PhD THESIS DECLARATION

I, the undersigned

FAMILY NAME	Cordova
NAME	Alessandro
Student ID no.	1317783

Thesis title:

A SCIENTIFIC APPROACH TO ENTREPRENEURIAL DECISION MAKING

PhD in	BUSINESS ADMINISTRATION AND MANAGEMENT
Cycle	29
Student's Advisor	CAMUFFO ARNALDO
Calendar year of thesis defence	2018

DECLARE

under my responsibility:

1) that, according to Italian Republic Presidential Decree no. 445, 28th December 2000, mendacious declarations, falsifying records and the use of false records are punishable under the Italian penal code and related special laws. Should any of the above prove true, all benefits included in this declaration and those of the temporary "embargo" are automatically forfeited from the beginning;

2) that the University has the obligation, according to art. 6, par. 11, Ministerial Decree no. 224, 30th April 1999, to keep a copy of the thesis on deposit at the "Bilioteche Nazionali Centrali" (Italian National Libraries) in Rome and Florence, where consultation will be permitted, unless there is a temporary "embargo" protecting the rights of external bodies and the industrial/commercial exploitation of the thesis;

3) that the Bocconi Library will file the thesis in its "Archivio Istituzionale ad Accesso Aperto" (Institutional Registry) which permits online consultation of the complete text (except in cases of temporary "embargo");

4) that, in order to file the thesis at the Bocconi Library, the University requires that the thesis be submitted online by the student in unalterable format to

Società NORMADEC (acting on behalf of the University), and that NORMADEC will indicate in each footnote the following information:

- PhD thesis: A SCIENTIFIC APPROACH TO ENTREPRENEURIAL DECISION MAKING;
- by CORDOVA ALESSANDRO;
- defended at Università Commerciale "Luigi Bocconi" – Milano in the year 2018;
- the thesis is protected by the regulations governing copyright (Italian law no. 633, 22nd April 1941 and subsequent modifications). The exception is the right of Università Commerciale "Luigi Bocconi" to reproduce the same, quoting the source, for research and teaching purposes;

- **only when a separate "Embargo" Request has been undersigned**: the thesis is subject to "embargo" for (indicate duration of the "embargo") .... months;

5) that the copy of the thesis submitted online to Normadec is identical to the copies handed in/sent to the members of the Thesis Board and to any other paper or digital copy deposited at the University offices, and, as a consequence, the University is absolved from any responsibility regarding errors, inaccuracy or omissions in the contents of the thesis;

6) that the contents and organization of the thesis is an original work carried out by the undersigned and does not in any way compromise the rights of third parties (Italian law, no. 633, 22nd April 1941 and subsequent integrations and modifications), including those regarding security of personal details; therefore the University is in any case absolved from any responsibility whatsoever, civil, administrative or penal, and shall be exempt from any requests or claims from third parties;

7) **choose 7a or 7b**:

7a) that the thesis is not subject to "embargo", i.e. that it is not the result of work included in the regulations governing industrial property; it was not written as part of a project financed by public or private bodies with restrictions on the diffusion of the results; is not subject to patent or protection registrations.

or

7b) that the PhD thesis is subject to "embargo" as per the separate undersigned "PhD Thesis Temporary "Embargo" Request".

Date 10/03/2018

A SCIENTIFIC APPROACH TO ENTREPRENEURIAL DECISION MAKING

Alessandro Cordova

ABSTRACT

Entrepreneurs create and develop new products and services in uncertain environments. In such a context of imperfect or incomplete information, where foreseeing the expected return of a given strategic action is inherently difficult, it often occurs that entrepreneurs fall trap of false positive or negative judgements. This is consistent with the high percentage of startup failures we see in our economies around the globe. Recent academic efforts have therefore concentrated on elaborating new decision-making process that help entrepreneurs make better decisions. The predominant new view prescribes a reduction in the use of planned-driven approaches to decision making that see in ex-ante commitment of resources and strategic planning the best way to achieve entrepreneurial success. Rather, entrepreneurs should be more experimental in their approach. Experiments allow to produce signals as to the expected return from following a given strategic path, in so doing minimizing the probability of wrong decisions and, eventually entrepreneurial failure. However, little discussion has been made as to how experiments should actually be conducted, which methods shall guide the design, execution and evaluation of experiments. In these pages, I advance the concept of *scientific* approach to entrepreneurial decision-making and argue that entrepreneurs who behave like scientists and elaborate and rigorously test well-articulated theories can produce more "precise" signals as to the expected return of their strategic actions. Consequently, they achieve higher economic performance. Overall this thesis aims at advancing practice guidelines for how new ventures can success under conditions of environmental uncertainty.

GENERAL INTRODUCTION

We already knew that a great majority of entrepreneurial projects failed, but what we did not know is what has been recently revealed by the international organization CBInsights: the top two reasons why failure happens are (i) there is no market need for the new entrepreneurial projects we launch, (ii) we deplete resources before a sustainable business model is identified. Why?

The environment in which entrepreneurs and managers interested in launching new products or services are called upon acting has changed. Increasing customer heterogeneity and competitor competition, fostered by globalization, and the faster rates of technological innovation, have spurred increasing uncertainty on which are the best strategies to follow for ensuring success. In turn, this has led practitioners to drive away from consolidated entrepreneurial practices, such as planning, and to seek new practices that allow to better cope with the increasing environmental uncertainty, so far with limited success. In fact, evidence seems to suggest business decision makers often pursue ideas with low expected returns but only realize it later on, when it is too late to change. This is where we are at today. We are looking for approaches that allow to reduce environmental uncertainty, help entrepreneurs and managers have more information regarding the expected return of their projects and consequently make better decisions.

A new wave of research has raised up to the challenge and has suggested that we move towards increasing experimentation and staggered investments. Streams of research such as real option theory, effectuation, disciplined entrepreneurship, have all explained why experimentation can benefit decision-makers well: if the ultimate goal is to reduce uncertainty, experiments can help generate, collect and analyze market signals about the expected return of a project. Early interviews to potential customers and minimum viable products are examples of such experiments.

In spite of this growing awareness, my sense is that we do not yet know how experimentation shall be conducted, what are the practices, rules and procedures that can assure entrepreneurs learn effectively. Our implicit assumption has been that it is just enough to experiment to reveal information about the future states of the world. That is not true. It is enough to walk one of the many startup events to see how entrepreneurs often lack an actual method to evaluate the value of their ideas. In literally all cases I have attended, for example, startups showed to the audience pie charts reporting the percentage of people that were interested in their product or service: it was consistently above 80%. We know this is inconsistent with the high rates of startup failure we observe every year. Many entrepreneurs struggle from the lack of a disciplined approach to experimentation: in interviews, for example, they often ask direct questions which do not allow to truly explore key customer problems and desires, and they often ask the feedback of family or friends who, perhaps unknowingly, contribute to foster confirmatory biases. Seemingly, entrepreneurs make changes to their products or to their marketing strategy evaluating their efficacy by comparing pre and post-performance, without taking into account the possibility that time fixed effects could have a role in explaining the difference in outcome. We know that a proper A/B test, instead, where a new product feature is provided first to a randomly selected segment of the customer base would minimize these biases.

Over time, observance of and reflection of these events rang a bell. Could the approach that we follow in academia, that is a rigorous process of theorizing, hypothesizing and testing, be applied to the entrepreneurial context?

In the pages that compose my thesis, I advance the concept of a *scientific* approach to entrepreneurial decision-making. Scientists develop theories, turn them in testable hypotheses and subject them to empirical test. In turn, this fosters learning. This same approach can be applied in the business practice. Entrepreneurs and managers who know how to craft precise, detailed, comprehensive, evidence-based theories of why their projects should be successful, who spend time in unearthing the assumptions that support their theory, that test their hypotheses by following rigorous procedures such as by using representative samples of customers, designing externally and internally valid tests, and that collect data in a systematic way, can be more precise in assessing the potential return of their ideas and, consequently, be more likely to avoid false positive and false negatives. If this is true, *scientific* decision-makers shall take better decisions and obtain better outcomes and, overall, contribute to a better allocation of resources at economy-wide level. This has the ambition that has driven the research efforts of my PhD days and whose considerations and contributions I share in these pages.

The sequence of papers that I hereafter present to you, retrace the same journey that I followed in discovering more about the concept of the scientific approach and its implications for decision makers. The first paper of the thesis is where I started off and where most of my energies have been spent on. I set up a field experiment that involved more than one-hundred early-stage startups which were offered, in my home university, a 4-month general training on business experimentation, and where one of two randomly divided groups was taught how to use a scientific approach to experimentation and decision making. My goal was to study the performance effect of using a scientific approach in the context of launching a new venture. Results have been enlightening and together with my two co-authors, Arnaldo Camuffo and Alfonso Gambardella, we did not only discover that the scientific approach enhances startup performance but we were also able to propose the mechanisms that explained this effect. At the time, I had a rough idea of how what will later be a more defined and comprehensive scientific approach could specifically be applied to startup practices and, in fact, the empirical analysis done in the first paper looked at the effect of the treatment on startup performance. This was what brought me to the next two chapters of my thesis.

The second chapter is a theoretical contribution which delves deeper into the concept of the scientific approach, into its origins and the reasons why we need it as a new construct in the field of entrepreneurial decision making. In this paper, I disentangled the scientific approach into four components and provide sixteen methodological characteristics that determine the level of "scientificness" of the approach carried out by entrepreneurs.

The third and final chapter of my thesis could not but be a scale validation study, which aims at testing the empirical validity of the theoretical representation of the scientific approach discussed in the second chapter of my thesis, by translating its concepts in a list of items subjectable to confirmatory factor analysis.

Overall, my desire is to contribute to the academic search for frameworks that help to understand the new entrepreneurial dynamics of our century. The new concept advanced on the scientific approach to entrepreneurial decision-making, its scale validation and a first empirical attempt to tease out its performance effect on decision-makers' behavior and outcomes, shall be therefore interpreted as a first bold attempt in this direction.

Sommario

PAPER I - A SCIENTIFIC APPROACH TO ENTREPRENEURIAL DECISION MAKING: EVIDENCE FROM A RANDOMIZED CONTROLLED TRIAL

Arnaldo Camuffo, Alessandro Cordova, Alfonso Gambardella

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.

We 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. We also thank 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.

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, and particularly 2 through 4, 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, typically with a button that users can click on 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 often 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 the ultimate 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 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 argue in favor of 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

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

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

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

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

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

where G_τ is the distribution function of the signal $\hat{r}_\tau$ received for any idea $\tau = t, t+1, t+2, \ldots$; $G_\tau^* \equiv G(r_\tau^*)$, 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^*, $\tau = t, t+1, t+2, \ldots$, 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, a word which in this specific case we use to represent a drift in θ, and we show that it 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-scientific entrepreneurs 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

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.

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 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 scientific 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 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 what we have for simplicity referred to as 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 evidence 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

[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.

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.

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

[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).

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 mobile-based (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.

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

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

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 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

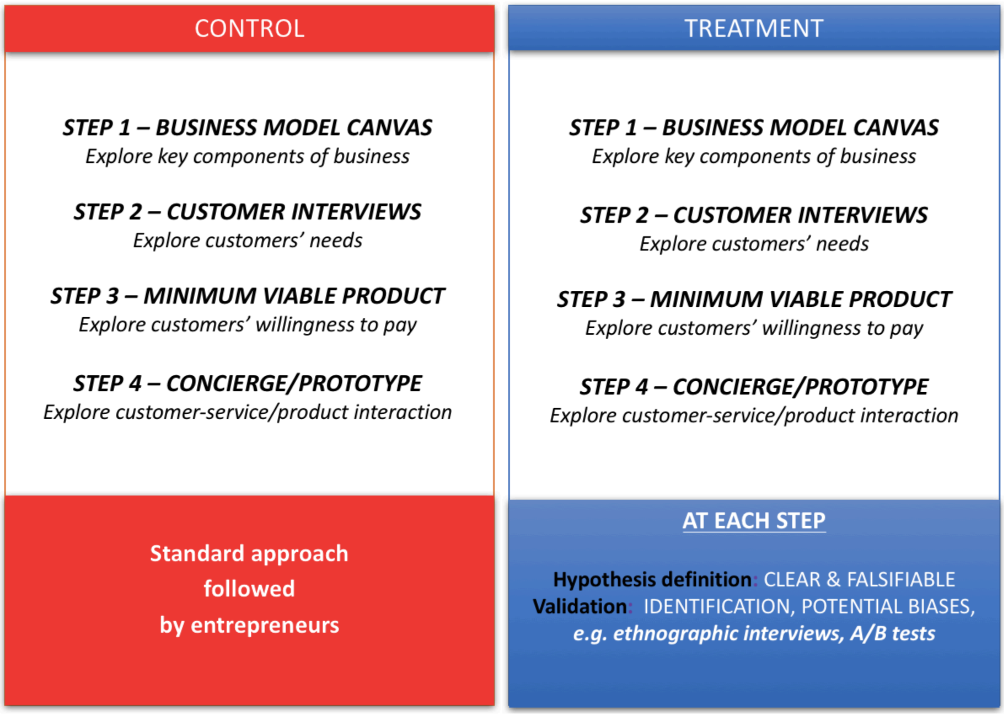
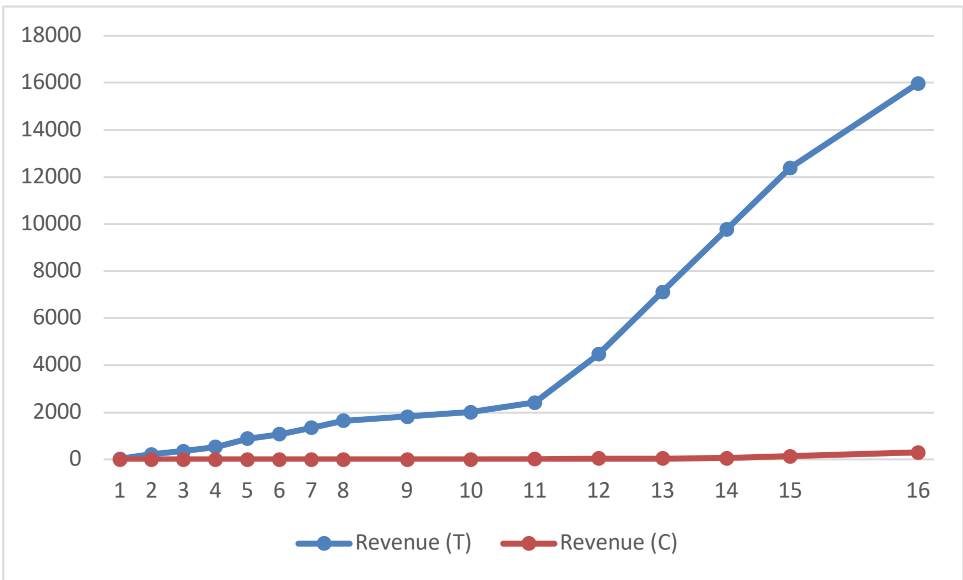


Figure 2: Average revenue over time (euros), treated and control startups



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).

PAPER II - A SCIENTIFIC APPROACH TO ENTREPRENEURIAL DECISION MAKING: TOWARDS A CONCEPTUALIZATION

Alessandro Cordova

Abstract

In recent years, both the practice of management, entrepreneurship and the scholarly debate have recognized that firms have to make decisions about new products or business ideas under growing uncertainty. This has encouraged firms, small and large alike, to be more experimental in their approach with the aim of reducing such uncertainty. To date, however, no specific method for guiding the process of experimentation has been advanced. In this paper, I treat the concept of "scientific approach" to business experimentation, which consists in a process of rigorous theory development, hypothesis making, empirical testing, and analysis. Because experimentation is about generating, collecting and analyzing signals as to the potential return of a given idea, the more rigorous entrepreneurs and managers are in theorizing, designing and evaluating their experiments, the more precise these signals will be, and so will the inferential power of their experiments. Overall, a more scientific approach should then allow entrepreneurs and managers to take better decisions relative to their projects, for example when to stop pursuing them, when to change course of action and when to just persevere, and, consequently, achieve better performance overall. In order to introduce the concept of scientific approach, I first clarify the role of uncertainty in the launch of new products and services; then, I discuss the role that *science* can play in this context. I articulate the concept of scientific approach into four distinctive components, each having four methodological properties, and along with it I provide examples that clarify its meaning. Finally, I offer a list of testable propositions for future empirical work.

