Learning from Data in Entrepreneurial Experimentation[1]

Hyunjin Kim

INSEAD

hyunjin.kim@insead.edu

July 31, 2024

[1] I thank Natalie Carlson, Ashley Craig, and participants in the Bayesian Entrepreneurship Workshop for valuable comments and conversations.

1 Introduction

Experimentation plays a pivotal role in entrepreneurial and strategic decision-making (Camuffo et al., 2020; Gans, 2023; Levinthal, 2017). The Bayesian entrepreneurship framework underscores this importance by positioning experimentation at the heart of entrepreneurial decisions, where the value of experimentation stems from its ability to update entrepreneurs' beliefs, thereby shaping their strategic choices (Agrawal et al., 2024). Through this lens, experiments encompass any activity that generates information for belief updating, be it feedback from customers, product launches, or A/B testing. In essence, the fundamental objective of an experiment is to produce informative data that facilitates learning for the entrepreneur.

Despite the recognized importance of experimentation, there has been less insight on how entrepreneurs design experiments to generate informative data and how they translate this data into learning. Much of the existing research has abstracted the data from experiments—both its design and its interpretation—into simplified signals. Studies on data-driven decision-making often represent information acquisition as a binary input, emphasizing the value of leveraging information for firm decisions (Nagaraj, 2022; Kim, 2024; Brynjolfsson and McElheran, 2016). In practice, however, entrepreneurs do not receive abstract signals or binary inputs. Instead, they must generate and derive insights from complex, multifaceted data. Yet, this process of data generation and interpretation has been largely overlooked.

The challenges of designing experiments that yield informative data and correctly interpreting them to extract learning are aptly illustrated by Netflix, the online streaming giant. As they noted: "Good businesses pay attention to what their customers have to say. But what customers ask for (as much choice as possible, comprehensive search and navigation tools, and more) and what actually works (a few compelling choices simply presented) are very different" (Gomez-Uribe and Hunt 2015, p.7). This insight underscores that experiments are not alike in the data they generate: in this case, an experiment asking customers what they want may not generate data as informative as launching product changes and observing real reactions. Moreover, the interpretation of experimental data presents its own set of challenges. Although Netflix has run experiments for decades, they emphasize that interpreting results "remains partly art" and requires "more sophisticated experiment design and analysis" to extract learning (Gomez-Uribe and Hunt 2015 p.11).

This highlights the importance of understanding the process of how entrepreneurs learn from data in experimentation, which this chapter aims to unpack. By anchoring on the premise that experimenting is fundamentally about purposefully acquiring data, it highlights that learning from experiments is far from straightforward. Entrepreneurs often face three challenges: (1) failing to collect informative data, (2) neglecting to pay sufficient attention to the data collected, and (3) making incorrect inferences even when attention is paid. Building on these observations, I propose that the value of experimentation is critically dependent on how entrepreneurs generate, attend to, and interpret data from their experiments. I then discuss the implications of these insights for entrepreneurial strategy and the potential role of artificial intelligence (AI) in the experimentation process.

The chapter is structured as follows: Section 2 provides an overview of research on entrepreneurial experimentation, reviewing existing definitions and key findings. It then introduces the Bayesian entrepreneurship framework and its emphasis on information and learning through experimentation. Section 3 explores why experiments may not lead to learning, highlighting three key frictions. Section 4 discusses the increasing value of the ability to learn from data in the context of recent advancements in artificial intelligence (AI). Finally, Section 5 concludes by summarizing insights and identifying open questions for future research.

2 Entrepreneurial experimentation

2.1 What is experimentation?

Experimentation has been a growing topic in entrepreneurship and strategy research. A significant literature emphasizes experimentation as a key part of the entrepreneurial learning process, helping early-stage ventures purposefully learn about the value of their idea or strategy (Murray and Tripsas, 2004; Eisenmann et al., 2012; Kerr et al., 2014).

The widely adopted lean startup approach highlights the importance of rapid iteration to test hypotheses, focusing on customer feedback to guide the experimentation process (Ries, 2011b; Eisenmann et al., 2012; Blank, 2013). Camuffo et al. (2020) emphasize the role of theory-driven hypotheses in improving entrepreneurial strategy and performance, while Gans et al. (2019) stress the importance of experimenting with different but equally viable alternatives to make a choice. This process extends beyond entrepreneurial firms to strategic decisions more broadly: the real options approach to strategy emphasizes how initial investments allow firms to collect signals about possible options (Adner and Levinthal, 2004). Moreover, Levinthal (2017) underlines an experimental approach to strategic decision-making, emphasizing the role of a "Mendelian" executive who explores different alternatives, tests them through experiments, and selects the best path.

A defining attribute of experimentation is that it is purposeful rather than simply a by-product of trial-and-error actions (Murray and Tripsas, 2004). Experimentation allows entrepreneurs to learn about the value of their idea or strategy (Agrawal et al., 2021), exploring aspects of their technology, market, or business model (Murray and Tripsas, 2004). This learning may not necessarily be low-cost: while the lean startup approach emphasizes small and low-cost experiments, Gans et al. (2019) note that what might appear to be a low-cost test may involve partial commitment to one plan over other equally viable alternatives. A growing body of work empirically evaluates the consequences of experimentation, finding it to result in more pivots and terminations (Camuffo et al., 2020), and positively correlated with product changes and higher performance measures conditional on survival (Koning et al., 2022).

The process of experimentation involves three key stages: (1) idea generation, (2) testing and learning from experiments, and (3) idea selection. Research has provided much insight on the first step of idea generation, which serves an important purpose: experiments require ideas to test, and not all possibilities can be explored. This work advocates for articulating scientific, falsifiable hypotheses that can be validated through experiments (Camuffo et al., 2020; Ries, 2011b; Blank, 2013), and emphasizes the value of experimenting with more uncertain hypotheses that inform the entrepreneur's broader theory (Camuffo et al., 2022). Gans et al. (2019) underscore the importance of considering more alternative hypotheses to avoid prematurely committing to a suboptimal course of action.

