DISCUSSION PAPER SERIES

DP12421

A SCIENTIFIC APPROACH TO ENTREPRENEURIAL DECISION-MAKING: EVIDENCE FROM A RANDOMIZED CONTROL TRIAL

Arnaldo Camuffo, Alessandro Cordova and Alfonso Gambardella

INDUSTRIAL ORGANIZATION

ISSN 0265-8003

A SCIENTIFIC APPROACH TO ENTREPRENEURIAL DECISION-MAKING: EVIDENCE FROM A RANDOMIZED CONTROL TRIAL

Arnaldo Camuffo, Alessandro Cordova and Alfonso Gambardella

Discussion Paper DP12421
Published 02 November 2017
Submitted 02 November 2017

Centre for Economic Policy Research
33 Great Sutton Street, London EC1V 0DX, UK
Tel: +44 (0)20 7183 8801
www.cepr.org

This Discussion Paper is issued under the auspices of the Centre's research programme in **INDUSTRIAL ORGANIZATION**. Any opinions expressed here are those of the author(s) and not those of the Centre for Economic Policy Research. Research disseminated by CEPR may include views on policy, but the Centre itself takes no institutional policy positions.

The Centre for Economic Policy Research was established in 1983 as an educational charity, to promote independent analysis and public discussion of open economies and the relations among them. It is pluralist and non-partisan, bringing economic research to bear on the analysis of medium- and long-run policy questions.

These Discussion Papers often represent preliminary or incomplete work, circulated to encourage discussion and comment. Citation and use of such a paper should take account of its provisional character.

Copyright: Arnaldo Camuffo, Alessandro Cordova and Alfonso Gambardella

A SCIENTIFIC APPROACH TO ENTREPRENEURIAL DECISION-MAKING: EVIDENCE FROM A RANDOMIZED CONTROL TRIAL

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 to conduct rigorous tests of their hypotheses very much like 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, pivot to a greater extent to a different idea, and do not drop out less 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 raises the odds of pursuing projects with false negative returns.

JEL Classification: L21, L26, M13, M21

Keywords: entrepreneurship, decision-making, scientific method, startup, randomized control trial

Arnaldo Camuffo - arnaldo.camuffo@unibocconi.it
Bocconi University

Alessandro Cordova - alessandro.cordova@phd.unibocconi.it
Bocconi University

Alfonso Gambardella - alfonso.gambardella@unibocconi.it
Università Bocconi and CEPR

Acknowledgements
We thank Mario Amore, Ashish Arora, Pierpaolo Battigalli, Raffaele Conti, JP Eggers, Teppo Felin, Eliana La Ferrara, Ramon Leucona, Fabio Maccheroni, Massimo Marinacci, and Todd Zenger, as well as seminar participants at Bocconi University, Insead, Harvard Business School, London Business School, New York University, Universty of North Carolina, for excellent comments and suggestions. We 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 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, Niola Lorenzo, Povelato Andrea, Rocca Eleonora, Teora Rocco, Tocchetti Gino, and Tonon Francesco. We also thank our research assistants, Gabriele Ceci, Nicola Coppo, Costantino Coco, and Venera Scuderi, for their precious work on data collection. All errors are ours.

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 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 in economics on this subject have emerged, such as discovery-driven planning (McGrath and McMillan, 1995 and 2009), real option strategies (McGrath, 1997; O'Brien et al., 2003; Adner and Levinthal, 2004; Mahoney, 2005; 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; Grandori, 2010; Felin and Zenger, 2009; 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 and 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 model or product 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 own approach and intuition to assessing the information they receive from the processes that we teach them in the program.

Firms may invest in projects that are less valuable than they think (false positives) or they may not invest in projects that are more successful than they believe (false negatives). While 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 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. We test these proposition in our RCT.

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 4-month training program: (1) business model canvas, (2) customer interviews, (3) minimum viable product, and (4) concierge or prototype. Figure 1 summarizes the training program contents. While we teach both treated and control startups about these four steps, we teach in particular the treated startups to elaborate a framework to understand the impact of their idea and to predict business performance, define clear hypotheses, design rigorous experiments to confirm or disconfirm them, and make decisions accordingly. This approach permeates all the steps of our training program, as summarized in Appendix Section A.

Figure 1 approximately here

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 through 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 through 4 focus on testing the potential of the founders' value proposition and its fit with the hypothesized market target, and that the approach they are learning is useful for testing aspects of the business that will be relevant later (e.g., the firm's revenue model).

Customers' interviews

We teach all startups how to interview customers in order to understand the firm's potential market, to segment it, to learn about the customers' needs, and to 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 to 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?", which 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 don't want to disappoint their peers, and (ii) this is a fictitious market setting where respondents do not use the service and therefore it is not costly to respond affirmatively. While 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 often 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, do 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 to identify the key areas requiring validation, which led to the articulation of four clear hypotheses: (a) tattooed people do not always use the same tattooist, (b) they choose new tattooists online, (c) this takes time and is painful, and (d) tattooed people can find online all the information 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 quasi-ethnographic 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 Section B reports the instructions for this quasi-ethnographic 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 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 quasi-ethnographic 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. Based on this information, Inkdome changed its business model from a search engine to a platform where users seek advice from experts.

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, with typically a button users can click to buy now, sign up for the free beta, or pre-order.

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 most 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. While 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 e-mail. 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 e-mail addresses that customers left) of two landing pages that were identical except that version A advertised that users would *receive advice via e-mail* from tattoo experts, and version B advertised that users would *chat* with tattoo experts. This experimental design allows 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 e-mail-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 e-mail addresses left on version A of the landing page. The test showed that version B produced 2.5 times more e-mails than version A. Inkdome therefore chose the chat-based system.

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 collected the descriptions of their tattoo idea 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 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.

Additional remarks

The Inkdome case study clarifies three relevant features of our framework and of 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 identify the problem in abstract ways, to formulate hypotheses, and to test these using rigorous experiments 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, initially the problem is not well defined, 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. Both in the lectures and in the one-to-one discussions, we teach and encourage the treated startups, during all four steps of our training, to collect this information, and to 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 our startups enter our RCT having a business idea. Inkdome, for example, began with its online search engine. However, none of the participant startups have 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 affect largely the ability of firms to evaluate the idea with which they enter the RCT. Over time, the information they collect can become useful for assessing modifications to this original idea or even radical departures from it to pivot to a new idea, as in Inkdome's case. Once again, this is true of both the treated and the control firms. However, the question is whether the 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 refer not 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, Lopez-Vega et al.'s (2016) study 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: (a) build-it-and-they-will-come, (b) waterfall planning, and (c) 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 is 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 (Rousseau, 2006; Pfeffer and Sutton, 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.) allows 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, experiments and tests of hypotheses) are one example of the actions that help to exercise real option opportunities.

4. Model

Our model, which builds on Arora and Gambardella (1994), focuses on how a scientific approach leads to more effective entrepreneurial decisions. A firm that explores a business idea must decide whether to pay k in order to observe a net revenue $r \in [0, R]$. When the firm decides whether to pay k, r is uncertain, but the firm observes a signal $\hat{r}$ of r, such that $F(r \mid \hat{r}, \theta)$ is the cumulative distribution of r conditional upon $\hat{r}$. It is natural to assume that F declines with $\hat{r}$; that is, that a higher signal makes higher levels of r more likely. The distribution F also depends on a parameter θ that captures the impact of the scientific method and that we discuss below.

The firm chooses an optimal threshold r^* such that the firm pays k if the signal $\hat{r}$ is greater than r^*. Thus, if $\hat{r} \geq r^*$, the firm pursues the current idea. If $\hat{r} < r^*$, the firm can drop out (and close the venture) or pivot to a new idea. If the firm pivots, it faces the same decision tree. It decides whether to pay a new k for the new idea based on a signal $\hat{r}$ correlated with the returns r of the new idea; the firm picks a new threshold r^* such that it pursues the new idea, drops out, or pivots following the same decision-logic of the first idea. In principle, the firm can pivot indefinitely, and further pivoting is only discouraged by a discount factor δ such that, other things being equal, the firm prefers to pursue an idea earlier rather than later. For simplicity, we assume that if the firm gives up an idea, and pivots to a new one, it can no longer exploit the abandoned idea at a later stage. This is consistent, for example, with Gans et al. (2017), who argue that once the firm commits to an idea, it loses the opportunity to exploit other ideas that it could have pursued.