1. Introduction

The standard approach to business strategy has for a long time been that managers rely on a plan that they formulate and execute (e.g., Ansoff, 1965; Andrews, 1971; Chandler, 1990). The plan is founded on a vision in which managers predict future contingencies and allocate resources accordingly. Nonetheless, today's environment has grown in complexity, fostered by global competition, heterogeneous customers' preferences and acquisition channels. This has made it harder for business decision makers to predict future contingencies (Cooper, 1993; HBR, 2012[9]; CBInsigths, 2016[10]), with the implication that ex-ante commitment of resources, even if only partially irreversible, is becoming an increasingly costly strategy. This is particularly true for entrepreneurial initiatives where managers and entrepreneurs wear the suit of creators of new opportunities and, as such, are open to a wide variety of alternatives which to choose among and in a condition of high resource constraint.

In response to increasing environmental uncertainty, both in the practice of management and in the strategy and entrepreneurship scholarly debate, it has been argued that firms should undertake more experimental approaches to investment decisions based on staggered investments, business experiments and flexible adaptations to environmental changes. Experimentation, which is not to be restrained to the concept of lab experiments (i.e. cases in which there is always a treated and control condition which perfectly allows to tease out cause-and-effect relationship) but rather referred to as any procedure which is carried out to support, refute or validate a hypothesis (thus also interviews, questionnaires, database analyses, etc.), is a means of learning about future contingencies. Before the 2000s, the entrepreneurship literature clearly examined learning in new firms, but characterized new ventures as continuously changing, fluid entities that engaged in unplanned "unscientific experimentation" and "opportunistic adaptation" (Bhide, 2000). Learning by experience (Argote, 1999) and learning by happenstance (Sorenson, 2003) remained the cornerstones of our understanding of how firms can learn about their new product/services and make sense of their profitability. In fact, the classical approach to collecting

[9] http://hbswk.hbs.edu/item/why-companies-failand-how-their-founders-can-bounce-back
[10] https://www.cbinsights.com/blog/startup-failure-post-mortem/

and elaborating information to make entrepreneurial decisions combines search heuristics such as trial-and-error processes (Nicholls-Nixon et al., 2000), effectuation (Sarasvathy, 2001), or confirmatory search (Shepherd et al. 2012). Not too long after, however, we began recognizing the active role of entrepreneurs in promoting learning for their organization. Sull (2004) and Tripsas (2004) made an important contribution in this direction. The former explained how new ventures suffer from resource constraints that limit their possibility of using diversification as an uncertainty-reducing strategy, and as such are urged to find direct ways to learn about the viability of their business. In this sense, iterative experimentation is seen as a key activity that disciplines the product-development process of an entrepreneur allowing her/him to accumulate valuable knowledge that makes her/him less uncertain: thanks to experimentation s/he can make sense of whether s/he had better continue on an explored path, change to another path or simply "pull the plug". Tripsas and Murray (2004) pushed the boundaries even further, associating the term "scientific" to that of experimentation, recognizing in the former the properties of a method, a mindful, analytic approach in which an entrepreneur identifies a problem or decision to be made (e.g. focus on Market A or B), builds a hypothesis as to the likely outcome and takes action to test the hypothesis through prototype development. This is the direction in which I want to contribute, explicating why we need entrepreneurs to behave like scientists in the strict sense of the word and elaborating on what actually constitutes a rigorous scientific approach applied to the launch of new products/services. In fact, the approach has been studied so far only to a very general level of analysis, described as consisting of testing hypotheses but without specifying how hypothesis testing should be conducted to be recognized as scientific, and practically confined to few specific aspects, such as whether firms are better off by conducting parallel as opposed to sequential experiments (Tripsas and Murray, 2004). In the journey to better specify the concept of a scientific approach to entrepreneurial decision-making, I first theorize the context in which entrepreneurs are called on acting in the next section. This helps to understand through what specific mechanism experimentation reduces uncertainty. This opens up to the concept of scientific approach to entrepreneurial decision-making, which I analyze in further details in the sections afterwards.

2. Uncertainty, experimentation and science

The typical context in which entrepreneurial managers, a word I use hereafter to identify interchangeably decision-makers who have the task of launching new products/services, operate is one of uncertainty and constrained resources. Uncertainty is usually described by scholars in the field of business experimentation as new information that is incomplete or imperfect (Sull, 2004), which may be linked to new technological shifts in the environment (Tripsas, 2004) or simply unpredictability of future states of the world (McGrath and MacMillan, 1995). While this helps explaining why planning is not a suitable strategy and experimentation is to be preferred– i.e. it is too risky for the decision maker to commit all the resources on a plan – it does not tell us how experimentation shall be conducted. In order to have an answer to this, it is necessary that a deeper notion of uncertainty is provided. Building on the recent work of Gans, Stern and Wu (2016), we can think of uncertainty as the situation in which the entrepreneurial manager has to choose among a multitude of options whose return is at least partially unknown to the decision maker. These options refer to the different possible paths the entrepreneur can follow in every decision s/he has to take: which of many customer problems s/he would be best in focusing on, which of many solutions (different products/services) s/he should offer to customers, which marketing channel s/he should bet on in order to optimize acquisition of new customers, which strategies s/he should enact in order to push new and old customers to buy her/his product/service, etc. In order to learn about the expected return of a given option the entrepreneurial manager has to commit to it, excluding the alternatives, and accept at least partial irreversibility. Given this condition, the entrepreneurial manager is confronted with the dilemma of what is the best path to explore. This is where experimentation kicks in. Experimentation is the process by which entrepreneurs generate, collect and analyze signals as to the potential return of a given option. But if this is true, then the process by which experimentation is conducted determines the precision that the entrepreneurial manager will have in predicting the potential return of her/his idea. In other words, an implicit assumption that has been made so far in the literature is that experimentation itself is enough to provide entrepreneurial managers with valuable knowledge, that is entrepreneurs need "just" to experiment to generate precise signals as to the potential return from following a given option. I argue this is not the case, that it is the method through which experimentation is conducted – i.e. the way alternative options are analyzed and confronted, the level of detail which hypotheses are framed with, the rigor which tests are carried out and the orderliness by which analysis of experimental tests are conducted - that explains the degree of precisions by which signals about the return of a given option are generated and therefore influence the entrepreneurs' ability to make better decisions.

The lack of method can actually render experimentation detrimental. Think for example to the very common situation where entrepreneurs participate to pitch competitions and walk the judges through their market research (e.g. interviews with potential customers) presenting pie charts that show very high percentages of people interested in using their products. We know this is inconsistent with the high rates of startup failure (Shook et al., 2003, Ghosh, 2012), which suggests entrepreneurs conduct ill-structured tests. It comes as no surprise that the top two reasons why startups fail today in the market are, one, that there is no market need for their product/service and, second, that entrepreneurs finish cash before having identified a sustainable business model (CB Insights, 2016). As long as firms do not have a method for conducting experimentation, they can be deceived by erroneous inference and make bad investments.

In this sense, it makes sense to borrow from the rigor of science and develop a more disciplined, purposeful or, as I call it, "scientific" approach to experimentation. Science is characterized by rules that need to be followed in order to generate learning. Scientists spend time in crafting their theories, writing down specific hypotheses, design experiments that allow to tease out cause-effect relationships, analyze data in a systematic way to evaluate the results of the experiments and update their theories. The extent to which an entrepreneur applies a scientific approach then can explain the extent to which s/he can generate precise signals about the returns of the options s/he is confronted with. In fact, precision is typical of scientists. In 1877, Charles Sanders Peirce characterized scientific inquiry not as the struggle to move from irritating, inhibitory doubts born of surprises, disagreements, and the like, and to reach a secure belief, belief being that on which one is prepared to act. In this sense we can even predict what the impact of a scientific approach will be on entrepreneurs' decisions and ultimately on the outcome of these decisions. If science improves precision, then we can expect entrepreneurs to be better able to avoid false positives and pursue false negatives (I will further elaborate on this point in the "Propositions for future research" section), at the same way we, as academicians, make most precise inference when we manage to reduce Type I and Type II errors in our researches.

3. A scientific approach to entrepreneurial decision-making

Experiments provide the foundation of the scientific method and have been widely used and refined by scientists since the seventeenth century. Scientists have used experiments to make systematic observations of the ways in which systems work, to test hypotheses, to falsify predictions, and to draw inferences (Tripsas and Murray, 2004). Great philosophers such as Karl Popper have conceptualized science as an iterative process of identifying an anomaly between existing theory and empirical data, forming a tentative hypothesis to explain the anomaly and then eliminating errors in the hypothesis by submitting it to logical scrutiny and empirical testing. This process creates new knowledge that the scientist uses to refine his understanding of the anomaly, which in turn stimulates further logical analysis and experimentation (Sull, 2004). I follow this view of science and argue that a scientific approach applied to the decision-making process followed by entrepreneurial managers is one in which they:

1) Articulate a theory
2) Break down the theory into a set of falsifiable hypotheses
3) Subject hypotheses to rigorous empirical testing
4) Evaluate information and data in a thorough way in order to update the theory and make sound decisions

Each of the components cited above is assumed to be equally important. The use of all of these 4 components is what actually distinguishes a scientific entrepreneurial manager from other typologies of business decision makers, such as what I call planners, effectuation entrepreneurs and lean entrepreneurs. The planners are those type of entrepreneurial managers who mostly trust their guts, who have a strong vision and are convinced that the plan execution and persistence will lead them to attain success. To the other side of the spectrum there is the effectuation entrepreneurs. Entrepreneurs following an effectuation approach assume the future cannot be predicted (Sarasvathy, 2001), thus that there is no "superior path" to look for. Rather the effectuation entrepreneur uses its means, start co-creating with customers and exploit contingencies as the future unfolds. The scientific entrepreneur, instead, thinks uncertainty can be reduced by means of a theoretically-driven hypothesis-testing approach that compares the potential return from following different paths. Specific tests can be designed to obtain information regarding the return of following one business development road rather than another (e.g. focusing on one customer segment rather than an another). In this sense, the approach is closer to the concept of "lean startup" (Ries, 2011). Lean entrepreneurs put minimum viable products in the hands of a real customer to yield validated learning, that is to test whether key hypotheses are falsified. This helps entrepreneurs decide whether it is best to "persevere, pivot or perish" (Eisenmann et al., 2015). However, there are two differences that mark the distinction between the lean and the scientific entrepreneurs. The former is the importance that the scientific entrepreneur lays on theorizing. In fact, the lean startup concept has often been criticized for adopting a "just do it" mentality, that is to spend not enough time on theorizing (Zenger, 2015; Gans et al., 2016). While I do not completely agree on this view, it is true that the focus of Ries' 2001 book "The lean startup" is on how to do tests for developing the product / service efficiently and less on how to figure out, for example, whether a customer problem that prompts that solution actually exists in the first place. What is missing is a more thorough theorization of why that product shall be successful, not only for how it is developed but also for whether it actually responds to a real customer problem, and for which marketing strategies shall be used to market it in an efficient way, etc. In this sense, I 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. The second, and most important key difference, is the rigor which the scientific entrepreneur puts in running experiments. In fact, a comprehensive construct of a scientific approach to decision-making cannot only require that the entrepreneurial manager adopts the four components of the approach, that is theory, hypotheses, empirical tests and analysis, but must require that also the methodical aspect of science is taken into account. Just think about the high research standards that we use as academics: we have endless of methodology papers, textbooks and courses which establish guidelines on how to craft theories, draft falsifiable hypotheses, run bias-free empirical tests, and analyze data. The same principles must be embedded in the scientific approach to business practices.

4. The four characteristics of the scientific approach

In this section I discuss in more details the concept of a scientific approach to entrepreneurial decision making. I will accompany my discussion of the four key constructs that comprise the scientific approach and their methodological characteristics with examples. For simplicity I will use one single fictitious business case as reference for these examples: an entrepreneur who wants to launch a web platform that helps people to search for a restaurant for dinner. Always for the sake of simplicity, while in the previous chapters I have mentioned that the scientific approach applies to all the decisions the entrepreneur makes over her/his lifecycle (which customer problem to focus upon, which solution is best to offer to the market, which acquisition and activation strategies it would be best to enact, etc.), I restrain the examples to the phase in which the entrepreneur is experimenting with his/her customer problem, that is s/he is trying to understand whether there actually exists a customer problem worth solving. Later on, in the section "Propositions for future research", I will better articulate how the scientific approach actually applies to all the different decisions the entrepreneur makes over time.

Theory

Often times, entrepreneurs have a vague idea of how their product/service will be successful, e.g. what exactly the customer problem is, why a given product feature should serve customers' needs, a given marketing strategy will generate more leads than another, etc. Instead, having a theory is essential for the entrepreneurial manager because s/he will know why certain things work and others do not and be more efficient in changing to more suitable strategies over time. In other words, having a well-defined theory guides the process of collection and assessment of market signals. Besides, because making experiments has a cost, conducting individual tests on any option the entrepreneur can choose among (e.g. interview all potential customer targets) will be quite demanding, while a theory can reduce the space of options for which it makes sense to experiment on. For this reason, Zenger (2015) defined pure trial-and-error experimentation as dispersive, time-consuming and, thus, low-value added. So, an important component of the scientific approach is the presence of a theory. I feel this component has not been stressed enough in previous research on experimentation, especially as the first and necessary pillar in the process of hypothesis testing, which is instead widely recognized as being the essence of the experimentation process (Sull, 2004). As previously emphasized, I refer to a scientific approach as one which incorporates the methodological rigor of science. For this reason, it is a necessary but not sufficient condition that entrepreneurial managers have a theory for them to be defined as using a scientific approach, but it is important that the theory is constructed methodically. “This is very much like scientific progress. Scientists design experiments to test a theory, and the more thought-through the theory is, the more likely it is to be validated. Experimentation around less-thought-through theories produces more failures” (Zenger, 2015). Building on this I define the characteristics that a scientific theory must have as follows:

1. *Clear*: as an illustrative example, consider the case of the entrepreneur who comes up with the idea to launch a web platform that helps people to search for a restaurant for dinner by using a search engine to filtrate by type of restaurant. If you ask the entrepreneur why this solution will be successful, s/he needs to be able to state the key reasons why it will be. For instance, saying that “the key problem for customers is the time they take to search for a restaurant online” makes it clear why a search engine may be a clever solution, because the filtering mechanism would speed up the search process. However, often times entrepreneurs cite several customer problems at the same time (the time it takes for them to search the restaurant, the impossibility to know the average meal price ex-ante, they cannot book in advance, etc.) and cite the many features the platform will have that will make it successful. A

clear theory implies that the entrepreneur is able to decompose all the various problems and associate the single solutions to those problems, rather than maintaining her/his value proposition to a more confusing general level.

2. *Elaborated*: the more the entrepreneur articulates, that is s/he provides details, her/his theory, the more s/he can reflect on the veracity of her/his statements and subject them to tests. Continuing the example above, the entrepreneur should wonder why people take time searching for restaurants. Is it because there is not enough information, information is hard to get or information available tends to be erroneous? Are all sources of information equally time-consuming (e.g. digital and not)? Do all people take the same time to search? Considering that the next steps of the experimentation phase will be to write testable hypotheses and subject them to empirical test (possibly interviews with potential customers), stating a theory which includes reflections on all these questions would help the entrepreneur set her/his expectations, understand area where s/he perceives greater uncertainty ("The Manager needs to know therefore how large his area of ignorance is" – Drucker, 1955) and think about the proper tests to conduct for each different aspect of the theory, etc.

3. *Comprehensive*: an important part of the process of theorizing on why choosing a given option will be successful, is to consider its likely success relative to alternatives. This is not as straightforward as one may think. A common limit that I have seen in my years of startup mentor is that the business idea that entrepreneurs have had has "hit" them so hard that they forget to analyze alternative options. This is often the case because entrepreneurs assume that solutions that work beautifully in other industries will do the trick in their industry too. For example, one may have had the idea of creating this engine for searching restaurants online because search engines for creating hotels are being successful. Needless to say, it may be that the problems customers have in searching for hotels are different than those people have when searching for a restaurant. For example, as the "job" (Christensen, 2016) of search engines is speeding up search, our entrepreneur may assume that people do have the problem of taking too

much time to search restaurants. However, this may not be the case. People could actually take little time to find a restaurant but have a hard time booking it last minute, such that the best product to launch on the market is not a search engine but a last-minute booking platform. More systematic analysis of the problem can help entrepreneurial managers consider options initially foregone and start theorizing on them too.

4. *Evidence-based:* it goes without saying that a scientist builds his own believes on evidence and an amount of evidence that is sufficient to come to a fair conclusion. This is not always the case for entrepreneurs in the customer problem phase. In fact, their project idea often comes from a personal experience which is considered to be enough, per se, to implicitly render this idea worth investing in. While the process through which evidence is collected and analyzed is discussed in the other three steps of the scientific approach, here I limit myself at saying that entrepreneurs using a scientific approach always try to base their opinions on significant data and reduce instead the cases in which they are confident about ideas that are simply perceived as intuitive or reasonable, but without having an empirical base. The literature on evidence-based management has pushed a lot in this direction (Briner, Denyer and Rousseau, 2009). Notice that this aspect of the scientific method is what makes the process of theory-hypotheses-empirical test-analysis circular and repetitive, possibly never ending. In other words, the scientific process is often not characteristic by just one cycle of experimentation but repeated ones where analysis of experimental results, described later in this section, plays an important role to provide the evidence on which new theory is generated and possibly re-tested.