There has also been substantial work on the final stage of idea selection. Idea selection can involve commitment to the idea or strategy, terminating the venture, or pivoting to new directions (Camuffo et al., 2020). One approach has been to see selection as an optimization process, identifying the option that yields the highest value relative to cost. However, Gans et al. (2019) and Levinthal (2017) emphasize the role of deliberate choice in this selection process.

Intriguingly however, there has been less insight on the actual process of testing and learning—the "experiment" itself.[1] This is surprising, given that this testing stage of experimentation is arguably its most distinctive aspect, with idea generation and selection forming core components of most strategic decision-making approaches (Newell and Simon, 1976). While the inputs (ideas and hypotheses) and outputs (choices and performance) of experimentation have received significant attention, there is less understanding of how the design of experiments results in different data and how this influences entrepreneurial learning.

In particular, there are two key gaps. First, "experiments" refer to a wide range of activities. They can be conceptual or real-world implementations (Agrawal et al., 2024). They can involve conversations with mentors (Cohen et al., 2019; Miller et al., 2024), upvotes on a forum (Cao et al., 2023), the release of a "minimimally viable product" (Ries, 2011b), or a formal A/B test (Thomke, 2020). In essence, experiments encompass any activity that might provide an opportunity to acquire information, which can be tightly or loosely tied to the hypotheses they are designed to test. Although these experiments generate different types of data, the literature often abstracts this data into simple signals. However, the experiments entrepreneurs undertake do not yield simple signals in practice. They must learn from the actual data generated, and we know less about this process.

[1]A key exception is Agrawal et al. (2021), which provides important theoretical insights on the tradeoffs between testing the value of an idea or strategy in isolation versus in combination, and Gans (2023) that formally highlights biases in experimental choice. Both have highlighted the importance of noise and bias in experiment design.

Second, learning from experiments is often implicitly assumed to be coupled with the act of undertaking the experiment itself. Since experiments are designed to acquire information, they generally result in some information. Yet, how entrepreneurs process and interpret this data and the conditions under which it translates into learning are much less understood. Akin to the broader organizational learning literature where practices associated with success are reinforced and those associated with failure avoided (Contigiani and Levinthal, 2019; Levitt and March, 1988; Argote, 2012), experimentation is often viewed informally as inherently resulting in learning, irrespective of the outcome, which overlooks complexities in how entrepreneurs process and learn from data to inform their decisions.

2.2 Experimentation through the Bayesian lens

The Bayesian entrepreneurship lens offers a framework that centers on the critical interplay between an entrepreneur's prior beliefs and how evidence from experimentation can update those beliefs (Agrawal et al., 2024). It provides a formal approach to assess how entrepreneurs form beliefs about an opportunity and how these beliefs evolve through experimentation—and the resulting implications for entrepreneurial strategy. This framework thus defines learning as updating entrepreneurs' beliefs and emphasizes that beliefs are central to both the inputs and outputs of experimentation: entrepreneurs' beliefs shape their choices on how to experiment, and their own or others' beliefs are also the outcomes of experimentation.

This provides three key implications for thinking about experimentation. First, experiments represent the acquisition of information that can meaningfully reduce uncertainty on entrepreneurial beliefs about their hypothesis, which can help refine our understanding of what represents informative experiments. Second, the design of experiments involves choice, and the informational value of experiments hinges on these explicit and implicit choices. Third, experiments may be informative, but whether this results in learning depends on the nature of belief updating.

3 Why might experiments not result in learning?

In this section, I highlight three key frictions that hinder learning from experiments, which interact in interesting and important ways when considered through the lens of Bayesian entrepreneurship.

3.1 Uninformative experiments

How to experiment is a fundamental choice that entrepreneurs must make in the experimentation process. Once they have generated their hypotheses, there are many ways to test them, among the wide range of activities referred to as experiments—all of which involve vastly different actions and data that result.

Much of the literature on how to experiment has focused on the fidelity of experiments as it relates to noise and bias, emphasizing that most experiments are noisy and biased, at times reinforcing the idea of information scarcity in early entrepreneurship (Agrawal et al., 2021; Gans, 2023; Agrawal et al., 2024). Moreover, it highlights that in many cases, lower-fidelity experiments, referring to those with smaller sample sizes or biased samples, may be more useful, suggesting that choosing fidelity often trades off the cost of experimentation (Azevedo et al., 2020).

However, entrepreneurs make many other choices on experiments that affect what data is generated, whether explicitly or implicitly, which may not necessarily come at the expense of higher costs, slower speed, or more difficult execution. These choices are often less salient but can have a profound impact on the informativeness of the experiment: (1) the type of activity to pursue, (2) how they operationalize their hypotheses (e.g., what features an MVP has, or what the A vs. B looks like in an A/B test), and (3) the outcomes they collect from the activity. Just as more data may not always improve predictions (Kim et al., 2024), if the experiment is uninformative, little learning can result, even with the most astute hypothesis (McGrath and MacMillan, 1995; Pfeffer and Sutton, 2006; Rousseau, 2006).

Putting this idea within the Bayesian entrepreneurship framework highlights that experiments must be designed to generate informative data about the hypothesis. If the experiment does not provide any new information about the hypothesis, it does not change the entrepreneur's posterior beliefs compared to their prior. Mathematically, this can be represented by the likelihood being constant for all hypotheses: $P(E|H) = P(E)$, where $P(E|H)$ is the probability of the evidence E given the hypothesis H, and $P(E)$ is the probability of the evidence across all possible hypotheses. If $P(E|H)$ is constant and equal to $P(E)$, the evidence E does not provide any additional information to differentiate between hypotheses. Similarly, when evidence E is independent of the hypothesis H (i.e., the occurrence of the evidence does not affect the probability of the hypothesis being true or false, as in $P(E|H) = P(E)$ and $P(H|E) = P(H)$), the evidence E does not change the distribution over the hypothesis H, leading to no update in beliefs.

