Management Science

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

A Scientific Approach to Entrepreneurial Decision Making: Evidence from a Randomized Control Trial

Arnaldo Camuffo, Alessandro Cordova, Alfonso Gambardella, Chiara Spina

informs
http://pubsonline.informs.org/journal/mnsc

MANAGEMENT SCIENCE
Vol. 66, No. 2, February 2020, pp. 564–586
ISSN 0025-1909 (print), ISSN 1526-5501 (online)

A Scientific Approach to Entrepreneurial Decision Making: Evidence from a Randomized Control Trial

Arnaldo Camuffo,[a] **Alessandro Cordova,**[a] **Alfonso Gambardella,**[a,b] **Chiara Spina**[a]

[a] Bocconi University, Invernizzi Center for Research on Innovation, Organization, Strategy, and Entrepreneurship ICRIOS, 20136 Milan, Italy; [b] Center for Economic and Policy Research CEPR, London, EC1V 0DX United Kingdom

Contact: arnaldo.camuffo@unibocconi.it, http://orcid.org/0000-0001-9039-1057 (ArC); cordova.ndr@gmail.com, http://orcid.org/0000-0001-6005-5167 (AlC); alfonso.gambardella@unibocconi.it, http://orcid.org/0000-0002-8714-5813 (AG); chiara.spina@unibocconi.it, http://orcid.org/0000-0002-2542-557X (CS)

Received: September 5, 2017
Revised: July 28, 2018
Accepted: October 16, 2018
Published Online in Articles in Advance: August 8, 2019

https://doi.org/10.1287/mnsc.2018.3249

Abstract. A classical approach to collecting and elaborating information to make entrepreneurial decisions combines search heuristics, such as trial and error, effectuation, and confirmatory search. This paper develops a framework for exploring the implications of a more scientific approach to entrepreneurial decision making. The panel sample of our randomized control trial includes 116 Italian startups and 16 data points over a period of about one year. Both the treatment and control groups receive 10 sessions of general training on how to obtain feedback from the market and gauge the feasibility of their idea. We teach the treated startups to develop frameworks for predicting the performance of their idea and conduct rigorous tests of their hypotheses, very much as scientists do in their research. We let the firms in the control group instead follow their intuitions about how to assess their idea, which has typically produced fairly standard search heuristics. We find that entrepreneurs who behave like scientists perform better, are more likely to pivot to a different idea, and are not more likely to drop out than the control group in the early stages of the startup. These results are consistent with the main prediction of our theory: a scientific approach improves precision—it reduces the odds of pursuing projects with false positive returns and increases the odds of pursuing projects with false negative returns.

History: Accepted by Marie Thursby, entrepreneurship and innovation.
Funding: The authors acknowledge financial support from the Innovation Growth Laboratory. The authors also acknowledge financial support on this project from the Invernizzi Center for Research on Innovation, Organization, Strategy, and Entrepreneurship (ICRIOS) of Bocconi University.
Supplemental Material: Data files are available at https://doi.org/10.1287/mnsc.2018.3249.

Keywords: entrepreneurship • decision making • scientific method • startup • randomized control trial

1. Introduction

In recent years, both the practice of management and the scholarly debate have recognized that firms must make decisions about new products or business ideas under growing uncertainty. This has discouraged firms from relying on heavy ex ante commitments of resources to specific business models or product features, and it has encouraged them to adopt more flexible approaches based on market feedback about early outlines of the idea, staggered investments, and adaptations to environmental changes. Not only have many firms adopted this approach (e.g., Brown 2008), but also, new theories in strategic management and economics on this subject have emerged, such as discovery-driven planning (McGrath and MacMillan 1995, 2009), real option strategies (McGrath 1997, O'Brien et al. 2003, Adner and Levinthal 2004, Li et al. 2007), effectuation (Sarasvathy 2001), design thinking (Martin 2009), and business experimentation (Kerr et al. 2014, Gans et al. 2017).

However, the academic literature and the practice of management have not deepened the question of whether there are different approaches to collecting and elaborating information to make these decisions. In this paper, we contrast two approaches. On one hand, firms can use search heuristics—like trial-and-error processes (Nicholls-Nixon et al. 2000), effectuation (Sarasvathy 2001), or confirmatory search (Shepherd et al. 2012). On the other hand, they can apply a more scientific approach to understand and test the mechanisms that affect the performance of their new products or ideas. Scholars and practitioners alike have explored this issue lately (e.g., Blank 2006, Felin and Zenger 2009, Grandori 2010, Ries 2011, Zenger 2016). However, it is worth exploring further how a scientific approach to entrepreneurial decision making affects performance, and we lack good evidence.

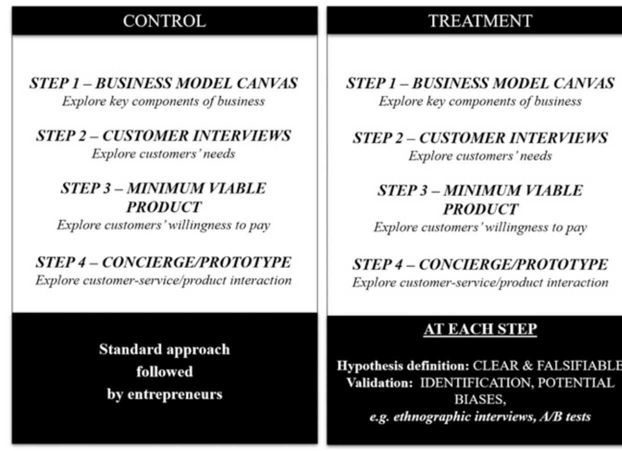
This study empirically tests the different performance effects of a scientific approach to the decision to launch a new business model or product idea compared with an approach based on heuristics and tries to explain this difference. It uses a randomized control trial (RCT) involving 116 Italian startup founders. We randomly assign these entrepreneurs to a treatment group or a control group, offer them a four-month entrepreneurship training program, and monitor the performance of the two groups over time. The program focuses on a set of managerial practices for making decisions about the viability of a new business model or product idea. We teach both the treated and control startups to search for, collect, and elaborate information about the feasibility of their idea before committing resources to it. We also teach them to run experiments to assess their business models or products and to modify them to increase performance if needed. The treatment consists of training the treated group to identify the problem, articulate theories, define clear hypotheses, conduct rigorous tests to prove or disprove them, measure the results of the tests, and make decisions based on these tools. Although we offer the same training to the treated and control groups, we do not provide these decision criteria to the control group. We let them follow their heuristics so that they use their own approach and intuition to assess the information that they receive from the processes that we teach them in the program.

Although our training program teaches all firms to collect signals about the value of entrepreneurial ideas, how entrepreneurs collect and elaborate information affects the interpretation of the signals, the quality of the inference that they make, and ultimately, their performance. We theorize that a scientific approach to entrepreneurial decision making leads to superior inferential power, because it reduces false positives and false negatives compared with the typical decision heuristics followed by entrepreneurs. In particular, because most entrepreneurial ideas eventually fail, entrepreneurs are more likely to face false positive than false negative ideas. As a result, a scientific approach to entrepreneurial decision making is more likely to encourage exits from current business ideas or pivots to new ideas. In our RCT, we first test these propositions on exit and pivot and then, test the effect of the scientific approach on performance.

2. Case Study—Inkdome

The case study of one of our treated startups, Inkdome, illustrates well our definition of a scientific approach to entrepreneurial decision making. When Inkdome entered our trial, its business idea was to create a search engine to help users to find the right tattooist for their style. We discuss Inkdome's behavior during the four steps of our four-month training program focused on market validation: (1) business model canvas, (2) customer interviews, (3) minimum viable product, and (4) concierge or prototype. Figure 1 summarizes the training program contents. Although we teach both treated and control startups about these

Figure 1. Training Program and Differences Between Treated and Control Startups



four steps, we teach the treated startups in particular to elaborate a framework for understanding the impact of their idea, predicting business performance, defining clear hypotheses, designing rigorous experiments to confirm or disconfirm them, and making decisions accordingly. This approach permeates all of the steps of our training program. Appendix A illustrates the differences between treatment and control in our training problem using the example of the first class.

2.1. Business Model Canvas

The business model canvas is an approach to business model design widely used in entrepreneurship education (Osterwalder and Pigneur 2009). It is a scaled-down representation of a generic business model that enumerates and illustrates its key components (customer segments, value proposition, etc.). Although the core of the training on the scientific method unfolds in Steps 2–4, the business model canvas is the starting context for treated startups to realize that their project relies on a set of hypotheses that they must test over time. In particular, we tell startups in the treated group that Steps 2–4 focus on testing the potential of the founders' value proposition and its fit with the hypothesized market target and that the approach that they are learning is useful for testing aspects of the business that will be relevant later (e.g., the firm's revenue model).

2.2. Customer Interviews

We teach all startups how to interview customers to understand the firm's potential market, segment it, learn about customers' needs, and collect feedback about the startup's idea. However, we further train the treated startups to collect and elaborate this information to develop general frameworks and formulate specific hypotheses about the behavior of customers.

We observed that startups in the control group conduct their customer interviews as an unstructured exploration. They typically create online questionnaires, which they post on their personal social media accounts, inviting their contacts to respond. A drawback of this approach is that the sampling is not representative of the population of customers. Also, questions are often direct, such as "Did you have problems finding tattooists online?"; this limits the ability to explore customers' experiences and derivate, abductively, their problems. They also ask for straight feedback on their idea, with questions like "Would you use our service?"; to which they often receive the following comments: "Yes, why not?! It seems a great idea." There are many reasons why this produces confirmation bias: (i) some questionnaire respondents are friends and do not want to disappoint their peers, and (ii) this is a fictitious market setting where respondents do not use the service; therefore, it is not costly to respond affirmatively. Although this approach sounds naïve, it is what typically happens, especially with novice entrepreneurs. For example, in many entrepreneurial pitches when entrepreneurs walk the judges through their ideas, they present pie charts showing high percentages of people who would use the product. These percentages are inconsistent with the high percentage of startups failing, suggesting that the typical startup, like the startups in our control group, does not conduct customer interviews rigorously and appropriately. The problem of collecting data or samples that tend to confirm prior hypotheses is common. For example, Clark and Wiesenfeld (2017) report cases of companies that make decisions based on biased samples that are more likely to corroborate the initial hypotheses or in which managers pursue their initial hypothesis even if the data suggest that it is unlikely to be supported.

Inkdome applied a different approach. First, it developed a framework to understand the mechanisms that can make the business idea feasible. This framework helped identify the key areas requiring validation, which led to the articulation of four clear hypotheses: (i) tattooed people do not always use the same tattooist, (ii) they choose new tattooists online, (iii) this takes time and is painful, and (iv) tattooed people can find online all of the information that they need to make their choice. Without a clear framework and clear hypotheses, entrepreneurs obtain generic feedback that can obscure important information about their business model or weigh equally components that contribute differently to value generation.

Second, Inkdome interviewed tattoo users or individuals as close as possible to their target audience—for example, they sought interviewees in Facebook groups of tattoo enthusiasts. Inkdome also asked open-ended questions: "When was the last time that you were tattooed? Did you know the tattooist? How did you choose him/her?" This quasiethnographic approach is an effective way to gather information to develop frameworks and to formulate and test hypotheses, especially when it involves knowledgeable sources of information, such as lead users (Von Hippel 1986). Appendix B reports the instructions for this quasiethnographic method that we handed to the treatment group. In particular, this approach enables the interviewer to collect facts with limited bias from customers' opinions (Kelley and Littman 2005).

Third, Inkdome defined clear metrics and set explicit decision rules. For example, it set a fraction of the customer interviews as a minimum threshold to support its hypotheses. In particular, Inkdome's decision rule is to reject a hypothesis if less than 60% of their interviews did not provide corroborating evidence (sample size of 50).