Hypothesis

Hypotheses represent the bridge scientific entrepreneurial managers construct to render their theory testable empirically. Before an actual test is carried out, translating the comprehensive theory in a set of hypotheses is essential for the entrepreneur to understand where her/his vision fails. The falsification of a hypothesis is what wakes up the spirit of inquiry of the scientist and makes him question her/his existing believes. Without hypotheses, the process of testing a theory is chaotic and reduces learning. As Geiger says: ''. . . the hypothetical spirit is the unique contribution scientific method can offer to human culture; it certainly is the only prophylactic against the authoritarian mystique so symptomatic of modern nerve failure'' (1950). I define the key characteristics of scientific hypotheses for entrepreneurial managers as:

1. *Explicit*: you'd be surprised to discover how few entrepreneurs explicit any hypotheses. Making hypotheses explicit is instead very important, because it allows the entrepreneurial manager to focus the experimentation process and helps her/him to reason on which experiment is best to run to validate that particular aspect of the entrepreneurial manager's theory. For example, suppose we make the following two hypotheses: one, people take time to search for a restaurant; two, quick information provided online reduces time to search. These two hypotheses actually require two different experiments to be tested. In the former case, customer interviews will do the trick. People would be asked about their experience with going out for lunch/dinner and the entrepreneur would see whether the time they take to search for a restaurant is indeed a relevant problem. However, one cannot precisely infer whether the future presence of online information will speed up search because customers do not know how they would interact with a product which does not exist yet. A proper test for the second hypothesis would rather be the release of a draft version of the web platform and observing how customers' actually use it. Overall, it is not secret that writing down things favor critical thinking.

2. *Coherent with the theory*: it is fundamental that hypotheses reflect exactly what theorized by the entrepreneur. Let me make an example. If our theory was that, among the various potential problems people have when searching for a restaurant, the most important one was the time people take to search, then the hypothesis should specifically refer to the concept of "time" – i.e. "people take time to search". Any similar concept, such as "stress" – i.e. "people get stressed in searching for a restaurant", would lead entrepreneurs to test alternative theories.

In fact, if it was a problem of stress, the web platform should not be as much concentrated on speeding up search but, perhaps, on simplifying the search.

3. *Falsifiable*: it means that it is possible to conceive of an observation which could negate the hypotheses (Popper, 1959). The classical example of a no-falsifiable hypothesis is "It will rain here in a million here", which is falsifiable in principle but not in practice. A falsifiable hypothesis is instead something like "all people are blonde" which would simply require one person to have a hair color different than blonde to disconfirm the hypothesis. While it is actually difficult to see entrepreneurs make no-falsifiable hypotheses, one less intuitive aspect of falsifiability is the specification by the entrepreneurial manager of a threshold which decrees whether the hypothesis is falsified or not. For instance, if I wanted to interview customers in order to understand whether they do indeed take time to search for a restaurant, then my hypothesis should be formulated as "7 people out of 10 take time to search" rather than more generally "People take time to search". The latter is in fact a hypothesis which is almost always falsifiable and does not produce any relevant learning for the entrepreneur. Eisenmann, Ries and Dillard (2015) also make this point very well, that is expressing a hypothesis in specific, quantifiable terms is critical for making it falsifiable. Suppose in fact that I did not use a threshold and found that 6 people out of 10 did take time to search a restaurant. The risk I have often witnessed is that non-scientific entrepreneurs end up reasoning in absolute terms, thus probably thinking that 60% of people is a high percentage. But is it? The correct answer is, of course, that it depends. It depends on how large the total addressable market is, what the margins from operating in this market are and what the aspiration levels of the entrepreneur are. Therefore, a scientific entrepreneur always tries to attach reasoned thresholds to their hypotheses so that they can be rejected or failed in a meaningful way.

4. *Precise*: it stands for the extent to which hypotheses are framed in a way that allows the entrepreneurial manager to test one thing at a time. This is important for the following reasoning. Assume that instead of distinguishing between the two hypotheses "It takes time to

search restaurants" and "It takes time to search restaurants online", the entrepreneurial manager specified only the second hypothesis and went out collecting data through a questionnaire which asked potential customers the question "Do you take time do to search restaurants online?". By this formulation, in case s/he found out that only few people reported they took time to search restaurants online, s/he would not know whether this is because people do not search restaurants at all or do not do it online. Simple mistakes like this can have profound implications, i.e. in case people did not search for restaurants at all (e.g. assume they always go to the same restaurant), then the overall project would be assessed as having low probability of being sustainable from an economic standpoint and it'd better for the entrepreneurial manager to drop out from the project; in the latter case, instead, that is people search restaurants but do not do it online, then the entrepreneurial manager could still find this an attractive idea to invest in but would be required to come up with a different solution.

Empirical testing

Once the theory has been delineated and the hypotheses defined the moment to experiment has finally arrived. Experiments is what allow companies to gather feedbacks on their ideas. Once again, notice I use a loose definition of experiment, such as any procedure which is carried out to support, refute or validate a hypothesis. In this sense, interviews, questionnaires, analysis of databases, are then all experiments, means by which entrepreneurs can test their ideas and collect signals about their veracity. How should experiments be conducted? A scientific entrepreneur would design tests which are:

1. *Coherent with the hypotheses*: this is similar to the concept of coherence between theory and hypotheses. I will temporarily push myself beyond the analysis of the customer problem to another phase of the startup lifecycle in order to provide a useful example here. Suppose I am in the phase of trying to validate my offer, i.e. I am trying to see whether

people respond positively to my offer of using my website for them to find a restaurant online. If I hypothesized that the key variable people look at in searching for a restaurant is how close the latter is from where users are geo-localized, then, my marketing campaigns should use a claim like "Use our website and find quickly the closest restaurants to you" in order to be coherent with the hypothesis stated. Differently, using a statement like "Use our website and find quickly the restaurants with the best deals" would not be coherent. Entrepreneurial managers need to be very careful at designing their experiments and make sure they are reflective of the theory and hypotheses they want to test if they want to achieve valuable learning.

2. *Externally valid*: I refer to this aspect to underlie the importance that tests are representative of the real situation the entrepreneur faces. As an example, suppose that the entrepreneurial manager wants to test whether the search engine will be successful in helping people find the restaurant more quickly. In order to reduce ex-ante commitment, i.e. paying a software developer to create a full-fledged website and search engine, s/he decides to create a draft version of the platform in which customers, once landed on the website, can chat with a dedicated person to whom they can ask any question they need to find their desired restaurant. While this test may help him/her understand whether providing information to customers helps them to choose faster, it does not really mimic the functioning of a search engine. What if asking questions to a human person is perceived by usera as much better than completing the search for the restaurants by themselves through a search engine? In this case the test has produced biased inference because the entrepreneur has assumed that experimental conditions are the same than real ones. Again, I do not intend to say that is always possible to make experiments that match 1-to-1 what happens in the real world but scientific entrepreneurs are aware of external validity, try to reproduce externally valid tests and always pay attention in extending their inference too far beyond the environmental conditions in which their experiment has taken place.

3. *Internally valid*: I define internal validity as the conditions the experiments must respect in order for a bias-free inference to be obtained. Because of their importance, I think it is of a certain value to distinguish between two broad categories of actions the scientific entrepreneurial manager shall pursue in order to make bias-free inference. The first is *sampling correctly*, that is to avoid making self-selection biases or carrying out tests with non-representative subjects. An example of self-selection is when entrepreneurs offer rewards for people to respond to their questionnaires without thinking that this may attract a special category of people who is different from the average targeted customer. One example of the sampling of non-representative subjects, instead, is the typical resort by entrepreneurs to interviews to individuals with unusually intense passion for the product category or sympathetic friends and family that tend to inflate confirmatory biases. The second important characteristic of internally valid experiments is the use of *rigorous testing procedures*. In the case of customer interviews, for example, a common mistake done by entrepreneurs is to ask direct questions like "Would you use my product/service?". The problem with this type of questions is that an interview is a fictitious market setting where respondents do not actually use the service and therefore it is not costly for them to respond affirmatively. Qualitative researchers especially know, instead, that open-ended questions better suit the purpose of exploring customers' real needs and problems because they allow them to express their real believes and opinions (Kelley and Littman, 2005). Each experiment has its own procedural rules that allow to make more precise inference. Scientific entrepreneurs know the rules and follow them to avoid making biased inference.

Analysis

Critical evaluation of experimental results is important to advance knowledge, especially if hypotheses have been falsified. Falsified theories are to be replaced by theories that can account for the phenomena that falsified the prior theory, that is, with greater explanatory power. For example, Aristotelian mechanics explained observations of everyday situations, but were falsified by Galileo's experiments, and were replaced by Newtonian mechanics, which accounted for the phenomena noted by Galileo (and others). In order to be able to extract valuable knowledge from information produced via experiments, I describe the process of analysis of a scientific entrepreneur as one which has the following characteristics:

1. *Data-driven*: the entrepreneur defines ex-ante specific metrics to evaluate the results of her/his experiments. These must be of course well-thought. For instance, as our restaurant entrepreneur goes making interviews to potential customers, how is s/he going to evaluate whether people do indeed take too much time searching for the restaurant? S/he may decide that the right metric to look at is the average number of hours customers spend searching for restaurants every month. More thoughtfully, s/he may realize that asking the percentage of times customers decided not to go to the restaurant because of how much time it took them to search for it, is a more relevant metric to test her/his theory as this translated into an action which has implications for business.

2. *Data is valid and reliable*: validity has to do with the extent to which the metric chosen by the entrepreneur actually measures what s/he intends to measure from a theoretical standpoint; reliability, instead, refers to the extent to which it produces similar results under consistent conditions. Going back to the example above, if the entrepreneurial manger decides to measure how many hours customers spend on average searching for a restaurant, this would be a valid measure for measuring the time people take to search for a restaurant. This would not be the case if s/he asked how stressful people find searching restaurants. Asking customers in an interview how many hours they spend in this task is instead a less reliable measure than, say, actually observing people search for a restaurant and counting how many minutes they take to complete the task. Again, the extent to which the entrepreneur applies this level of details to her/his experimental approach is a choice, but one who implies a greater scientific approach and lead to more or less precise inference.

3. *Data is collected systematically*: rarely I have seen entrepreneurs collect and monitor their data on a truly systematic fashion. For example, a great deal of digital entrepreneurs are not properly familiar with automatic data collection methods or tools such as google analytics, nor they more generally establish routines to monitor data a regular time intervals (Croll and Yoskovitz, 2013). Seemingly, I have seen several entrepreneurs interviewing potential customers without taking any notes, or recording them. They may tend to write down things occasionally, but they surely do not do it rigorously. Instead, scientists use "field journals" to make sure information are lost or forgot. These can help avoiding inferential biases. For instance, it is acknowledged that when we make interviews, we are subject to a cognitive bias called "primacy effect", that is the tendency to remember more the beginning of conversations than their ends. Then, recording and listening back to an interview can help rendering the experimental evaluation process less biased. Seemingly, rotation of questions reduces the same effect in surveys (Edwards, Thomas, Rosenfeld and Booth-Kewley, 1997).

4. *Analysis is explicative*: it is not sufficient that the entrepreneurial manager collects data. S/he must be able to draw implications from it. At the extreme, a scientific entrepreneur is able to comment on any piece of information s/he has collected and derive implications for her/his overarching theory. For instance, if by interviewing potential customers the entrepreneurial manager finds that 60% of customers take so much time in searching for restaurants that they decide not to go out for lunch/dinner, but also that there is 40% of people who do not search for restaurants because they prefer using delivery food services, s/he must acknowledge that: yes, there is a "time problem" but it applies only to a certain customer base. Also, s/he may realize food delivery services can be competitors to her/his web platform. For that reason, s/he may want to know what characteristics make people differ in their choice of going for a restaurant rather than using delivery services, whether it is a matter of different users or same users in different situations. S/he would then start exploring this path. Because the knowledge development is often based on sequential discoveries, it is essential that each piece

of information is taken into consideration and analyzed thoroughly before being discarded as irrelevant, and relevant as well, for the sustainability of the entrepreneurial managers' business.

5. Propositions for future research

Empirical propositions

So far, I have described the concept of a scientific approach to entrepreneurial decision making and illustrated its key characteristics. In this section I argue what its implications are in terms of the outcome of the decision making process of the entrepreneurial manager.

In the context of uncertainty which normally pervades the launch of innovative products and/or services, entrepreneurial managers are better off by gathering market signals. The use of experiments allows decision makers to generate, collect and analyze such signals in order to have a better estimate of the expected return of their idea and reduce the likelihood of failure from early commitment. Failure is defined as the impossibility for the entrepreneurial manager to continue investing in her/his project as resources for product/service development are exhausted. In this context, scientific experimentation helps decision makers collect more precise signals as to the distribution of returns of their initiative. The key characteristics of this approach impact on the precision of the signals produced by the experimentation process: theorizing allows to envision all possible alternative ideas and then to reduce them to only a relevant group that is worth testing; the use of clearly stated falsifiable hypotheses makes sure the entrepreneurial manager identifies the reasons why the idea may or may not be successful and understand more clearly how to possibly change; an externally and internally valid empirical testing procedure reduces the potential biases that are made in making inference on the value of the decision maker's project; a systematic, valid, reliable analysis of data guarantees theory is improved and reiterated. Overall, we should see that a scientific entrepreneurial manager makes better decisions. How can we infer they are making better decisions? If entrepreneurial managers are better at recognizing whether their projects are high or low value (to make things simple), we shall observe they are more likely to drop out (that is to quit their projects) when they face a false positive project, while they are less likely to dropout when they face false negative projects. Dropout is however not the only decision entrepreneurs may resort to. In fact, an alternative is to pivot, that is to make a radical change to their project (e.g. addressing a different problem, changing customer target, elaborating a different solution, etc.). If we accommodate for this possibility, in general we shall observe that entrepreneurial managers on the whole (jointly) drop out and pivot more in case of bad projects and drop out or pivot less in case of good projects.

Proposition 1: *in uncertain environments in which ideas are randomly distributed across startups and the actual likelihood of startup success is low, scientific startups are going to do on the whole more dropouts and pivots than non-scientific entrepreneurs. Seemingly, in uncertain environments in which ideas are randomly distributed across startups and the actual likelihood of startup success is high, scientific startups are going to do on the whole fewer dropouts and pivots than non-scientific entrepreneurs.*

From the beginning of the proposition, it is worthwhile to clarify some important points of the above proposition. Firstly, I use of the term “uncertain environments” because if there was no uncertainty, it would equally possible for scientific and non-scientific entrepreneurs see the expected return of their idea and perfectly choose the optimal decision to carry out. Secondly, the fact new product/service ideas are to be distributed randomly between scientific and non-scientific startups is a key assumption that must hold for us to observe the proposed path. In fact, if scientific startups tended to have the best ideas, they would never dropout or pivot. Another way of saying the same is that proposition 1 refers to the ceteris paribus effect of scientificness on the rate of dropout and pivot. Thirdly, while it would be more appropriate to define a threshold for low and high success, this is not as straightforward. Theoretically it would be enough that the likelihood of startup success was below 50% for observing more dropouts and pivots by scientific entrepreneurs and above 50% to observe fewer. Nonetheless, this would be empirically observed only if all startups were highly scientific and would never make false positives and negatives. In other words, it is more likely to see the proposed trend in environments where entrepreneurial ideas are clearly skewed towards positive or negative outcomes. Fourthly, it should be clear by now that the extent to which an entrepreneur is defined as scientific depends on her/his adoption of the 4 aspects of the approach (theory, hypotheses, empirical tests, data analysis) and adopts them rigorously, that is following their methodological characteristics (e.g. theory must be clear, detailed, comprehensive, evidence-based). Empirically speaking, Cordova (2017) proposes a validated scale for measuring the use of a scientific approach on a 1 to 5 Likert scale. Finally, I purposefully said scientific entrepreneurs do "on the whole" more dropouts and pivots because it is not easy to predict whether they will either do more pivots and also more dropouts. In fact, if pivots and dropouts are at least partial substitutes, in the sense that when falsifying hypotheses entrepreneurial managers can equally decide between dropping out or changing something about their business, then a higher pivoting rate will entail a lower dropout rate.