The expression for v_t, the expected value of the firm's t^{th} idea, is

$$v = E_{\Omega}[-k + \int_0^R r dF(r|\hat{r} \geq r^*, \theta)](1 - G(r^*)) \qquad (1)$$

where we dropped the subscript t for simplicity, G is the cumulative distribution of the signal $\hat{r}$, and E_{Ω} indicates expectation conditional upon Ω, where Ω is a shorthand notation for the knowledge set of the firm at t. The set Ω and θ are related, and we discuss them below. Expression (1) says that conditional upon observing a signal higher than the threshold, the firm pays k and obtains an expected return equal to the expected value of r conditional upon $\hat{r} \geq r^*$. Using the fact that $F(r \mid \hat{r} \geq r^*) = \frac{\int_{r^*}^R F(r \mid \hat{r})}{1 - G(r^*)}$, and after integrating by parts, we rewrite (1) as

$$v = (R - k)\big(1 - G(r^*)\big) - E_\Omega \int_0^R \int_{r^*}^R F\,(r|\hat{r},\theta)dGdr \qquad (2)$$

The objective function of the firm working on its t^{th} idea is then

$$V_t = E_{\Omega_t}(v_t + G_t^*\delta v_{t+1} + G_t^* G_{t+1}^*\delta^2 v_{t+2} + G_t^* G_{t+1}^* G_{t+2}^*\delta^3 v_{t+3} + \ldots) = E_{\Omega_t}(v_t + G_t^*\delta V_{t+1})$$

where G_τ is the distribution function of the signal $\hat{r}_\tau$ received for any idea $\tau = t$, $t+1$, $t+2$, . . . ; $G_\tau^* \equiv G(r_\tau^*)$, E_{Ω_t} denotes expectation conditional upon Ω_t; and δ is the discount factor mentioned earlier. This objective function says that when the firm does not pursue the t^{th} idea, which happens with probability G_t^*, it can pivot to a new idea whose value is v_{t+1}, and it can do the same at $t+1$, $t+2$, ... The problem of the entrepreneur is to pick the optimal thresholds r_τ^*, $\tau = t$, $t+1$, $t+2$, . . . , that maximize V_t.

Before we discuss these optimal choices, the parameter θ reduces F, which means that higher θ is desirable. We posit that the scientific method enables the firm to predict θ more precisely, and in this respect the shorthand notation Ω_t captures the difference between the knowledge set of a firm exposed to the scientific method and one not exposed to it. In other words, Ω_t simply denotes that the firm exposed to the scientific method picks the optimal r^* using a different knowledge basis that enables the decision maker to rely on a more precise estimate of θ. Also, each idea (t, $t+1$, $t+2$, etc.) corresponds to a different parameter θ_t, θ_{t+1}, θ_{t+2}, and so on. For now, we assume that there is no drift of θ over time: the parameters θ unfold randomly, and they can be higher or lower as the firm pivots to new ideas. This enables us to focus our theoretical discussion on the effects of the scientific method on the precision with which the entrepreneurs estimate the value of their ideas. Later, we explore the implications of learning, that is, a drift in θ, and we show that learning does not change the substance of our argument. From the point of view of our entrepreneurs, our assumption means that when they pivot to a new idea, they do not expect the new idea to be better. They are equally uncertain about it, and the switch only mirrors the benefits of making another draw from the distribution of returns.

The predictions of our model rest on two assumptions. First, the scientific approach enables the entrepreneur to predict the current θ, that is, θ_t, with greater precision. Falsifiable hypotheses and rigorous tests corroborate or reject the theory, providing better information about the true θ. In other words, the scientific approach provides the conditioning set for a Bayesian update of the entrepreneur's prior distribution of θ. This update generates a higher probability mass around the true value of θ. The scientist entrepreneur then observes a distribution $F(r \mid \hat{r}, \theta)$ closer to the true distribution F, in terms of a smaller error or distance from it. As we discuss below, the assumption is that this leads to the choice of an optimal threshold r^* for the signal $\hat{r}$ closer to the optimal choice that the decision maker would make if she observed the true θ. Of course, the non-scientist may have other rules to update her prior distribution of θ, but we posit that the update provided by the scientific approach is more precise.

Second, when evaluating future ideas, the scientist entrepreneur does not predict the future θ, that is, θ_{t+1}, θ_{t+2}, . . . , better than the non-scientist entrepreneur. This is because when the scientist entrepreneur is assessing the t^{th} idea, she has not yet worked on the future ideas. She has not formulated a theory about it and has not tested it with her rigorous experiments. However, unlike the non-scientist entrepreneur, she knows that when she evaluates these future ideas, the scientific method will help her pick a better optimal threshold than the control because she will have more information. Specifically, she will be able to see a θ closer to the true θ, very much like in the current period. As a result, even though she can only make the same prediction as the non-scientist entrepreneur about the future θ, she expects to know it more precisely if it comes to making that decision. The better optimal threshold will generate a higher expected return, which is why the scientist entrepreneur predicts a higher V_{t+1} than the non-scientist entrepreneur.[1]

Our entrepreneurs choose r_t^* to maximize $V_t = E_{\Omega_t}(v_t + G_t^* \delta V_{t+1})$, whose first order condition (foc) is $E_{\Omega_t}\left(\frac{\partial v_t}{\partial r_t^*} + g_t^* \delta V_{t+1}\right) = 0$, where g_t^* is the density of G_t^*. Using (2), $\frac{\partial v_t}{\partial r_t^*} = -(R-k)g^* + \int_0^R F(r|r^*, \theta) g^* dr$, where again we do not use subscripts for simplicity. The foc becomes

$$E_{\Omega_t}[-(R-k) + \int_0^R F(r|r^*, \theta) dr + \delta V_{t+1}] = 0 \qquad (3)$$

Moreover, since F declines with r^*, the second order condition is satisfied.

[1] A simple intuition is the following. You can be in a state of nature, which occurs with probability p, that yields an objective $f(x, z_1)$, or in a state of nature, which occurs with probability $1 - p$, that yields $f(x, z_2)$. Suppose that you do not know in which state you are. You then pick x to maximize $pf(x, z_1) + (1 - p)f(x, z_2)$. Suppose instead that you know in which state you are. You pick x_1 that maximizes $f(x, z_1)$ if you are in state z_1, and x_2 that maximizes $f(x, z_2)$ if you are in state z_2. If you are not yet there, but you know that you will be there, the expected value is $pf(x_1, z_1) + (1 - p)f(x_2, z_2)$. Compared with the previous case, $f(x_1,z_1) \geq f(x, z_1)$ and $f(x_2, z_2) \geq f(x, z_2)$ because x_1 maximizes $f(x, z_1)$ and x_2 maximizes $f(x, z_2)$.

The key differences between scientist and non-scientist entrepreneurs are Ω_t and the fact that scientist entrepreneurs expect a higher V_{t+1}. First, as noted, scientist entrepreneurs predict θ closer to the true θ, which enables them to make a superior choice of the optimal r^*, in the sense of a value of r^* that generates a higher V_t than non-scientist entrepreneurs.[2] This implies that scientist entrepreneurs achieve higher performance. To highlight the mechanisms that generate this higher performance, we must preliminarily clarify that, as widely known, most new entrepreneurial ideas are not profitable. For example, Fairlie and Miranda (2017) show that 84.4% of U.S. startups fail within 7 years. (See Table 1A of their NBER working paper.) For our model, this means that it is more likely that a scientist entrepreneur, who is more precise, realizes that θ is lower than does a non-scientist entrepreneur – that is, the scientific method is more likely to reveal false positives. If so, in most cases the scientist entrepreneurs predict a higher F, which, combined with a higher V_{t+1}, implies that scientists-entrepreneurs are more likely to pick a higher r^* and therefore to pivot more.

In addition, a reasonable assumption is that entrepreneurs drop out when they observe V_t smaller than a threshold (e.g. zero). This implies that the dropout rate of scientist- versus non-scientist entrepreneurs is ambiguous. On one hand, because most ideas are bad, scientist entrepreneurs are more likely to predict a lower θ_t and therefore a lower v_t; on the other hand, they predict a higher V_{t+1}. Therefore, we cannot predict whether $V_t = E_{\Omega_t}(v_t + G_t^* \delta V_{t+1})$ is higher or lower for one or the other type of entrepreneur. This prompts two clarifications. First, scientist entrepreneurs choose a superior optimal r^*, which yields a higher V_t; however, this is the "true" V_t. Because they have poorer information, the non-scientists do not predict a V_t as close to the true V_t as the scientists do, and they may well perceive a higher V_t. In this study, the notion of dropout is different from that of failure, which occurs if a firm pays k and later realizes that actual profits are negative.[3] Second, if scientist entrepreneurs predict a very low θ_t, the optimal r^* increases, making v_t close to zero, and V_t close to δV_{t+1}. However, whether this makes V_t for the scientists higher

[2] All we need for this assumption is that V_t is smooth and concave in r^*, and when the predicted θ is closer to the exact θ, the optimal r^* is closer to the optimal r^* computed with the exact θ. The maximum of V_t obtains when the firm observes the exact θ and chooses the optimal r^* accordingly. A smooth and concave function for the optimized V_t implies that any choice of r^* closer to the optimal value computed using the exact θ yields a higher V_t.

[3] In our RCT some firms dropped out, but we lack a sufficient window for observing whether some firms fail, particularly some of the control firms that have not dropped out. However, this is not crucial for our analysis because we employ information on whether they drop out, and we do not use information on whether they fail.

or lower than that for the non-scientists depends on functional forms, and thus we cannot make unambiguous predictions.

The following proposition summarizes the predictions of the theory that we test in our RCT.