Given this threshold, the customers' interviews corroborated Inkdome's first three hypotheses but not the fourth one. Inkdome also collected stories and examples from many interviewees that suggested that the problem was not finding a tattooist but was evaluating the tattooist's skills. Without a clear set of hypotheses and a rigorous method for testing them, they might have collected less useful feedback, made wrong inferences, and probably continued with their business idea. The scientific approach gave Inkdome a clear decision rule: pursue the original idea if all four hypotheses are corroborated; otherwise, abandon the idea of launching a startup or investigate alternative solutions (pivot). In this specific case, the founders saw a new opportunity and pivoted. Thanks to the quasiethnographic approach to customers' interviewing, they learned that the most satisfied interviewees knew tattoo experts (e.g., a friend with several tattoos inked at different locations) who helped them find the right tattooist for their idea. This example not only shows that the scientific approach enabled Inkdome to understand why its idea was unlikely to work but also, suggested where to pivot more successfully. Based on the information gathered through open-ended interviews guided by precise hypotheses, Inkdome changed its business model from a search engine to a platform where users seek advice from experts.

2.3. Minimum Viable Product

Minimum viable product is another widely used concept in entrepreneurship education. We taught all entrepreneurs that, before committing to a final product or service, it is advisable to create a preliminary basic version of the offering with just enough features to let customers experience it and assess their willingness to pay for it. Most of our companies created a web page describing and advertising the new product or service, typically with a button that users can click to buy now, sign up for the free beta, or preorder.

Assume, counterfactually, that Inkdome was a startup in the control group. How would it design and release its landing page? Based on what we observed of firms in the control group, first, Inkdome would not formulate clear hypotheses to understand how to design and release the page but would simply design and release it to begin testing. Second, Inkdome would begin promoting the page on its personal social networks, opening up to feedback mainly from friends or acquaintances. Third, it would not specify an evaluation criterion, a valid and reliable metric, or a decision rule to assess whether the landing page is a successful vehicle for the product. As time elapsed, it might learn and eventually improve the platform and service based on a sequence of trial-and-error attempts. However, this process has limitations similar to those highlighted in the case of customers' interviews. The lack of clear hypotheses renders the startup search process chaotic; similarly, a lack of rigorous testing is likely to generate mistakes and induce bad inferences—for example, control startups mostly make sequential revisions to the landing page (or multiple changes simultaneously) rather than running parallel A/B tests.

Because of the treatment, Inkdome instead began by eliciting its implicit hypotheses. Although it was clear that customers sought contact with tattoo experts, there are different ways to induce this contact. Inkdome initially considered collecting experts' advice and sending it to users via email. Thus, Inkdome developed alternative versions of its landing page and tested them by conducting split (A/B) tests. Inkdome accurately monitored the comparative performance (number of email addresses that customers left) of two landing pages that were identical, except that version A advertised that users would receive advice via email from tattoo experts and version B advertised that users would chat with tattoo experts. This experimental design allowed Inkdome to tease out the different effects of the two design options on performance.

Finally, Inkdome used clear thresholds to corroborate its hypothesis: that an expert-user chat system would outperform the email-based advice system, because users trust conversations with experts more. However, creating a chat system requires substantial resources (technology and tattoo experts) that imply a substantial commitment. Therefore, Inkdome set a sufficiently challenging threshold to justify the investment in the chat option: twice the number of email addresses left on version A of the landing page. The test showed that version B produced 2.5 times more emails than version A. Inkdome, therefore, chose the chat-based system.

2.4. Concierge or Prototype

The term concierge (for services) or prototype (for products) is typically used to denote the delivery of a basic product or service to a small group of customers. Inkdome created a website section where customers could find descriptions of tattoo ideas and put them in contact with the experts. The scientific approach implied, again, that Inkdome asked the right questions (problem identification and hypotheses formulation) and conducted meaningful, rigorously designed experiments (hypothesis test). A control startup would concentrate instead on monitoring general customers' opinions through some type of customer satisfaction survey right after they received the advice of an expert. The control startup also would most likely provide the service by using as an expert one of the company founders to minimize resources and effort. Among other things, the use of a company expert is likely to reinforce a confirmation bias.

A startup following the scientific approach acknowledges that a valid and reliable metric for monitoring the success of the experiment is not what customers say in a customer satisfaction survey but what they do, and in this case, the success factor is the time between receiving expert advice and getting a tattoo. Inkdome realized that, consistent with its hypotheses (online search is painful and time consuming), its service had to reduce the time needed for users to search and evaluate a tattooist online. Inkdome then monitored the time that customers spent to decide where to get tattooed through their service compared with the benchmark average time in the market by calling its users at regular intervals. At the same time, Inkdome realized that it should involve external experts, because founders are biased by their implicit belief or motivation that a venture is successful. The use of external experts reduces the risk of accepting false positives.

2.5. Additional Remarks

The Inkdome case study clarifies three relevant features of our framework and our RCT.

First, we do not give the control group a lighter treatment that makes them less productive than the treated startups. As we will also see when we discuss our data and results, we offer the control group the same number of hours of training and spend the same time teaching them content relevant to the four steps. The only difference is that we do not teach them to frame the problem building a theory, to formulate hypotheses, and to test these using rigorous experiments with valid and reliable metrics and setting thresholds for these metrics to make decisions.

Second, our notion of scientific approach is not a straight deductive method beginning with abstract frameworks that percolate down to hypotheses definition and testing. As shown by Inkdome, the problem is not well defined initially, and the decision makers lack a good idea of the problem itself and of what they are looking for. Discussions within the team or with the customers help them clarify the questions and the problem and then, formulate frameworks and hypotheses in forms that are falsifiable and testable. As we explain in Section 5, our intervention is composed of lectures and one-to-one mentorship. In both the lectures and the one-to-one discussions, we teach and encourage the treated startups during all four steps of our training to collect this information and define the problem and the key issues so that they can elaborate a framework and formulate clear hypotheses to test. Most often, the control startups keep the problem ill defined and neither clarify the questions nor formulate as clearly as the treated group what must be decided or the context or implications of their decisions.

Third, all of our startups entered our RCT having a business idea. Inkdome, for example, began with its online search engine. However, none of the participant startups had developed or tested the idea to a significant extent. Indeed, they were selected to be fully prepared to absorb our approach (whether in the treatment or control group) without any prior commitment to a particular idea. As a result, the initial weeks of training largely affected the ability of firms to evaluate the idea with which they entered the RCT. Over time, the information that they collected became useful for assessing modifications to this original idea or even radical departures from it to pivot to a new idea, such as in Inkdome's case. Once again, this is true of both the treated and the control firms. However, the question is whether the treated firms evaluate their original idea or develop new ideas more effectively than the control group.

3. Science in Entrepreneurial Decision Making: Literature Background

When we say that the behavior of managers or entrepreneurs ought to incorporate aspects of the scientific method, we do not refer to the findings of science but to a general method of thinking about and investigating problems. This idea is not new. It was central in the early studies of management as a discipline as exemplified by Drucker (1955) and Bennis (1962). However, it has been "lost in translation" in management theory (Freedman 1992).

More recently, strategy and entrepreneurship research has elaborated on this idea, emphasizing different components of the scientific attitude (e.g., Sarasvathy and Venkataraman 2011, Venkataraman et al. 2012). Felin and Zenger (2009), in particular, see entrepreneurs as theory developers, engaged in deliberate problem framing and solving, and Zenger (2015) suggests that strategies cannot be mere trial-and-error search processes. Similarly, the problem-finding and problem-solving perspective argues that entrepreneurs and firms create value as they formulate, identify, and solve problems (Hsieh et al. 2007, Felin and Zenger 2015). Building on Grandori (2010), who suggests that managers and entrepreneurs can resort to rational heuristics for better decision making, the study by Lopez-Vega et al. (2016) on open innovation search paths suggests that the scientific search path leads to the discovery of theories and models that birth predictions and hypotheses to be tested by entrepreneurs and managers.

This squares with the notion of business experimentation. Sull (2004) was the first to model the entrepreneurial process as a Popperian process of hypotheses falsification, suggesting that entrepreneurs conduct experiments to test hypotheses around a hypothesized gap in the market that can be filled profitably by a novel combination of resources. Eisenmann et al. (2013) further show the superiority of adopting a scientific approach to business experimentation vis-à-vis three other typical entrepreneurial approaches: (i) build it and they will come, (ii) waterfall planning, and (iii) just do it. Kerr et al. (2014) maintain that entrepreneurship is fundamentally about experimentation, because the knowledge required to succeed cannot be known in advance or deduced from some set of first principles. At the same time, experimenting always implies at least partial strategic commitment, and commitment implies forgoing options (Gans et al. 2017). Hypothesis testing and experimentation are also the basis of a leading approach in entrepreneurial practice today: the lean startup method (Ries 2011). Moreover, there is growing attention to data-driven management decisions from the evidence-based management literature (Pfeffer and Sutton 2006, Rousseau 2006, Briner et al. 2009) to the more recent work of Brynjolfsson and McElheran (2016). Overall, we follow Zenger (2016), who parallels scientists and entrepreneurs/managers conceiving strategy as a corporate theory to be thoroughly considered, soundly tested through experiments, and eventually validated.

This line of reasoning echoes the application of real option theory to strategy (McGrath 1999, Adner and Levinthal 2004) and complements the discovery-driven approach to strategic planning (McGrath and MacMillan 1995). Running experiments can be thought of as buying (cheap) real options. If well designed and conducted (i.e., according to the scientific method), they provide both useful signals about courses of action (the business hypotheses under test) and helpful information about other courses of action (other hypotheses). Through experiments, entrepreneurs and managers can affect outcomes and variances and avoid the problems due to uncertainty resolution becoming endogenous to their own activity. Designing and conducting rigorous experiments (clear counterfactuals, valid and reliable metrics, evidence-based decisions, etc.) allow entrepreneurs to avoid "option traps" that might hinder dropout and/or generate escalation and overcommitment. In this respect, our approach, like the other approaches in strategy (particularly Adner and Levinthal 2004), marks the difference between real options in strategy vis-à-vis finance. In strategy, the resolution of the uncertainty associated with real options does not just rest on the mere elapse of time: it depends on actions. We then posit that the actions of a scientific approach (definition of problems, formulation of frameworks, and experiments and tests of hypotheses) are one example of the actions that help to exercise real option opportunities.

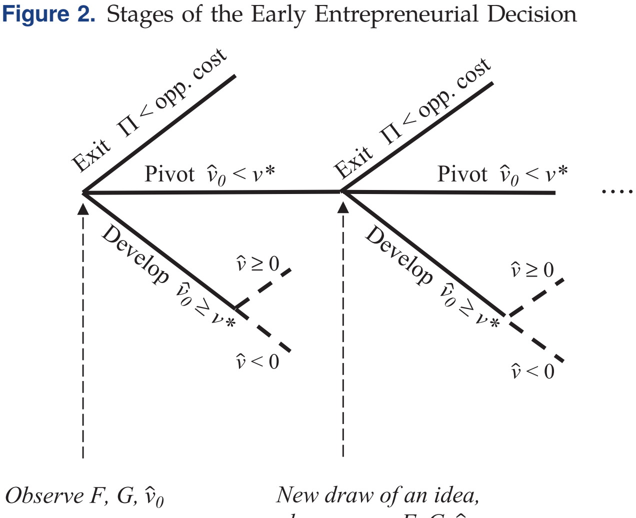
4. Framework

4.1. Setup

We study the early-stage decision of entrepreneurs who consider whether to develop a business idea. The value of the business idea is $v = \bar{v} + \varepsilon$, where $\bar{v}$ is the expected value of the idea and ε is a stochastic term with zero mean, finite variance, and support and distribution that we do not need to define to keep notation at minimum. The term ε captures factors exogenous to the actions and choices of the entrepreneur, such as states of demand, competition, technology, or other environmental conditions.

The decision process of the entrepreneurs unfolds as represented in Figure 2. We focus on the decision at the initial node of Figure 2. At this stage, the entrepreneur has three choices: develop the idea, pivot to a new idea, or exit. If the entrepreneur develops the business idea, she will eventually observe the true value. However, at this initial stage, she predicts a value of the business idea equal to $\hat{v}$, which is different from v. Specifically, $\hat{v} = \hat{v}_0 + \omega$, where $\hat{v}_0$ is a signal observed at the initial stage with support $(-\infty, A]$ and

Figure 2. Stages of the Early Entrepreneurial Decision



cumulative distribution G and ω is an unobserved stochastic factor with zero mean, finite variance, support $(-\infty, a]$, and cumulative distribution F. If the entrepreneur exits, she abandons the firm and earns a positive opportunity cost. If she pivots, she enters a new stage in which she faces the same problem: she draws a new idea, observes a new signal $\hat{v}_0$, and has to decide whether to develop, pivot, or exit. For simplicity, we assume that there are infinite opportunities to pivot. If she develops the idea, she moves to a subsequent stage as shown in Figure 2.