Although theoretically intuitive, analyzing empirically the above proposition is not immediate. In fact, identifying conditions of higher or lower probability of success is not easy. From an empirical standpoint, this would imply that, given a sample of startups, if we were able to assess the "goodness" of their projects, we could observe that they drop out or pivot more than non-scientific startups in the case of bad projects and viceversa in the case of good projects. In fact, for the same reason I entrepreneurial managers cannot know the return of their project ex-ante and need to collect signals from experimentation, it is not reasonable that researchers assign an objective value to entrepreneurial projects (even if experts were asked to assess their potential instead). A solution would be to select on specific entrepreneur level characteristics which may be correlated with the extent to which they perceive environments to offer higher or lower probability of success than they actually offer. As an example, one could sample novice entrepreneurs, which tend to be more optimistic than seasoned entrepreneurs (Fraser and Greene, 2006) and expect that those of them who receive a scientific treatment are more likely to dropout and/or pivot than their counterparts. This is what Camuffo et al. (2017) did in their paper. Inversely, one could sample more pessimistic entrepreneurs – one way to do so is to use a questionnaire to measure their level of independence, perception of control, creativity and risk aversion, all of which are negatively correlated with pessimism (Liang & Dunn, 2010) – and expect a lower level of dropouts and/or pivots from those who received a more scientific training.

Overall, if the scientific approach allows entrepreneurs to make better choices, it should impact performance positively. Indeed, if firms can avoid future failure by dropping or pivoting out earlier on, the average performance, at the net of firm expenditures, would be higher for scientific entrepreneurs: they spare losses in the case of bad ideas and avoid foregoing positive cash flows in the case of good ideas. Empirically speaking, observing costs is important to see significant differences in performance between scientific and non-scientific entrepreneurs. In fact, suppose one could only observe revenues and we were in an uncertain environment where the likelihood of startup success was low. Then we should observe scientific startups to pivot or dropout more than their counterparts. In that case, performance for scientific startups that have recently dropped out or pivoted would remain flat, increasing the variance within scientific startups. To make this even more straightforward, assume majority of scientific entrepreneurial managers dropped out very early in the lifecycle of their startup, then we would observe a lots of zero-revenues among the scientific startups, while non-scientific entrepreneurs would not recognize their ideas are going to fail in the future and keep investing, generating at least some positive revenues. The increasing variance within the scientific startups would increase the standard error of regression coefficients and reduce their significance. This is one of the limitations of Camuffo et al. (2017) which regress the effect of the scientific approach on the revenues of startups. This would of course not happen in the case we had a record of startup expenditures as it would likely be the case that non-scientific startups record negative profits over time as opposed to the zero profits of the scientific ones[11]. For this reason my proposition refers to profits rather than revenues.

Proposition 2*: In uncertain environments, scientific entrepreneurs record higher profits than*

[11] Observing significant differences in profits still require that a large enough sample is collected, given that a majority startups, if sampled after a limited time since inception (up to a year after they began working on their project), show zero revenues and expenditures. This is a problem Cordova (2017) runs into in its regressions to validate the scale for the scientific approach.

non-scientific entrepreneurs.

Once again, the decision to specify the environment must be uncertain derives from the fact that when uncertainty is null, scientific and non-scientific entrepreneurs would perform the same given that the efficacy of their decision-making approach would be the same, ceteris paribus. Notice that I refer to profits and do not include other business outcome measures of better decision making, such as the likelihood of external funding or firm exit because this also depends on other characteristics of the firm. Acquiring companies may for example be looking for startups that can be a strategic asset to the firm, regardless of their experimentation method. It is true that, ceteris paribus, scientific entrepreneurs should also be more likely to achieve better outcomes of these type; however, one does not know whether the kind of rational and rigorous experimental approach followed by scientific entrepreneurs affect audiences' perceptions or whether the time scientific entrepreneurial managers spend experimenting rigorously crowd out other activities, such as networking, which may be beneficial for firms to attain other outcomes, such as fundraising. Profits remain the more direct evidence of the outcome of optimal decision-making process entrepreneurial managers follow to understand the strategic direction to have their startup follow over time. For this reason, it is preferred to other measures on which, however, future empirical research is warranted. Finally, notice I did not differentiate between environments with a majority of false positives or negatives because, regardless of that, scientific entrepreneurs should perform better, either because they avoid failure or avoid foregoing rewarding initiatives.

Another set of predictions I want to advance relate to the antecedents of the scientific approach. In addition to being aware of the method, which is something Camuffo et al. (2017) find to matter significantly, there are characteristics of the entrepreneur which may influence her/his propensity to adopt the approach. In order to think about antecedents, I find it useful to go back and think about the simple concept I advanced in this paper: uncertainty requires a method for reducing uncertainty and the method used matters in explaining how much uncertainty is actually reduced by the entrepreneur. This implies that (i) the level of perceived uncertainty incentivizes people to look for a method to reduce uncertainty (ii) the skills the entrepreneur has in using the scientific method explain her/his ability to reduce uncertainty. Empirical research could therefore try to assess the level of perceived uncertainty surrounding the likelihood of an entrepreneur's project and collect some indicators of her/his skills or propensity to use the scientific approach and look at the correlation between these constructs. Among these I propose two variables. The first is an indicator of whether the entrepreneur had a scientific background in her/his education. Entrepreneurs who have been taught the principle of scientific thinking, such as those who have studied biology, chemistry, physics, mathematics, are likely more used to develop and test theories and therefore more likely to use it in business decisions as well. The second is the extent to which entrepreneurs adopt a more analytical rather than intuitive reasoning in taking decisions. Kahneman (2003) dual system theory of reasoning is useful in this sense. The underlying assumptions regarding the use of the two systems are that System 2 reasoning requires a greater use of appropriate information and analysis (Kahneman 2003) and that a greater use of System 2 or logic-based reasoning by the decision maker will result in better solutions to more complex problems than a greater use of intuitive reasoning (Stanovich and West, 2002). Summarizing, I propose that:

Proposition 3: *the higher the level of the entrepreneur's perceived uncertainty regarding the value of her/his project, the more likely s/he is to use the scientific approach.*

Proposition 4: *entrepreneurs with a scientific education are more likely to use the scientific approach*

Proposition 5: *entrepreneurs who use more the system 2 analytical approach to decision making, as opposed to the System 1 intuitive decision making, are more likely to use the scientific approach*

Further theoretical developments

In the introduction I referred to the fact uncertainty in the context of launching new products/services pervades all the decisions the entrepreneurial manager has to undertake: which customer problem to focus on, which solution to offer to the market, which marketing, activation, growth, funding, strategy to adopt. Then, for the sake of maintaining the core arguments simple, I maintained the application of the scientific method to a general level. The empirical propositions advanced do indeed refer to how the different signals scientific entrepreneurs generate, collect, analyze about the expected return of their business idea collectively contribute to her/his decision-making process. That said, I now want to bring the attention back to the different decisions the entrepreneur has to take over the life cycle. This is important for researchers and practitioners alike. Understanding if and to what extent the scientific approach adapts to each core decision the entrepreneurial manager is called upon undertaking, would strengthen the practical contribution of our prescriptions to practitioners and, at the same time, improve our theoretical understanding of the sequence and potential interdependence of decision-making at different levels of the business. Let see this in more details.

The reason why the scientific approach can in principle be applied to each of these decisions is because they can all be described by the same framework: there is uncertainty regarding the expected return of a given option (one of the possible paths the entrepreneur can choose). For example, I already discussed how the entrepreneur who wants to launch the search engine for restaurants could focus on alternative customer problems, i.e. speed up search but also make search simpler, augment search results, provide more information, etc. Seemingly, in the phase of solution creation, the entrepreneur could opt for a search engine or, totally on the other side of the spectrum, on a telephone-based service. Also, each of these two different products could be made with different features, e.g. the search engine may or may not include previous customers' feedbacks. Again, when the entrepreneur arrives at the phase of product launch, s/he can use different channels to promote her/his activity, either online or offline. For example, in order to give visibility to the search engine solution, the entrepreneur could invest in facebook advertising, google adwords, google display, search engine optimization, etc. Once customers begin visiting the website, they may be incentivized to use the service by means of different user experience designs, discount or affiliation plans, etc.

In principle, a scientific approach can help entrepreneurs choose among the different options for each of these decisions. For instance, when evaluating the introduction or not of the customer feedback system in the platform, the entrepreneur could use well-thought A/B tests. While non-scientific entrepreneurs would offer the new product feature to all the customers at the same time and compare their pre and post customer satisfaction, without taking into account that time fixed effects may play a role in explaining differential outcomes, scientific entrepreneurs would create a reliable counterfactual. They would split the customer base in two random halves (or stratifying on some relevant characteristic) and offer the new product feature to one of the two, while continuing offering the old product to the other half. The A/B test allows to better discern the effect of the new product feature. Additionally, a scientific entrepreneur will have a theory of why the new product feature represent a performance-improving opportunity and will try to study the mechanism that explains the differential performance of the two product versions in order to validate her/his theory. For example, s/he may theorize that showing customer feedbacks would increase users' trust in the platform and ultimately its usage. Then, the experiment will not only compare the number of times the new website version will be used as opposed to the old one, but will monitor whether users have viewed the comments and engaged with them. For example, the entrepreneur could track whether users starting with a low-feedback restaurant are more likely to change their initial choice as opposed to when they start reading about a high-feedback restaurant. These conditions may even be fictitiously created by the entrepreneur, forcing the first restaurant in the list of those the user can choose among to be high or low feedback. A similar distinction between the approach followed by a scientific and non-scientific entrepreneur apply to other decisions.

Bringing the application of the scientific method to this lower level of decision making is important for improving our understanding of entrepreneurs' behavior and enhance the scope of our prescriptions. The positive implication of this more comprehensive model would be that academics may study the effect of the scientific approach by studying firms that are at any step of the entrepreneurial life cycle. This is important because it widens our opportunity to study the impact of the approach on entrepreneurial managers' performance: we would need not be restricted in our research to monitor the performance of startups for a long time period (say from inception to two years afterwards) but rather it would be possible to focus on one phase of the lifecycle, one that concerns a specific decision the entrepreneur has to take –e.g. on which marketing strategy to follow – and then measure the extent to which the application of the scientific approach on that decision impacts the outcome of that decision.

On the other hand, the challenge is to model the application of the scientific approach to each of these decisions. In fact, we do not know whether these decisions require the same level of scientific approach or the same use of the 4 constructs of the approach. As an example, what if earlier decisions in the lifecycle of the entrepreneur proportionately required entrepreneurs to do be more rigorous in the process of theorizing and making hypotheses than decisions that came later in the lifecycle? The phase of identifying the key customer problem is in fact very uncertain and complex, the entrepreneurial manager has to analyze different combinations of customer targets and problems. Additionally, the cost of experimenting is potentially quite high in this phase: one would have to interview different representative samples of customers and each interview could take quite some time. Besides, there is no clear metric to look at in order to easily identify customer problems, as much as there are simple metrics to measure, let's say, how many people are visiting a website. In this complex system, spending quality time on theory, that is on ranking alternatives and understanding which are the best signals to assess customers' problems, can be extremely essential. This recalls the importance of "theory-driven structures" (Walsh, 1995) or cognitive representations (Gavetti & Levinthal, 2000, p.121) of highly complex solution landscapes in speeding up problem solving and selecting trials that maximize the probability of discovering a high-value solution (Nickerson and Zenger, 2004). Differently, when entrepreneurial managers get to the point of launching their product on the market and experiment with different marketing campaigns, the situation is less complex. The cost of experiment is likely to be moderate, especially for a digital startup: one could run two low-budget campaigns on facebook and google simultaneously, and easily compare the relative performance of the two, by looking at very straightforward metrics. In this context, entrepreneurs could be observed spending less time on theorizing and more on testing. This is possibly why the so-called discovery-phase, when entrepreneurs concentrate on identifying a sustainable customer problem, tend to last longer than the validation phase, when startups focus on promoting their service and generate their first leads (Genome Report, 2011[12]). In this later-stage phases, inversely, the rigor of testing is of outmost importance. While interviewing customers produce at least some learning, even if the entrepreneur does not pose questions in the most rigorous way possible, a badly run A/B test could produce zero or misleading results.

Overall, these examples suggest that even if applying all the steps of the scientific approach may be value adding for any decision, entrepreneurs may tend to overuse or underuse some aspects of the scientific method across decision types. This is an important question, either from a theoretical and empirical point of view, as we could think of different optimal "levels" of the scientific approach along the lifecycle of the entrepreneur, where theory, hypotheses, empirical testing and analysis assume different weights depending on the characteristics of the decision the entrepreneur has to undertake – the number of options available, the ex-ante predictability of their return, the cost of experimenting with one option, the irreversibility of experimenting, the need to focus on one option rather than having the possibility to focus on more than one simultaneously (e.g. the entrepreneur can pursue only one customer problem but s/he may use more marketing channels at the same time), etc.

The challenge would continue if we were interested in analyzing how signals gathered on each of these decision levels ultimately contribute to the entrepreneur's overarching decision as to whether to persevere with her/his business or rather pivot or drop out. In this case, we would need to understand the relative importance the entrepreneur associated to these different decisions. For instance, we could sensibly assume that because every decision that comes later in the lifecycle depends on the right identification by the entrepreneur of the key customer problem, the latter has a bigger weight in the decision of the entrepreneur as of whether to drop out from the project or continue. For the same reason, the probability of dropout would decrease over time, while pivoting gradually becomes a more valuable "option".

Overall, I think that a more fine-grained model of how the scientific approach apples to lower

[12] https://s3.amazonaws.com/startupcompass-public/StartupGenomeReport1_Why_Startups_Succeed_v2.pdf

levels decisions would greatly improve our understanding of how entrepreneurs behave and increase the external validity of our prescriptions. This would be a central topic for feature research.

6. Conclusions

When entrepreneurs and managers are called upon launching new products and services, they most often face uncertain environments where decision making entails choosing among alternative options whose expected return is at least partially unknown. In this context, experimentation is needed to generate, collect and analyze signals as to the distribution of returns of these options, which helps managers reduce risky commitment. To date, while the importance of experimenting has been increasingly recognized as fundamental for achieving success in contexts of high uncertainty, no guidance has been offered as to how experimentation should be conducted to minimize commitment and risk of failure. In this paper, I advance the concept of a scientific approach to experimentation and decision making. Scientific entrepreneurs and managers adopt the rigorous hypothesis testing approach typical of scientists, and gather the most precise signals possible as to the expected return of the many possible strategies they could adopt. This helps them to take optimal decisions as along the lifecycle of the startup. Overall, what discussed in this paper is a first attempt to renovate an interest in management as a science and a discipline, a concept recently lost in translation in recent years (Freedman, 1992). Future research should contribute to further develop a model of scientific experimentation and in empirically testing the implications of this approach at all level of entrepreneurial decision making, ultimately contributing to make sure science "arm the manager's imagination" and "supply him with the vision needed to make rational decisions in respect to the business enterprise," and should not serve as a substitute for decision and judgment but should "supply methods for making possible more effective decisions and more informed judgment" (Drucker, 1955 p. 123).

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.

Ansoff, H. I. (1965) *Corporate Strategy: An Analytic Approach to Business Policy for Growth and Expansion*, McGraw-Hill, New York NY.

Andrews, K. R. (1971) *The Concept of Corporate Strategy*, Richard D. Irwin, Homewood IL. 30

Argote L., 1999. *Organization learning: creating, retaining, and transferring knowledge*. Boston, MA: Kluwer Academic Press.

Bhide A.V., 2000. *The origin and evolution of new business*. New York, NY: Oxford University Press.

Briner R.B., Denyer D., Rousseau D.M., 2009. "Evidence-Based Management: Concept Cleanup Time?" *Academy of Management Perspective*, 23(4), 19–32.

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

Camuffo A., Cordova A., Gambardella A., (2017). "A scientific approach to entrepreneurial decision making: evidence from a randomized controlled trial". Early draft available at https://events.insead.edu/doriot-entrepreneurship-conference/documents/A-Scientific-Approach-to-Entrepreneurial-Experimentation_%20Evidence%20from%20a%20Randomized%20Control%20Trial.pdf

Chandler G.N., DeTienne D.R., McKelvie A., Mumford T.V., 2011. "Causation and effectuation processes: A validation stud". Journal of Business Venturing, vol. 26, pp.375-390.

Christensen C. M., 2016. *Competing against luck*. *The story of innovation and customer choice*. Harper Business, section 3.

Cooper, A.C., Gimeno-Gascon, F.J., Woo, C.Y., (1994). "Initial human capital and financial capital as predictors of new venture performance". *Journal of Business Venturing* 9, 371–396.

Cordova A., (2017). *A scientific approach to entrepreneurial decision making: scale validation through a confirmatory factor analysis*. Thesis Chapter.

Croll A., Yoskovitz B., (2013). *Lean Analytics: Use Data to Build a Better Startup Faster*. The Lean Series, Oreilly.