This enables us to increase precision on what constitutes informative experiments. It emphasizes that the defining feature of experiments from other forms of information acquisition is its focus on a specific hypothesis (Murray and Tripsas, 2004; Agrawal et al., 2021), and crafting informative experiments requires making deliberate choices that ensure that the resulting data informs the hypothesis under investigation. This illuminates two pivotal choices for designing informative experiments that must be made in the "testing" stage following hypothesis generation.

The first key choice involves operationalizing the hypothesis in the experiment such that the activity accurately reflects its core idea. Consider the case of Dropbox. In its early days, CEO Drew Houston and his team faced a significant challenge in growing their company, which sought to offer a cloud-based file-sharing and synchronization service. Their hypothesis was that providing a seamless customer experience would increase demand for file synchronization. However, many people struggled to grasp the concept of storing files online and accessing them seamlessly across multiple devices (Ries, 2011a). Determined to validate their assumption that a superior user experience would attract customers, the Dropbox team experimented with an MVP explainer video that has since become renowned. Narrated by Houston himself, the video displayed what the seamless user experience would look like, demonstrating a problem many were unaware they had.

The explainer video was a resounding success, driving "hundreds of thousands of people to the website" and expanding their beta waiting list from "5,000 people to 75,000 people literally overnight" (Ries, 2011a). This success help validate their hypothesis and provided valuable insights into their target audience.

This example highlights that the activity entrepreneurs choose to pursue and how they design it to operationalize their hypothesis is crucial. The hypothesis Houston had was not an easy one to test through simply any experiment: just as potential customers had difficulty understanding the concept, describing it to investors or mentors for advice made it challenging to gather informative data. Houston had many meetings with venture capital investors, but was met with skepticism, as they could not imagine his vision. Although Houston believed that "if the software 'just worked like magic', customers would flock to it", investors pointed out that the market was crowded with existing products that had yet to generate substantial profits – suggesting that file synchronization was not an actual problem that customers had (Ries, 2011a).[2] Moreover, another experiment of even the same activity type may not have generated informative data. Had Dropbox chosen to experiment with a buggy MVP that was launched to early users rather than a video showcasing what the seamless experience could be like, it may have resulted in very different data because the experiment was independent from the hypothesis. By focusing on a video that could showcase the seamless user experience—the core idea behind their hypothesis—Dropbox was able to collect data that enabled them learn about their hypothesis.

The second key experiment choice involves identifying and selecting experiment metrics to generate data that can evaluate the hypothesis. When experiments collect "wrong" metrics that are independent of the hypothesis, they cannot be informative. Consider

[2]This example also highlights how the data collected might involve correlated signals, which relates to the argument made in Section 3.3: in this case, the advice from different investors are likely to be correlated, as they all observe the same evidence (existing file synchronization products that were failing). Research has documented how individuals fail to properly account for the correlation between different pieces of information, which is referred to as "correlation neglect" (Enke and Zimmermann, 2019; Enke, 2020). This may lead entrepreneurs to learn incorrectly from informative evidence, and highlights how experiment choices can also interact with biases in learning: advice may not be as valuable an activity to pursue when advisors' signals are correlated.

Coca-Cola's well-known experiment with New Coke. In the early 1980s, its rival Pepsi-Cola gained market share through their "Pepsi Challenge" campaign, where consumers taking blind taste tests learned to their surprise that they preferred the flavor of Pepsi. Hypothesizing that the taste of Coke was driving declines in market share, Coca-Cola undertook an experiment to inform their choice, conducting 190,000 blind product tests between Old Coke and a new, sweeter formula (Klein, 2023). The experiment found that more than half of the participants preferred New Coke over both the original formula and Pepsi. Armed with the results from the experiment, Coca-Cola introduced New Coke to the market and withdrew the original formula. However, this resulted in decreased market performance and widespread customer outrage, with thousands of angry phone calls a day and protests from grassroots groups such as "Old Cola Drinkers of America" (Klein, 2023). It turns out that Old Coke had an emotional value for customers, evoking personal memories, and changing the formula did not result in higher sales or market share. What went wrong? Coca-Cola chose an activity that could inform their hypothesis, and operationalized it by testing two formulas that directly mapped to its core idea. However, the metric they chose for the experiment was taste, not sales: they asked whether tasters preferred the taste of New Coke to Old Coke, not whether they would buy the New Coke over the original. In this case, the metric they chose, taste, turned out to be independent of sales and market share – which was the key outcome critical to evaluate their hypothesis. While this example centers on choosing the wrong metric, it is worth noting that once we have chosen the right metric for the hypothesis, considerations of noise and bias can also become important – if tasters were a biased sample of never-tried-Coke drinkers, one could imagine that even with better metrics, the resulting data may have been less informative, especially if executives were not able to take into account the potential bias.[3]

[3]For example, Cao et al. (2023) shows that female-focused products may observe biased feedback from early user forums like Product Hunt, whose user base is composed of a high proportion of male users. When such products launch on days when more women are active on the platform, this bias is reduced, suggesting that choices on experiment bias can help them gather more informative data. Furthermore, research on the use of algorithmic predictions in decision-making shows that even experienced managers may not be able to discern the existence of bias, let alone its size and direction (Biermann et al., 2022; Kim et al., 2024)—suggesting that finding ways to develop this capability may be another approach to address this problem.

Uninformative experiments that do not result in learning may seem like a rather benign problem. However, it is worth highlighting when this may have much more severe consequences. In the Bayesian framework, priors can influence experiment choices. For example, Gans (2023) shows that if an entrepreneur has a strong conviction about a particular hypothesis, they may be more likely to choose an experiment biased toward positive news rather than one that may be more likely to challenge their beliefs. This can lead to a self-reinforcing cycle where prior beliefs shape experimental choices, which in turn provide confirming evidence, further entrenching the initial assumptions. This underscores that focusing solely on generating hypotheses and theories—without thinking carefully about how experiment choices affect the data that are generated from it—may distort their learning.