The prediction $\hat{v}$ differs from v for two reasons. First, the expected value of $\hat{v}$ may differ from the expected value of v. Because ε and ω have zero mean, this amounts to saying that the expected value of the signal $\hat{v}_0$ is systematically different from $\bar{v}$. Second, the distributions of ω and ε may differ. These differences depend on the inability to identify exactly the determinants of v or the probability distributions of their realizations, and they affect the distributions G or F. Some factors may be more likely to affect the signal, whereas others may affect mainly the perceived spread of $\hat{v}$. For example, biases such as overconfidence or optimism are likely to be of the first type (e.g., Åstebro 2003, Dushnitsky 2010), whereas biases such as representativeness, ignorance of sample size, or conjunction fallacy are more likely to be of the second type (e.g., Busenitz and Barney 1997). We remain agnostic on which bias affects what, because it is hard to link precisely one type of bias or effect to signal or spread. However, the point that we want to make is that the scientific approach can correct both. A proper formulation of the market problem, a theory of how the hypothesized solution to that problem will generate value, and rigorous tests of such hypotheses provide objective anchors that reduce biases or help us understand the contingencies that affect value. This can make $\hat{v}_0$ closer to $\bar{v}$ or reduce the difference between the distributions of ω and ε.

Because the entrepreneur knows that she will eventually observe the value of the idea, at the initial node in Figure 2, she knows that she will only obtain nonnegative values, because if she observes a negative value, she will stop and earn zero. However, because she predicts that the true value is $\hat{v}$ and not v, the condition is $\hat{v} \geq 0$ or $\omega \geq -\hat{v}_0$. We assume for simplicity that, if she pursues the idea and the idea is ultimately unprofitable, she earns zero and cannot pivot or open a new firm. Therefore, given $\hat{v}_0$, if the entrepreneur develops her idea, she expects to earn

$$\mathrm{E}(\hat{v}, \hat{v} \geq 0, \hat{v}_0) = \int_{-\hat{v}_0}^{a} \hat{v}_0 + \omega dF = \hat{v}_0 + a - \int_{-\hat{v}_0}^{a} F d\omega, \quad (1)$$

where the second equality stems from integration by parts. It is easy to see that a higher signal predicts higher expected returns—that is, (1) increases with $\hat{v}_0$. The entrepreneur compares this return with the returns that she will obtain if she pivots. Clearly, the returns from pivoting depend on future signals on the future ideas that she draws and do not depend on the current signal $\hat{v}_0$ associated with the current idea. As a result, a higher signal $\hat{v}_0$ on the current idea raises (1) relative to the returns from pivoting. This implies that there is a threshold v^* such that the entrepreneur does not pivot if $\hat{v}_0 \geq v^*$. The expected returns from developing the idea before the current signal is observed are then

$$\pi \equiv \int_{v^*}^{A} \hat{v}_0 + a - \int_{-\hat{v}_0}^{a} F d\omega dG, \tag{2}$$

and the overall predicted value of the entrepreneur is

$$\prod = \pi + \delta G^*\pi + (\delta G^*)^2\pi + (\delta G^*)^3\pi + \ldots = \frac{\pi}{1-\delta G^*}, \tag{3}$$

where δ is a discount factor. At each round, the firm draws a new idea with probability G^*, which is the probability that, at the previous round, the firm observed a signal $\hat{v}_0 < v^*$. We assume for simplicity that the expected stream of future π at each round is the same as the current π that appears as the first term of (3). At each stage, the entrepreneur exits if Π is smaller than her positive opportunity cost, and if Π is larger than this opportunity cost, she develops the current idea if $\hat{v}_0 \geq v^*$; otherwise, she pivots. Note that the entrepreneur can exit at any stage, because at each new stage, she has new information about the distributions F and G and develops a new set of expected values that update the stream π and therefore, Π.

The first-order condition (*foc*) of this problem is

$$-(v^* + a) + \int_{-v^*}^{a} F d\omega + \frac{\delta\pi}{1-\delta G^*} = 0. \tag{4}$$

The second-order condition for a maximum is satisfied, because the derivative of the first term with respect to v^* is negative and the derivative of the last term with respect to v^* is equivalent to the first-order condition, which is equal to zero.

4.2. Implications of the Scientific Approach

We represent the mismatch between expected and true value by the notion of first-order stochastic dominance applied to the distribution G and the notion of second-order stochastic dominance (mean-preserving spread) applied to the distribution F, respectively. We begin by comparing the distributions of the true and predicted values of the idea: v and $\hat{v}$. For reasons that we explain below, we focus on the case in which the signal $\hat{v}_0$ is systematically higher than $\bar{v}$ and there is a higher mean-preserving spread of ω with respect to ε. The former implies that the distribution G evaluated at any realization of the signal $\hat{v}_0$ is smaller than the distribution of $\bar{v}$ evaluated at the same realization, which increases the probability that the entrepreneur observes a higher signal. The latter implies that the entrepreneur attributes a higher probability that ω falls in the left or right tail of F. For example, this is a good representation of the fact that the entrepreneurs cannot discern clearly the specific contingencies that affect value. They are "confused" in the sense that they perceive that value is affected by more contingencies than the few key contingencies revealed by a simple and clean theory. As a result, they are likely to perceive ω to be the sum of a larger number of stochastic factors, which produces a distribution with fatter tails.

To streamline our discussion, we begin with the case of a higher systematic signal $\hat{v}_0$, assuming that the error terms ω and ε are the same. We then consider the case in which the expected value of the signal $\hat{v}_0$ is equal to $\bar{v}$ and the error terms ω and ε differ. Also, we simplify the logic of our predictions by assuming that δ is sufficiently small such that, in *foc* (4), the impact of the term $\frac{\delta\pi}{1-\delta G^*}$ on the optimal v^* is negligible. This amounts to assuming that the decision of the entrepreneur depends on the current signal that she observes and her perceived distribution F rather than expectation about future returns from drawing another idea, which is captured by the term $\frac{\delta\pi}{1-\delta G^*}$ in (4).

The scientific approach can reduce the expected value of the signal, possibly until the point where it is equal to $\bar{v}$. The higher G implies that the scientist entrepreneur perceives lower profits Π than the nonscientist entrepreneur.[1] As a result, the scientist entrepreneur is more likely to exit. In addition, in (4), G appears only in the term $\frac{\delta\pi}{1-\delta G^*}$. If this term is negligible, the bias has a negligible effect on the optimal threshold v^*. However, a scientist entrepreneur draws her signals $\hat{v}_0$ from a higher distribution G, which implies that she is more likely to obtain lower signals $\hat{v}_0$, making $\hat{v}_0 \geq v^*$ less likely. The scientist entrepreneur is then more likely to pivot.

Consider now the case of lower mean-preserving spread, which by definition of second-order stochastic dominance, implies higher $\int_{-\hat{v}_0}^{a} F d\omega$ or $\int_{-v^*}^{a} F d\omega$ in (2) and (4). In (2), this reduces the perceived expected return π, which reduces Π in (3). As a result, the scientist entrepreneur is more likely to exit. In (4), the higher $\int_{-v^*}^{a} F d\omega$ raises the marginal value of v^*, suggesting that the entrepreneurs who adopt a scientific approach are more likely to pivot. The higher π raises $\frac{\delta\pi}{1-\delta G^*}$, which also appears in (4), countering in part this effect; however, as noted earlier, we assume that the discount factor δ is sufficiently small such that the contemporaneous effect dominates.

As a result of this discussion, regardless of whether a scientific approach reduces the expected value of the signal or lowers the mean-preserving spread, it makes exit or pivot more likely. We are agnostic on whether the effect is predominately one or the other. Because both are plausible and yield the same result, we do not commit to one of these two explanations. Interestingly, reduction of the expected value of the signal yields exit and pivot, because the scientist entrepreneur obtains a lower signal $\hat{v}_0$, whereas the optimal threshold is unaffected; with lower mean-preserving spread, the signal is the same, but the optimal threshold is higher. We summarize this discussion in Proposition 1 below, which is the main proposition that we test in our RCT.

Proposition 1. *An entrepreneur who adopts a scientific approach is more likely to exit and pivot.*

4.3. Additional Remarks and Implications for Performance

First, we illustrate the mechanism of the framework. The effect of a higher signal is straightforward. The nonscientist entrepreneur perceives a higher profit or receives a higher signal and therefore, is less likely to exit or pivot. As far as the effect of the spread is concerned, second-order stochastic dominance implies that the expected value of ω does not change. However, the entrepreneur's representation of reality is such that she perceives that she will never earn $\hat{v} < 0$ or (which is the same) that she will earn returns from her business idea only if $\omega \geq -\hat{v}_0$; otherwise, she earns zero. Therefore, a lower mean-preserving spread shrinks the right tail of the distribution. It also shrinks the left tail, but this is less relevant for the decision, because the entrepreneur perceives that she will never incur values of ω smaller than $-\hat{v}_0$. In other words, the scientist entrepreneur perceives that she is less likely to enjoy the high returns in the right tail of the distribution of ω. Conversely, the nonscientist considers them more likely. This implies that, although the expected value of ω does not change, the expected value of ω of the scientist entrepreneur, conditional on $\omega \geq -\hat{v}_0$, decreases.

Second, we clarify the relationship between the scientific approach and performance. To streamline this discussion, consider the extreme case in which, in assessing profits (2) and in *foc* (4), the scientist entrepreneur uses the exact distributions G and F used by an entrepreneur who observes v. If so, the scientist entrepreneur makes the optimal decision about v^* and observes the right signal such that her decisions about exit or pivot are the same as the optimal decisions of an entrepreneur who observes the exact distribution of v. In contrast, the decisions of the nonscientist entrepreneurs are clearly suboptimal, because they base their optimizations on the wrong G and F. In reality, the scientist entrepreneur may not make predictions that overlap exactly with the true value v. However, we assume that, by getting closer to it, she will enjoy higher performance than a nonscientist entrepreneur. Specifically, the scientist entrepreneurs exhibit higher performance than nonscientist entrepreneurs, because they exit some unprofitable ventures that the nonscientist entrepreneurs do not abandon; in addition, for some business ideas the scientist entrepreneurs pivot correctly to other ideas with higher real expected value, whereas the counterfactual nonscientist entrepreneurs do not.

Third, although we focus on the case of overestimation of the value of the ideas or a narrower distribution of the phenomenon compared with the perception of the entrepreneur, the opposite case is also possible. The entrepreneur may be a "contrarian" and perceive systematically lower signals $\hat{v}_0$, or the phenomenon may exhibit fatter tails than perceived by the entrepreneurs. The scientific approach could also help in this case. It can restore the more positive perspective of the value of the idea, or it can show that the phenomenon depends on more contingencies than the few perceived by a naïve entrepreneur. For example, as noted by Fleming (2001) and Murray and Tripsas (2004), scientific capabilities help one to see a larger number of modules of knowledge, and they help to connect these modules. This could explain, for instance, the case of very successful entrepreneurial ventures that capture opportunities at the far-right tail of the distribution.

In this case, we have opposite predictions on exits and pivots. If a scientist entrepreneur corrects underestimation upward, she will perceive higher profits Π, and thus, she will be less likely to exit; she will also receive higher signals $\hat{v}_0$, and with roughly the same optimal v^*, she will be less likely to pivot. The terms $\int_{-\bar{v}}^{a} F d\omega$ or $\int_{-v^*}^{a} F d\omega$ in (2) and (4) will be lower than for the nonscientist entrepreneurs. Again, this implies higher Π, making the firm less likely to exit, and lower v^*, making pivot (for given $\hat{v}_0$) less likely. The scientist entrepreneur enjoys higher performance. However, it now stems from the fact that the nonscientist exits or pivots when it is not profitable to do so. Simply put, in the previous case, the scientist entrepreneurs do not incur false positives, whereas in this case, they pursue false negatives.