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

Edwards J.E., Thomas M.D., Rosenfeld P., Booth-Kewley S., (1997). *How to conduct organization surveys*. Sage Publications, page 48.

Eisenmann, T., Ries, E. and Dillard, S. (2014) "*Experimenting in the entrepreneurial venture*", Harvard Business Publishing.

Fraser, S., Greene, F., 2006. "The effects of experience of entrepreneurial optimism and uncertainty". Economica 73, 169–192

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

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

Gavetti G.D., Levinthal D.E., (2000). "Looking forward and looking backward: cognitive and experiential search. *Administrative Science Quarterly*, 45 113-137.

Geiger, G. (1950). Values and Social Science, *J. Soc. Issues*, vol. VI, no. 4, pp. 8–16, 1950.

Ghosh S., (2012) *The Venture Capital Secret: 3 Out of 4 Start-Ups Fail*. The Wall Street Journal 20/09/2012.

Kahneman, D. (2003). A perspective on judgment and choice. *American Psychologist*, 58(9), 697-720.

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.

Liang C. K., Dunn P., 2010. “Entrepreneurial characteristics, optimism, pessimism, and realism—correlation or collision”. Journal of Business and Entrepreneurship, Vol. 22, No. 1.

McGrath, R. G. and MacMillan, I. C. (1995) *Discovery Driven Planning*. Harvard Business Review, 73 (4), 44-54.

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.

Nickerson J.A., Zenger T. R., (2004). “A knowledge-based theory of the firm – the problem-solving perspective”. *Organization Science*.

Popper, Karl 1959. *The Logic of Scientific Discovery*. New York, NY: Basic Books.

Ries E., (2011). *The Lean Startup: How Today’s Entrepreneurs Use Continuous Innovation to Create Radically Successful Businesses*, Crown Pub edition.

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.

Sanders C.P. 1877. *The Fixation of Belief*. 12. Wikisource. pp. 1–15

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.

Shook, C. L., R. L. Priem, J. E. McGee. (2003). Venture creation and the enterprising individual: A review and synthesis. *Journal of Management*, 29 360-379.

Stanovich, K. E., & West, R. F. (2002). Individual differences in reasoning: Implications for the rationality debate. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), Heuristics and biases (pp. 421-440). New York: Cambridge University Press.

Sull, D.N. (2004) “Disciplined Entrepreneurship,” *MIT Sloan Management Review*, 46 (1), 71.

Tripsas M., Fiona M., (2004). *The Exploratory processes of entrepreneurial firms: The role of purposeful experimentation*, in Joel A.C. Baum, Anita M. McGahan(ed.) Business Strategy over the Industry Lifecycle (Advances in Strategic Management, Volume 21) Emerald Group Publishing Limited, pp.45 - 75

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.

Walsh J.P., (1995). “Managerial and organizational cognition. Notes from a trip down memory lane”. *Organization Science*, 12(3) 325-338

PAPER III - A SCIENTIFIC APPROACH TO ENTREPRENEURIAL DECISION MAKING: SCALE VALIDATION THROUGH CONFIRMATORY FACTOR ANALYSIS

Alessandro Cordova

Abstract

Scholars and practitioners alike are increasingly becoming promoters of experimentation in the context of innovative product/service launches. Experimentation is a means through which market signals as to the expected return of new projects are generated, offering an opportunity to better estimate their likely success. A key question that is being recently addressed by some, is how experimentation should be conducted in order to minimize the uncertainty pervading the expected return of new projects. One proposition in this direction has been that entrepreneurs and managers need to experiment by following a scientific approach, which consists in the following four steps: they need to theorize on why their project will be successful, break down their theory into a series of falsifiable hypotheses, test them via well-designed experiments, analyze the results of these experiments to learn about the potential value of their project. This paper aims at constructing and validating through confirmatory factor analysis an empirical measure of the scientific approach.

1. Introduction

After conception of a new product/service to be launched on the market, entrepreneurs and managers (hereafter entrepreneurial managers) need to validate the likelihood success of the project in order to evaluate the opportunity of investing in it. Because the project could be realized in different ways, by focusing on solving different customer problems, being characterized by different product/service features, being promoted through different channels and incentive mechanisms, the entrepreneur faces an uncertain distribution of returns. In other words, the entrepreneur does not know ex-ante what is the best path to follow among the different options available in order to optimize her/his returns (Cordova, 2017).

Scholars in the field of strategic entrepreneurship (e.g., Sull, 2004; Tripsas and Murray, 2004; Gans, Stern and Wu, 2016) have argued in favor of an experimental rather than planning approach (e.g., Ansoff, 1965; Andrews, 1971; Chandler, 1990). In fact, the former allows to reduce uncertainty by means of partial commitments that help reveal the potential value of a given path, while the latter remains a very risky approach to the launch of new projects, because assumptions on which projects rely on, remain untested (McGrath and MacMillan, 1995). Experiments are to be interpreted as any type of test that allows to make inference on the likelihood of success of a given path, such as interviews to potential customers, aimed at eliciting whether the entrepreneurial managers' envisioned solution matches with their key problems, desires and needs, or a minimum viable product, that is a draft version of the final product/service which helps to gather feedback on the product-customer interaction.

Recent research papers in this area (Cordova, 2017) have made us notice that the experimental approach which is followed to generate, collect and analyze information can make a difference for entrepreneurial managers. For example, customer interviews lacking theoretical guidance may be dispersive and time-consuming; the lack of clearly-specified hypotheses that include proper thresholds for their validation may lead to cases of wrong falsification; choosing non-perfectly representative customers may limit the impact of the entrepreneurial manager's discoveries to a limited portion of his/her actual target market; asking questions in a completely structured way with no room for more open-ended questions may reduce learning scope; non-systematic codification of interview data may lead entrepreneurs fall victims of primacy effect and other cognitive biases. The same applies to experiments conducted on product development or marketing strategies. Comparisons of new product features or new promotion campaigns with older ones do not take into account the fact that time may play a role in affecting relative performance and may lead entrepreneurial managers to elect as more efficacious paths those which are actually not. Seemingly, election of invalid or unreliable measures of performance may spoil the election of the winning strategy. For this reason, in order to maximize the precision of experiment-generated signals, entrepreneurial managers should adopt a scientific approach, that is a more thorough and methodologically rigorous approach which is typical of scientists, whose rational heuristics of hypothesis generation and testing minimize measurement errors, biased inferences, acceptance of false positives or rejection of true ones (Grandori, 2013).

The scientific approach is characterized by four key characteristics which, in turn, can be applied to a more or less scientific level. The four characteristics are: the use of theories prior to running experiment, the declination of theories into hypotheses, the empirical test of these hypotheses, the analysis of experimental data that lead to theories evaluation and re-evaluation. Entrepreneurial managers that do not follow all of the four steps in their experimental approach cannot be defined as scientific, while those that do so can have different degrees of scientificness depending on the extent to which the four characteristics have rigorous methodological properties (Cordova, 2017). In particular, the scientificness of theories is evaluated based on the extent to which these are clear, detailed, comprehensive and evidence-based; hypotheses must instead be explicit, coherent with the theory, falsifiable and precise; scientific empirical tests should be coherent with the hypotheses, externally valid and internally valid; finally, evaluation of experimental results should be data-driven, based on valid and reliable measures, data shall be collected systematically, and the analysis ought to be explicative. The higher the extent to which the scientific approach follows these properties, the more precise the signals produced by experiments should be and, as a consequence, the decisions the entrepreneurial managers should end up taking. Better decisions can be inferred through the pivoting and dropping out behaviors of entrepreneurs and, ultimately, on their profits attained.

The potential implications of this approach are relevant. For example, to date, 70% of startups are reported to fail within few years (Shook et al., 2003, Ghosh, 2012)) especially because there is no demand for the new products/services launched by entrepreneurs (CB Insights, 2016[13]). This suggests that the latter could benefit from improving their ability to forecast the likely success of their projects. In fact, in a field experiment that involved more than 100 early stage startups receiving a different training on how to conduct experimentation, Camuffo, Cordova and Gambardella (2017) found that those that were treated with a training fostering a more scientific approach attained on average higher revenues. However, in their empirical work, the authors ran an ITT regression, that is they used as independent variable the dummy categorizing whether startups were treated or not with the scientific approach, rather than a direct measure of the extent to which they actually adopted the approach. In fact, at the time the experiment was run, there was not the same in-depth conceptualization of the scientific method and the 4 factor model advanced in this paper was not developed yet. The next step for empirical research that wants to test the effect of a scientific approach on entrepreneurial managers' business choices and ultimately their performance is thus the construction and validation of a measure of the scientific approach. This is the aim of this paper. The work is organized as follows: in the next section I define the construct domain and generate the list of items that are reflective of Cordova (2017)'s theoretical argumentation on the scientific method, then I analyze the result of a pre-test with a panel of startup mentors who were asked to analyze the content validity of the scale, next I show the results of a confirmatory factor analysis for a 4-factor model and run tests of reliability and construct, convergent and discriminant validity. Results confirm that the scientific approach is a one dimension concept made up of four constructs highly correlated among each other. Finally, I compose a scientific score for each startup in the sample and further analyze concurrent, predictive and discriminant validity of the measure, by looking at the correlation between the scientific score and the rate of dropout and pivot of startups and their level of profitability (for predictive validity), the entrepreneurs' spending attitude (for concurrent validity), the team average education and years of experience working in startup and

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

managerial positions endowed with the task of launching new products and services (for discriminant validity). I conclude by advancing implications for future empirical research.

2. Construct domain and measure development

The objective of this paper is to propose a scale measuring the scientific approach that mirrors the four constructs and, more specifically, their underlying 16 items described by Cordova (2017). The first step in this direction is to translate the 16 items into operational measures that can be used in assessing entrepreneurial managers actual use of the method. However, this requires that the method of assessment is specified (questionnaire, interview, online exercise, etc.).

At this regard, I propose that phone interviews are the most appropriate means to objectively measure the adoption of a scientific approach. In fact, interviews conducted by researchers that try to elicit entrepreneurial managers to recount their actual behavior, although more time-consuming, allow to tease out the actual quality/intensity of the approach followed by the interviewee. For example, if one wants to know how entrepreneurs have theorized their customer problem, talking directly to the entrepreneurs and listening to them articulating their answer allows to better understand the extent to which they have made detailed theories, considered alternatives and based them on hard facts. For obtaining the same result with surveys, one should ask respondents to write down their thoughts in an online form, which may reduce the response rate. Alternatively, in order to solve this problem, online surveys would require the formulation of short questions and short answers. However, the risk is that this would push the researcher to ask direct questions where the entrepreneurial manager is indirectly induced to provide responses which are more scientific by construction. For example, the research would need to ask directly which hypotheses an entrepreneur has done and ask to list them, or to ask how many experiments s/he has done in given time period. This would be likely to push the entrepreneurial manager to reflect and give a response which is not representative of what s/he did for real. Instead, in interviews, open ended questions like "Tell me about the customer problem you are trying to solve. Why do you think this is a relevant problem which to focus upon?" would allow the entrepreneurial manager to talk about his/her actual approach and the researcher could evince whether s/he actually had a theory, nailed it down into hypotheses, tested them rigorously, etc. Therefore, the way the 16 items were generated was by formulating a description whose response could then be assessed from interviews.

In constructing the scale of the scientific method, a second important step was to create the scoring system. I purposefully used a behaviorally anchored rating scale (BARS). More specifically, I followed the approach used by Bloom and Van Reenen (2007) which used a "practice evaluation tool" to score the extent to which firms adopted high or low-quality management practices by scoring their responses from one (worst practice) to five (best practice). I follow the same approach and provide guidelines for scores 1, 3 and 5. The main idea is that the degree of scientificness increases with the extent to which the entrepreneur is precise and rigorous in using the corresponding item. In fact, as explained by Cordova (2017), a distinctive characteristic of entrepreneurs is exactly the precision by which s/he uses an experimental approach, which then determines the precision of the signals collected via experimentation. For this reason, a score of 1 is representative of a poorly scientific use of the item, a score of 3 as a somewhat good scientific use of the item and 5 as a perfectly scientific use of the item. As an example, the entrepreneur obtains a score of 1 on the item clear_theory if s/he poorly specifies the core of her/his theory (in the sense that it sounds confusing and logically fallible); s/he obtains a score of 3 if the theory is not confusing but it is not completely clear either and a score of 5 if it is perfectly clear. For example, if I want to open up a hair-saloon in my area, then I will have a poorly articulated theory if I state that, let's say, hair-saloon are generally a good investment and for this reason my hair-saloon will also be successful; I will instead be assigned a score of 3 if I state that in my area a hair-saloon is missing and so this explains why it may be rewarding to make this investment; I will get a score of 5 if I articulate even further that there is no hair-saloon in this area and people living in and passing by the neighborhood are observed to go in other neighborhoods to get their hair done exactly because there is no hair-saloon in theirs. Scores of 2 is given when the interviewer perceives that the score is more skewed towards a poorly scientific use of the item and, seemingly, a score of 4 is given when the interviewer perceives that the score is more skewed towards a perfectly scientific use of the item. Note that the scoring system also allows to give a score of 0 for those cases in which the entrepreneurs are recorded as not using a given construct of the scientific approach. This is because, as discussed by Cordova (2017), in order for an entrepreneur to be defined as scientific, s/he must show to experiment by using all of the four key characteristics of the scientific approach: theory, hypotheses, empirical test, data analysis. In fact, this is what distinguishes a scientific from other types of approaches such as planning, effectuation, lean startup or others. For example, planning is characterized by very low levels of empirical testing and high levels of theorizing, which are required to plan ahead the project execution. The lean approach is, on the contrary, very much focused on executing tests but less so in crafting comprehensive theories (Zenger, 2015; Gans et al., 2016). The effectuation approach does consider the importance of experimenting but associates another meaning to it, that is entrepreneurs experiment by remaining flexible and leveraging contingencies rather than using a top-down approach in which hypotheses are identified and tests are run with the intent of causally inferring market signals that reveal information about the potential return of their project. This aspect of the method has an important implication for our scale development: a startup that receives a score of 0 on at least one construct will be scored 0 overall in the extent to which it uses the scientific approach. Inversely, those entrepreneurs that are scored no-zero values in the four constructs can be defined as scientific entrepreneurs and will therefore get at least a value 1 in the scientific scale. The extent to which they are more or less scientific will depend on the extent to which the four key characteristics of the scientific process follow the methodological characteristics typical of the rigor of science (e.g. theory must be precise, detailed, comprehensive and evidence-based), that is it depends on the score obtained on the sixteen methodological items comprising the scientific approach. This implies that the level of scientificness for those entrepreneurs that can be defined as scientific is going to be measured in a continuum that goes from 1 to 5. I discuss the implications of such a scoring system for the confirmatory factor analysis and the analysis of the scale validity of the scientific approach in more details in the "Empirics" section.

In the Table 1 below, the description and scoring guideline for each of the 16 items described is provided.

Table 1 approximately here

The first step I followed in constructing Table 1 was to write down the 16 methodological characteristics of the scientific approach which were listed in Cordova (2017). Worth noticing is the construction of the "empirical test" construct. Cordova (2017) defines three overarching methodological characteristics of empirical testing which are coherence with theory, external and internal validity. I chose to list four characteristics rather than three by splitting internal validity into two components: the use of representative samples and rigorous testing procedures. This decision was motivated by Cordova (2017) argument that these two aspects are worth distinguishing between, given their individual importance. In fact, cases in which bad inference is generated because, let' say, interviews are done to friends and family, are not few among entrepreneurial cases; seemingly the procedure utilized in running interviews or A/B tests is often mistaken by entrepreneurs.

Then the question is how the description of the 16 characteristics were constructed, as the wording used in the description is important in guiding researchers interviewing process of entrepreneurial managers. Once again, the core idea was to try collapsing Cordova's (2017) more thorough descriptions for each characteristic to shortened versions, but without excluding important meanings from such descriptions. Certainly, one could envisage different sentences to express the same concept. Therefore, I decided that the most important thing would be to ensure the interpretation of the descriptions I proposed was straightforward and not confusing. For ensuring that, I interviewed a panel of seven different startup mentors, which are expert in the field of entrepreneurship and took part to the field experiment conducted by Camuffo, Cordova and Gambardella (2017) and were therefore trained in the concept of the scientific approach although were not exposed directly to its theoretical codification, which was done after the experiment was conducted. This somewhat ensured that they had independent opinions on what constituted a proper scientific approach. Notice that, an alternative, would have been to interview other startup mentors, but the risk was that they were not familiar with the concept and their lack of knowledge and expertise made them non-suitable for checking the face validity of the construct. The interview consisted in providing them with the table shown above and asking them to provide three score for each item in terms of (i) the extent to which it belonged to and was reflective of the key characteristic of the scientific method, (ii) the extent to which the description was clear and reflective of that characteristic, (iii) the extent to which the scoring guideline were effective in helping researchers score entrepreneurs. I also asked mentors whether they thought there were other items they felt that were worth including in the list of items to subject to CFA. The average value for each of the three scores was never below 6.4 (on a scale from 1 to 7 where 1 was the lowest value and 7 the highest) with an average of 6.7, and the mean standard deviation was never above 0.9 with an average of 0.4, suggesting the face validity test was satisfactory, i.e. their specification was perceived as clear and representative of the underlying construct. Only one additional item was proposed to be added, which was "speed of testing" in the third construct - empirical testing- which was proposed by two of the seven panelists. Consequently, I conducted a larger interview involving all of the seven panelists together and discussed this possibility of including this item in the final list. In the end, it was agreed that the suggested item should not be included in the final list because speed is not a necessary condition for a scientific approach. In fact, one could trade a faster test with a slower test, that allows to test a theory more precisely. Once again, this is reflective of the difference between the lean startup and the scientific approach. The latter pursues inferential precision potentially even at the expense of development speed.