Moreover, the fact that priors influence the choice of experiment is especially concerning when considering that individuals often do not follow precise Bayesian updating processes in practice. For example, when there is noninformative evidence, they should not update their beliefs. However, Kieren and Weber (2020) show that we tend to process noisy information signals in a reference-dependent manner based on our prior beliefs. They provide experimental evidence that instead of recognizing that uninformative signals do not carry information, individuals update their beliefs based on the valence of the signal relative to their prior expectations, distorting learning.

Together, these insights highlight that poor choices on experiments can result in uninformative experiments that impede learning. In addition to considerations of noise and bias, entrepreneurs make choices on the type of experiment, how hypotheses are operationalized, and what metrics are collected – all of which are key determinants of whether experimentation results in learning about the chosen hypotheses.

3.2 (In)attention to data from experiments

Even when entrepreneurs have selected an informative experiment, they may fail to learn from it due to problems of attention. This attention problem can manifest in two key ways: inattention and selective attention to data.

First, entrepreneurs may struggle to identify and attend to key insights. For example, in an experiment with entrepreneurial firms, Kim (2024) found that managers knew competitor prices were important and believed they already knew this information, but 58% held outdated knowledge and could not state competitors' current prices. When they became aware of up-to-date prices, they changed their pricing and quality decisions, and their performance measures increased by 8-15%. Applied to the context of entrepreneurial experimentation, this suggests that entrepreneurs may examine data at a specific point in time but fail to continuously monitor it as the experiment progresses. For example, an entrepreneur might design an appropriate experiment, collecting customer and competitor reactions, but fail to track competitors after the initial analysis, assuming they remain static. McGrath and MacMillan (1995) illustrate that Polaroid fell into this exact pitfall, resulting in a $200 million loss from Polavision instant movies. In this case, Polaroid assumed that a $7 three-minute cassette would compete effectively against a $20 half-hour videotape, failing to attend to the rapidly declining costs of videotaping and playback technologies.

A related but distinct issue is that entrepreneurs may be selectively attentive. Hanna et al. (2014) highlight the critical role of noticing in learning, finding that 85% of seaweed farmers do not know the size of their seaweed pods, a key determinant of performance, despite being knowledgeable about many other aspects of their business. Even when provided with informative data, they fail to learn from it without being provided with a summary that explicitly highlights the relationship between pod size and performance. The key difference between this and inattention described in the first point is that inattention highlights the failure to continue attending to data, while selective attention emphasizes the failure to notice critical insights even when attending to the data.

As an example, consider the case of Kozmo. In the early 2000s, Kozmo was heralded as the epitome of the ambitious startup, aiming to revolutionize the way people received goods by offering one-hour deliveries of various items including food, basics, movies, and games. Kozmo ran many experiments, testing and iterating on its offerings to optimize its business model: they tested leasing out warehouse space to deliver orders by bike (Moser, 2021), equipping couriers with motorized scooters, and even expanded their offering to sell cigarettes (Bensinger, 2012). For these experiments, they collected data on delivery times, order volumes, customer satisfaction, and more. In recent years, the world has come to resemble Kozmo's vision: DoorDash was valued at $72 billion on its stock market debut (Hall, 2020), Uber acquired alcohol delivery service Drizly for 1.1 billion to help "consumers get almost anything—from food to groceries to alcohol" (Morris, 2024), and retail giants Amazon, Walmart, and Target are in a shipping war to deliver the fastest delivery speeds (Hadero et al., 2023). Yet, Kozmo is nowhere to be found. What prevented it from becoming a trailblazer of the delivery space?

While there may be many reasons for its failure, a key factor stands out: although Kozmo ran experiments and collected a wide range of data, it failed to pay attention to a crucial metric: average order size (Sandoval, 2002). Many of Kozmo's deliveries were for low-value items, with its average order size at $12, when labor and overhead alone was $10 (Cassidy, 2002). While their data from experiments showed how low their average order size was, Kozmo paid selective attention to growth metrics on customer and order numbers, neglecting the fundamental unit economics of each order, which led them to continue investing in marketing and expansion (Sandoval, 2002). In the last few months before it shuttered, it finally instituted minimum order sizes and delivery fees—but it was too late. Ultimately, despite raising over $250 million in funding and achieving significant brand recognition, Kozmo shut down in April 2001, laying off its 1,100 employees (Bensinger, 2012; Sandoval, 2002). Kozmo's story highlights the dangers of selective attention: despite running experiments and collecting data, they failed to notice the crucial insight that their average order size was too low to support their business model.

In a Bayesian framework, these problems of attention can lead to a failure to learn, even when the evidence is new and informative. Inattention and selective attention can both be modeled by ignoring a portion of the evidence, leading to slower or incomplete belief updating. Alternatively, inattention can be modeled as adding noise to the likelihood function, $P(H|E) = \frac{(P(E|H)+\epsilon)P(H)}{P(E)}$, reflecting the idea that not all relevant aspects of the evidence are considered or are perceived incorrectly. The noise term ϵ, dilutes the influence of actual evidence on the posterior belief. Similarly, selective attention may introduce a parameter that limits the amount of new evidence that can be considered at each update. New information may be weighted less heavily in the updating process, reflecting a conservative adjustment to prior beliefs if the amount of evidence exceeds the processing capacity.

While the Bayesian framework allows us to highlight the implications of these attention problems, it is important to emphasize that the Bayesian process itself can exacerbate them. Priors are likely to guide attention (Felin and Zenger, 2017; Zellweger and Zenger, 2023), and if an entrepreneur believes that an important dimension is unlikely to matter, they may be especially likely to not notice it to be able to learn whether it actually does matter. For example, Kozmo's founder-CEO, Joseph Park, seems to have believed that allowing customers to order anything they wanted was key to success, which may have reinforced their blind spot on order size (Bensinger, 2012). Additionally, overconfident entrepreneurs may be less receptive to new information that challenges their views and fail to pay attention to it, effectively leading them to discount the evidence. Over time, the cumulative effect of these factors can lead to a significant divergence between the learning an inattentive entrepreneur takes away and the true state of the world – even when they have designed informative experiments.