We do not attempt to predict whether a scientific approach solves the problem of false positives or negatives, which would require introducing additional assumptions. We leave this to future investigation and use a practical rationale to focus our study on the effect discussed in the previous section. It is well documented that the main problems faced by many entrepreneurs are that they perceive higher returns than they eventually obtain (e.g., Åstebro 2003, Galasso and Simcoe 2011), and many startups fail, suggesting that entrepreneurs overestimate opportunities. For example, Fairlie and Miranda (2017) show that 84.4% of U.S. startups fail within seven years (see also table 1A in Fairlie and Miranda 2016). This suggests that a scientific approach, enabling better predictions, will mostly help to uncover false positives by reducing overestimation or optimism or by making the perceived distribution of the phenomenon more precise. It will then encourage entrepreneurs to exit or pivot, because it reveals that the current idea that they are pursuing is probably less valuable than they think.[2]

Finally, the scientific approach could produce an alternative mechanism, learning, which can be represented by a first-order stochastic dominant effect on the distribution of G. However, learning implies that the higher expected value of the signal $\hat{v}_0$ mirrors an increase in $\bar{v}$—that is, that the entrepreneurs who adopt the scientific approach draw their ideas from a better distribution—than a mere increase in the expected value of the signal. Because a decrease in G raises Π, learning increases the last term in *foc* (4) and therefore, the optimal v^*. Thus, learning increases pivot, like our hypothesized effect, and the mechanism is that it makes future draws more attractive. However, a higher Π also reduces exit. Thus, if we observe that the scientific approach does not reduce exit, we cannot rule out our main hypothesized effect that the scientific approach reduces a positive bias on the signal or an unjustified amplification of the spread.

5. Research Design, Data, and Method

5.1. RCT Design

We partnered with two institutions that train startups and that have pioneered the use of approaches close to the scientific approach that we discuss in this paper: the Doers and the Lean Startup Machine. The Doers have developed a long-term module for startups to learn the method of validated learning and provided in-class lectures to our startups. We built on their teaching material to tailor it to our experiment: for the treatment group, we further stressed the scientific component of the validated learning process that the Doers normally teaches, but we kept the program as it was for the control group. In this way, we offered a meaningful learning experience to startups in both groups while ensuring that the only difference was related to the scientific method. The Lean Startup Machine operates worldwide, offering two-day workshops that teach entrepreneurs the process for validating business ideas. They provided us with a network of mentors who provided coaching sessions to ensure that the startups in our training followed what the Doers taught in class.

We promoted our own training program to nascent startups. We focused on these firms, because they are neither established startups with past experience that could affect the experiment nor people who are only remotely evaluating the possibility of becoming entrepreneurs and therefore, more likely to drop out for lack of commitment. We did not restrict to particular industries. We advertised the course through digital channels as a general course covering the important aspects of new venture creation—market sizing, business model creation and analysis, how to create a landing page, relevant startup data analytics and accounting, and so forth. This helped us attract many startups and avoid self-selection by those only interested in some aspects of the training. To encourage the participation of qualified and motivated startups, we advertised that the training would end with a private event where participant startups could meet with investors. The course was free to ensure participation of firms with limited financial resources. The call was launched on November 2015 and remained open until mid-January 2016. We received 202 applications.

Before beginning the training, we asked the startups to sign a document, approved by the Ethical Committee of Bocconi University, stating that Bocconi University was investigating the determinants of the success of startups and that we were providing management advice and training to firms and collecting performance data. In other words, they knew that they were participating in an activity in which we were offering a free service in exchange for monitoring their actions for educational and research purposes. We also told them that there were two groups of startups and that there were some differences in the content of the training program. However, they did not know whether they were part of the treatment group or the control group.

Startups received 10 sessions of training at Bocconi University in Milan, Italy. Five sessions were frontal lectures lasting four hours, and five were one-hour sessions per startup with mentors for both treated and control firms. As discussed in Section 2, the duration and content of the intervention were the same for both groups. However, treated startups were taught in each of the four steps of the process to frame, identify, and validate the problem; formulate falsifiable hypotheses; and test them in a rigorous fashion (using data and experiments), including defining valid and reliable metrics and establishing clear thresholds for concluding whether a hypothesis is corroborated or not. "Scientific" problem framing and identification, hypothesis formulation, and rigorous testing were integrated into both the content of the frontal lectures and the feedback that mentors provided to the treated firms during the one-to-one meetings—for example, mentors encouraged startups to think about the broader framework of their idea and the customers' problem that they were trying to solve, formulate falsifiable hypotheses, and test them rigorously. This encouragement was not offered to the control group, where startups received, during both the lectures and the one-to-one meetings, general instructions about the importance of keeping their business models or products flexible, seeking and eliciting customer feedback, and using this information to experiment with different solutions before choosing a final business model or product. This approach encouraged them to conduct these activities based on their own intuitions, heuristics, and approaches.

We offered the same number of hours of training to both groups to ensure that there was no other effect in the treatment than a scientific approach to entrepreneurial decision making. To maintain the same hours of training for treated and control startups despite the difference in content, we implemented the following design choices. The treatment group was taught market validation for 70% of the time and the scientific approach for 30% of the time during each lesson. The control group instead was taught market validation 100% of the time during each lesson. We told our mentors and instructors to stick to these percentages in both the frontal classes and the one-to-one interactions. We did not provide additional content to the control group; therefore, whereas the treatment group was taught about the scientific method, the control group reviewed the class material about market validation and had the opportunity to ask questions of the instructors. Despite this difference, we did not provide a significantly better learning opportunity to the control group. The most natural reason is that there are diminishing returns to learning. Training entrepreneurs an additional 30% of time on market validation is likely to produce a small amount of learning on top of the bulk of the class (70%) on the same topic. Conversely, the 30% on the scientific approach has a notable effect, because it is the only 30% taught on the topic. Moreover, the conceptual tools, the course materials, the exercises and simulations, and the homework as well as the instructors were the same (except for the differences described above) and might be considered "state-of-the art" in entrepreneurship education in the sense that they include a set of conceptual tools (such as business model canvas, customer interviews, and minimum viable products) widely used worldwide to teach entrepreneurs how to develop an idea. This experimental design aims at a good balance between the need to build an effective and sufficiently strong treatment and the need to create a good counterfactual. Indeed, getting the marginal 30% of hours of training deepening the common content for control startups did not represent a significant addition to their learning. At the same time, dedicating 30% of the training program to the scientific approach represented a significant treatment for the experimental group.

The training program consisted of 10 sessions: 5 lectures lasting 3.5 hours each for both treatment and control groups and 5 individual coaching sessions lasting 1 hour each in which mentors provided advice and coaching to each startup. The program was offered on Saturdays, alternating the five frontal lectures with the individual coaching sessions every other Saturday. The same instructor taught the five frontal lectures. Each startup was randomly assigned to a mentor who provided advice and mentorship during the five one-hour individual coaching sessions. Overall, 21 mentors were involved. Each mentor supported three startups from the treatment group and three from the control group. Both the instructor (frontal lectures) and the mentors had significant mentorship experience. The authors designed and conducted "training the trainers" activities about the scientific approach and standardized the teaching materials within and across the experimental groups and the coaching process across mentors.

Our research team coordinated the activities and ensured that the learning modules and mentoring activities conducted by the instructor and mentors were balanced between treated and control startups. To avoid contamination between the two groups, the research team ensured that the 10 sessions were held at different times of the same day (morning and afternoon) and kept all communication to the two groups of startups distinct and segregated. This required creating two separate groups on Facebook publicized to no one but the teams in the relevant group. We systematically monitored startups' learning and performance by collecting data via telephone interviews from March to November. We conducted telephone interviews, because we could assess the actual use of a scientific approach only by knowing the activities in which the startups were engaged when they were in their locations, away from the training sessions. We provide additional details about data gathering in Section 6.

5.2. Sample and Randomization

Before beginning the training program, we asked all applicant startups to send us a pitch for their business idea and the vitae of their founders. Using this information, we categorized them across development stages, industries, and regions of origin. We defined their stage of development as "idea" when the startups only had a business project in mind, "development" when they had begun to work on their product/service, "prerevenue" when the product/service was out in the market but the firm had yet to earn revenue, and "startup" when it had earned revenue. As mentioned, we focused on early ventures—that is, initiatives at the idea and development stages—because a scientific approach to entrepreneurial decision making is more difficult and costlier to adopt when firms have incurred sunk costs. Also, startups at more advanced stages are more likely to be self-selected, because they have survived the earlier phases. Of the 202 applicants for the program, 164 startups were in the idea (105) and development (59) groups, and 38 were in the prerevenue (16) and startup (22) phases. Given our resource constraints (instructors, mentors, research team, and funds), we capped enrollment in the training program at 116 startups randomly selected from the 164 startups in the first stages. To classify firms across industries, we used the classification suggested by CB Insights, a startup-dedicated database that reports European and American angel and venture capital investments in startups.[3] From the vitae of each startup team, we inferred its region/location.

We opted for pure randomization with balance tests, because it is, in our case, a better strategy than stratified randomization. Several relevant variables could be used as strata, such as whether startups offer products and/or services that are business to consumer rather than business to business or whether they joined the training after beginning work on their project or with just an idea in mind. Choosing the appropriate strata among these variables to implement stratified randomization and allocate the 116 selected startups to the treatment and control groups was not obvious from a theoretical standpoint, and it was practically unmanageable.

To check the soundness of our sampling and randomization choices, we proceeded as follows. First, to ensure that the 116 selected startups did not differ significantly on any meaningful attribute from those not included in the training program, we followed Gelber et al. (2016) and ran reduced form ordinary least squares (OLS) regressions of startup characteristics before entering the program on a dummy for selection into the training.[4] Second, we ran similar OLS regressions of startup characteristics on a dummy for the allocation to the treatment or control group. We define all of the variables used in the balance tests in Appendix C.

Most firms in our final sample of 116 are internet-based companies (55) followed by furniture (29) and retail (10). The others are spread across diverse sectors, such as leisure, food, healthcare, and machinery. This is a fair representation of the distribution of Italian startups, because it reflects a mix of internet-based origins and Italian industries. Most of our firms come from Lombardy, the region of Milan (61); the others come largely from the Italian north (34), and the rest come from the center and the south. Although Lombardy is overrepresented, largely because of geographic proximity to where the experiment was conducted, the distribution between north and south mirrors the distribution of industrial activities in Italy. Moreover, this breakdown by industry and region mimics the breakdown in the original 164 firms as well as in the original 202 applicants.

Table 1 reports some randomization checks. We show the average effects of available variables for the 164 firms with respect to selection into the training program. We checked for idea stage versus development, the three main sectors of our sample of firms (internet, furniture, and retail), the main region of origin (Lombardy), and the size of the founding team. Consistent with the validity of the randomization, none of these variables are significantly related to selection into the program. The 116 startups selected were then randomly assigned to the treatment ($n = 59$) and control ($n = 57$) groups. We conducted balance tests using as dependent variables the same covariates from the previous check and as the independent variable the dummy for selection into the treatment group (1 = treatment group, 0 = control group). Once again, estimated p-values show no statistically significant difference between the groups. For the 116 selected firms, we gathered additional information on experience, education, and work. As shown by the last column of Table 1, none of these variables are significantly associated with selection into the treatment group, increasing our confidence in the robustness of the RCT design.

To summarize, the startups selected into the training program are mostly digital, early-stage companies with two or three team members. They have on average 2.5 years of experience in the industry in which they launched their startup, slightly less managerial experience, and much less experience working with and inside startups (on average less than 1 year). On average, their team members have completed college education, and more than one-half were employed at the beginning of the program. Overall, the sample is composed of teams with low levels of industry, managerial, and entrepreneurial experience. From our conversations with the mentors and other practitioners, it seems that the sample characteristics well represent the broader Italian entrepreneurial community.