3. Empirics

After face validity of the proposed scale was analyzed, I proceeded to its validation.

The objective was to run a confirmatory factor analysis where the 16 formative items would factor in 4 constructs. This would mirror the theoretical model by Cordova (2017), who describes the scientific approach as characterized by four main constructs each of which has four methodological characteristics.

Given my earlier study in Camuffo, Cordova and Gambardella (2017), I had to opportunity to sample startups from the group of more than 100 startups that were subject to a field experiment. The startups in the study were randomly divided into two halves which were subject to four-months general training on entrepreneurial experimentation (that is the importance of running tests that allow firms to gather signals as the likely success of their product/service) with the only difference that one group was also taught how to run experimentation by following some scientific guidelines, such as how to develop consistent theories and avoid biases in empirical testing. Using this sample represented an opportunity because, after confirmatory factor analysis is conducted, one can further test the construct validity of the scale by checking that there is a positive correlation between the scientific score and the dummy that identifies whether a startup had been treated or not. Exploiting the startups used by Camuffo et al. (2017) also has another advantage: startups which participated in Camuffo et al. (2017) were all early-stage startups at the beginning of the program, that is they had not launched their project on the market yet. The use of a scientific approach can be more easily detected in early-stage startups than later stage ventures, given that as time elapses firms are more likely to have developed sunk costs which makes it difficult to adopt a scientific approach. In fact, from Cordova (2017) we know that the scientific approach helps firms make better decisions such that, in cases of environments with higher likelihood of startup failure, for example, firms can be more likely to avoid false positives, thus dropout or pivot more, thus reduce the probability of failure than their counterparts. However, when firms have made sunk costs (e.g. paid a software agency to develop a full-fledged platform for their service), it is harder to leverage the results of scientific experiments because the cost of dropping out or pivoting increases.

I sent an email to all of the startups which participated in the Camuffo et al. (2017) study and asked their representative to take part to a 30-minutes interview which had the objective to re-trace their entrepreneurial path, the choices that they had made over time as this was of research interest for the University and for assessing the impact of the training that startups were provided with. Initially, 36 startups gave their availability, 16 in the treatment and 20 in the control. Unfortunately, the sample size was too small, especially considering that, for the purpose of conducting confirmatory factor analysis, startups with a scientific score of 0 in at least one of the four constructs of the scientific approach could not be used for CFA. In fact, a confirmatory factor analysis that tries to test a model of the scientific approach cannot include those observations that belong to non-scientific entrepreneurs, because the factor loadings will be affected by non-representative observations. For this purpose, I tried to oversample treated startups which were more likely to apply the scientific approach. I therefore cold-called all of the startups in the treatment group to obtain their permission for the interview. The final sample consisted of 47 startups, 26 treated and 21 in the control group. The description of the descriptive characteristics of the final sample and their sample statistics are displayed Table 2. These are the same baseline characteristics Camuffo et al. (2017) used in their study.

Table 2 approximately here

Oversampling on the treated does not lead to any bias for the CFA and its validity and reliability analysis, as the latter only looks at whether the sixteen items load well on the specified four factors and the overarching model fits the data well, but it may cause biases in the subsequent validity analysis of the scientific score. In fact, suppose that treated startups, which are more likely to be scientific, also have baseline conditions which makes them different from sampled control startups. Then when I analyze the correlations between the scientific approach score with, for example, dropout, pivot or performance (measures I will use for testing the predictive validity of the score), these correlations could be contemporaneously affected by the level of scientificness of startups but also the different baseline conditions among treated and control startups. For this reason, I ran balance checks between the oversampled treated startups and control startups. I follow Gelber et al. (2016) and ran reduced-form ordinary least squares (OLS) regressions of startup characteristics on a dummy for selection into the training. The results of the regressions are displayed in Table 3. None of the variables is correlated with treatment status, suggesting the two groups are balanced in observables.

Table 3 approximately here

Startups did not know they were scored so that their answer was not biased by the scoring grid, that is anchored toward those answers that the respondent would expect the interview to think are correct (Bertrain and Mullainathan, 2001). I personally conducted five recorded interviews which I then used as the basis to train five different research assistants who then conducted the remaining interviews. Their inter-rater reliability was checked on a random 20% of the cases and it was never recorded to be below 0,8. The researcher assistants were not aware of whether startups were treated during the field experiment conducted by Camuffo et al. (2017) nor they knew their financial performance, so that they could bias their scores based on ex-ante perceptions of the firms they were interviewing. In other words, the interview was double-blind, i.e. startups did not know they were being scored and interviewers did not know the performance or treatment status of the startups. Because the objective was to score the entrepreneur on each of the 16 items of the scientific approach, the interview questions were constructed with the aim to (i) understand how the entrepreneur behaved and decided how their business should proceed (ii) gather a score for each of the four constructs, that is to assess whether and how entrepreneurs developed their theory, whether and how they broke the theory down into a set of testable hypotheses, whether they experimented in a scientific manner and whether and how they analyzed results of experiments. In other words, the questions partially guided the entrepreneur in touching upon each of the four constructs. This was deemed necessary because of the degree of details that were necessary to provide a score to some of the items (e.g. the internal validity of the tests, whether the metrics used for analyzing experimental results were valid and reliable, the extent to which hypotheses were coherent with the theory, etc.), that made necessary to focus the attention of the entrepreneur on the most salient behaviors. Notice that the alternative would have been to ask more general questions and record whether the entrepreneur freely discussed about the four constructs. However, upon confrontation with the seven startup mentors, who were consulted for assessing the validity of the scale, it was unanimously agreed that this was going to be dispersive and that it would not have allowed to investigate the topic at hand.

The questions used to elicit desired responses are listed below:

1. Tell me about your startup idea and why you think it should be successful?
2. How did you have your startup idea?
3. When is it that you realized that this was worth investing in and why?
4. Did you run any test to make sure your idea, your product and your strategies were right? If yes, can you tell me about the last one you did? What is that you wanted to find out? What did you find out?

Proceeding from the top, the first question aims at elaborating what is the overarching theory of the entrepreneur, that is to why s/he thinks her/his idea should be successful. In this answer, the interviewer can particularly assess the extent to which theory is clear and detailed. The extent to which it is comprehensive, if not brought up autonomously by the entrepreneur, was investigated by means of ancillary questions such as "Were you focused on solving this customer problem all along or were there other problems you were thinking about solving for your customers?".

The second question tries to assess whether her/his theory is evidence-based or has remained limited to personal experience and partial evidence. Scientific entrepreneurs should discuss how the initial idea, which could have certainly raised from a personal experience, was then investigated further with the aim of understanding whether the problem existed for a wider audience of potential customers. Resort to official statistics or experiments, such as interviews or field observations, that provide evidence of the existence of the problem at a wider level, why it exists, whether people are actively looking for a solution to solve it, would provide support to the theory of the entrepreneur. For instance, if her/his idea is to create a service similar to Uber, an app to request the arrival of a car driver, and s/he theorizes that this is going to be successful because in her/his city taxes are too expensive, then s/he should look for, let's say, statistics on the comparative taxi fares in her/his city relative to representative other cities or, even better, conduct an interview in which s/he tries to find out how many people in her/his city are discouraged from using taxi services because of their excessive prices.

The third and fourth questions shall lead the entrepreneur to talk about the experiments s/he has done and how these contributed to her/his decision to invest or not in her/his business. The interviewer here will go in the details of how experiments were run. For example, following up on the previous Uber example, suppose the interviewee did customer interviews to test her/his theory. Then the interviewer would investigate how interviews were conducted, who were the sampled respondents, what kind of questions were asked, what it is that the entrepreneur was trying to learn (here the interviewer will assess whether the entrepreneur makes explicit hypotheses about customers' behavior, in which case further questions on how these were framed will be asked).

The extent to which an entrepreneur is data-driven in her/his analysis of experimental results and the development of her/his business is elicited instead from the last part of the fourth question ("What did you find out?"). In the Uber example, the interviewer would ask what it is that the entrepreneur felt s/he learnt from the interviews s/he did and high scientific scores would respond, for example, in the following way: "I was tracking the % of people in my city that, conditional on wanting to use car transportation services, which was my target customer, were discouraged from using taxi services for price-related reasons. So, I asked them to tell me about those situations in which they needed a car transportation service and to tell me what they did, if the used a taxi or not and if not why. I asked the interviewees to be recorded during the interview, so that I could listen to the conversation again later on. Then I made a list and ranking of all the top cited reasons why people were not using taxi services. Excessive price was indeed the top reason why people were not using these services. The average fare would cost 30 euros and customers felt they would pay a maximum of 25. From my calculation, drivers in my app, who are not professional taxi-drivers would charge 20% less than normal taxi fares because they do not have to repay their purchased taxi-license. At that point I made further tests, I spoke to about 10 potential taxi drivers for my app and asked them what it is the minimum they would accept charging for the average fair and they actually confirmed they would be willing to make a 25% discount". From this response, you can see that the entrepreneur had some specific metric s/he was targeting (% of people who found taxis to be too expensive), the metric which s/he recorded by generally investigating why potential customers did not use taxi services was valid (because it was indeed measuring whether people were discouraged by high taxi fares) and reliable (possibly because the interviewee is asking the customer about the last time s/he was looking for a taxi service, this may not be represented of her/his common behavior, so a follow-up question like “Does it occur frequently” would have increased the reliability of the answer – for this reason the interviewer may decide not to give top scores on the validity and reliability item of the scientific scale). The entrepreneur does a great job in collecting data and record them in a systematic way (the interviewer could further investigate this by asking the interviewee to show the file s/he has worked on in order to assess the actual orderliness of the data-collection and analysis) and even more in explaining the results of her/his tests by showing that s/he even analyzed the severity of the actual vs. desired fare gap currently observed in the market.

Confirmatory Factor Analysis

CFA enables an examination of the overall fit of a measurement model to a dataset. The model I advance is one where the 16 items factor on four constructs. Factor 1 is called Theory and includes the extent to which theory is clear, detailed, comprehensive and evidence-based. Factor 2 is called Hypotheses and includes the extent to which there are explicit hypotheses, hypotheses are coherent with the theory, specific and falsifiable. Factor 3 is called Empirical Test and refers to how much tests run by the entrepreneur are consistent with her/his theory, externally valid, refer to a representative sample and follow a rigorous procedure. Factor 4 includes the items referred to the extent to which analysis of experiments and firm progress is data-driven, data used in the analysis is valid and reliable, data is systematically collected and interpreted in an explicative way. Descriptive statistics on the 16 items collected via phone interviews are displayed below in Table 4.

Table 4 approximately here

As stated before, I exclude from the CFA analysis the observations that refer to non-scientific startups, that is those that obtained a score of 0 on at least one of the 4 constructs of the scientific approach, because the CFA attempts to verify the fit of the data to the pre-specified model of scientific approach to decision making and, as such, must include only observations that are relative to scientific entrepreneurs. Overall there are 15 non-scientific observations that are excluded from the final sample of 47 interviewed startups, therefore the CFA is run on 32 cases. I used Stata version 15.1 to perform the CFA. Results of the CFA are shown in Table 5 which reports standardized factor loadings and their standard error and the main statistics for goodness of fit of the model.

Table 5 approximately here

Notice that the results refer to a 4-factor model which also takes into account that there is a path between the error terms of three pairs of items: clear_theory and detailed_theory, systematic_data_collection and explicative_data_analysis, explicative_data_analysis and evidence_based_theory. The addition of this correlation to the model was motivated by the results of the analysis of the modification indices, which suggests additional paths that could be specified that could improve the fit of the model (Brown, 2006). Out of all suggested paths, which can be seen in Table 6 below, I considered the inclusion of only those for which it was theoretically sound to assume an error correlation.

Table 6 approximately here

In fact, it seems reasonable that the extent to which a theory is clear is akin to the extent to which it is detailed, because more details help rendering a theory clearer. Seemingly, if a startup collects data in a systematic and orderly way it automatically becomes easier to extract information from data (therefore the error terms of systematic_data_collection and explicative_data_analysis can be correlated) and, ultimately, use this information to update one's theory (thus it is sensible to assume there is also a positive correlation between the error terms of explicative_data_analysis and evidence_based_theory). Inversely, it is not clear why, for example, error terms of the items alternative_theories and valid_and_reliable_metric should be correlated, that is why a shock to the ability of seeing alternative theories shall be somewhat correlated with that of choosing valid and reliable metrics to monitor the results of a test. By adding these terms, the goodness of fit of the model slightly improves. The chi-square statistic falls from 127,84 (p.value=0,0231) to 115,5 (p.value=0,075), the Comparative Fit Index (CFI) raises from 0,92 to 0,94, the root mean square error of approximation (RMSEA) falls from 0,098 to 0,082, the standardized root mean square residual (SRMR) from 0,08 to 0,079 and, finally, the Akaike's Information Criterion (AIC) from 1049 to 1043.

Next, I follow Hamann, Schiemann, Bellora, and Guenther (2013) procedure for analyzing the validity and reliability of the 4-factor model. I present results from construct validity, reliability, convergent validity and discriminant validity.

Construct Validity

In order to analyze the construct validity, that is the extent, to which the 4-factor model theorized fits the data, I start by comparing the goodness of fit statistics of the retained model with well-defined cutoff criteria. As reported by Hamann, Schiemann, Bellora, and Guenther (2013), because goodness-of-fit indices are sensitive to sample size, definite cutoff criteria may yield a high Type I error (i.e., rejecting acceptable misspecified models) if they are too conservative (Marsh, Hau, & Wen, 2004). Therefore, I use the cutoff criteria for acceptable fits of the measurement model to the data. So, starting from the Chi-square statistics, the acceptable cutoff criteria is that the p-value is above 5%, which leads to failure to reject the null that there is no difference between the patterns observed in these data and the model specified. Therefore, according to the p-value of 7.5% (Table 5), the model seems to fit the data. However, small sample sizes may affect the p-value upwards *(*Schlermelleh-Engel et al. 2003*,* Vandenberg 2006*)*. For this reason, analysis of additional fit statistics was carried out. Always from Table 5, SRMR is observed to be below the acceptance threshold of 0,08, while the RMSEA is slightly above the line. However, the latter is influenced by small sample size as well, this time in the opposite sense –i.e. smaller sample sizes may lead to rejection of valid models (Chen, Curran, Bollen, Kirby, Paxton, 2009). On the last instance, the CFI index is above the 0,9 thresholds and takes the value of 0,94. Overall results seems to suggest there is a fit between the data and the 4-factor model.

Reliability

Item Reliability

Item reliability refers to the extent the single item is contributing in explaining the factor it refers to. To assess item reliability, the R^2 value that is associated with each item to factor equation is analyzed. This criterion measures the strength of the linear relationship between an indicator and its latent factor (Bagozzi & Baumgartner, 1994, p. 402). The level of acceptance reliability is an R^2 above 0,4. Table 7 reports the R^2 for each item.

Table 7 approximately here

There are only two items that do not meet the criterion: clear_theory and test_with_representative_sample. The explanation seems to lie in the fact that, as shown in Table 4, average scores for these items are slightly higher than the rest (although this is true for consistent_test also). By re-analyzing the content of the interviews, it appears that most entrepreneurs can indeed explain the core of their theory and test with a somewhat[14] representantive sample. This may render these items less unique to a scientific approach relative to the other items. For making sense of whether the model would be improved by exclusion of one or both of these items, I compared the model fit statistics of the original model with three nested models: the first excludes the clear_theory from the analysis, the second excludes the test_with_representative_sample, the latter both of them. A priori, I set as decision rule for approving the removal of these items, that all of the goodness of fit statistics should improve relative to the original model. This is because the items used in this paper are to be interpreted as

[14] I use the term "somewhat" to refer to the fact that we are still talking about an average score of 3 on the 1-5 likert scale.

formative measures, that is they combine to form the latent constructs (as opposed to reflexive measures, where each item is views as an imperfect reflection of the underlying construct). In these cases, even low factor loadings may not justify the exclusion of proposed items because this may result in the elimination of precisely those items that are most likely to alter the empirical and conceptual meaning of the construct (MacKenzie, Podsakoff P., Podsakoff N., 2011). The established condition was not met, the alternative models did not perform better than the original one, a fact which led to the retention of the original model. The statistics of the four models are shown in Table 8.