Proposition. *A scientific approach to entrepreneurial decision making yields higher performance because the scientist-entrepreneur avoids false positives and false negatives. If most entrepreneurial ideas are not profitable, it induces more pivots and has an ambiguous effect on the rate of dropout.*

The gist of our story is that scientist entrepreneurs perform better because they are more likely to detect false positives, which occur more frequently, and therefore place greater value on pivoting. The intuition of our model is that if the scientist entrepreneur predicts a lower θ than a non-scientist entrepreneur, and such that it is closer to the true value, then she chooses a higher optimal r^*. Using (3), the marginal projects that received a signal $\hat{r}$ between the higher threshold r^* chosen by the scientist entrepreneurs and the lower threshold chosen by the non-scientist entrepreneurs yield, as expected, negative returns. The non-scientists pick these projects because they do not predict θ as precisely as the scientists do. While we stress that, in practice, a lower θ is the more common case, the scientist will also predict, correctly, a higher θ when this is the case. If so, she will set a lower r^* than the non-scientists, such that all the projects with signals $\hat{r}$ between the lower threshold r^* of the scientist entrepreneurs and the higher threshold of the non-scientist entrepreneurs yield, as expected, positive returns. Again, the non-scientists do not pick them because they do not predict θ as precisely as the scientists do.

So far we have ignored learning, and particularly the fact that the scientific method can produce a drift of θ over time. In such cases, a straight implication would be that the mechanism through which the scientific approach affects performance is not just pivot; it would also directly affect performance. This is easy to see from our model because, irrespective of pivoting, a drift in θ increases both v_t and V_{t+1}, and therefore V_t. This ought to reduce the dropout rate because the scientist entrepreneur predicts a higher V_t. The effect on pivoting is instead ambiguous depending on the relative effect of the drift on F and V_{t+1} in (3). A natural assumption is that the learning effect exhibits diminishing returns over time. If so, as t increases, the effect of F in (3) dominates that of V_{t+1}. As a result, r^* is likely to decline as the firm pivots, making it likely that a firm adopting the scientific method makes fewer pivots after the first pivots.

In terms of our empirical strategy, our RCT tests whether a scientific approach yields higher performance and induces more pivots, whereas we make no prediction for the dropout rate. We cannot test that pivot is the mechanism through which the scientific method affects performance, as predicted by our theory. This would require another treatment for pivot, which we lack in this study. However, we can provide evidence consistent with the mechanism by showing that the treatment yields higher performance and more pivots. We cannot rule out that, along with performance, the scientific method provides learning, in the sense discussed above. However, we can exclude that there is only a learning and no precision effect. Learning implies that the treated firms are less likely to drop out. Thus, if along with greater performance and more pivoting the treated firms do not drop out less than the control firms, we have evidence consistent with a precision effect. Further evidence for a precision effect is that the treated firms do not reduce their pivots after they pivot a few times. As noted, a simple assumption of diminishing returns to learning suggests that if there is only learning, treated firms pivot less after some initial pivots.

Finally, greater variance in the performance of the treated firms compared with the control firms would further evince a precision effect. We theorize that some firms adopting the scientific method see a high θ and correctly pursue profitable opportunities that the control firms do not see or that are not available to other treated firms that were not equally lucky and observed a low θ. Thus, control firms will perform more similarly because their behavior is more homogenous than that of treated firms, in that they all see similar θ around the expected value. In contrast, treated firms see different θ, which maps onto different behavior – that is, higher or lower optimal r^*, which implies that for some of them performance is higher because they do not pursue bad opportunities that the control firms do pursue, whereas the treated firms that see a higher θ perform better because they earn a higher revenue. Moreover, the variance in the performance of the treated firms is likely to increase over time. Since most ideas are not profitable a priori, at the beginning all the treated firms earn no profits, either because they have not yet found the right opportunity or because they are in the gestation period before the revenues of a good opportunity take off. Over time, some of these firms are still seeking the good opportunity because, thanks to the scientific method, they have discarded many false positives, while others have actually found such opportunities, and their revenues are growing.

To summarize, we cannot rule out that the scientific method has a learning effect. However, we can provide evidence suggesting that, apart from a learning effect, the scientific method provides greater precision – in particular, we provide evidence for a precision effect: if the scientific method does not produce a higher rate of dropout, it does not reduce pivoting after the initial pivots, and the variance in the performance of the treated firms is higher than that of the control firms, and possibly increases over time.

5. Research design, data, and method

Randomized control trial design

We partnered with two institutions that train startups and that have pioneered the use of approaches close to the scientific approach we discuss in this paper: the Lean Startup Machine and the Doers. The Lean Startup Machine operates worldwide, offering 2-day workshops that teach entrepreneurs the process for validating business ideas. They provided us with a network of mentors to ensure that the startups in our training followed what our second partner taught in class. 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 promoted our training program to nascent startups. We focused on these firms because they are neither established startups, whose past experience 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.[4] 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, so that we were providing management advice and training to firms and collecting performance data. In other words, they knew 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 or the control group.

Startups received 10 sessions of training at Bocconi University, Milan. Five sessions were frontal lectures lasting four hours, and five were one-hour sessions per startup with mentors for both treated and control firms.[5] As discussed in Section 2, the duration and content of the intervention was 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; to formulate falsifiable hypotheses; and to test them in a rigorous fashion, 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 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 they were trying to solve, to formulate falsifiable hypotheses, and to 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.

[4] The reader can infer how we advertised the training from our website: www.thestartuptraining.com

[5] We provide some pictures taken during the training sessions in Appendix Section C.

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. The same instructors taught the classes for both treatment and control groups. We ensured that each mentor followed three startups from the treated and three startups from the control group, and the instructors were randomly assigned to the startups. The Bocconi University research team coordinated the activities and ensured that the learning modules and mentoring activities conducted by the research partners 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. This separation 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 phone 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.

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, as "development" when they had begun to work on their product/service, as "pre-revenue" when the product/service was out in the market but the firm had yet to earn revenue, and as "startup" when it had earned revenue. As mentioned, we focused on early ventures, that is, on initiatives at the idea and development stages, because a scientific approach to entrepreneurial decision making is more difficult and costly 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 pre-revenue (16) and startup (22) phases. Given our resource constraints (instructors, mentors, research team, 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 CBInsights, a startup-dedicated database that reports European and American angel and venture capital investments in startups.[6] From the vitae of each startup team, we inferred its region.

We opted for pure randomization with balance tests, as 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 (B2C) rather than business-to-business (B2B), or whether they join 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 to allocate the 116 selected startups to the treatment and control groups was not obvious from a theoretical standpoint and 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.[7] Second, we ran similar OLS regressions of startup characteristics on a dummy for the allocation to the treatment or control group. We define all the variables used in the balance tests in Appendix Section D.

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, as 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 the experiment, 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.

[6] https://www.cbinsights.com/

[7] This is a sort of 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).

Table 1 reports some randomization checks. First, 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), main region of origin (Lombardy), and size of the founding team. Consistent with the validity of the randomization, none of these variables is 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 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 is significantly associated with selection into the treatment group, increasing our confidence in the robustness of the RCT design.

Table 1 approximately here

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 a year). On average their team members have completed college education, and more than half are 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 appears 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 research assistants. Overall we collected 16 observations 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 4-month training period, we collected data biweekly after each mentoring session (phone interviews took place within 3 days). After the training period we collected data monthly, but the last observation (16th data point) was collected 2 months after the 15th observation. The different frequencies are not an issue in our empirical analysis, as we employ 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.

Research assistants attended the entire training program themselves and underwent specific training on the research protocol, on how to conduct phone interviews to get the required data and, when necessary, on 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, in order 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 2 weeks. These narratives gave us grounds for evaluating the level of adoption of a scientific approach to decision making, as research assistants employed, as a coding scheme, the themes described in the theory and Inkdome case study sections. These themes are reported and summarized in Appendix Section B. Because the startups did not know 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 & Van Reenen, 2007). In part 2 of the phone interviews, we asked startups to report their performance, particularly their revenue.

All regressions are based on 1,612 observations. This is fewer observations than 116 × 16 = 1,856 because we exclude firms after they drop out. Table 2 reports the descriptive statistics for the key variables described below. Table 3 shows their correlations. During our time frame, 17 firms earn positive revenue (9 in the treatment and 8 in the control group), 44 drop out (24 in the treatment and 20 in the control group), and 30 pivot at least once to a main new idea (19 in the treatment and 11 in the control group.) Overall, 75 firms in our sample take one or more of these actions; 41 take no action. This is in line with expectations and suggests that the startups in our sample were not just formed and left inactive. If you include firms that received at least one e-mail from potential customers interested in the firm's product (a variable we do not use in our regressions), 93 of our 116 firms took one of these four actions. 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.

Table 2 and 3 approximately here

Dependent variable

Revenue. Our main dependent variable is the cumulated euro amount of firm revenue. The 17 firms with positive revenue in our sample correspond to 107 of our 1,612 observations: 85 from the 9 firms in the treatment group versus 22 from the 8 firms in the control group. We also checked whether our regression results depend largely on one outlier firm. All results are robust to the exclusion of any of the firms with non-zero revenues in the treatment group. Moreover, we run all our regressions using firm fixed effects, which implies that all our estimates are within-firm estimators over the longitudinal dimension of our sample. Finally, the average revenue for the 85 non-zero observations in the treatment group is about 31,000 euros; the 22 observations in the control group earn about 1,000 euros.

Dropout. This is a binary variable that takes value 0 until the firm drops out (they abandon the program and cease the startup), 1 in the time period in which the firm drops out, and a missing value thereafter. To avoid attrition biases, we checked that the entrepreneurs that informed us of their decision to discontinue their initiative truly abandoned their activity. Following our earlier discussion, all firms that dropped out from our sample had not yet made heavy investments in their company. Using our terminology in Section 3, they are genuine dropouts and not failures.