6. Data

We collected the data during the training program, which lasted from March to June, and after it ended, from June to April. The program entailed in-class lectures on Saturday followed by mentoring sessions the next Saturday. The data sources are phone interviews conducted by five purposely trained research assistants. Overall, we collected 16 observations

Table 1. Randomization Checks

Variables	Applicant startups' characteristics with respect to selection in training program	Selected startups' characteristics with respect to assignment to control or treatment group
Idea stage	0.021 (0.795)	−0.220 (0.807)
Internet sector	−0.064 (0.460)	−0.068 (0.467)
Furniture sector	0.091 (0.206)	0.009 (0.920)
Retail sector	0.003 (0.980)	0.031 (0.549)
Lombardy	−0.064 (0.460)	−0.081 (0.366)
Team size	0.193 (0.470)	0.128 (0.606)
Number of observations	164	116
Industry experience		−0.010 (0.991)
Management experience		0.810 (0.190)
Experience working with startups		−0.001 (0.980)
Experience working in startups		0.590 (0.110)
Currently employed		−0.043 (0.570)
Currently studying		−0.085 (0.249)
Level of education		0.216 (0.190)
Number of observations		116

Notes. OLS regressions use variables as the dependent variable and dummies for selected/nonselected or treatment/control as regressors; coefficients are differences between means (*p*-values are in parentheses).

per firm over time for the firms that never dropped out and for the other firms, up to the period in which they dropped out. During the four-month training period, we collected data biweekly after each mentoring session (phone interviews took place within three days). After the training period, we collected data monthly, but the last observation (16th data point) was collected two months after the 15th observation. The different frequencies are not an issue in our panel and cross-section regressions, because we use time dummies. Moreover, the coarser frequencies after the training enabled us to collect information over a longer period without bothering the firms with too many data requests. In our survival regressions, we set time in a chronological fashion and counted each period as the cumulation of biweekly periods. Thus, we counted time as 1–8 for the first eight fortnights, then 9–21 for the next seven periods (where 9–21 are the midpoints between the monthly collections), and 24 for the last period, which is the midpoint for the two months between fortnights 22 (next to the last data collection) and 26 (last data collection). We used midpoints as our best guess of the exact time of activation, acquisition, or revenue that was recorded at one data collection point but not in the previous one. However, our results are robust to different timing of the event between the two collections.

Research assistants attended the entire training program themselves and underwent specific training on the research protocol, how to conduct phone interviews to get the required data and when necessary, how to code interview content using thematic analysis. Through the phone interviews, we gathered a variety of data from startup performance data to specific actions and behaviors during the observation period to evaluate the extent to which the teams adopted a scientific approach to decision making. Each research assistant interviewed the same set of startups over time to ensure that she became acquainted with their business model and could spot significant events in each startup's life. Periodically, the research assistants and in some cases, the mentors and authors independently conducted thematic analysis of a small subset of the same phone interviews, coded them, and checked the extent to which coding was aligned. This allowed us to build and maintain over time high levels of interrater reliability. Phone interviews lasted about 45 minutes and were open-ended conversations with the entrepreneurs. As part of the phone interview protocol, we asked entrepreneurs to report what they had done for the past two weeks. These narratives gave us grounds for evaluating the level of adoption of a scientific approach to decision making, because research assistants used as a coding scheme the themes described in the theory and Inkdome case study sections. These themes are reported and summarized in Appendix B. Because the startups did not know that they were being scored, scoring reflected the interviewer's evaluation of the firm's practices rather than the entrepreneur's perceptions or the interviewer's impressions (Bloom and Van Reenen 2007). In the second part of the phone interviews, we asked startups to report their performance, particularly their revenue.

We run both cross-sectional regressions with our 116 observations and panel regressions using 116 × 16 = 1,856 observations, except for the exit regression, where we only use observations up to the date on which the firm exits. In this case, we have 1,606 observations. We use observations for firms even after they exit to ensure that our results are not altered by systematic differences of firms that did not exit.

However, the results are robust to the case in which we exclude these firms from our panel regressions after they exit.

During our timeframe, 44 firms exit (24 in the treatment group and 20 in the control group), and 30 firms pivot at least once to a main new idea (19 in the treatment group and 11 in the control group); we have 41 pivots in total. Interestingly, seven firms pivot more than once in the treatment group (five pivot twice, one pivots three times, and one pivots four times), whereas only one firm pivots twice in the control group. Also, 17 firms earn positive revenue (9 in the treatment group and 8 in the control group), and overall, 75 firms in our sample take one or more of these three actions (exit, pivot, or revenue). If we include firms that received at least one email from potential customers interested in the firm's product (acquisition) or asking to try the product (activation), 98 of our 116 firms took one of these five actions. This is in line with expectations and suggests that the startups in our sample were not just formed and left inactive. As noted, most firms in our sample were formed just before March 2016, when we began the training program. Because our last data collection was in April 2017, we are not surprised to see the rate of activities just described over a period slightly longer than one year.

The analysis uses the following variables. Table 2 reports their descriptive statistics, and Table 3 shows their correlations.

6.1. Exit

In the panel regressions, this is a binary variable that takes the value zero until the firm exits (abandons the program and ceases the startup), one in the time period in which the firm drops out, and a missing value thereafter. In the cross-section regressions, this is a dummy equal to one if the firm exits at any point and zero otherwise. To avoid attrition biases, we checked that the entrepreneurs who informed us of their decision to discontinue their initiative truly abandoned their activity. All firms that exit from our sample had not yet made heavy investments in their company. In terms of our framework in Section 3, they are firms that abandon their idea before they commit to developing it and discover later on that it does not work. In other words, they are genuine exits in the terminology of our framework and not failures.

6.2. Pivot

In the panel regressions, this is a binary variable that takes the value one in the exact time period in which the firm made a major change to its business model and zero whenever the firm does not pivot. Firms that made more than one such major change will take the value one at any time in which a major change occurs. In the cross-section, this variable is equal to the total number of major changes of the firms during our timeframe. We defined a change as major by analyzing whether the entrepreneur moved from the original idea to another idea that changed the core value proposition of the product or service offered or its target customers.[5] For example, a major change was Inkdome's decision to pivot from a search engine platform to one where users contact tattoo experts.

6.3. Revenue

In the panel regressions, this is the firm's revenue flow in euros between any 2 of our 16 time periods. The 17 firms with positive revenue in our sample correspond to 107 of our 1,865 observations: 85 observations of the 9 firms in the treatment group versus 22 observations of the 8 firms in the control group. In the cross-section, we use the cumulative revenue over the 16 time periods of our sample. To control for outliers, along with absolute value we show results using revenue winsorized at 99% and the revenue's hyperbolic sine transformation. The results using $\log(1 + Revenue)$ are practically identical to those obtained using the hyperbolic sine transformation. Finally, the average and median revenue for the 85 nonzero observations in the treatment group are about 7,800 and 1,300 euros, respectively; for the 22 nonzero observations in the control group, they are about 900 and 500 euros, respectively.

6.4. *Time_to_Exit*, *Time_to_Acquisition*, *Time_to_Activation*, and *Time_to_Revenue*.

In the survival regressions, these variables measure the weeks in which the startups exit, acquire, activate their first client, or begin earning revenue. To account for censored observations, the survival regressions also rely on the failure dummies equal to one for firms that experience exit, acquisition, activation, or revenue and zero for the firms that do not experience the event. As noted earlier, we use time chronologically, and thus, the weeks correspond to the actual time in which the event has occurred. Acquisition and activation are standard performance measures that account for early signals of performance (Blank and Dorf 2012, Jackson et al. 2015, Ripsas et al. 2015).

6.5. Intervention

In the panel regressions, this is a binary variable equal to one for all of the observations of the treated firms and zero for all of the other observations. In the cross-section, it is equal to one for the treated firms. In robustness checks, we use *Cumulative_intervention* and *Postintervention*. Bloom et al. (2013) also use these variables as alternative specifications to the

Table 2. Descriptive Statistics

Variables	*N*	Mean	SD	Min	Max
Panel					
Exit[a]	1,606	0.027	0.163	0	1
Pivot[b]	1,856	0.022	0.147	0	1
Revenue (euros)[c]	1,856	344.8	4,503.5	0	103,474.5
Revenue (winsorized 99%)[c]	1,856	98.51	603.6	0	5,000
Revenue (hyperbolic sine transformation)[c]	1,856	0.388	1.737	0	12.24
Intervention	1,856	0.509	0.500	0	1
Cumulative_intervention	1,856	3.179	3.558	0	8
Postintervention	1,856	0.254	0.436	0	1
Cross-section					
Exit	116	0.379	0.487	0	1
Pivot[b]	116	0.353	0.701	0	4
Revenue (euros)[c]	116	5,517.3	41,812.1	0	437,474.5
Revenue (winsorized 99%)[c]	116	1,530.4	7,349.1	0	44,200
Revenue (hyperbolic sine transformation)[c]	116	1.284	3.217	0	13.68
Intervention	116	0.509	0.502	0	1
Time_to_exit[d]	44	13.02	6.947	3	24
Time_to_acquisition[d]	67	7.254	6.029	1	24
Time_to_activation[d]	53	8.660	6.699	1	24
Time_to_revenue[d]	17	13.71	9.265	1	24
Cross-section (intervention = 1)					
Exit	59	0.407	0.495	0	1
Pivot[b]	59	0.492	0.858	0	4
Revenue (euros)[c]	59	10,630.2	58,410.7	0	437,474.5
Revenue (winsorized 99%)[c]	59	2,791.7	10,167.0	0	44,200
Revenue (hyperbolic sine transformation)[c]	59	1.467	3.624	0	13.68
Time_to_exit[d]	24	12.58	6.700	3	24
Time_to_acquisition[d]	33	6.667	6.158	1	24
Time_to_activation[d]	26	7.500	7.106	1	24
Time_to_revenue[d]	9	8.222	9.176	1	24
Cross-section (intervention = 0)					
Exit	57	0.351	0.481	0	1
Pivot[b]	57	0.211	0.453	0	2
Revenue (euros)[c]	57	224.9	665.4	0	3,000
Revenue (winsorized 99%)[c]	57	224.9	665.4	0	3,000
Revenue (hyperbolic sine transformation)[c]	57	1.094	2.752	0	8.700
Time_to_exit[d]	20	13.55	7.480	5	24
Time_to_acquisition[d]	34	7.824	5.94	1	21
Time_to_activation[d]	27	9.778	6.21	2	24
Time_to_revenue[d]	8	19.88	4.224	13	24

Note. SD, standard deviation.
[a]Missing after exit.
[b]In the panel, pivot is a binary variable = 1 only at the time of pivot; in the cross-section, it is the number of pivots.
[c]In the panel, revenue is flow between periods; in the cross-section, it is cumulated revenue in the last period.
[d]Only firms that experience exit, acquisition, activation, or revenue; time is 1–8, 9–21 (only odd numbers), or 24 (see text for explanation).

intervention dummy. *Cumulative_intervention* takes values from one to eight for the treated startups during the first eight periods in which the firms underwent their training. It is equal to eight for these firms for the other eight observations after the training, and it is equal to zero for the control startups during the entire period. *Postintervention* is a binary variable equal to one for the treated startups after the training (i.e., periods 9–16) and zero for all of the observations of the control group. *Cumulative_intervention* takes into account the capabilities accrued during the training period (cumulative learning), and we use it together with *Postintervention* to account for different effects in the posttraining periods.

Finally, we carefully selected entrepreneurs with limited entrepreneurial experience and an initial business idea that had not begun any activity. We can fairly say that all of these firms were at a baseline level. Therefore, our analysis is de facto a difference in difference in which any variable regarding these firms before the intervention is at a baseline level of zero, making the difference across firms before the intervention

Table 3. Correlations

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
(1) *Exit*[a]	1									
(2) *Pivot*	–0.003	1								
(3) *Revenue*	–0.014	–0.008	1							
(4) *Revenue (winsorized 99%)*	–0.028	0.001	0.578***	1						
(5) *Revenue (hyperbolic sine transformation)*	–0.027	0.015	0.472***	0.842***	1					
(6) *Intervention*	0.016	0.060*	0.072**	0.138***	0.160***	1				
(7) *Cumulative_intervention*	0.043	0.036	0.092***	0.149***	0.172***	0.878***	1			
(8) *Postintervention*	0.049*	–0.029	0.105***	0.117***	0.117***	0.574***	0.791***	1		
(1) *Exit*	1									
(2) *Pivot*	–0.014	1								
(3) *Revenue*	–0.102	–0.064	1							
(4) *Revenue (winsorized 99%)*	–0.154	–0.086	0.734***	1						
(5) *Revenue (hyperbolic sine transformation)*	–0.222*	0.047	0.484***	0.687***	1					
(6) *Intervention*	0.058	0.201*	0.125	0.175	0.058	1				
(7) *Time_to_exit*[b]	–0.783***	0.076	0.080	0.120	0.173	–0.072	1			
(8) *Time_to_acquisition*[b]	0.195*	–0.185*	–0.155	–0.261**	–0.413***	–0.002	–0.277***	1		
(9) *Time_to_activation*[b]	0.197*	–0.150	–0.202*	–0.327***	–0.489***	–0.030	–0.237*	0.809***	1	
(10) *Time_to_revenue*[b]	0.090	–0.020	–0.495***	–0.760***	–0.819***	–0.182	–0.072	0.366***	0.447***	1

Note. All panel and cross-section variables are defined as in Table 2.
[a]Correlations with exit based on 1,606 observations.
[b]Set to 24 for firms that do not experience the event.
$^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

equal to zero. In addition, in the panel regressions, *Cumulative_intervention* and *Postintervention* enable us to use firm fixed effects, and thus, we can check the robustness of our results using a within-firm estimator.