3.3 Model bias in interpreting data from experiments

Even when entrepreneurs have selected an informative experiment and attended to the key data, they may still make incorrect inferences, hindering learning.

Much evidence suggests that priors bias how individuals update beliefs (Benjamin, 2019). This is highlighted in experiments on belief polarization, where people with different priors who observe the same evidence draw opposite conclusions and move further apart in their beliefs. For example, Lord et al. (1979) show that proponents and opponents of capital punishment report further divergence in beliefs after reading the same account of a study on capital punishment – a finding that has been replicated in diverse contexts including consumer beliefs of brand quality (Russo et al., 1998).

While this evidence of updating toward one's priors is often interpreted as evidence of a bias relative to Bayesian updating, this may in fact be consistent with and even exacerbated by Bayesian reasoning (Dixit and Weibull, 2007; Andreoni and Mylovanov, 2012; Benoît and Dubra, 2018). For instance, Baliga et al. (2013) attribute this to ambiguity aversion, and Benoît and Dubra (2018) show how this may manifest when people have private information about an "ancillary matter" that does not directly bear on the issue of interest but matters for the interpretation of evidence. One proposed mechanism that may be especially relevant for entrepreneurial experimentation is the lack of consideration of alternative hypotheses. Fischhoff and Beyth-Marom (1983) propose that this updating behavior may be explained by the fact that people assess how consistent a signal is with their favored hypothesis, without considering its consistency with alternative hypotheses—highlighting the importance of considering multiple alternatives (Gans et al., 2019).

Furthermore, Schwartzstein and Sunderam (2021) show that a wrong model can be more compelling than the truth, especially when the data are quite random under the true model—which biases learning and leads to suboptimal decisions. They consider a framework where persuaders cannot change the observed data but propose models for interpreting this data to a decision-maker, who chooses the model that makes the observed data most likely given their prior. A key assumption of their framework is that a proposed model is more compelling than an alternative if it makes the observed data most likely given their prior. This assumption corresponds to various ideas in the strategy literature on what people might find persuasive, such as narratives that help with "sensemaking" (Weick, 1995; Chater and Loewenstein, 2016; Pillai et al., 2024), which entrepreneurs may also fall prey to.

While Schwartzstein and Sunderam (2021)'s model focuses on persuading others, this model can be applied to persuading oneself, as illustrated by the xkcd comic strip on "curve-fitting methods".[4] It highlights that scientists might fit the same data with differ-

[4]https://www.explainxkcd.com/wiki/index.php/2048

ent curves, which can result in them persuading themselves with a particular narrative, as well as persuading others. Similarly, research in strategy and management shows how managerial beliefs are resistant to change in the face of contradictory facts (Argyris and Schön, 1997; Levitt and March, 1988; Denrell and March, 2001; Pillai et al., 2024) and subject to systematic myopias that impede learning (Levinthal and March, 1993; Levitt and March, 1988; Tripsas and Gavetti, 2000; Menon and Yao, 2024).

Applied to entrepreneurial experimentation, this work suggests that even with informative experiments and attention to data, learning from experiments in ways that improve strategies and firm performance may be difficult. Such examples abound, particularly among startups and their beliefs about their business models. Jennifer Hyman, the CEO of Rent the Runway, a clothing rental subscription service, has discussed how the belief that their core hidden business is dry-cleaning—rather than high-technology or fashion as one might assume—was key to unlocking their growth, as it dramatically influenced their learnings and decisions (Hoffman, 2024). Without it, her early experiments may have led the company to outsource their drycleaning operations or solve customer problems through technology. Instead, her belief led her to focus on their human operations to improve the customer experience, investing in quality control functions that could only be monitored by humans. This belief led her to learn from her experiments to invest in specialized labor in their warehouses—expert seamstresses—and experiment with providing corporate benefits to these hourly workers that could help retain them and improve the business.

A more cautionary tale is Framebridge, a venture-backed startup that sold custom framing for photos online. Despite burning through $20 million dollars in losses and struggling to achieve economies of scale, they saw themselves as a technology business and believed they would eventually achieve scale efficiency like Zappos, Facebook, and Uber given their growth numbers (Koning and Dadlani, 2022). This belief led its founder, Susan Tynan, to invest in robots for their warehousing and pursue growth driven by automation, even as she saw her experiments result in losses. Only later did she realize they were misinterpreting their data due to beliefs about their business being similar to technology unicorns. They shifted their model to see themselves as an on-demand manufacturing business, which changed how they interpreted their data and led them make decisions that allowed them to reach profitability.

4 Learning from experimentation in the age of AI

The previous section highlighted the frictions that impede learning from experiments, emphasizing the importance of designing informative experiments, attending to data, and drawing correct inferences when using experimentation to improve strategies.

Recent advancements in AI suggest that it may play an increasingly important role in various domains, including entrepreneurial experimentation. A growing body of research demonstrates AI's superior performance relative to human experts across diverse tasks, including making predictions (Agrawal et al., 2018; Cowgill, 2019; Kim et al., 2024), writing (Noy and Zhang, 2023; Doshi and Hauser, 2023), and business ideation and mentoring (Girotra et al., 2023; Dell'Acqua et al., 2023; Otis et al., 2023).