Pivot. This is the cumulative number of times that a startup made a major change to its business model. We defined a change to be 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. For example, a major change was Inkdome's decision to pivot from a search engine platform to one where users contact tattoo experts.

Independent variables

Intervention, postintervention, and cumulative_treatment. We employ three main independent variables in our analysis. *Intervention* is a dummy variable taking the value 1 for a firm in the treated group during all 16 periods in which we collected firm data, and 0 otherwise. *Postintervention* is a dummy taking the value 1 for all firms in the treated group after completing the treatment, and 0 for all firms in the control group and for the treated firms before completing the treatment. Because the training lasted for 8 of our 16 periods (with frequency every fortnight, approximately 4 months in total) and began right after we enrolled the firms in the program, postintervention takes the value 1 for the treated firms starting with time period 9 and ends in time period 16; it is 0 in the first 8 periods of the treated firms and for any observation belonging to the control group. *Cumulative_treatment* takes the value 0 for the control startups for the entire period, and is equal to the number of periods into the treatment for the treated startups. It is then equal to 1 in the first period, 2 in the second, and so on, until it takes the value 8 from the eighth period until the end of our training. We noted that startups learned how to use the approach progressively over the 8 periods of training rather than all upfront. Our estimates are robust to different functional forms of the dynamic treatment, for example logarithmic and quadratic.

Bloom et al. (2013) use the same three variables: a dummy equal to 1 for the treated group during the treatment, a dummy equal to 1 for the treated group after the treatment, and the same cumulative treatment variable that we use. Like them, we employ, alternatively, all three variables in our analysis and show that our results are robust to the various variables we use. Compared with Bloom et al. (2013), we do not have a diagnostic period in which we observe the firms and measure their data before the intervention. We called participants to a training initiative, and it would have been hard to keep them in the program, and to collect data, for a few months without giving them the training. However, as noted, we were careful to select firms that had an initial business idea but that had not begun any activity. We can fairly say that all these firms were at a baseline level, and that therefore any effect observed as they move into the program is *de facto* a difference-in-difference because we can set any variable regarding these firms before the intervention at a baseline level of 0, making the difference across firms before the intervention equal to 0. As we will see, the effects of both *intervention* and *postintervention* are meaningful, suggesting that we find an effect irrespective of whether we look at the interim period before the intervention ends or focus on the effect after the intervention.

Scientific_approach. We also measure the adoption of a scientific approach to decision making using a scale from 1 to 10, where 1 is the lowest and 10 the highest level. We code the content of the episodes narrated during the phone interviews. The phone interviews asked questions like "Can you narrate the most significant events that happened during the last two weeks?", "Can you tell what you spent most of your time on in the last 2 weeks?", "What were your main results?", "Did you change anything in your strategy?", and "If yes, why?" As described above, we assessed the adoption of a scientific approach based on whether and to what extent their narratives included specific references to the creation of a framework, formulation and testing of hypotheses, the setup of rigorous experiments, and evidence-based decision making. In addition to the intention-to-treat (ITT) regressions that employ *intervention*, *postintervention*, and *cumulative_treatment* as alternative regressors, we use *scientific_approach* as an endogenous regressor, identified alternatively by the three ITT instruments, to provide support for our mechanism. As shown in Table 3, the average levels of this variable for treatment and control groups across all 1,612 observations are 3.71 and 2.74. Interestingly, the difference is even more marked for the 85 non-zero revenue observations in the treatment group versus the 22 non-zero revenue observations in the control group, 4.65 versus 2.73 ($p < 0.01$).

7. Empirical results

In all our regressions we use all firms in all periods, removing firms after they drop out, and we employ time fixed effects. When we use *intervention* as the independent variable for our treatment, we cannot use firm fixed effects because *intervention* does not change over time and thus overlaps with the firm fixed effects. We employ firm fixed effects in all our regressions using *postintervention* and *cumulative_treatment*. In the regressions using *intervention* we include dummies for the mentors who worked with the companies in the one-to-one interviews. Companies were allocated randomly to mentors, and mentors attended, randomly, companies in the treatment or control group. Since mentors do not change over time, we do not need mentor dummies when we employ firm fixed effects. Interestingly, in all the regressions below, the mentor dummies, whenever we used them, are largely insignificant, suggesting that the mentors acted fairly homogenously. We also show our results using standard errors clustered by firms.

Figure 2 illustrates the average revenues for treated and control firms. The figure scales the time periods by actual length, that is, periods 9 through 15 is twice the length of periods 1 through 8 (4 vs. 2 weeks), and four times that between periods 15 and 16 (2 months). Table 4 reports our results using the three independent variables *intervention*, *postintervention*, and *cumulative_treatment*. As the table shows, the effect of the treatment is sizable. From Table 2, the average revenue in our sample is 1,649.7. The estimated impacts of our three variables in Table 4 are respectively 3,092.2, 5,520.2, and 7,2120.0 – where the latter effect is the estimated impact of cumulative_treatment (901.5) times 8, which is the final value of the cumulated variable. As a result, the estimated impacts of the treatment imply, respectively, an increase in revenue by 87%, 235%, and 437%. These impacts are big also because we begin from a basis of zero revenue. Nonetheless, they suggest that the estimated impact of the treatment is not negligible.

Figure 2 and Table 4 approximately here

Interestingly, all the estimated impacts when we use firm and time fixed effects without clustering standard error by firms show p-values smaller than 5% or even 1%. However, the standard errors increase when we cluster by firm. This is consistent with our story in that we predict that the treatment raises precision and, thus, increases revenue but also enables firms to recognize that they are on a bad track and therefore either exert little effort, pivot to a new idea, or drop out. This implies that, as time elapses, the wedge between high and low performers within the treatment group increases. The direct implication of this phenomenon is an increase in the standard error of the regression. However, the standard error of the regression increases the standard errors of the estimated coefficients, which is what we observe in Table 4.

The high-average/high-variance impact of the treatment is a natural outcome of our theory; therefore, we want to provide additional evidence for it. First, the variance of the impact of the treatment unfolds over time because there is a natural gestation period before some treated firms find good opportunities. Table 5 reports the same revenue regressions in Table 4 using data up to periods 10, 12, and 14. The standard error of the regression, and therefore the standard error of the treatment effect, ought to be smaller in these earlier periods. As Table 5 shows, the standard errors of the treatment are indeed smaller, and the p-values of the effect of the treatment are below 10% in most cases.

Table 5 approximately here

We also show more direct evidence that the variance of performance is higher for the treated firms, and that the increase is more pronounced over time. Fleming and Sorenson (2004) addressed the same problem by regressing the squared residuals of their main regression onto variables of interest. The first three columns of Table 6 report the differences in the means of the squared residuals obtained using *intervention*, *postintervention*, and *cumulative_treatment* as regressors in our Table 4. The differences are sizable and statistically significant for p-values smaller than 5% or even 1%. The last three columns of Table 6 check whether this increase in variance is more pronounced in later periods. We checked this effect in several ways (e.g. by interacting time dummies with any of the treatment effects), and they are all robust. In Table 6 we use *intervention* and *postintervention* as regressors and show that the significant difference between the means occurs later, in the postintervention period. As predicted by our theory, the treated firms appear to exhibit greater variance in performance, particularly later in time.

The greater variance in the performance of the treatment group is important for another reason. The effects of our treatment variables in the ITT regressions may stem from factors other than our hypothesized mechanism. We are confident that our RCT carefully gives the treated group greater ability to frame, define, validate, and test their business problem in a scientific way as opposed to other potential effects. For instance, as discussed, we gave both groups the same content and hours of training, and we made the classes for the control group as exciting, energetic, and informative as the classes for the treated group. At the same time, any other factor we can think of, other than our mechanism, would raise the average effect of the treatment but not its variance. For example, if we provided the treated group with greater excitement, energy, or content, we ought to observe an increase in average performance but not necessarily in the variance. Indeed, the increase in variance, as also documented below for the dropout rate, makes us confident that the treatment captures the proposed mechanism rather than other factors.

Table 6 approximately here

To provide further support for our mechanism, in Appendix Section E we report our results using *scientific_approach* as independent variable instrumented by, alternatively, *intervention*, *postintervention*, and *cumulative_treatment*. As noted, we already found a sizable and statistically significant difference between the averages of *scientific_approach* for treated and control firms, which is even more marked for the treated and control firms that earn some revenue. Appendix Section E shows that the estimated impacts of *scientific_approach* are sizable. For example, when *postintervention* is the independent variable, the impact of scientific_approach on revenue is 13,593.3 euros. For one standard deviation away from the mean (2.11, Table 2), this corresponds to an increase in revenue of nearly 30,000 euros, well above the 1,649.7 average revenue in our sample. Again, as shown in Appendix E, standard errors increase when we also cluster by firm, consistent with our theory, as discussed. Appendix E also shows the analog of Table 6: correlations between the squared errors of these instrumental variable regressions, on *intervention*, *postintervention*, and *cumulative_treatment*, taking also into account potential differences in the post-intervention period. Again, we find evidence that the treated firms exhibit greater variance, particularly in the later phase of the RCT.