7. Empirical Results

The empirical results of our analysis are in line with the predictions of our framework. We find that the intervention does not reduce the probability that startups exit; if anything, it increases it. This is consistent with our prediction that, because a priori, most entrepreneurial ideas are unprofitable, a scientific approach makes entrepreneurs more cautious about the profitability of their ideas, which raises the odds that a startup exits. We find clear evidence that the intervention increases pivots. Our framework suggests that the scientific approach reduces the expected value of the signal $\hat{v}_0$ or increases the threshold v^*, which raises the odds that a startup pivots to a new business idea.

The intervention increases revenue. Our framework predicts that, if a scientist entrepreneur develops an idea, it has a higher expected value because of the more stringent conditions for developing an idea (higher threshold or the perception of a lower expected value). We also take into account that it takes time before startups attain monetary performance. In fact, only 17 startups in our sample earn positive revenues in our one-year timeframe. This suggests that the effect of the intervention on revenue is produced by the firms that discovered fairly early a value higher than the threshold. Conversely, the startups that discover that they have an idea with value that is below the threshold and particularly, the startups in the treated group pivot. This means that many firms that pivot and then discover in the future that they are working on a valuable idea do not earn revenue in the timeframe of our sample. Because the acquisition or activation of customers represents early signals of future monetary returns, the firms that during our timeframe acquire or activate customers without earning revenue are likely to be those that obtain an initial value v smaller than the threshold v^* and pivot successfully to a new idea. Although we only have 17 firms earning revenue, there are 67 and 53 startups that acquire or activate customers, respectively.

We find that, after pivoting, the treated firms in our sample are more likely to acquire or activate customers. This is in line with our framework. Using the notation in Section 4, G^* is the probability of pivoting, and therefore, $(G^*)^n(1 - G^*)$ is the probability of finding a successful new idea after n pivots. This expression increases with G^* for reasonably low levels of G^*. Our framework predicts that scientist entrepreneurs exhibit a higher G^*, because they are more conservative about the signal and because they set a higher v^*. Thus, in line with our empirical results, a treated entrepreneur is more likely to find a valuable idea after she pivots. We also find that treated startups are more likely to earn revenue earlier, and in our empirical results, this is unrelated to pivoting. This finding is again consistent with our framework. As noted, many firms that pivot and that earn revenue not now but in the future exhibit the same value of the dependent variable as firms that do not pivot and will unsuccessfully stay at the lowest level of the dependent variable in the future. In other words, pivot does not capture, in this case, the effect on performance.

Table 4. Exit Regressions, Dependent Variable = *Exit*

Variables	Linear probability	Survival (Cox)	Panel
Intervention	0.035**	0.101	0.003
	(0.045)	(0.567)	(0.716)
Constant	0.316		−0.008
	(0.533)		(0.585)
Observations	116	116	1,606
R^2	0.183		
Dummies for mentors	Yes	Yes	Yes
Time fixed effects	—	—	Yes
Clustered errors	Intervention	Intervention	Firms
Number of identification			116

Note. Robust p-values are in parentheses.
$^{**}p < 0.05$.

All of our regressions include dummies for the mentor who supervised the firm in the interactive sessions. This ensures that our results are not affected by the vagaries associated with individual instructors. For simplicity, we do not show the mentor dummies in our regressions. However, it is interesting that the mentor dummies are largely insignificant, suggesting that we promoted a standardized approach to the training program. We cluster errors by *Intervention*. This controls for the important concern that the startups that followed the same model of training might exhibit correlated errors. In the panel regressions, we also cluster errors by firms.

Table 4 reports the results of our *Exit* regressions. The cross-section regression shows a positive and significant effect of *Intervention* with respect to *Exit*. In all of the other cases (survival and panel), the effect is positive but statistically insignificant. The scientific approach affects exit but does not affect the time to exit. Table D.1 in Appendix D shows that using *Cumulative intervention*, *Postintervention*, and firm fixed effects instead of *Intervention* do not change the results of the panel regression in Table 4. As noted in our framework section, *Exit* plays an important role in our story, because a natural alternative consequence of the scientific approach is that it provides learning opportunities. However, unlike our precision mechanism, learning implies that scientist entrepreneurs are less likely to exit. The results in Table 4 do not rule out that the scientific approach could provide learning. However, they provide sufficient evidence that the scientific approach does not provide learning *only*. To reinforce our confidence in the mechanism of the scientific approach, we have sought additional evidence of its effect on exit. In April 2017, when we collected our last set of results, we then asked all of the firms that survived or that had just exited in that period (81 firms) the following question: "Given what you learnt in the course, if you had to launch a second startup, how confident would you be in making drastic decisions such as abandoning your startup?" Respondents answered on a one to seven Likert scale, where one equals not at all and seven equals very confident. The average score of treated firms was 4.4, and for the control group, it was 3.2 ($p < 0.01$).

Table 5 shows that treated firms are more likely to pivot. This is robust across our estimations. We find a positive and significant effect of *Intervention* with respect to *Pivot* in both our cross-section regressions

Table 5. Pivot Regressions, Dependent Variable = *Pivot*

Variables	Linear regression	Negative binomial	Panel
Intervention	0.261**	0.803***	0.016**
	(0.027)	(0.000)	(0.033)
Constant	0.536	−0.944	0.011
	(0.529)	(0.244)	(0.737)
Observations	116	116	1,856
R^2	0.105	—	—
Dummies for mentors	Yes	Yes	Yes
Time fixed effects	—	—	Yes
Clustered errors	Intervention	Intervention	Firms
Number of identification			116

Notes. Robust p-values are in parentheses. In the linear regression and the negative binomial regression, *Pivot* is equal to the total number of major changes of the firms during our timeframe. In the panel regressions, *Pivot* is a binary variable that takes the value one in the exact time period in which the firm made a major change to its business model and zero whenever the firm does not pivot.
$^{**}p < 0.05$; $^{***}p < 0.01$.

Table 6. Revenue Regressions (Cross-section), Dependent Variable = *Revenue*

Variables	Linear regression	Winsorized 99%	Hyperbolic sine transformation
Intervention	10,799.5**	2,666.8***	0.396*
	(0.010)	(0.008)	(0.065)
Constant	−4,899.7	−833.4	1.252
	(0.594)	(0.754)	(0.613)
Observations	116	116	116
R^2	0.220	0.276	0.178
Dummies for mentors	Yes	Yes	Yes
Clustered errors	Intervention	Intervention	Intervention

Note. Robust *p*-values are in parentheses.
$^*p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

(linear and negative binomial), where our dependent variable is equal to the total number of major changes of the firms during our timeframe and in our panel regression where we operationalize *Pivot* as a binary variable. Table D.2 in Appendix D shows that this result is robust to the use of *Cumulative intervention*, *Postintervention*, and firm fixed effects instead of *Intervention*.

Table 6 and 7 report our results for revenue. Both the cross-section and the panel show that the *Intervention* has a positive impact with respect to *Revenue*. The results are robust to winsorizing *Revenue* or a hyperbolic sine transformation. They are also robust to the use of *Cumulative intervention*, *Postintervention*, and firm fixed effects as shown in Table D.3 in Appendix D.

Table 8 shows Cox regressions for *Time_to_acquisition*, *Time_to activation*, and *Time_to_revenue*. In all of these regressions, *Pivot* is always the number of pivots before the time of the focal event. The effect of the interaction between *Intervention* and *Pivot* with respect to *Time_to_acquisition* and *Time_to activation* suggests that pivoting is more effective when associated with the intervention. As shown earlier, this is an implication of our framework in that a more stringent rule for developing an idea (higher G^*) raises the odds of picking an idea higher than the threshold after n pivots. This is an intriguing result in that we corroborate empirically that being conservative raises the odds that you eventually pick a more valuable idea. At the same time, we cannot exclude that the effect of pivoting is produced by learning in the sense that the scientific approach enables the firms to pivot to a better distribution. However, as noted when we discussed our framework, learning implies that exit is less likely, which is not what we find. Thus, although we do not rule out that the scientific approach provides learning, the results of Table 8 on the implications of pivoting for *Time_to_acquisition* and *Time_to activation* of the treated firms associated with our evidence that the scientific approach does not reduce exit are suggestive of the mechanisms associated with the scientific approach that we envisage.

As also shown in Table 8, *Pivot* and *Intervention* × *Pivot* do not accelerate *Time_to_revenue*, and the collinearity of this regression hides the impact of *Intervention* as well. As discussed earlier, *Pivot* does not have traction, because quite a few firms that pivot and earn revenue in the future are not taken into account in this estimation. When we exclude *Pivot* and *Intervention* × *Pivot* in the last column of Table 8, the effect of *Intervention* becomes sizable and statistically significant. This squares with our prediction that a startup that earns revenue earns more revenue than the counterfactual firm in the control group, because it develops an idea only under more stringent conditions.

Table 7. Revenue Regressions (Panel), Dependent Variable = *Revenue*

Variables	Linear regression	Winsorized 99%	Hyperbolic sine transformation
Intervention	674.968	172.505**	0.565**
	(0.127)	(0.037)	(0.018)
Constant	−637.271	−139.716	−0.510*
	(0.180)	(0.131)	(0.059)
Observations	1,856	1,856	1,856
Number of identification	116	116	116
Dummies for mentors	Yes	Yes	Yes
Time fixed effects	Yes	Yes	Yes
Clustered errors	Firms	Firms	Firms

Note. Robust *p*-values are in parentheses.
$^*p < 0.1$; $^{**}p < 0.05$.

Table 8. Time to Acquisition, Activation, Revenue, Survival (Cox) Regressions

Variables	*Time_to_acquisition*	*Time_to_activation*	*Time_to_revenue*	*Time_to_revenue*
Intervention	−0.120	−0.070	0.041	0.253**
	(0.642)	(0.794)	(0.740)	(0.020)
Pivot	−0.067	−0.385	−0.742	—
	(0.735)	(0.108)	(0.186)	
Intervention × *Pivot*	0.415*	0.577**	0.881	—
	(0.093)	(0.039)	(0.228)	
Observations	116	116	116	116
Dummies for mentors	Yes	Yes	Yes	Yes
Clustered errors	Intervention	Intervention	Intervention	Intervention

Notes. Robust *p*-values are in parentheses. In the survival regressions, *Pivot* is always the number of pivots before the time of the focal event.
$^{*}p < 0.1$; $^{**}p < 0.05$.

For example, this also explains why we have approximately the same number of firms earning revenue in the treatment and control groups (nine versus eight, respectively). Because our firms enter the RCT with an idea for which value is randomly distributed between treated and control firms, the more stringent conditions imply that fewer firms that adopt the scientific approach launch an idea right away. However, conditional on the early start, the idea is more valuable than the counterfactual nonscientific firm as implied by the higher expected revenue.