Table 8 approximately here

Construct reliability

Continuing with the retained model, I analyze construct reliability, which represents the proportion of systematic variance in the set of items that belong to the same factor and the acceptance level is a value greater than 0,6 (Bagozzi & Baumgartner, 1994, p.403). All values for scale reliability are above the threshold: Theory = 0,84, Hypotheses=0,86, Empirical Test=0,86, Analysis=0,84. Results are shown in Table 9, which in addition to providing the scale reliability coefficient (which in the Table corresponds to the alpha number for the test scale, that is the Cronbach alpha of the construct), it also decomposes it in its different components, that is the item-test correlation, item-rest correlation and average interitem covariance.

Table 9 approximately here

Average variance extracted

Average variance extracted measures the amount of variance in a set of indicators that is accounted for by the latent factor in the model (Fornell & Larcker, 1981, pp. 45-46) and the acceptable level is above 0,5. This is calculated by dividing the sum of the squared factor loadings for the number of items in the single factor. AVEs for Theory, Hypotheses, Empirical Test and Analysis are respectively: 0,56, 0,62, 0,63, 0,74. As explained by Hamann, Schiemann, Bellora, and Guenther (2013), this is calculated by taking the sum of the squared factor loadings.

Convergent Validity

Convergent validity, which measures the extent to which the items actually correlate with their factor, is assessed by looking at whether the standardized factor loadings have all the theoretically predicted sign and have an estimate above 0,5. From Table 5, we can see all items correlated positively with their factor, which we expect given that the scale was constructed in a way that a higher score for the single item would imply a higher value for the factor it refers to, e.g. if an entrepreneur is more detailed in the way s/he expresses her/his theory, the higher the overall construct of THEORY; seemingly, a higher value for, let's say, systematic data collection should increase the overall score of ANALYSIS. Additionally, all loadings are above 0,5.

Discriminant validity

Last but not least, we can assess the discriminant validity of the constructs which aim at analyzing the extent to which the factors do indeed measure different constructs. For this purpose, I use the Fornell-Larcker criterion which compares the average variance extracted for a factor to the squared correlations among factors. If the former is greater than the squared correlations in all cases, then this is a strong indicator of discriminant validity (Fornell & Larcker, 1981). Results are displayed in Table 10.

Table 10 approximately here

Not surprisingly, none of the 4 factors show full discriminant validity. In fact, correlations among factor is quite high, ranging from 0,76 to 0,94. This does not come as a surprise because, as specified in Cordova (2017), the scientific approach is the approach by which entrepreneurs generate, collect and interpret market signals, i.e. the entrepreneur uses all of the 4 constructs in a unique process of hypothesis testing. For the same reasoning, entrepreneurs who do not use one of the four factors were excluded from the CFA analysis and were recorded with a scientific score of 0. Thus, I conclude that the scientific approach is a one-dimension consisting of 4 constructs, pretty much like other researchers have concluded in their scale validation, an example of which is Narver and Slater (1990). In their paper, they found the three behavior components of market orientation, that is customer orientation, competitor orientation and interfunctional coordination, were indeed all positively correlated, with correlations ranging from 0,72 up to 0,91, such that they ultimately use one average measure of market orientation as representative of the higher-level construct. I use the same approach to obtain a score of scientific approach.

4. Scientific approach score

The scientific approach score is one measure reflective of the extent to which entrepreneurs behave scientifically across the four constructs of the scientific scale, that is the extent to which s/he uses the 16 items of the scale in a more or less scientific way. An important question in obtaining a score for the scientific approach is whether a simple average value for the four constructs that make up the scientific approach would be a sensible measure. While it is true that each of the four factors are equally important for the scientific approach, as pointed out in Cordova (2017), it is also true that we need to take into account value-dispersion. In other words, suppose you compare two entrepreneurs who are both scientific, in the sense they use all of the four constructs of the scientific approach, but in different ways: one has average levels for each of the four constructs, while another has high levels for some of the constructs and low levels for the others –i.e. the former is equally good in crafting theories, defining hypotheses, running experiments and analyzing their results, while the latter is better than the former at, let's say, theorizing, while less good at running experiments. Then, which of the two entrepreneurs should be recognized as more scientific? For the same principle that a scientific entrepreneur must use all of the four constructs of the approach to be defined as such, I propose that entrepreneurs with a lower variance across the 4 constructs, are to be viewed as more scientific. For this reason, I construct a weighted average of the four constructs, which is equal to the simple average divided by the standard deviation of the score of the four constructs (this is basically the inverse of the coefficient of variation). The subscript i refers to the fact there is one score for each startup in the sample.

$$scientific\ simple\ average_i = \frac{theory_i + hypotheses_i + empirical\ test_i + analysis_i}{4}$$

$$\boldsymbol{scientific\ weighted\ average_i} = \frac{scientific\ simple\ average_i}{standard\ deviation\ of\ (theory_i + hypotheses_i + empirical\ test_i + analysis_i)}$$

While I already tested the validity and reliability of the 4-factor model, I proceeded with further validity analysis on the single measure of the scientific score. More specifically, from Cordova (2017), we know that in uncertain environments in which ideas are randomly distributed across startups and the actual likelihood of startup success is low, more scientific startups are going to do on the whole more dropouts and pivots. Also, in uncertain environments scientific entrepreneurs record higher profits. Therefore, we expect that a measure of the scientific approach will be positively correlated with the rate of dropout and pivot, and the profitability of firms. During the interviews to startups I therefore collected data on whether startups were still active at the time of the interview, the number of core changes to their business model they had done over time and the cumulative amount of profits accrued to date and used this data to test the predictive validity of the scale of the scientific approach. During the same interview I also collected data for testing concurrent and discriminatory validity. For the latter, I collected three variables: the average level of education of the startup team, the team average years of working experience in previous startups and in managerial positions tasked with launching new products/services. The idea is that the scientific approach should be something different than general ability and experience in managing innovation. The hypothesis is therefore that there is no correlation between these three scores and the scientific score of a startup. For testing concurrent validity, instead, I measured the entrepreneurs' spending attitude, that is their self-report of how much they believe startups need investment to generate demand as opposed to the belief that demand can be generated with little investment if the entrepreneur knows how demand behaves. Startups who are more scientific should believe investment is a less needed condition for growing their business. Finally, I also control that the level of scientificness is positively correlated with whether startups were treated in Camuffo et al. (2017), which we should expect given the treatment incentivized people to follow a more scientific approach. The complete description and measurement of all the variables used in the validity analysis used and their expected correlation with the scale of the scientific approach are displayed in Table 11, which also presents their descriptive statistics.

Table 11 approximately here

I use reduced form least squares regression equations where the scientific score is used as independent variable and the variables used for the validity analysis are used as dependent variables. Results are shown sequentially in Table 12. Notice that while I use the weighted average score of the scientific approach, results are consistent with the simple average specification.

Table 12 approximately here

Predictive validity is confirmed: the scientific approach significantly increases the number of pivots while there is no statistically significant difference in terms of dropout. As previously explained, it is not necessary that scientific firms do both more pivots and dropouts because pivots are a partial substitute of dropouts. Profits are higher for firms adopting a scientific approach, however the difference is only significant at a 11.2% level. This is possibly because the number of observations is low and the fact that scientific startups have made more pivots could delay their time to profits, especially if we can only observe performance of startups after a limited time since the beginning of their entrepreneurial projects, which is our case (just about one year and a half have elapsed since their inception). That said, more evidence is warranted in this respect. Concurrent validity is confirmed at a 3.5% significance level: scientific startups tend to be more likely to believe that investment is not as much a necessary condition for success than knowledge of customer demand. Analysis of discriminant validity confirms that the scientific approach is something different than general ability of experience as it is neither correlated with education nor with experience in innovative contexts. Finally, the positive and significant correlation with the dummy treatment provides further support for the construct validity of the scientific score.

5. Conclusions

Empirical research on experimentation and, more specifically, the benefits of a scientific approach to decision making and ultimately firm performance, requires the use of a validated scale for the scientific approach. This paper aimed at testing the theoretical 4-factor model described in Cordova (2017) by means of a confirmatory factor analysis. Analysis of goodness of fit statistics and robustness checks showed that the 4-factor model fits the data. Discriminant validity of the model showed that the 4 constructs are highly correlated among each other, suggesting that they do indeed reflect a higher order dimension, that of the scientific approach, which prescribes the use of all of the 4 factors in the process of generating, collecting and interpreting signals as to the potential return of an entrepreneurial project. A one summary score for the scientific approach was then calculated that weights the average score for each of the four factors by the standard deviation among them, a way to "value more", in terms of scientificness, those entrepreneurs who are shown to be equally good in using all of its four constructs. Further analysis of the one summary score shown that the scientific approach predicts entrepreneurial behaviors in terms of pivot, dropout and profits, consistently with the existing literature, and that it is different from general ability or experience. Additionally, the results also provide hints to potentially interesting future research. For example, as shown in Table 4, entrepreneurs seem to be better at conducting some methodological characteristics of the method rather than others. This is interesting as it may have either theoretical and practical implications. From the former standpoint, one could wonder why this is the case. Perhaps, some aspects of the scientific method are less understood or have higher costs. From an empirical standpoint, we ought to make sure entrepreneurs absorb the approach thoroughly on all its characteristics, especially if our view is that scientific entrepreneurs should be equally good at the following the different characteristics of the method. Alternatively, one could think that when we move away from the individual entrepreneurial context to the team level, skill heterogeneity may help achieve the same purpose. Another interesting direction for future research is also dig deeper in the concept and measurement of non-scientific startups. In the current study, in fact, startups which showed not to use at least one of the constructs of the scientific approach were assessed to be non-scientific. This is consistent with the theoretical idea that is exactly the use of whole the 4 constructs that differentiate a scientific entrepreneur from other entrepreneurs. However, it would be interesting to classify the different non-scientific entrepreneurs in different typologies, measure their own scale and compare their decision-making process and efficacy with that of scientific entrepreneurs. Finally, of course, there are limitations to the current study that could be addressed in future research, in particular the limited sample size on which the CFA relied. Alternatively, other scholars could further confirm the predictive validity of the scientific score on firm profitability and on variables which are reflective of the antecedents to the scientific approach, which would allow to further test the consistency of the scientific score validated in this paper to the theory and propositions advanced in Cordova (2017).

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.

Ansoff, H. I. (1965) *Corporate Strategy: An Analytic Approach to Business Policy for Growth and Expansion*, McGraw-Hill, New York NY.

Andrews, K. R. (1971) *The Concept of Corporate Strategy*, Richard D. Irwin, Homewood IL. 30

Bagozzi, R. P., & Baumgartner, H. (1994). *The evaluation of structural equation models and hypothesis testing*. In R. P. Bagozzi (Ed.), Principles of marketing research (pp. 386-422). Cambridge, MA: Blackwell.

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.

Bertrand, Marianne, and Sendhil Mullainathan, (2001) "Do People Mean What They Say? Implications for Subjective Survey Data," *American Economic Review Papers and Proceedings*, 91 (2), 67–72.

Brown T. A., (2006). *Confirmatory Factor Analysis for Applied Research.* The Guildford Press.

Camuffo A., Cordova A., Gambardella A., (2017). "A scientific approach to entrepreneurial decision making: evidence from a randomized controlled trial". Early draft available at https://events.insead.edu/doriot-entrepreneurship-conference/documents/A-Scientific-Approach-to-Entrepreneurial-Experimentation_%20Evidence%20from%20a%20Randomized%20Control%20Trial.pdf

Chen F, Bollen KA, Curran PJ, Paxton P, Kirby J., (2001) "Improper Solutions in Structural Equation Modeling: Causes, Consequences and Strategies". *Sociological Methods and Research*, 29:468–508

Cordova A., 2017. *A scientific approach to entrepreneurial decision-making: towards a conceptualization*. Thesis Chapter.

Fornell, C., & Larcker, D. F. (1981). "Evaluating structural equation models with unobserveable variables and measurement error". *Journal of Marketing Research*, 18(1), 39-50.

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

Ghosh S., (2012) *The Venture Capital Secret: 3 Out of 4 Start-Ups Fail*. The Wall Street Journal 20/09/2012.

Grandori, A. (2010) "A Rational Heuristic Model of Economic Decision-Making," *Rationality and Society* 22 (4): 1–28.

Hamann, P. M., Schiemann, F., Bellora, L., & Guenther, T. W. (2013). Exploring the dimensions of organizational performance: a construct validity study. Organizational Research Methods, 16(1), 67-87. http://dx.doi.org/10.1177/1094428112470007

MacKenzie, S. B., Podsakoff, P.M., & Podsakoff, N. P. (2011). "Construct measurement and validation procedures in MIS and behavioral research: integrating new and existing techniques". *MIS Quarterly*, 35(2), 293-334.

Marsh, H. W., Hau, K.-T., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler's (1999) findings. *Structural Equation Modeling*, 11(3), 320-341. doi:10.1207/s15328007sem1103_2

McGrath, R. G. and MacMillan, I. C. (1995) *Discovery Driven Planning*, Harvard Business Review, 73 (4), 44-54

Narver J. C., Slater S. F., 1990. "The effect of a market orientation on business profitability". *Journal of Marketing*. Vol. 54, No. 4 (Oct., 1990), pp. 20-35

Schermellehengel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the Fit of Structural Equation
Models: Tests of Significance and Descriptive Goodness-of-Fit Measures. *Methods of Psychological Research Online,* 8(2), 23–74

Shook, C. L., R. L. Priem, J. E. McGee. 2003. "Venture creation and the enterprising individual: A review and synthesis". *Journal of Management*, 29 360-379.

Sull, D.N. (2004) "Disciplined Entrepreneurship," *MIT Sloan Management Review*, 46 (1), 71.

Tripsas M., Fiona M., (2004). *The Exploratory processes of entrepreneurial firms: The role of purposeful experimentation*, in Joel A.C. Baum, Anita M. McGahan(ed.) Business Strategy over the Industry Lifecycle (Advances in Strategic Management, Volume 21) Emerald Group Publishing Limited, pp.45 – 75 Vandenberg, R. J. (2006), "Statistical and Methodological Myths and Urban Legends", *Organizational Research Methods*, Vol. 9, No. 2, pp. 194-201.