Tables 7 and 8 report our results using *dropout* and *pivot* as dependent variables. Simple and convincing evidence that our treatment does not reduce dropout is that 24 firms in our treated group drop out versus 20 in our control group. Table 7 confirms that the treatment does not reduce the dropout rate for the treated firms. The estimated impacts of *intervention*, *postintervention*, and *cumulative_treatment* are positive and statistically insignificant. This evidence is consistent with our mechanism. To strengthen evidence in favor of our mechanism, in April 2017, when we collected our last set of results, we also asked all the firms that survived or just dropped out 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 feel in taking drastic decisions such as abandoning your startup?” Respondents answered on a

1-to-7 Likert scale, where 1 = not at all and 7 = very confident. The average score of treated firms was 4.4 and for the control group was 3.2 ($p < 0.01$).[8]

Tables 7 and 8 approximately here

Table 8 shows that, on average, treated startups pivot more than those in the control group. The results are robust to the use of all independent variables, *intervention*, *postintervention*, or *cumulative_treatment*. This is consistent with our theory. In addition, as discussed in Section 4, if the only effect of a scientific approach was to increase the ability of startups to draw ideas from better distributions, we ought to observe that startups pivot to a lesser extent after the initial pivot because they sit on better distributions in the subsequent steps. Of the 19 firms in the treated group that pivot at least once, five pivot a second time, one pivots a third time, and one pivots a fourth time; of the 11 firms in the control group that pivot at least once, only one pivots a second time. Treated firms do not appear to sit on better distributions after their first pivot. Moreover, the treated firms' higher propensity to pivot suggests that for these firms pivoting is a more valuable alternative, which offsets their higher propensity to drop out, and explains why we do not observe that a scientific approach produces, unambiguously, a higher dropout rate.

We provide some final overarching evidence of our theory by running a competing risk regression model. This model enables us to take into account the time sequence of events by checking at each point whether a given firm drops out, pivots, or begins to earn revenue. Thus, for each period, our dependent variable takes the value 0 if the firm performs no action, 1 if it drops out, 2 if it pivots, and 3 if it begins earning revenue. We discard all observations after the firm drops out or begins earning revenue. The reason for ignoring observations after dropout is straightforward; we ignore the data after the firm earns revenue to focus on the event in which the firm begins earning revenue. One firm earns revenue and after three periods drops out: we ignore the three interim observations, but we include both the period in which it begins earning revenue and the period in which it drops out. For comparison, we also show the results for revenue as the failure event when we include the observations after the firm has begun to

[8] We are not concerned about biases in this answer since, as we saw, dropout is not affected by the treatment.

earn revenue. We do not stop observations after a firm pivots, because it can pivot more than once; we set the dependent variable to 2 on the date of pivoting (whether the first or subsequent pivot) and 0 otherwise. No firm pivots and drops out or begins earning revenue on the same date. Our time dimension follows the chronological elapse of time with a period of 1 weeks as the unit of time: it takes values 1 through 8 in the first 8 fortnights, then monthly occurrences (10–22 for periods 9–15), and finally a bimonthly occurrence in the final period (26).

Table 9 reports odds ratios for the event in the column against the baseline event in which there is no action and the dependent variable takes the value 0. For each event in the column, the other two events represent competing events. The table's results are consistent with the results shown so far. The *intervention* does not have a significant effect on dropout but does have a significant effect on pivot. At each moment, treated firms are not more likely to drop out, but they are more likely to pivot. Treated firms are not more likely to begin earning revenue at each point, again in line with our story. The scientific method enables these firms to see both good and bad opportunities. Therefore, part of their greater performance depends on the fact that they do not start a business that is likely to be a false positive. As a result, some of our treated firms begin earning revenue while others wait because they have not yet found the right opportunity. Since we do not observe the future failures of the control firms that pursued the false positives, we do not have all the information that would enable us to observe the positive performance of all the treated firms – that is, of those that begin earning revenues, and of those that do not pursue false positives that eventually produce negative profits. Moreover, the fact that most entrepreneurial ideas are unprofitable suggests that many of our firms take advantage of this ability to predict false positives rather than that they have found a good idea to pursue. As a result, we can only expect that the likelihood that treated firms begin earning revenue is not pronounced. However, a pivot is an early sign that a firm recognizes a false positive and moves to a different idea, and the significant impact of our treatment on pivot, throughout our empirical analysis, provides robust evidence consistent with our theoretical mechanisms.

Table 9 approximately here

Moreover, the last column of Table 9 shows that when we include all observations in which the firms earn revenue, the probability that a treated firm earns revenue at any moment becomes sizable and significant. This suggests, once more, that not all treated firms earn revenue, but when they do so, earning revenue becomes a persistent event. This squares with the results in Table 4, where we find a high average impact of the treatment but also a high variance, and it is consistent with our interpretation of the impact of the scientific method. If the scientific method produced only learning, we should observe not a high variance, or that only some treated firms begin earning revenue at each date, but instead more homogenous patterns. A reason consistent with the heterogeneity that we observe across treated firms is that they produce bad and good ideas, and because they can recognize them, they are more likely to pursue the good ones and leave the bad ones behind. The control group, instead, has a fuzzier view of the potential of their ideas; it is less capable of screening them and therefore exhibits more homogeneous behaviors.

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, whose importance 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 when they analyze market signals (Shepherd et al., 2014; Hayward et al. 2006), reducing the likelihood of incurring false positives and false negatives.

The limitations of our paper raise natural questions for future research. We observed that, in spite of our heavy treatment, only 15% of the treated startups in our sample reached a score of 7 or more out of 10 on our scale measuring the adoption of the scientific approach. This raises, first, a question of whether we can improve our measurement of the adoption of a scientific approach. Our measure is based on codified answers to codified questions. Still, the codification could be more precise. In addition, while we observe that the treated startups use the method to a greater extent, the lack of high values in our scale suggests that some barriers exist. Making decisions according to the scientific approach requires rigorous thinking and disciplined behavior that might not come naturally to individuals outside the scientific world and that might be difficult to sustain over time. In this paper we have not explored these processes. Moreover, while we have produced evidence that a scientific approach provides predictive capability, we have not established whether it provides learning. If the approach only provided predictive capability, it should focus on decisions under uncertainty, whereas learning also makes it useful for decisions with no uncertainty. This is important to understanding the breadth of application of the method for practical entrepreneurial decisions. The time span of our RCT did not allow us to test whether some firms in the control group eventually fail and thus some firms in the treated group perform better because they fail faster without incurring high costs.

We have focused on a particular decision – profitability of the business idea – in which there are many false positives. However, the scientific approach can be applied to several decisions – from the set of decisions required to launch a new product or service (e.g., what customer problem to focus upon, what solution to offer, which marketing and product development strategy to follow) to decisions like employee selection or fundraising strategies. Some of these decisions may face mostly false negatives. For example, in a market with many potential bright collaborators, a scientific approach applied to employment decisions can help an entrepreneur hire individuals who would be false negatives if the entrepreneur's bias is toward hiring someone whom she knows or trusts based on gut feelings. As she faces mostly good candidates, the scientific approach enables her to find a good employee early in the hiring process rather than to pivot many times until she finds someone "she likes". Similarly, there are biases against novelty in science (Stephan et al., 2017), which may well extend to larger firms that often do not pursue projects that do not conform to their expertise and domain (Gambardella et al., 2015). On the theory side, we addressed very simple firms, and even slightly more complex organizations make many decisions simultaneously. This raises questions about how to handle correlations among signals – particularly, how higher- and lower-level decisions concur about whether to pivot, dropout, or continue with a project, or how the signal on a project influences decisions about parallel projects. Again, we need a full understanding of these issues to offer a thorough and valuable framework for practitioners that differentiates behavioral prescriptions depending on the type of decision. Moreover, as this discussion suggests, a scientific approach can help larger firms make decisions, but we have not provided any clues about how this would play out within their complex organizations.

REFERENCES

Adner, R. and Levinthal, D. (2004) "What Is Not a Real Option: Considering Boundaries for the Application of Real Options to Business Strategy," *Academy of Management Review*, 29 (1), 74–85.

Arora, A. and Gambardella, A. (1994) "Evaluating Technological Information and Utilizing it," *Journal of Economic Behavior and Organization*, 24, 91-114.

Aspelund, A., Berg-Utby, T. and Skjevdal, R. (2005) "Initial Resources' Influence on New Venture Survival: A Longitudinal Study of New Technology-Based Firms," *Technovation*, 25 (11), 1337-1347.

Bennis, W. G. (1962) "Towards a 'Truly' Scientific Management: The Concept of Organization Health", *MIT Sloan Management Review* (formerly *Industrial Management Review*), 4.1, Fall: 1-19.

Blank S.G., (2006) *The Four Steps to the Epiphany*, K&S Ranch, 2nd edition.Briner, R. B., Denyer, D. and Rousseau, D. M. (2009). "Evidence-Based Management: Concept Cleanup Time?," *The Academy of Management Perspectives*, 23 (4), 19-32.

Bloom N. and Van Reenen J., (2007) "Measuring and 'Explaining' Management Practices Across Firms and Countries", *The Quarterly Journal of Economics*, 22 (4): 1351-1408.