Most relevantly for entrepreneurial experimentation, emerging research suggests that these advancements may impact all three stages of the experimentation process: idea generation, testing and learning from experiments, and idea selection. Csaszar et al. (2024) highlights that AI may change the strategic decision-making process by enhancing the generation and evaluation of strategies. They provide evidence from a leading European startup accelerator that Large Language Models (LLMs) can generate entrepreneurial strategies at a level comparable to, or even surpassing, those of human entrepreneurs. This raises the intriguing possibility that AI could expand the search space for hypotheses and potentially generate superior ideas. Corroborating this potential, Ludwig and Mullainathan (2024) and Mullainathan and Rambachan (2024) demonstrate machine learning algorithms' capacity to generate scientific theories through data-driven pattern recognition and anomaly detection, often identifying those overlooked by humans. While these findings raise the possibility that hypothesis generation may no longer be the exclusive domain of human cognition, questions remain about AI's ability to develop truly innovative theories, especially in the realms of strategy and entrepreneurship (Boussioux et al., 2023; Camuffo et al., 2022).

AI's potential may also shape the design and implementation of informative experiments. Horton (2023) and Brand et al. (2023) introduce the concept of using LLMs to simulate human samples in real-world experiments. Their findings indicate that these AI-generated samples produce responses consistent with economic theory predictions and real-world willingness-to-pay estimates. This suggests that AI could dramatically expand the scope and reduce the cost of experimentation, potentially providing more informative data and facilitating easier implementation of complex experimental designs.

Finally, the last phase of experimentation may also be affected by AI. Csaszar et al. (2024) show that LLMs can provide evaluations of entrepreneurial business plans that correlate with those of experienced angel and venture capital investors. These AI evaluations exhibit higher inter-rater reliability than that observed across human investors, suggesting their potential value in the idea selection and evaluation process. That said, whether these evaluations contain true informative signal and what the equilibrium effects might look like remain key open questions.

Together, these findings suggest that AI may significantly aid in the entire process of experimentation. If so, this raises important questions about the evolving role of human entrepreneurs and the capabilities that may become increasingly valuable complements to AI-assisted experimentation.

One possibility is that the ability to learn effectively from data may become an increasingly scarce and valuable entrepreneurial skill. While AI can assist in expanding the range of possible experiments and improving their design, the ultimate responsibility for choosing which experiments to conduct and updating prior beliefs based on the resulting data remains with the entrepreneur. The entrepreneur's capacity to synthesize AI-generated insights with their own domain knowledge and strategic intuition may also become a key differentiator in successful entrepreneurial experimentation.

Moreover, the role of prior beliefs may gain increased importance in this AI-augmented landscape. Different entrepreneurs may derive vastly different insights from the same set of (AI-assisted) experiments, depending on their initial beliefs and their ability to integrate new information. This raises the possibility that the capacity to develop novel, contrarian beliefs with causal logic may remain a crucial entrepreneurial skill (Camuffo et al., 2022; Zellweger and Zenger, 2023).

In sum, while AI advancements have the potential to transform entrepreneurial experimentation, the ability to choose informative experiments, make meaningful inferences from the data, and integrate these insights into novel and coherent strategies may remain critical skills for entrepreneurs. Future research should explore how AI might shape the process and landscape of entrepreneurial experimentation, and how entrepreneurs can best leverage these technologies to enhance their learning processes and strategic decision-making capabilities.

5 Discussion and conclusion

This chapter explores the critical role of testing in entrepreneurial experimentation and highlights how its effectiveness depends on a range of factors relating to how entrepreneurs design and learn from information.

These insights raise many more questions than they answer, and highlight new avenues for research on entrepreneurial experimentation. First, if learning from data is a complement to theory generation and choice, then we need more nuance on how entrepreneurs learn from data. The insights in this chapter suggest that paying attention to more theory-free data in the world may be in and of itself more valuable. Might too much emphasis on theory lead us to possibly underplay this value? As Ludwig and Mullainathan (2024) suggest, machine learning algorithms can often outperform human experts in identifying data-driven patterns and anomalies without relying on explicit theories. This suggests that entrepreneurs may benefit from a more data-driven approach to experimentation, using AI tools to identify promising opportunities and refine their strategies.

Second, we may need to teach entrepreneurs not just to be scientific in theory generation (Camuffo et al., 2020; Felin and Zenger, 2017), but also in the steps of inference. We often assume that entrepreneurs have the tools to measure and learn what they need, but might this be a larger friction than we realize? As an analogy, while scientists need to carefully design the theories they are testing, if they use a broken microscope, they will not learn about the hypothesis. What, then, are the microscopes in entrepreneurial experimentation? It may be that entrepreneurs need to develop a range of "experimentation capabilities," including the ability to design informative experiments, interpret data, and make correct inferences from the available evidence. Identifying and addressing the specific tools and skills that entrepreneurs need to conduct effective experiments – especially when they hold on to specific theories – may be an important area for future research and education.

Lastly, this capability to design informative experiments and learn from data may explain why we see successful businesses that arrived there without clear theories. Maybe they lucked into a great theory, or maybe they were great at learning from data. It raises the question: what is the binding constraint on better entrepreneurship? Is it the ability to generate good theories, to learn from experiments effectively, or to make the right choices based on the available evidence? The answer likely lies in a combination of factors, including the entrepreneur's cognitive abilities, the richness of the available data, and the complexity of the problem space. Understanding how these factors interact and shape entrepreneurial success is a key challenge for future research.

References

Adner, R. and Levinthal, D. A. (2004). What is not a real option: Considering boundaries for the application of real options to business strategy. *Academy of Management Review*, 29(1):74–85.

Agrawal, A., Camuffo, A., Gambardella, A., Gans, J. S., Scott, E. L., and Stern, S. (2024). Bayesian entrepreneurship. *Working paper*.

Agrawal, A., Gans, J., and Goldfarb, A. (2018). *Prediction Machines: The Simple Economics of Artificial Intelligence*. Harvard Business Press, Cambridge, MA.

Agrawal, A., Gans, J. S., and Stern, S. (2021). Enabling entrepreneurial choice. *Management Science*, 67(9):5510–5524.

Andreoni, J. and Mylovanov, T. (2012). Diverging opinions. *American Economic Journal: Microeconomics*, 4(1):209–232.

