagermann, H. (2008). Reinventing your business model. Harvard Business Review, December, 59–68.

Johnson-Cramer, M. E., Parise, S., & Cross, R. L. (2007). Managing change through networks and values. *California Management Review, 49*(3), 85–109. https://doi.org/10.2307/41166396.

Kim, A. J., & Ko, E. (2012). Do social media marketing activities enhance customer equity? An empirical study of luxury fashion brand. *Journal of Business Research, 65*(10), 1480–1486. https://doi.org/10.1016/j.jbusres.2011.10.014.

King, A. A., & Baatartogtokh, B. (2015). How useful is the theory of disruptive innovation? *MIT Sloan Management Review, 57*(1), 77–90.

Laney, D. (2001). 3D data management: Controlling data volume, velocity and variety. *META Group Research Note, 6*, 70.

MacDiarmid, Moukanas, H., & Nehls, R. (1998). Reaping the fruits of business design. *Mercer Management Journal, 10*, 57–68.

Mahmoud, A. (2016). The impact of Airbnb on hotel and hospitality industry. Retrieved on December 23, 2016 from http://www.hospitalitynet.org/news/4074708.html.

Makridakis, S., Hogarth, R. M., & Gaba, A. (2009). Forecasting and uncertainty in the economic and business world. *International Journal of Forecasting, 25*(4), 794–812. https://doi.org/10.1016/j.ijforecast.2009.05.012.

Manyika J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Byers, A.H. (2011). Big data: The next frontier for innovation, competition and productivity. Retrieved on May 31, 2016 from http://www.mckinsey.com/business-functions/business-technology/our-insights/big-data-the-next-frontier-for-innovation.

March, J. G. (1991). Exploration and exploitation in organizational learning. *Organization Science, 2*(1), 71–87. https://doi.org/10.1287/orsc.2.1.71.

March, J. G. (2011). *The ambiguities of experience*. Cornell: Cornell University Press.

Mayer-Schönberger, V., & Cukier, K. (2013). *Big data: A revolution that will transform how we live, work and think*. Canada: Eamon Dolan/Houghton Mifflin Harcourt.

McAfee, A., & Brynjolfsson, E. (2012). Big data: The management revolution. Harvard Business Review, October, 3–9.

McKinsey & Company. (2017). Setting an agenda for organic growth in the digital age. Retrieved on October 13, 2017 from https://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/setting-an-agenda-for-organic-growth-in-the-digital-age#_=_

Miles, R. E., Snow, C. C., Mathews, J. A., Miles, G., & Coleman Jr., H. J. (1997). Organizing in the knowledge age: anticipating the cellular form. *Academy of Management Executive, 11*(4), 7–20.

Musteen, M., Rhyne, L., & Zheng, C. (2013). Asset or constraint: Corporate reputation and MNCs' involvement in the least developed countries. *Journal of World Business, 48*(3), 321–328. https://doi.org/10.1016/j.jwb.2012.07.016.

Neus, A., Buder, F., & Galdino, F. (2017). Are you too successful to digitalize? How to fight innovation blindness. *Marketing Intelligence Review, 9*(1), 30–35.

New Vantage Partners. (2016). *Big Data Executive Survey*, Retrieved on March 23, 2017 from http://newvantage.com/wp-content/uploads/2016/01/Big-Data-Executive-Survey-2016-Findings-FINAL.pdf.

New York Times. (2012). The age of big data. Retrieved on December 24, 2015 from http://www.nytimes.com/2012/02/12/sunday-review/big-datas-impact-in-the-world.html.

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology, 2*(2), 175–220. https://doi.org/10.1037/1089-2680.2.2.175.

Nohria, N., & Gulati, R. (1996). Is slack good or bad for innovation? *Academy of Management Journal, 39*(5), 1245–1264. https://doi.org/10.2307/256998.

O'Reilly, C. A., & Tushman, M. L. (2008). Ambidexterity as a dynamic capability: Resolving the innovator's dilemma. *Research in Organizational Behavior, 28*, 185–206. https://doi.org/10.1016/j.riob.2008.06.002.

O'Reilly, C. A., & Tushman, M. L. (2013). Organizational ambidexterity: Past, present, and future. *Academy of Management Perspectives, 27*(4), 324–338. https://doi.org/10.5465/amp.2013.0025.

Osterwalder, A., & Pigneur, Y. (2010). *Business model generation: A handbook for visionaries, game changers, and challengers*. New Jersey: Wiley.

Owens, T., & Fernandez, O. (2015). The lean enterprise: How corporations can innovate like startups. Audible Studios on Brilliance Audio.