Bloom N., Eifert B., Mahajan A., Mckenzie D., and Roberts J. (2013) "Does Management Matter? Evidence from India", *The Quarterly Journal of Economics* 128 (1): 1-51.

Brynjolfsson, Erik and Kristina McElheran (2016) "The Rapid Adoption of Data-Driven Decision-Making," American Economic Review 106 (5): 133-139.

Brown, T. (2008) "Design Thinking," *Harvard Business Review*, 86 (6), 84-92.

Clark, T. and Wiesenfeld, D. (2017) "3 Things are Holding Back Your Analytics, and Technology Isn't One of Them," *Harvard Business Review*, June.

Drucker, P. F. (1955) "Management Science and the Manager," *Management Science*, 1 (2), 115-126.

Eisenmann, T., Ries, E. and Dillard, S. (2013) "Hypothesis-driven Entrepreneurship: the Lean Startup," *Harvard Business School Note* 9-812-095

Fairlie, R. and Miranda, J. (2017) "Taking the Leap: The Determinants of Entrepreneurs Hiring their First Employee," *Journal of Economics and Management Strategy*, 26 (1), 3-34. (See also NBER WP #22428.)

Felin, T. and Zenger, T.R. (2009) "Entrepreneurs As Theorists: On the Origins of Collective Beliefs and Novel Strategies," *Strategic Entrepreneurship Journal*, 3, 127–146.

Felin, T. and Zenger, T.R. (2015) "CROSSROADS—Strategy, Problems, and a Theory for the Firm," *Organization Science*, 27 (1), 222-231.

Fleming, L. and Sorenson, O. (2004) "Science as a Map in Technological Search," *Strategic Management Journal*, 25, 909-928.

Freedman, D. H. (1992) "Is Management Still a Science?," *Harvard Business Review*, 70(6), 26.

Gambardella, A., Ganco, M. and F. Honorè (2015) "Using What You Know: Patented Knowledge and Incumbent Fims in Employee Entrepreneurship," *Organization Science*, 26 (2), 456-474.

Gans, J., Stern, S. and Wu, J. (2017) *Foundations of Entrepreneurial Strategy,* available at SSRN: https://ssrn.com/abstract=2844843 or http://dx.doi.org/10.2139/ssrn.2844843.
Gelber, A., Isen, A., Kessler, and Judd B. (2016) *The Effects of Youth Employment: Evidence from New York City Lotteries*, NBER Working Paper No. 20810.
Gimmon, E. and Levie, J. (2010). "Founder's Human Capital, External Investment, and the Survival of New High-Technology Ventures", *Research Policy*, 39 (9), 1214-1226.
Grandori, A. (2010) "A Rational Heuristic Model of Economic Decision-Making," *Rationality and Society* 22 (4): 1–28.
Hansen B.B. and Bowers, J., (2008). "Covariate Balance in Simple, Stratified and Clustered Comparative Studies", *Statistical Science*, 23(2), 219-236.
Hayward M.L.A., Shepherd D.A., and Griffin D., (2006). "A Hubris Theory of Entrepreneurship", *Management Science*, 52(2), 160-172.
Hsieh, C., Nickerson, J.A. and Zenger, T.R. (2007) "Opportunity Discovery, Problem Solving and a Theory of the Entrepreneurial Firm," *Journal of Management Studies*, 44(7), 1255-1277.
Kelley, T. and Littman, J. (2005). The Ten Faces of Innovation: IDEO's Strategies for Beating the Devil's Advocate & Driving Creativity Throughout Your Organization, New York: Currency/Doubleday.
Kerr, W.R., Nanda, R. and Rhodes-Kropf, M. (2014) "Entrepreneurship as Experimentation", *Journal of Economic Perspectives*, 28 (3), 25–48.
Korunka, C., Frank, H., Lueger, M., and Mugler, J. (2003). "The Entrepreneurial Personality in the Context of Resources, Environment, and the Startup Process – A Configurational Approach". *Entrepreneurship Theory and Practice*, 28 (1), 23-42.
Li, Y., James, B., Madhavan, R., and Mahoney J.T. (2007) "Real Options: Taking Stock and Looking Ahead", *Advances in Strategic Management*, 24, 31–66.
Lopez-Vega, H., Tell, F. and Vanhaverbeke, W. (2016) "Where and How to Search? Search Paths in Open Innovation." *Research Policy*, 45 (1), 125-136.
Mahoney, J.T. (2005) *Economic Foundations of Strategy*, SAGE Publications, Thousand Oaks, CA.
Martin, R. (2009) *The Design of Business: Why Design Thinking is the Next Competitive Advantage*, Harvard Business School Press, Boston MA.
McGrath, R. G. (1997) "A Real Options Logic for Initiating Technology Positioning Investments," *Academy of Management Review*, 22 (4), 974–996.
McGrath, R.G. (1999) "Falling Forward: Real Options Reasoning and Entrepreneurial Failure," *Academy of Management Review*, 24 (1), 13-30.
McGrath, R. G. and MacMillan, I. C. (1995) "Discovery Driven Planning," *Harvard Business Review*, 73 (4), 44-54.
McGrath, R. G. and MacMillan, I. C. (2009) *Discovery Driven Growth: A Breakthrough Process to Reduce Risk and Seize Opportunity*, Harvard Business Publishing, Boston MA.
Mitchell R.K., Busenitz L., Lant T., McDougall P.P., Morse E.A., and Smith J.B., (2002). "Toward a Theory of Entrepreneurial Cognition: Rethinking the People Side of Entrepreneurship Research". *Entrepreneurship Theory and Practice*, 27 (2), 93–104.
Nickerson, J.A., Yen, C.J and Mahoney, J.T. (2011), "Exploring the Problem Finding and Problem Solving Approach for Designing Organizations," *Academy of Management Perspectives*, 27, 52-72.
Nicholls-Nixon, C., Cooper, A. and Woo, C. (2000), "Strategic Experimentation: Understanding Change and Performance in New Ventures," *Journal of Business Venturing*, Vol. 15(5-6), 493-521.
Osterwalder, A. and Pigneur, Y (2009). *Business Model Generation: A Handbook for Visionaries, Game Changers, and Challengers*, Self-Published.
O'Brien, J.P., Folta, T.B. and Johnson, D.R. (2003) "On Entrepreneurial Entry in the Face of Uncertainty," *Managerial and Decision Economics*, 24, 515–533.
Pfeffer, J. and Sutton, R.I. (2000) *The Knowing-Doing Gap. How Smart Companies Turn Knowledge into Action*, Harvard Business School Press, Boston.
Pfeffer, J. and Sutton, R. I. (2006) *Hard Facts, Dangerous Half-truths, and Total Nonsense: Profiting from Evidence-Based Management*, Harvard Business Press, Boston MA

Ries E., (2011). *The Lean Startup: How Today's Entrepreneurs Use Continuous Innovation to Create Radically Successful Businesses*, Crown Pub edition.
Rousseau, D. M. (2006) "Is There Such A Thing As 'Evidence-Based Management'?," *Academy of Management Review*, 31 (2), 256-269.
Sarasvathy, S.D. (2001) "Causation and Effectuation: Toward a Theoretical Shift from Economic Inevitability to Entrepreneurial Contingency", *Academy of Management Review*, 26 (2), 243–263.
Sarasvathy, S.D. and Venkataraman, S. (2011) "Entrepreneurship as Method: Open Questions for an Entrepreneurial Future," *Entrepreneurship Theory and Practice*, 35 (1), 113-135.
Shepherd, D.A., Haynie, J.M. and McMullen, J.S. (2012) "Confirmatory Search as a Useful Heuristic: Testing the Veracity of Entrepreneurial Conjectures." *Journal of Business Venturing* 27(6): 637-651.
Shepherd, D.A., Williams, T.A. and Patzelt, H. (2014) "Thinking about Entrepreneurial Decision Making: Review and Research Agenda." *Journal of Management* 41(1): 11-46.
Stephan, P., Veugelers, R. and Wang, J. (2017) "Reviewers Are Blinkered by Bibliometrics," *Nature*, 544, 411-412.
Sull, D.N. (2004) "Disciplined Entrepreneurship," *MIT Sloan Management Review*, 46 (1), 71.
Taylor, F.W. (1914). *The Principles of Scientific Management*. Harper.
Venkataraman, S., Sarasvathy, S.D., Dew, N. and Forster, W.R. (2012) "Reflections on the 2010 AMR Decade Award: Whither the Promise? Moving Forward with Entrepreneurship as a Science of the Artificial," *Academy of Management Review*, 37 (1), 21-33.
Von Hippel, E. (1986). "Lead Users: A Source of Novel Product Concepts", *Management Science*, 32 (7), 791-805.
Zenger T.R. (2015) "Trial and Error Is No Way to Make Strategy," *Harvard Business Review*, April 24, 2015.
Zenger T.R. (2016). *Beyond Competitive Advantage: How to Solve the Puzzle of Sustaining Growth While Creating Value*, Harvard Business School Press, Boston MA.

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

OLS regressions using variables as the dependent variable and dummies for selected/non-selected or treatment/control as regressors; coefficients are differences between means.