8. Conclusions

In explaining the high rates of startup failure, the entrepreneurship literature has emphasized several factors, such as the size and characteristics of the founding team or the technology (e.g., Korunka et al. 2003, Aspelund et al. 2005, Gimmon and Levie 2010). In this paper, we focus instead on the role of entrepreneurial decision making, the importance of which in affecting new venture performance has become increasingly central in the stream of research that links entrepreneurship and strategic management (Mitchell et al. 2002, Gans et al. 2017). We have shown that entrepreneurial decision making can benefit from the use of a scientific approach. This approach increases firm performance, because entrepreneurs can recognize when their projects exhibit low or high returns or when it is profitable to pivot to alternative ideas. In other words, entrepreneurs with thoroughly considered, validated theories of their business and hypotheses about what customers want that are then soundly tested through experiments can better mitigate their biases or imprecisions when they analyze market signals (Hayward et al. 2006, Shepherd et al. 2014), reducing the likelihood of incurring false positives and false negatives.

The limitations of our paper raise natural questions for future research. The results presented here are based on an intention to treat rather than an exact measure of the adoption of the scientific approach. The natural next step of this research would be to identify the key mechanism underlying our results. In this study, we present a model that shows that a scientific approach enables better predictions, and this is why we observe different performance outcomes between treated and control entrepreneurs. However, we are unable to directly rule out that a scientific approach improves operations conducted by entrepreneurs. Nevertheless, what we observe in this study provides evidence consistent with the predictions of our framework. In particular, the intervention does not reduce exit, which is consistent with our framework that the scientific method provides predictive capabilities. Moreover, we have not established whether the scientific method provides learning. Also, the time span of our RCT does not allow us to test whether the firms in our sample eventually fail and thus, whether the treated firms fail faster without incurring high costs. Another intriguing question for future research is that the context that we have studied—fairly standard businesses—implies that the scientific method will address the problem of false positives. It will be interesting to study whether the scientific approach can also correct our inability to pursue false negatives. For example, there are biases against novelty in science (Stephan et al. 2017) that may extend to the entrepreneurial or innovation decisions of firms. Last but not least, a scientific approach can help larger firms make decisions, but we have not provided any clue about how this would play out within their complex organizations.

Acknowledgments

The authors thank Rajshree Agarwal, Mario Amore, Ashish Arora, Pierpaolo Battigalli, Luca Berchicci, Raffaele Conti, J. P. Eggers, Teppo Felin, Pooyan Khashabi, Eliana La Ferrara, Ramon Lecuona, Fabio Maccheroni, Massimo Marinacci, John Shook, Scott Stern, and Todd Zenger, as well as seminar participants at Bocconi University, Insead, Harvard Business School, King's College, London Business School, New York University, Politecnico di Milano, Rotterdam School of Management, and the University of North Carolina for excellent comments and suggestions. They also thank Lean Startup

Machine–Italy, represented by Umberto Canessa, and the Doers, represented by Cosimo Panetta and Irene Cassarino, for delivering the treatment of this randomized control trial and all of the mentors for their precious work: Abruzzese Gianluca, Archibugi Monica, Barbieri Giulia, Barboni Luca, Bellini Leonardo, Canalicchio Massimo, Carozza Giacomo, Cocchi Andrea, Detomatici Giulia, Folonorari Giuseppe, Magnifico Roberto, Michetti Francesca, Morandi Marco, Nicola Lorenzo, Povelato Andrea, Rocca Eleonora, Teora Rocco, Tocchetti Gino, and Tonon Francesco. The authors also thank their research assistants Gabriele Ceci, Nicola Coppo, Costantino Coco, and Venera Scuderi for their work on data collection.

Appendix A. Content of Training Steps

We provide an illustration of the key differences in the training provided to treatment and control groups. Table A.1 represents the content of the training of the first lesson, which focuses on an introduction to the course and the business model canvas. The lecture lasts for four hours for both the treatment group and the control group, and it covers five topics for the control group and six topics for the treatment group. The first five topics are identical for the two groups, resulting in 225 slides of content being taught to each group. The treatment group, however, receives additional training (87 slides) on the scientific method (topic 6) and is taught to use the business model canvas as a tool to represent a theory and articulate hypotheses that will be tested at a later stage. The control group is not provided with this training and uses the business model canvas as an overview of the business. To ensure that the training lasts for four hours for both groups, we provide more in-depth explanations to the control group and offer them the opportunity to ask questions to the instructor.

We adopt a similar logic for all of the lessons—Table A.2 presents key differences between treatment and control that were implemented during each training module. Moreover, all of the mentors were instructed to keep roughly this 70:30 ratio for standard versus scientific content in the one-to-one interaction sessions, where the 30 represents scientific training or additional training in the standard material.

Table A.1. Example of Differences Between Treatment and Control Groups, First Lecture

Duration of lecture (4 hours for both treatment and control groups)	
Treatment group	Control group
(1) What is a startup	(1) What is a startup
(2) Innovation and uncertainty	(2) Innovation and uncertainty
(3) Business models	(3) Business models
(4) Business model canvas	(4) Business model canvas
(5) Segmentation, targeting, and positioning	(5) Segmentation, targeting, and positioning
(6) How to deal with uncertainty	(6) Review of the material; question and answer

Table A.2. Overview of the Training with Key Differences Between Treatment and Control Groups

Training	Control	Treatment
Step 1. Business model canvas (BMC): Explore key components of business	1. Does not recognize BMC as overarching theory 2. Does not see individual blocks as representing hypotheses to validate 3. Does not see blocks as being interdependent (as one is falsified, others are too)	1. Aware that BMC is the overarching theory of the firm 2. Sees every block as containing one or more hypotheses that require validation 3. Sees blocks as being interdependent
Step 2. Customer interviews: Explore customers' needs	1. Does not define key hypotheses 2. Poor identification strategy • Interview friends and family • Ask confirmatory questions • Argue in favor of one's idea 3. No clear threshold to direct decision making	1. Define key hypotheses on why customers need your product/service 2. Good identification strategy • Interview potential customers • Ask open-ended questions 3. Use thresholds to falsify hypotheses
Step 3. Minimum viable product: Explore customers' willingness to pay	1. Does not define key hypotheses 2. Poor identification strategy • Does not try parallel variations of the product/service to evaluate improvement • Change more than one thing of the product/service at a time 3. No clear threshold to direct decision making	1. Define key hypotheses on what makes customers most willing to pay 2. Good identification strategy • A/B tests • Change only one thing at a time to identify cause-effect relationships 3. Use thresholds to falsify hypothesis
Step 4. Concierge/prototype: Explore customer service/product interaction	1. Does not define key hypotheses 2. Poor identification strategy • Use available resources to deliver the product/service • Focus on very short-term measure of success 3. No clear thresholds to direct decision making	1. Define key hypotheses on what makes the business sustainable 2. Good identification strategy • Deliver the product/service with the resources that will be used at regime • Focus on longer-term measure of success 3. Use thresholds to direct decision making

Appendix B

Table B.1. Content of Customer Interviews

1. Plan the interview
a. Define learning goal for the interviews
b. Define key assumptions about the customer persona
c. Create a screener survey of simple questions that will identify if the potential interviewee matches your target customer persona. Here is a nice article on screener questions from Alexander Cowan
1. What is the hardest part about (problem context)?
2. Can you tell me about the last time that happened?
3. Why was that hard?
4. What, if anything, have you done to solve that problem?
5. What do you not love about the solutions that you have tried?
d. Make an interview guide (not a write and strictly follow script). If you do not know where to start, check out some questions from Justin Wilcox or Alexander Cowan. Something like this
e. Prepare a handy template to put your notes in afterward or check on the tools to record your interview (check first on legal restrictions that may apply to recordings)
f. Prepare any thank you gifts (e.g., gift cards)
Potential biases
Confirmation bias: The interviewer can be prompted to sell his or her vision in the case that the interviewee's vision differs drastically; the interviewee is tempted in his or her turn to adjust answers to the interviewer's expectations due to personal sympathy
Order bias: Sometimes, the order in which you ask questions can affect the answers that you get; try to run questions in different orders in different interviews

Appendix C

Table C.1. Definition of Variables Used in Balance Tests

Variables	Measurement	Data source
Idea stage	Takes value one if the startup has only a business idea in mind; takes value zero if the startup has started working on the project but has not launched it on the market yet	Project pitch: Research assistants' assessment of the stage of development of the startup based on the milestones achieved by the latter
Internet sector	Takes value one if the startup operates in the internet sector (i.e., provides a service that can be "consumed" online from a computer); takes value zero otherwise	Project pitch: Research assistants' assessment of the sector in which the startup operates based on the product/service offered and the hypothesized channels of sales
Mobile sector	Takes value one if the startup operates in the mobile sector, i.e., provides a service which can be consumed online, from a mobile and/or tablet; takes value zero otherwise	Project pitch: Research assistants' assessment of the sector in which the startup operates based on the product/service offered and the hypothesized channels of sales
Retail sector	Takes value one if the startup operates in the retail sector (i.e., sells a product that is either commercialized via a physical shop or the largest commercial distribution); takes value zero otherwise	Project pitch: Research assistants' assessment of the sector in which the startup operates based on the product/service offered and the hypothesized channels of sales
Lombardy	Takes value one if the majority of team members come from the Italian region of Lombardy; takes value zero otherwise	Team members' vitae: Retrieved from city of domicile
Team size	It is the absolute number of team members of the startup	Team members' vitae: We count the number of vitae sent by the team
Industry experience	It is the average number of years of experience of the team in the industry in which the startup operates before entering the training	Project pitch and team members' vitae: We match the Standard Industrial Classification (SIC) codes (at the 83 two digit-level major groups) of the startup (assessed by the research assistant) and the firms in which the founders previously held a job position as described in their vitae
Management experience	It is the average number of years of managerial experience of the team before entering the training	Team members' vitae: We look at the years that each team member had in a managerial job position as described in their vitae; the count includes both higher- and lower-level managerial positions and all four managerial functional roles (Barbero et al. 2011)
Experience working with startups	It is the average number of years of experience of the team working with/for startups other than the one that the team members intend to launch before entering the training	Team members' vitae: We look at the years that each team member had as either founder or employee in a startup (this should have been defined so by the team member itself in the vitae)

Management Science, 2020, vol. 66, no. 2, pp. 564–586, © 2019 INFORMS

Table C.1. (Continued)

Variables	Measurement	Data source
Experience working in startups	It is the average number of years of experience of the team within startups other than the one that the team members intend to launch before entering the training	Team members' vitae: We look at the years that each team member had as mentor and/or consultant to a startup (this should have been defined so by the team member itself in the vitae)
Currently employed	It is the proportion of team members employed at the time of entry into the training	Team members' vitae: We record a team member as currently employed if any of his or her job positions described in the CV do not show an ending time (e.g., "from 15 Feb 2004 to present")
Currently studying	It is the proportion of team members enrolled in an education program at the time of entry into the training	Team members' vitae: We record a team member as currently studying if any of his or her enrollments in an educational program described in the vitae do not show an ending time (e.g., "from 15 Feb 2004 to present")
Level of education	It is the level of education of the team in the industry in which the startup operates	Team members' vitae: We look at the educational titles achieved by each team member, and we record them as following: one is for high school, two for Bachelor's degree, three for Master's degree, four is for MBA, and five is for Ph.D.

Appendix D. Robustness Check Regressions (*Exit*, *Pivot*, and *Revenue*)

Table D.1. Exit Regressions, Dependent Variable = *Exit*

Variables	Panel
Cumulative_intervention	−0.002
	(0.682)
Postintervention	0.025
	(0.297)
Constant	−0.019**
	(0.019)
Observations	1,606
Number of identification	116
R^2	0.055
Time fixed effects	Yes
Firm fixed effects	Yes
Clustered errors	Firms

Note. Robust p-values are in parentheses.
$^{**}p < 0.05$.

Table D.2. Pivot Regressions, Dependent Variable = *Number of Pivots*

Variables	Poisson	Panel
Cumulative_intervention	—	0.007*
		(0.060)
Postintervention	—	−0.051**
		(0.016)
Intervention	0.794***	—
	(0.000)	
Constant	−0.879	−0.003
	(0.136)	(0.492)
Observations	116	1,856
R^2		0.030
Dummies for mentors	Yes	
Time fixed effects	—	Yes
Clustered errors	Intervention	Firms
Number of identification	—	116
Firm fixed effects	—	Yes

Note. Robust p-values are in parentheses.
$^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

Table D.3. Revenue, Panel Regressions

Variables	Linear regression	Winsorized 99%	Hyperbolic sine transformation
Cumulative_intervention	16.502	18.459*	0.097**
	(0.105)	(0.064)	(0.020)
Postintervention	896.542	−13.644	−0.403**
	(0.324)	(0.839)	(0.039)
Constant	5.400	4.404	0.020
	(0.982)	(0.913)	(0.875)
Observations	1,856	1,856	1,856
R^2	0.015	0.020	0.039
Number of identification	116	116	116
Firm fixed effects	Yes	Yes	Yes
Time fixed effects	Yes	Yes	Yes
Clustered errors	Firms	Firms	Firms

Note. Robust p-values are in parentheses.
$^*p < 0.1$; $^{**}p < 0.05$.