Argote, L. (2012). *Organizational learning: Creating, retaining and transferring knowledge*. Springer Science & Business Media.

Argyris, C. and Schön, D. A. (1997). Organizational learning: A theory of action perspective. *Reis*, (77/78):345–348.

Azevedo, E. M., Deng, A., Montiel Olea, J. L., Rao, J., and Weyl, E. G. (2020). A/b testing with fat tails. *Journal of Political Economy*, 128(12):4614–000.

Baliga, S., Hanany, E., and Klibanoff, P. (2013). Polarization and ambiguity. *American Economic Review*, 103(7):3071–3083.

Benjamin, D. J. (2019). Errors in probabilistic reasoning and judgment biases. *Handbook of Behavioral Economics: Applications and Foundations 1*, 2:69–186.

Benoît, J.-P. and Dubra, J. (2018). When do populations polarize? an explanation.

Bensinger, G. (2012). In kozmo.com's failure, lessons for same-day delivery. https://www.wsj.com/articles/BL-DGB-25584 Accessed: 2024-07-30.

Biermann, J., Horton, J. J., and Walter, J. (2022). Algorithmic advice as a credence good. *ZEW-Centre for European Economic Research Discussion Paper*, (22-071).

Blank, S. (2013). Why the lean start-up changes everything. *Harvard Business Review*, 91(5):63–72.

Boussioux, L., N Lane, J., Zhang, M., Jacimovic, V., and Lakhani, K. R. (2023). The crowdless future? how generative ai is shaping the future of human crowdsourcing. *The Crowdless Future*.

Brand, J., Israeli, A., and Ngwe, D. (2023). Using GPT for market research. Available at SSRN: https://ssrn.com/abstract=4395751.

Brynjolfsson, E. and McElheran, K. (2016). The rapid adoption of data-driven decision-making. *American Economic Review*, 106(5):133–39.

Camuffo, A., Cordova, A., Gambardella, A., and Spina, C. (2020). A scientific approach to entrepreneurial decision making: Evidence from a randomized control trial. *Management Science*, 66(2):564–586.

Camuffo, A., Gambardella, A., Pignataro, A., et al. (2022). Framing strategic decisions in the digital world. *Strategic Management Review*, 4(2):127–160.

Cao, R., Koning, R., and Nanda, R. (2023). Sampling bias in entrepreneurial experiments. *Management Science*.

Cassidy, J. (2002). *Dot. con: The greatest story ever sold*. HarperCollins Publishers.

Chater, N. and Loewenstein, G. (2016). The under-appreciated drive for sense-making. *Journal of Economic Behavior & Organization*, 126:137–154.

Cohen, S. L., Bingham, C. B., and Hallen, B. L. (2019). The role of accelerator designs in mitigating bounded rationality in new ventures. *Administrative Science Quarterly*, 64(4):810–854.

Contigiani, A. and Levinthal, D. A. (2019). Situating the construct of lean start-up: Adjacent conversations and possible future directions. *Industrial and Corporate Change*, 28(3):551–564.

Cowgill, B. (2019). Bias and productivity in humans and machines. Available at SSRN: https://ssrn.com/abstract=3433737.

Csaszar, F., Ketkar, H., and Kim, H. (2024). Artificial intelligence and strategic decision-making: Evidence from entrepreneurs and investors. Available at SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4913363.

Dell'Acqua, F., McFowland, E., Mollick, E., Lifshitz-Assaf, H., Kellogg, K. C., Rajendran, S., Lisa, K., Candelon, F., and Lakhani, K. R. (2023). Navigating the jagged technological frontier: Field experimental evidence of the effects of AI on knowledge worker productivity and quality. Working Paper 24-013, Harvard Business School.

Denrell, J. and March, J. G. (2001). Adaptation as information restriction: The hot stove effect. *Organization science*, 12(5):523–538.

Dixit, A. K. and Weibull, J. W. (2007). Political polarization. *Proceedings of the National Academy of sciences*, 104(18):7351–7356.

Doshi, A. R. and Hauser, O. (2023). Generative artificial intelligence enhances creativity. Available at SSRN: https://ssrn.com/abstract=4535536.

Eisenmann, T. R., Ries, E., and Dillard, S. (2012). *Hypothesis-driven entrepreneurship: The lean startup*. SSRN.

Enke, B. (2020). What you see is all there is. *The Quarterly Journal of Economics*, 135(3):1363–1398.

Enke, B. and Zimmermann, F. (2019). Correlation neglect in belief formation. *The review of economic studies*, 86(1):313–332.

Felin, T. and Zenger, T. R. (2017). The theory-based view: Economic actors as theorists. *Strategy Science*, 2(4):258–271.

Fischhoff, B. and Beyth-Marom, R. (1983). Hypothesis evaluation from a bayesian perspective. *Psychological review*, 90(3):239.

Gans, J. S. (2023). Experimental choice and disruptive technologies. *Management Science*, 69(11):7044–7058.

Gans, J. S., Stern, S., and Wu, J. (2019). Foundations of entrepreneurial strategy. *Strategic Management Journal*, 40(5):736–756.

Girotra, K., Meincke, L., Terwiesch, C., and Ulrich, K. T. (2023). Ideas are dimes a dozen: Large language models for idea generation in innovation. Available at SSRN: https://ssrn.com/abstract=4526071.

Gomez-Uribe, C. A. and Hunt, N. (2015). The netflix recommender system: Algorithms, business value, and innovation. *ACM Transactions on Management Information Systems (TMIS)*, 6(4):1–19.

Hadero, H., D'Innocenzio, A., and the Associated Press (2023). Amazon, walmart and target are speeding up deliveries as they fight challengers shein and temu. Accessed: 2024-07-30.

Hall, C. (2020). Doordash closes first day of trading with $72b valuation. https://news.crunchbase.com/venture/doordash-stock-pops-on-first-day-of-trading/ Accessed 2024-08-01.