Pauwels, K. H. (2014). *It's not the size of the data, it's how you use it: Smarter marketing with analytics and dashboards*. New York: AMACOM.

Pauwels, K., & van Ewijk, B. (2013). Do online behavior tracking or attitude survey metrics drive brand sales? An integrative model of attitudes and actions on the consumer boulevard. *Marketing Science Institute, Report*, 13–11.

Pauwels, K., Demirci, C., Yildirim, G., & Srinivasan, S. (2016). The impact of brand familiarity on online and offline media synergy. *International Journal of Research in Marketing, 33*(4), 739–753. https://doi.org/10.1016/j.ijresmar.2015.12.008.

Peters, K., Chen, Y., Kaplan, A. M., Pauwels, K., & Ognibeni, B. (2013). Social media metrics: A framework and guidelines for managing social media. *Journal of Interactive Marketing, 27*(4), 281–298. https://doi.org/10.1016/j.intmar.2013.09.007.

Pfeffer, J., & Sutton, R. I. (2006). Evidence-based management. *Harvard Business Review, 84*(1), 62–74, 133.

Popper, K. (1959). *The logic of scientific discovery*. London: Hutchinson.

Ries, E. (2009). Pivot, don't jump to anew vision, Retrieved on May 16, 2017 from http://www.startuplessonslearned.com/2009/06/pivot-dont-jump-to-new-vision.html.

Ries, E. (2011). *The lean startup: How today's entrepreneurs use continuous innovation to create radically successful businesses*. New York: Crown Business.

Saleh, T., Lenhard, E., Goldon, N., & Opolon, D. (2015). How to avoid the big bad data trap, BCG perspectives. Retrieved on June 24, 2017 from https://www.bcgperspectives.com/content/articles/big-data-digital-economy-how-to-avoid-big-data-trap/.

Smith, W. K., Binns, A., & Tushman, M. L. (2010). Complex business models: Managing strategic paradoxes simultaneously. *Long Range Planning, 43*(2-3), 448–461. https://doi.org/10.1016/j.lrp.2009.12.003.

Sorescu, A., Frambach, R. T., Singh, J., Rangaswamy, A., & Bridges, C. (2011). Innovations in retail business models. *Journal of Retailing, 87*, S3–S16. https://doi.org/10.1016/j.jretai.2011.04.005.

Soyer, E., & Hogarth, R. M. (2015). Fooled by experience. *Harvard Business Review, 93*(5), 72–77.

Staw, M. (1976). Knee deep in the big muddy: A study of escalating commitment to a chosen course of action. *Organizational Behavior and Human Performance, 16*(1), 27–44. https://doi.org/10.1016/0030-5073(76)90005-2.

Sterman, J. D. (2011). Communicating climate change risks in a skeptical world. *Climatic Change, 108*(4), 811–826. https://doi.org/10.1007/s10584-011-0189-3.

Sull, D., Homkes, R., & Sull, C. (2015). Why strategy execution unravels – and what to do about it. *Harvard Business Review, 71*, 59–66.

Taleb, N. N. (2010). *The black swan: The impact of the highly improbable fragility* (Vol. 2). New York: Random House.

Teece, D. J. (2010). Business models, business strategy and innovation. *Long Range Planning, 43*(2/3), 172–194. https://doi.org/10.1016/j.lrp.2009.07.003.

Tirunillai, S., & Tellis, G. J. (2014). Mining marketing meaning from online chatter: analysis of big data using latent dirichlet allocation. *Journal of Marketing Research, 51*(4), 463–479. https://doi.org/10.1509/jmr.12.0106.

Tushman, M. L., & O'Reilly, C. A. (1996). Ambidextrous organizations: Managing evolutionary and revolutionary change. *California Management Review, 38*(4), 8–30. https://doi.org/10.2307/41165852.

Vinekar, V., Slinkman, C. W., & Nerur, S. (2006). Can agile and traditional systems development approaches coexist? An ambidextrous view. *Information Systems Management, 23*(3), 31–42. https://doi.org/10.1201/1078.10580530/46108.23.3.20060601/93705.4.

Weiblen, T., & Chesbrough, H. W. (2015). Engaging with startups to enhance corporate innovation. *California Management Review, 57*(2), 66–90. https://doi.org/10.1525/cmr.2015.57.2.66.

Wiesel, T., Pauwels, K. P., & Arts, J. (2011). Marketing's profit impact: quantifying online and offline funnel progression. *Marketing Science, 30*(4), 604–611. https://doi.org/10.1287/mksc.1100.0612.

Zott, C., & Amit, R. (2010). Business model design: an activity system perspective. *Long Range Planning, 43*(2/3), 216–212. https://doi.org/10.1016/j.lrp.2009.07.004.