Table 2: Descriptive Statistics

VARIABLES	Mean	Sd	min	max	Treatment Mean	Treatment sd	Control Mean	Control sd	Diff p-value
Revenue	1649.7	16924.7	0	437474.5	3278.0	23860.6	29.4	227.8	0.000
Intervention	0.499	0.500	0	1	1	0	0	0	n/a
Postintervention	0.220	0.414	0	1	0.440	0.497	0	0	n/a
Cumulative_treatment	2.980	3.461	0	8	5.975	2.472	0	0	n/a
Scientific_approach	3.224	2.116	1	10	3.711	2.318	2.739	1.766	0.000
Dropout	0.027	0.163	0	1	0.030	0.170	.025	.155	0.530
Pivot	0.203	0.525	0	4	0.272	0.648	.134	.351	0.000

N. of obs. (total) = 1,612; N. of obs. (treated) = 804; N. of obs. (control) = 808.

Table 3: Correlations

VARIABLES	Revenue	Intervention	Postintervention	Cumulative_ treatment	Scientific_ Experimentation	Dropout	Pivot
Revenue	1						
Intervention	0.096***	1					
Postintervention	0.153***	0.532***	1				
Cumulative_treatment	0.133***	0.864***	0.770***	1			
Scientific_approach	0.058*	0.230***	0.200***	0.293***	1		
Dropout	-0.016	0.016	0.049*	0.043	-0.062*	1	
Pivot	-0.036	0.132***	0.183***	0.209***	0.277***	0.044	1

N. of obs. = 1,612.

Table 4: Performance Regression, Dependent variable = Revenue

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)
Intervention	3092.2**			3092.2**		
	(0.047)			(0.046)		
Postintervention		5520.2***			5520.2	
		(0.000)			(0.151)	
cumulative_treatment			901.5***			901.5
			(0.003)			(0.116)
Constant	-2934.5	75.5	-362.2	-2934.5*	75.5	-362.2
	(0.424)	(0.955)	(0.789)	(0.071)	(0.934)	(0.761)
Observations	1612	1612	1612	1612	1612	1612
R-squared	0.021	0.030	0.026	0.021	0.030	0.026
Number of id	116	116	116	116	116	116
Dummies for mentors	Yes	No	No	Yes	No	No
Time FE	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	No	Yes	Yes	No	Yes	Yes
Clustered Errors by Firms	No	No	No	Yes	Yes	Yes

OLS regression. P-value in parentheses, *** p<0.01, ** p<0.05, * p<0.1. In (1) and (4) *intervention* implies that we cannot use of firm FE. In (2), (3), (5), (6) firm FE implies that we cannot use dummies for mentors

Table 5: Performance Regression, Dependent variable = Revenue, different periods

VARIABLES	Up to Period 10	Up to Period 10	Up to Period 10	Up to Period 12	Up to Period 12	Up to Period 12	Up to Period 14	Up to Period 14	Up to Period 14
Intervention	908.6*			1233.7**			2007.4**		
	(0.091)			(0.044)			(0.025)		
postintervention		1094.9*			1788.7**			3461.7	
		(0.062)			(0.047)			(0.107)	
cumulative_ treatment			247.6*			339.9*			579.9*
			(0.090)			(0.051)			(0.072)
Constant	-923.1	29.3	-95.9	-1264.9*	37.1	-134.7	-2001.7**	53.2	-234.1
	(0.118)	(0.919)	(0.790)	(0.068)	(0.915)	(0.754)	(0.045)	(0.920)	(0.733)
Observations	1089	1089	1089	1276	1276	1276	1447	1447	1447
R-squared	0.027	0.038	0.051	0.027	0.042	0.043	0.022	0.032	0.029
Number of id	116	116	116	116	116	116	116	116	116
Dummies for mentors	Yes	No	No	Yes	No	No	Yes	No	No
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Clustered Errors by Firms	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

OLS regression. P-value in parentheses, *** p<0.01, ** p<0.05, * p<0.1. In (1) and (4) *intervention* implies that we cannot use of firm FE. In (2), (3), (5), (6) firm FE implies that we cannot use dummies for mentors

Table 6: Variance of Performance, Dependent variable = squared residuals of the regressions in Table 4

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)
Intervention	377.7**			90.5	127.6	127.9
	(0.044)			(0.682)	(0.531)	(0.532)
Postintervention		642.7***		652.2**	560.8**	560.3**
		(0.002)		(0.014)	(0.023)	(0.023)
cumulative_treatment			67.2***			
			(0.007)			
Constant	5.1	48.0	-10.5	5.1	2.4	2.9
	(0.969)	(0.623)	(0.927)	(0.969)	(0.984)	(0.981)
Observations	1612	1612	1612	1612	1612	1612
R-squared	0.003	0.006	0.004	0.006	0.006	0.006

OLS regression. P-value in parentheses, *** p<0.01, ** p<0.05, * p<0.1. Values in 10^6. In Table 4, (1) & (4), (2) & (5), (3) & (6) generate the same residuals. In this table, they correspond, respectively, to columns (1) & (4), (2) & (5), (3) & (6).

Table 7: Dropout Regression, Dependent variable = Dropout

VARIABLES	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Intervention	0.003			0.003		
	(0.704)			(0.703)		
Postintervention		0.019			0.019	
		(0.246)			(0.258)	
cumulative_treatment			0.002			0.002
			(0.601)			(0.611)
Constant	-0.008	-0.020	-0.021	-0.008	-0.020**	-0.021**
	(0.721)	(0.173)	(0.157)	(0.592)	(0.011)	(0.010)
Observations	1612	1612	1612	1612	1612	1612
R-squared	0.055	0.062	0.061	0.055	0.062	0.061
Number of id	116	116	116	116	116	116
Dummies for mentors	Yes	No	No	Yes	No	No
Time FE	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	No	Yes	Yes	No	Yes	Yes
Clustered Errors by Firms	No	No	No	Yes	Yes	Yes

OLS regression. P-value in parentheses, *** p<0.01, ** p<0.05, * p<0.1. In (1) and (4) *intervention* implies that we cannot use of firm FE. In (2), (3), (5), (6) firm FE implies that we cannot use dummies for mentors

Table 8: Pivot Regression, Dependent variable = Pivot

VARIABLES	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
intervention	0.149*			0.149*		
	(0.071)			(0.057)		
postintervention		0.159***			0.159**	
		(0.000)			(0.030)	
cumulative_treatment			0.043***			0.043**
			(0.000)			(0.020)
Constant	0.133	-0.002	-0.023	0.133	-0.002	-0.023
	(0.474)	(0.927)	(0.379)	(0.702)	(0.956)	(0.642)
Observations	1612	1612	1612	1612	1612	1612
R-squared	0.131	0.148	0.162	0.131	0.148	0.162
Number of id	116	116	116	116	116	116
Dummies for mentors	Yes	No	No	Yes	No	No
Time FE	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	No	Yes	Yes	No	Yes	Yes
Clustered Errors by Firms	No	No	No	Yes	Yes	Yes

OLS regression. P-value in parentheses, *** p<0.01, ** p<0.05, * p<0.1. In (1) and (4) *intervention* implies that we cannot use of firm FE. In (2), (3), (5), (6) firm FE implies that we cannot use dummies for mentors

Table 9: Competing Risk Analysis of Dropout, Pivot, Revenue

VARIABLES	Event type = dropout	Event type = pivot	Event type = revenue	Event type = revenue (all obs.)
Intervention	1.21	2.74***	1.22	3.95**
	(0.552)	(0.008)	(0.684)	(0.10)
Observations	1522	1522	1522	1612
# events	42	38	17	107
# competing events	55	59	80	80

Competing risk regressions. P-value in parentheses, *** p<0.01, ** p<0.05, * p<0.1. In all regressions errors are clustered by firms. Event types: 0 = censored; 1 = dropout; 2 = pivot; 3 = revenue. Each column reports the odd ratio of the corresponding event at each moment in time taking into account the other two competing events. Odd ratios higher than 1 imply that for the treated firms the event is relatively more likely. In parenthesis p-values of differences from 1. In the first three columns observations exclude both firms from the period after they drop out and firms from the period after they start earning revenue. In the last columns observations include periods after the firm starts earning revenue.