Hanna, R., Mullainathan, S., and Schwartzstein, J. (2014). Learning through noticing: Theory and evidence from a field experiment. *The Quarterly Journal of Economics*, 129(3):1311–1353.

Hoffman, R. (2024). The secret business behind the business. Podcast Episode https://mastersofscale.com/jenn-hyman-the-secret-business-behind-the-business/. Masters of Scale.

Horton, J. J. (2023). Large language models as simulated economic agents: What can we learn from Homo Silicus? arXiv:2301.07543.

Kerr, W. R., Nanda, R., and Rhodes-Kropf, M. (2014). Entrepreneurship as experimentation. *Journal of Economic Perspectives*, 28(3):25–48.

Kieren, P. and Weber, M. (2020). Expectation formation under uninformative signals. *Available at SSRN 3971733*.

Kim, H. (2024). The value of competitor information: Evidence from a field experiment. *Forthcoming, Management Science*.

Kim, H., Glaeser, E. L., Hillis, A., Kominers, S. D., and Luca, M. (2024). Decision authority and the returns to algorithms. *Strategic Management Journal*, 45(4):619–648.

Klein, C. (2023). Why coca-cola's 'new coke' flopped. https://www.history.com/news/why-coca-cola-new-coke-flopped. Accessed: 2024-07-29.

Koning, R. and Dadlani, A. (2022). Framebridge (a): Reimagining custom framing. Harvard Business School Case 723-352. Revised November 2023.

Koning, R., Hasan, S., and Chatterji, A. (2022). Experimentation and start-up performance: Evidence from a/b testing. *Management Science*, 68(9):6434–6453.

Levinthal, D. A. (2017). Mendel in the c-suite: Design and the evolution of strategies. *Strategy Science*, 2(4):282–287.

Levinthal, D. A. and March, J. G. (1993). The myopia of learning. *Strategic management journal*, 14(S2):95–112.

Levitt, B. and March, J. G. (1988). Organizational learning. *Annual review of sociology*, 14(1):319–338.

Lord, C. G., Ross, L., and Lepper, M. R. (1979). Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *Journal of personality and social psychology*, 37(11):2098.

Ludwig, J. and Mullainathan, S. (2024). Machine learning as a tool for hypothesis generation. *The Quarterly Journal of Economics*, 139(2):751–827.

McGrath, R. G. and MacMillan, I. C. (1995). Discovery-driven planning recognizes that planning. *Harvard business review*, 45.

Menon, A. and Yao, D. (2024). Rationalizing outcomes: Interdependent learning in competitive markets. *Strategy Science*.

Miller, A., O'Mahony, S., and Cohen, S. L. (2024). Opening the aperture: Explaining the complementary roles of advice and testing when forming entrepreneurial strategy. *Organization Science*, 35(1):1–26.

Morris, C. (2024). Uber's big $1 billion bet on alcohol delivery is getting shuttered after less than 3 years. https://fortune.com/2024/01/16/uber-shuts-down-drizly-alcohol-delivery-1-billion-acquisition/ Accessed: 2024-07-30.

Moser, W. (2021). Why dot-com disaster kozmo never became instacart. https://marker.medium.com/5283563dc79b Accessed: 2024-07-30.

Mullainathan, S. and Rambachan, A. (2024). From predictive algorithms to automatic generation of anomalies. Technical report, National Bureau of Economic Research.

Murray, F. and Tripsas, M. (2004). The exploratory processes of entrepreneurial firms: The role of purposeful experimentation. In *Business strategy over the industry lifecycle*, pages 45–75. Emerald Group Publishing Limited.

Nagaraj, A. (2022). The private impact of public data: Landsat satellite maps increased gold discoveries and encouraged entry. *Management Science*, 68(1):564–582.

Newell, A. and Simon, H. A. (1976). Computer science as empirical inquiry: Symbols and search. *Communications of the ACM*, 19(3):113–126.

Noy, S. and Zhang, W. (2023). Experimental evidence on the productivity effects of generative artificial intelligence. *Science*, 381(6654):187–192.

Otis, N., Clarke, R. P., Delecourt, S., Holtz, D., and Koning, R. (2023). The uneven impact of generative AI on entrepreneurial performance. Available at SSRN: https://ssrn.com/abstract=4671369.

Pfeffer, J. and Sutton, R. I. (2006). Evidence-based management. *Harvard Business Review*, 84(1):62.

Pillai, S. D., Goldfarb, B., and Kirsch, D. (2024). Lovely and likely: Using historical methods to improve inference to the best explanation in strategy. *Strategic Management Journal*.

Ries, E. (2011a). How dropbox started as a minimal viable product. https://techcrunch.com/2011/10/19/dropbox-minimal-viable-product/ Accessed: 2024-07-30.

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

Rousseau, D. M. (2006). Is there such a thing as "evidence-based management"? *Academy of management review*, 31(2):256–269.

Russo, J. E., Meloy, M. G., and Medvec, V. H. (1998). Predecisional distortion of product information. *Journal of Marketing Research*, 35(4):438–452.

Sandoval, G. (2002). Kozmo to shut down, lay off 1,100. *CNET News*. https://www.cnet.com/tech/tech-industry/kozmo-to-shut-down-lay-off-1100/.

Schwartzstein, J. and Sunderam, A. (2021). Using models to persuade. *American Economic Review*, 111(1):276–323.

Thomke, S. H. (2020). *Experimentation works: The surprising power of business experiments*. Harvard Business Press.

Tripsas, M. and Gavetti, G. (2000). Capabilities, cognition, and inertia: Evidence from digital imaging. *Strategic Management Journal*, 21(10-11):1147–1147.

Weick, K. E. (1995). *Sensemaking in organizations*, volume 3. Sage publications Thousand Oaks, CA.

Zellweger, T. and Zenger, T. (2023). Entrepreneurs as scientists: A pragmatist approach to producing value out of uncertainty. *Academy of Management Review*, 48(3):379–408.