Zenger T.R. (2015), *Trial and Error Is No Way to Make Strategy*, Harvard Business Review, April 24, 201

TABLE 1

The 16 items of the scientific approach, their description and scoring system

The 4 key characteristics of the scientific approach	The 12 methodological characteristics of the approach	description of the items	scoring 0	1	2	3	4	5
THEORY	clear	one can easily understand the theory of the entrepreneur, it is not logically fallible or confusing	the entrepreneur explains what he wants to do but does not articulate why her/his market solution is needed and will be successful	the entrepreneur poorly explains the core of his theory (i.e. it is confusing)		the entrepreneur explains the core of her/his theory in a somwhat good way (i.e. it is not confusing but also not perfectly clear)		the entrepreneur explains the core of her/his theory in a perfectly scientific way (i.e. it is fully clear)
	detailed	it is well articulated, it contains a lot of details		the entrepreneur exposes his theory in a very poor way (i.e. few details)		the entrepreneur exposes his theory in a somewhat good way (i.e. a few details but not all)		the entrepreneur exposes his theory in a perfectly scientificway (i.e.full details)
	comprehensive	the entrepreneur consideres and reasons on alternative options		the entrepreneur consideration of alternative theories is very poor (i.e. few alternative theories)		the entrepreneur consideration of alternative theories is somewhat good (i.e. a few alternative theories but not all)		the entrepreneur consideration of alternative theories is perfectly scientific (i.e. all alternative theories are convincingly analyzed)
	evidence-based	it is based on hard facts, population statistics or results of experiments conducted by the entrepreneur over time. Personal experience is not sufficient		the entrepreneur's theory relies on poor data (i.e. anecdotal evidence only)		the entrepreneur's theory relies on somewhat good data (i.e. uses little or partial data)		the entrepreneur's theory relies on perfectly scientific data (i.e. such official statistics or results of rigorously run experiments)
HYPOTHESES	explicit	theory is delineated in a set of hypotheses. The entrepreneur is able to cite them	the entrepreneur does not have explicit hypotheses	the entrepreneur has poorly explicited hypotheses (i.e. misses key ones)		the entrepreneur has somewhat good explicited hypotheses (i.e. considers key ones but not all)		the entrepreneur has perfectly good explicited hypotheses (i.e. considers all key ones)
	coherent with theory	hypotheses and theory are aligned, a set of hypotheses can be retraced back to one theory only		hypotheses are specified in a way that poorly reflects the theory of the entrepreneur (i.e. most hypotheses are not coherent with the theory)		hypotheses are specified in a way that reflects the theory of the entrepreneur in a somewhat good way (i.e. not all hypotheses are coherent with the theory)		hypotheses are specified in a way that reflects the theory of the entrepreneur in a perfectly scientific way (i.e. all hypotheses are coherent with the theory)
	precise	allow to test one thing at a time without confounding effects		hypotheses are defined in a poorly precise way (i.e. most are defined in a complex way so that it is difficult to generate learning from their test)		hypotheses are defined in a somewhat precise way (i.e. some hypotheses are too complex)		hypotheses are defined in a perfectly precise way (i.e. all hypotheses are precise and allow to generate learning)
	falsifiable	it can be proven wrong, it contains thresholds that make it falsifiable		the entrepreneur has defined poor thresholds for falsifiability (many hypotheses do not have thresholds and/or are created with no precise reasoning)		the entrepreneur has defined somewhat good thresholds for falsifiability (few hypotheses do not have thresholds and/or are created with no precise reasoning)		the entrepreneur has defined perfectly scientific thresholds for falsifiability (all hypothese have thresholds and/or are created with precise reasoning)
EMPIRICAL TEST	coherent with theory	the test chosen actually allows to test the specified hypothesis	the entrepreneur executes her/his own plan without any experimental approach	tests run by the entrepreneur poorly test her/his theory (i.e. most times other tests should have been used to test that theory)		tests run by the entrepreneur test her/his theory in a somewhat good way (i.e. few times other tests should have been used to test that theory)		tests run by the entrepreneur test her/his theory in a perfectly scientific way (i.e. all times the right tests are used to test that theory)
	externally valid	the environment chosen for the test is representantive of the external reality		the environment chosen by entrepreneurs for testing is poorly representantive of the real external environment (i.e. the environment is clearly different from what happens in reality and this is not taken into account by adjusting results for such potential biases)		the environment chosen by entrepreneurs for testing is somewhat representantive of the real external environment (the environment has some differences with what happens in reality and not all these are taken into account by adjusting results for such potential biases)		the environment chosen for testing by entrepreneurs is representantive of the real external environment in perfectly scientific way (there is no difference between the environment and reality or these are taken into account by adjusting reults for such potential biases)
	based on representative sample	the sample of people on which the test is run is representantive of the target population		the sample is poorly representative of the target population (i.e. key characteristics of the sample are different from that of the population and they are not taken into account by adjusting results for potential biases)		the sample is somewhat representative of the target population (i.e. some characteristics of the sample are different from that of the population and they are not always taken into account by adjusting results for potential biases)		the sample is representative of the target population in a perfectly scientific way (i.e. all characteristics of the sample are the same as that of the population or if there are differences, they are taken into account by adjusting results for potential biases)
	rigorous	the test has a clear identification strategy and the entrepreneur minimizes biases		tests are ill-designed and produce evident biases (i.e. the identification strategy does not allow to tease our the desired effect)		tests are well-designed but there are some biases (i.e. the identification strategy does not completely allow to tease our the desired effect)		tests are designed in a perfectly scientific way and produce no biases (i.e. the identification strategy allows to tease our the desired effect)
ANALYSIS	data-driven	the entrepreneur specifies metrics that are used to evaluate the falsification of her/his theory	the entrepreneur gathers no data to evaluate the success of her/his actions	the entrepreneur chooses poor metrics for validating her/his theory (i.e. key metrics are missing for explaining the theory of the entrepreneur)		the entrepreneur chooses somewhat good metrics for validating her/his theory (i.e. some metrics are missing for explaining the theory of the entrepreneur)		the entrepreneur chooses the metrics for validating her/his theory in a perfectly scientific way (i.e. all metrics are for explaining the theory of the entrepreneur are specified)
	based on valid and reliable metrics	chosen metrics are reflective of the theoretical construct the entrepreneur wants to measure and they produce similar results under consistent conditions		elected metrics are poorly valid and reliable (i.e. most metrics do not measure what they tend to measure and/or provide consistent results across conditions)		elected metrics are somewhat valid and reliable ((i.e. some metrics do not measure what they tend to measure and/or provide consistent results across conditions)		elected metrics are perfectly valid and reliable (i.e. all metrics measure what they tend to measure and provide consistent results across conditions)
	based on systematic data collection	the entrepreneur has systems/routines in place to collect and monitor data at regular time intervals		the entrepreneur collects the data in a poorly systematic and organized way (i.e.data collection is very disordered and does not cover all metrics)		the entrepreneur collects the data in a somewhat good systematic and organized way (i.e.data collection is not fully ordered and does not cover all metrics)		the entrepreneur collects the data in a perfectly scientific systematic and organized way (i.e.data collection is very ordered and covers all metrics)
	explicative	the entrepreneur draws meaningful implications for her/his theory from the data s/he collects		data is analyzed in a poorly rigorous and meaningul way (i.e. only few aspects of the experimental results are analyzed and clearly analyzed)		data is analyzed in a somehat good rigorous and meaningul way (i.e. Not all aspects of the experimental results are analyzed and clearly analyzed)		data is analyzed in a perfectly scientific rigorous and meaningul way (i.e. all aspects of the experimental results are analyzed and clearly analyzed)

TABLE 2

Definition of descriptive variables

descriptive variable names	description	measurement
treatment	it defines whether the startup had been subject to a treatment that favors the use of a scientific approach	it is a dummy variable taking value 1 if the startup was treated in the field experiment of Camuffo et al. (2017); 0 if it was in the control arm
team size	it is the size of the startup founding team	it is the absolute number of team memebrs of the startup
lombardy	it identifies where the majority of the founders come from geographically speaking	it is a dummy that takes value 1 if the majority of team founders comes from the italian region of Lombardy, which was phisically closer to the place where the field experiment was held; 0 otherwhise
internet_or_mobile	it identifies the sector in which the startup operates	it is a dummy that takes value 1 if the startup operates in either an internet or mobile sector as opposed to any other sector. Internet & mobile represent the great majority of projects in Camuffo et al. (2017)
idea_stage	it describes the stage of development of the startup upon inception of the field experiment	it is a dummy that takes value 1 if the statup started the field experiment by Camuffo et al. (2017) by havnig only a business idea in mind; 0 if it had already started workin on the project but still had not launched it on the market yet. Startups that take value 0 were slightly ahead with their work than those take value 1

Sample statistics

Descroptive statistics of sampled startups					
	obs	**mean**	**Std.Dev**	**Min**	**Max**
treatment	47	0,55	0,50	0	1
team size	47	0,29	0,13	1	7
lombardy	47	0,34	0,48	0	1
internet_or_mobile	47	0,72	0,45	0	1
idea_stage	47	0,62	0,49	0	1

TABLE 3

Balance checks between treated and control startups sampled from Camuffo et al. (2017)	
descriptive characteristic	**estimated beta value and standard error from reduced form OLS regression of each of the descriptive characteristics on the treatment dummy**
team size	0,53 (0,57)
lombardy	0,01 (0,14)
internet_or_mobile	0,08 (0,14)

idea_stage	0,1
	(0,13)
n.obs	47

TABLE 4

Descriptive statistics of the sixteen items of the scientific approach for scientific startups only N.obs=32				
Variable	**Mean**	**Std.Dev.**	**Min**	**Max**
clear_theory	3	0,72	2	4
detailed_theory	2,81	0,78	2	4
comprehensive_theory	1,88	0,83	1	4
evidence_based_theory	2,63	1,18	1	5
explicit_hypotheses	2,34	0,60	2	4
coherent_hypotheses	2,09	0,86	1	4
precise_hypotheses	1,72	0,81	1	4
falsifiable_hypotheses	1,63	0,91	1	4
consistent_test	3	0,84	2	5
externally_valid_test	2,72	0,99	1	5
tests_with_representantive_sample	3	0,80	2	5
rigorous_procedure	2,16	0,85	1	5
data_driven	2,50	0,72	1	4
valid_and_reliable_metrics	2,41	0,95	1	5
systematic_data_collection	2,72	1,30	1	5
explicative_data_analysis	2,56	1,22	1	5

TABLE 5

4 FACTOR MODEL CFA results		
		STANDARDIZED FACTOR LOADINGS
THEORY		
	clear_theory	0,56***
	detailed_theory	0,73***
	comprehensive_theory	0,97***
	evidence_based_theory	0,68***
HYPOTHESES		
	explicit_hypotheses	0,73***
	coherent_hypotheses	0,74***
	precise_hypotheses	0,89***
	falsifiable_hypotheses	0,78***
EMPIRICAL TEST		
	consistent_test	0,81***
	externally_valid_test	0,8***
	tests_with_representantive_sample	0,55***
	rigorous_procedure	0,97***
ANALYSIS		
	data_driven	0,86***
	valid_and_reliable_metrics	0,95***
	systematic_data_collection	0,64***
	explicative_data_analysis	0,64***
		FIT STATISTIC
	CHI-SQUARE (p.value>0,05)	115,5 (0,075)
	CFI (> 0,90)	0,95
	RMSEA (< 0,08)	0,082
	SRMR (< 0,08)	0,079

*** p<0.01, ** p<0.05, * p<0.1

TABLE 6

Modification indices **(only those that reduce chi-square by at least 3 are shown)**	**Reduction in Chi Square statistic**
cov(e.detailed_theory,e.explicit_hypotheses)	3,58
cov(e.comprehensive_theory,e. data_driven)	3,42
cov(e. comprehensive_theory,e.valid_and_reliable_metric)	4,28
cov(e. comprehensive_theory,e.coherent_hypotheses)	6,39
cov(e. comprehensive_theory,e.explicative_dataanalysis)	4,01
cov(e.explicit_hypotheses,e.consistent_test)	3,303
cov(e.precise_hypotheses,e. data_driven)	4,37
cov(e.precise_hypotheses,e.systematic_datacollection)	5,84
cov(e.hypotheses_with_thresholds,e. data_driven)	3,97
cov(e.tests_with_representativesample,e. data_driven)|	3,79
cov(e. data_driven,e.valid_and_reliable_metric)	8,55
cov(e. data_driven,e.systematic_datacollection)	5,25
cov(e.systematic_datacollection,e.explicative_dataanalysis)	6,16

TABLE 7

R-square from each item to factor equation		
		R^2
THEORY		
	clear_theory	0,31
	detailed_theory	0,54
	comprehensive_theory	0,94
	evidence_based_theory	0,46
HYPOTHESES		
	explicit_hypotheses	0,53
	coherent_hypotheses	0,55
	precise_hypotheses	0,8
	falsifiable_hypotheses	0,62
EMPIRICAL TEST		
	consistent_test	0,66
	externally_valid_test	0,64
	tests_with_representantive_sample	0,3
	rigorous_procedure	0,93
ANALYSIS		
	data_drivens	0,74
	valid_and_reliable_metrics	0,91
	systematic_data_collection	0,41
	explicative_data_analysis	0,41

TABLE 8

COMPARATIVE FIT STATISTICS				
	original 4 factor model	**4 factor model without clear_theory**	**4 factor model without test_with_representative_sample**	**4 factor model without both clear_theory and test_with_representative_sample**
CHI-SQUARE (p.value>0,05)	115,5 (0,075)	102,76 (0,0603)	102,77 (0,052)	90,42 (0,043)
CFI (> 0,90)	0,945	0,942	0,94	0,0939
RMSEA (< 0,08)	0,082	0,089	0,092	0,098
SRMR (< 0,08)	0,079	0,074	0,077	0,071
AIC	1043,1	972,2	981,2	910,2

TABLE 9

Construct reliability		**Obs**	**Sign**	**item-test correlation**	**item-rest correlation**	**average interitem covariance**	**alpha**
THEORY							
	clear_theory	32	+	0,77	0,64	0,55	0,83
	detailed_theory	32	+	0,83	0,72	0,49	0,79
	comprehensive_theory	32	+	0,86	0,76	0,45	0,77
	evidence_based_theory	32	+	0,88	0,71	0,37	0,82
	test scale					*0,46*	***0,84***
HYPOTHESES							
	explicit_hypotheses	32	+	0,8	0,7	0,48	0,85
	coherent	32	+	0,84	0,7	0,38	0,83
	precise_hypotheses	32	+	0,88	0,77	0,37	0,8
	hypotheses_with_thresholds	32	+	0,87	0,4	0,35	0,82
	test scale					*0,4*	***0,86***
EMPIRICAL TEST							
	consistent_test	32	+	0,86	0,74	0,47	0,82
	externally_valid_test	32	+	0,89	0,78	0,4	0,8
	tests_with_representantive_sample	32	+	0,71	0,52	0,59	0,9
	rigorous_procedure	32	+	0,91	0,84	0,42	0,78
	test scale					*0,47*	***0,86***
ANALYSIS							
	data_driven	32	+	0,76	0,66	0,83	0,83
	valid_and_reliable_metrics	32	+	0,87	0,77	0,64	0,76
	systematic_data_collection	32	+	0,86	0,69	0,56	0,8
	explicative_data_analysis	32	+	0,85	0,7	0,58	0,79
	test scale					*0,65*	***0,84***

TABLE 10

DISCRIMINANT VALIDITY AVEs on the diagonal, correlations below the diagonal				
	THEORY	HYPOTHESES	EMPIRICAL TEST	ANALYSIS
THEORY	**0,56**			
HYPOTHESES	0,86	**0,62**		
EMPIRICAL TEST	0,57	0,87	**0,63**	
ANALYSIS	0,69	0,88	0,74	**0,74**

TABLE 11

Description of the variables used for validity analysis of the scientific scale

validity variable names	description	measurement	expected correlation with scientific approach
pivots	number of core chances to the business model	it is a count a variable that takes value 1 for each change of focus by the entrepreneur on a different customer problem or customer target	positive
dropout	operative status of the firm at the time of the interview	it is a dummy variable which takes value 1 in case firms have stopped working at their project at the time of the interview, 0 otherwise	positive or null
profits	it is the financial performance recorded by the firm at the time of the interview	it is the difference between the cumulative revenue and expenditures sustained by the firm from the inception up to the time of the interview. All costs are included	positive
education	it is the average level of education achieved by the team	the interviewed startup representantive is asked to report the highest school achievement of the team founders to whom a score is assigned: 1 for middle school, 2 for high school, 3 for bachelor, 4 for master, 5 for MBA and 6 for PhD. The team average is calculated.	null
exp_startup	it is the average working experience in startups	the interviewed startup representantive is asked to report the years of work experience of team founders in other startups other than the one under interview. The team average is then calculated.	null
exp_manager	it is the average working experience in managerial positiions linked to the launch of innovative products and services	the interviewed startup representantive is asked to report the work experience that team founders had in any managerial position where their task was to coordinate the launch of new products and services. The team average is then calculated.	null
spending attitude	it is the entrepreneurs' attitude towards spending	the interviewed startup representantive is asked to state the extent to which s/he agrees or disagrees (on a scale from 1 to 7 when 1 is strongly disagree and 7 strongly agree) with the following statements. 1) An entrepreneur must embrace risk and invest to generate demand. There is no investment without demand; 2) An entrepreneur has to be cautious and invest only after having understood whether there exists a demand, how big this is, how customers behave. If s/he knows the demand well, little investment is needed. Then the difference between the second and the first response is calculated.	negative

Descriptive statistics for variables used in the validity analysis of scientific score				
Variable	Mean	Std. Dev.	Min	Max
dropout	0,41	0,49	0	1
pivot	9,42	0,74	0	3
profit	-6.756	33057	-120000	150000
education	4	0,84	2	6
exp_start	1	3	0	18
exp_manager	2	5	0	20
spending_attitude	-0,98	2	-6	3

TABLE 12

Reduced OLS Variables in column are the dependent variables, the weighted average scientific score is the independent variable								
	dropout	pivot	profit	spending attitude	average education	average startup exp	average mgmt exp	treament
beta coefficient	-0,0029037	0,0919122	2126	-0,21	0,017	0,056	-0,08	0,2
standard error	0,05	0,03	1310,4	0,09	0,03	0,13	0,19	0,07
p-value	0,955	0,001	0,112	0,035	0,0624	0,067	0,665	0,003
N	47	47	47	47	47	47	47	47
dropout, pivot, profit used for predictive validity; spending attitude used for concurrent validity; average education, average startup exp and average mgmt exp used for discriminant validity; treatment used for further construct validity NB: all OLS regressions except for dropout and treatment which are dummies and for which a probit model was used								