Figure 1: Training program and differences between treated and control startups

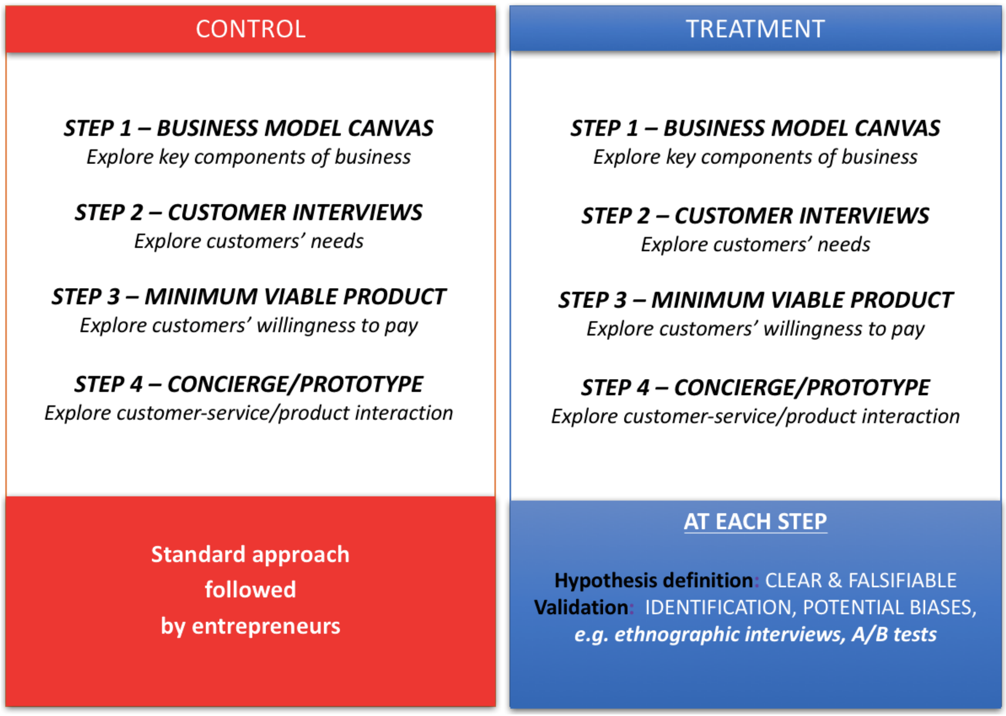
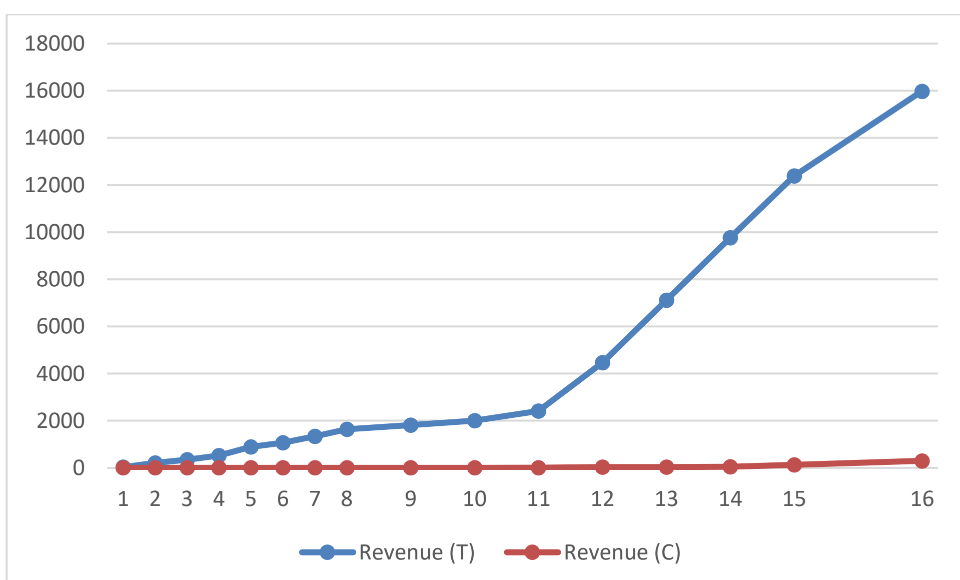


Figure 2: Average revenue over time (euros), treated and control startups



Time line legend: 16 periods corresponding to actual time gaps (2 weeks for periods 1-8, 4 weeks for periods 8-15, 8 weeks for periods 15-16)

APPENDIX

Section A: Content of training steps

Training	Control	Treatment
STEP 1 ***BUSINESS MODEL CANVAS*** *Explore key components of business*	1. Don't recognize BMC as overarching theory 2. Don't see individual blocks as representing hypotheses to validate 3. Don't 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. Don't 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 1. Interview potential customers 2. Ask open-ended questions 3. Use thresholds to falsify hypotheses
STEP 3 ***MINIMUM VIABLE PRODUCT*** *Explore customers' willingness to pay*	1. Don't define key hypotheses 2. Poor identification strategy • Don't try parallel variations of the product/service to evaluate improvement • Change more than 1 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 1. A/B tests 2. Change only 1 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. Don't define key hypotheses 2. Poor identification strategy • Use available resources to deliver the product/service • Focus on very short-term mesure 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 used at regime • Focus on longer-term measure of success 3. Use thresholds to direct decision making

Section B: 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's a nice article on screener questions from Alexander Cowan.

      1. What's 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 don't you love about the solutions you've tried?

   d. Make an interview guide (not a write-and-strictly-follow script). If you don't 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 afterwards or check on the tools to record your interview (check first 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/her vision in case the interviewees vision differs drastically. The interviewee is tempted in his/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 you get. So try to run questions in different order in different interviews.

Section C: Classes & mentoring

Section D: Definition of variables used in balance tests

VARIABLES	Measurement	Datasource
Idea stage	takes value 1 if the startup has only a business idea in mind, takes value 0 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 1 if the startup operates in the internet sector, i.e. provides a service which can be "consumed"online from a computer; takes value 0 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 1 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 0 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 1 if the startup operates in the retail sector, i.e. sells a product that is either commercialized via a physical shop or the larget commercial distribution; takes value 0 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 1 if the majority of team members come from the italian region of Lombardy; takes value 0 otherwise	Team members' CV: retrieved from city of domicile
Team size	it is the absolute number of team members of the startup	Team members' CV: we count the number of CVs 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 prior to entering the training	Project Pitch & Team members' CV: we match the SIC codes (at the 83 2-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 CV
Management experience	it is the average number of years of managerial experience of the team prior to entering the training	Team members' CV: we look at the years each team member had in a managerial job position as described in their CV. The count includes both higher and lower levels 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 the team members intend to launch prior to entering the training	Team members' CV: we look at the years 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 CV)
Experience working in startups	it is the average number of years of experience of the team working within startups other than the one the team members intend to launch prior to entering the training	Team members' CV: we look at the years each team member had as either mentor and/or consultant to a startup (this should have been defined so by the team member itself in the CV)
Currently employed	it is the proportion of team members employed at the time of entry into the training	Team members' CV: we record a team member as currently employed if any of his/her job positions described in the CV does 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' CV: we record a team member as currently studying if any of his/her enrollments in an educational program described in the CV does 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' CV: we look at the educational titles achieved by each team member and we record them as following: 1 is for high school, 2 for bachelor, 3 for master, 4 for MBA and 5 for PhD.

Section E: IV Regression

Table E1: Performance Regression (IV), Dependent variable = Revenue

VARIABLES	IV= Intervention (1)	IV= Postintervention (2)	IV= cumulative_ treatment (3)	IV= Intervention (4)	IV= Postintervention (5)	IV= cumulative _ treatment (6)
scientific_approach	3408.9	13593.3**	9970.3**	3409.2*	13593.3	9970.3
	(0.104)	(0.019)	(0.026)	(0.072)	(0.334)	(0.266)
Constant	-7335.9	-28066.8**	-20569.5**	-7335.4	-28066.8	-20569.5
	(0.192)	(0.021)	(0.030)	(0.112)	(0.344)	(0.286)
Observations	1612	1612	1612	1612	1612	1612
Number of id	116	116	116	116	116	116
Dummies for mentors	Yes	No	No	Yes	No	No
Time FE	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	No	Yes	Yes	No	Yes	Yes
Clustered Errors by Firms	No	No	No	Yes	Yes	Yes

IV regression. P-value in parentheses, *** p<0.01, ** p<0.05, * p<0.1. In (1) and (4) *intervention* implies that we cannot use of firm FE. In (2), (3), (5), (6) firm FE implies that we cannot use dummies for mentors

Table E2: First Stage Regression, Dependent variable = Scientific_approach

VARIABLES						
Intervention	0.880***			0.880***		
	(0.004)			(0.002)		
postintervention		0.406***			0.406	
		(0.002)			(0.190)	
cumulative_treatment			0.0904***			0.0904
			(0.001)			(0.113)
Constant	1.332*	2.070***	2.027***	1.332*	2.070***	2.027***
	(0.054)	(0.000)	(0.000)	(0.084)	(0.000)	(0.000)
Observations	1612	1612	1612	1612	1612	1612
R-squared	0.144	0.149	0.150	0.144	0.149	0.150
Number of id	116	116	116	116	116	116
Dummies for mentors	Yes	No	No	Yes	No	No
Time FE	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	No	Yes	Yes	No	Yes	Yes
Clustered Errors by Firms	No	No	No	Yes	Yes	Yes

P-value in parentheses, *** p<0.01, ** p<0.05, * p<0.1.

Table E3: Variance of Performance, Dependent variable = squared residuals of the regressions in Table E1

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)
intervention	380.9**			128.3	123.6	126.1
	(0.031)			(0.537)	(0.567)	(0.553)
postintervention		735.6***		573.8**	656.2**	619.2**
		(0.001)		(0.022)	(0.012)	(0.016)
cumulative_treatment			70.7***			
			(0.007)			
Constant	19.0	304.2***	129.0	19.0	260.0**	140.6
	(0.879)	(0.003)	(0.278)	(0.879)	(0.044)	(0.268)
Observations	1612	1612	1612	1612	1612	1612
R-squared	0.003	0.007	0.005	0.006	0.007	0.007

OLS regression. P-value in parentheses, *** p<0.01, ** p<0.05, * p<0.1. Values in 10^6. In Table E1, columns (1) & (4), (2) & (5), (3) & (6) generate the same residuals. In this table, they correspond, respectively, to columns (1) & (4), (2) & (5), (3) & (6).