Endnotes

[1] To see this, define (1) as z, and after integrating (2) by parts, $\pi = z(A) - z(v^*)G^* - \int_{-v^*}^{A} z_{\hat{v}_0} G d\hat{v}_0$, where the subscript denotes derivative. From *foc* (4), $z(v^*) = \frac{\delta^* \pi}{1-G^*}$. Replace in π, and after bringing $z(v^*)G^*$ to the left-hand side, obtain $\frac{\pi}{1-G^*} = z(A) - \int_{-v^*}^{A} z_{\hat{v}_0} G d\bar{v}$, which decreases as G increases.

[2] An intriguing implication of our framework is that, if the scientific approach helps to reduce overestimation and make the entrepreneur aware of a wider distribution, exit or pivot will be ambiguous, because both the signal $\hat{v}_0$ and the threshold v^* decrease. This suggests, for example, that sometimes biases or misperceptions could lead to ideal outcomes. We leave this to future research.

[3] See https://www.cbinsights.com.

[4] This is a t test, which is preferred to running a logit/probit regression of selection into the training (or treatment) on all covariates simultaneously. In small samples, running the regression with all covariates simultaneously can reduce the significance of coefficient estimates (Hansen and Bowers 2008).

[5] We were careful in distinguishing a minor change in target customers from a major change. The former—the so-called zoom in or zoom out—consists of more narrowly or more broadly, respectively, identifying the same potential market without a change in target customers. The latter consists of a drastic change of target customers for a product or service (Ries 2011, Crilly 2018).

References

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

Aspelund A, Berg-Utby T, Skjevdal R (2005) Initial resources' influence on new venture survival: A longitudinal study of new technology-based firms. *Technovation* 25(11):1337–1347.

Åstebro T (2003) The return to independent invention: Evidence of unrealistic optimism, risk seeking or skewness loving? *Econom. J.* 113(484):226–239.

Barbero JL, Casillas JC, Feldman HD (2011) Managerial capabilities and paths to growth as determinants of high-growth small and medium-sized enterprises. *Internat. Small Bus. J.* 29(6):671–694.

Bennis WG (1962) Toward a "truly" scientific management: The concept of organization health. *Indust. Management Rev.* 4(1):1–19.

Blank S, Dorf B (2012) *The Startup Owner's Manual: The Step-by-Step Guide for Building a Great Company* (K&S Ranch, Pescadero, CA).

Blank SG (2006) *The Four Steps to the Epiphany: Successful Strategies for Products That Win* (K&S Ranch, Pescadero, CA).

Bloom N, Van Reenen J (2007) Measuring and "explaining" management practices across firms and countries. *Quart. J. Econom.* 22(4):1351–1408.

Bloom N, Eifert B, Mahajan A, Mckenzie D, Roberts J (2013) Does management matter? Evidence from India. *Quart. J. Econom.* 128(1):1–51.

Briner RB, Denyer D, Rousseau DM (2009) Evidence-based management: Concept cleanup time? *Acad. Management Perspect.* 23(4):19–32.

Brown T (2008) Design thinking. *Harvard Bus. Rev.* 86(6):84–92.

Brynjolfsson E, McElheran K (2016) The rapid adoption of data-driven decision-making. *Amer. Econom. Rev.* 106(5):133–139.

Busenitz LW, Barney JB (1997) Differences between entrepreneurs and managers in large organizations: Biases and heuristics in strategic decision-making. *J. Bus. Venturing* 12(1):9–30.

Clark T, Wiesenfeld D (2017) 3 things are holding back your analytics, and technology isn't one of them. *Harvard Bus. Rev.* (June 8), https://hbr.org/2017/06/3-things-are-holding-back-your-analytics-and-technology-isnt-one-of-them.

Crilly N (2018) "Fixation" and "the pivot": Balancing persistence with flexibility in design and entrepreneurship. *Internat. J. Design Creativity Innovation* 6(1–2):52–65.

Drucker PF (1955) Management science and the manager. *Management Sci.* 1(2):115–126.

Dushnitsky G (2010) Entrepreneurial optimism in the market for technological inventions. *Organ. Sci.* 21(1):150–167.

Eisenmann T, Ries E, Dillard S (2013) Hypothesis-driven entrepreneurship: The Lean Startup. Note 90-812-095, Harvard Business School, Boston.

Fairlie R, Miranda J (2016) Taking the leap: The determinants of entrepreneurs hiring their first employee. NBER Working Paper No. 22428, National Bureau of Economic Research, Cambridge, MA.

Fairlie R, Miranda J (2017) Taking the leap: The determinants of entrepreneurs hiring their first employee. *J. Econom. Management Strategy* 26(1):3–34.

Felin T, Zenger TR (2009) Entrepreneurs as theorists: On the origins of collective beliefs and novel strategies. *Strategic Entrepreneurship J.* 3(2):127–146.

Felin T, Zenger TR (2015) Crossroads—Strategy, problems, and a theory for the firm. *Organ. Sci.* 27(1):222–231.

Fleming L (2001) Recombinant uncertainty in technological search. *Management Sci.* 47(1):117–132.

Freedman DH (1992) Is management still a science? *Harvard Bus. Rev.* 70(6):26.

Galasso A, Simcoe TS (2011) CEO overconfidence and innovation. *Management Sci.* 57(8):1469–1484.

Gans J, Stern S, Wu J (2017) Foundations of entrepreneurial strategy. Working paper, University of Toronto, Toronto. Accessed June 24, 2018, https://ssrn.com/abstract=2844843.

Gelber A, Isen A, Kessler JB (2016) The effects of youth employment: Evidence from New York City lotteries. NBER Working Paper No. 20810, National Bureau of Economic Research, Cambridge, MA.

Gimmon E, Levie J (2010) Founder's human capital, external investment, and the survival of new high-technology ventures. *Res. Policy* 39(9):1214–1226.

Grandori A (2010) A rational heuristic model of economic decision-making. *Rationality Soc.* 22(4):1–28.

Hansen BB, Bowers J (2008) Covariate balance in simple, stratified and clustered comparative studies. *Statist. Sci.* 23(2):219–236.

Hayward MLA, Shepherd DA, Griffin D (2006) A hubris theory of entrepreneurship. *Management Sci.* 52(2):160–172.

Hsieh C, Nickerson JA, Zenger TR (2007) Opportunity discovery, problem solving and a theory of the entrepreneurial firm. *J. Management Stud.* 44(7):1255–1277.

Jackson WT, Scott DJ, Schwagler N (2015) Using the business model canvas as a methods approach to teaching entrepreneurial finance. *J. Entrepreneurship Ed.* 18(2):99–112.

Kelley T, Littman J (2005) *The Ten Faces of Innovation: IDEO's Strategies for Beating the Devil's Advocate & Driving Creativity throughout Your Organization* (Currency/Doubleday, New York).

Kerr WR, Nanda R, Rhodes-Kropf M (2014) Entrepreneurship as experimentation. *J. Econom. Perspect.* 28(3):25–48.

Korunka C, Frank H, Lueger M, Mugler J (2003) The entrepreneurial personality in the context of resources, environment, and the startup process—A configurational approach. *Entrepreneurship Theory Practice* 28(1):23–42.

Li Y, James B, Madhavan R, Mahoney JT (2007) Real options: Taking stock and looking ahead. *Adv. Strategic Management* 24: 31–66.

Lopez-Vega H, Tell F, Vanhaverbeke W (2016) Where and how to search? Search paths in open innovation. *Res. Policy* 45(1): 125–136.

Martin R (2009) *The Design of Business: Why Design Thinking Is the Next Competitive Advantage* (Harvard Business School Press, Boston).

McGrath RG (1997) A real options logic for initiating technology positioning investments. *Acad. Management Rev.* 22(4):974–996.

McGrath RG (1999) Falling forward: Real options reasoning and entrepreneurial failure. *Acad. Management Rev.* 24(1):13–30.

McGrath RG, MacMillan IC (1995) Discovery driven planning. *Harvard Bus. Rev.* 73(4):44–54.

McGrath RG, MacMillan IC (2009) *Discovery Driven Growth: A Breakthrough Process to Reduce Risk and Seize Opportunity* (Harvard Business Publishing, Boston).

Mitchell RK, Busenitz L, Lant T, McDougall PP, Morse EA, Smith JB (2002) Toward a theory of entrepreneurial cognition: Rethinking the people side of entrepreneurship research. *Entrepreneurship Theory Practice* 27(2):93–104.

Murray F, Tripsas M (2004) The exploratory processes of entrepreneurial firms: The role of purposeful experimentation. Baum J, McGahan A, eds. *Business Strategy over the Industry Lifecycle* (Emerald Group, Bingley, UK), 45–75.

Nicholls-Nixon C, Cooper A, Woo C (2000) Strategic experimentation: Understanding change and performance in new ventures. *J. Bus. Venturing* 15(5–6):493–521.

Osterwalder A, Pigneur Y (2009) *Business Model Generation: A Handbook for Visionaries, Game Changers, and Challengers* (Alexander Osterwalder & Yves Pigneur, Amsterdam).

O'Brien JP, Folta TB, Johnson DR (2003) A real options perspective on entrepreneurial entry in the face of uncertainty. *Managerial Decision Econom.* 24(8):515–533.

Pfeffer J, Sutton RI (2006) *Hard Facts, Dangerous Half-truths, and Total Nonsense: Profiting from Evidence-Based Management* (Harvard Business Press, Boston).

Ries E (2011) *The Lean Startup: How Today's Entrepreneurs Use Continuous Innovation to Create Radically Successful Businesses* (Crown Business, New York).

Ripsas S, Schaper B, Tröger S (2015) A startup cockpit for the proof-of-concept. Faltin G, ed. *Handbuch Entrepreneurship* (Springer Gabler, Wiesbaden, Germany), 1–16.

Rousseau DM (2006) Is there such a thing as "evidence-based management"? *Acad. Management Rev.* 31(2):256–269.

Sarasvathy SD (2001) Causation and effectuation: Toward a theoretical shift from economic inevitability to entrepreneurial contingency. *Acad. Management Rev.* 26(2):243–263.

Sarasvathy SD, Venkataraman S (2011) Entrepreneurship as method: Open questions for an entrepreneurial future. *Entrepreneurship Theory Practice* 35(1):113–135.

Shepherd DA, Haynie JM, McMullen JS (2012) Confirmatory search as a useful heuristic: Testing the veracity of entrepreneurial conjectures. *J. Bus. Venturing* 27(6):637–651.

Shepherd DA, Williams TA, Patzelt H (2014) Thinking about entrepreneurial decision making: Review and research agenda. *J. Management* 41(1):11–46.

Stephan P, Veugelers R, Wang J (2017) Reviewers are blinkered by bibliometrics. *Nature* 544(7651):411–412.

Sull DN (2004) Disciplined entrepreneurship. *MIT Sloan Management Rev.* 46(1):71.

Venkataraman S, Sarasvathy SD, Dew N, Forster WR (2012) Reflections on the 2010 AMR Decade Award: Whither the promise? Moving forward with entrepreneurship as a science of the artificial. *Acad. Management Rev.* 37(1):21–33.

Von Hippel E (1986) Lead users: A source of novel product concepts. *Management Sci.* 32(7):791–805.

Zenger TR (2015) Trial and error is no way to make strategy. *Harvard Bus. Rev.* (April 24), https://hbr.org/2015/04/trial-and-error-is-no-way-to-make-strategy.

Zenger TR (2016) *Beyond Competitive Advantage: How to Solve the Puzzle of Sustaining Growth While Creating Value* (Harvard Business School Press, Boston).