DISCUSSION PAPER SERIES

DP15972
(v. 3)

A Scientific Approach to Innovation Management: Theory and Evidence from Four Field Experiments

Arnaldo Camuffo, Alfonso Gambardella, Danilo Messinese, Elena Novelli, Emilio Paolucci and Chiara Spina

ORGANIZATIONAL ECONOMICS

CEPR

ISSN 0265-8003

A Scientific Approach to Innovation Management: Theory and Evidence from Four Field Experiments

Arnaldo Camuffo, Alfonso Gambardella, Danilo Messinese, Elena Novelli, Emilio Paolucci and Chiara Spina

Discussion Paper DP15972
First Published 27 March 2021
This Revision 27 March 2022

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

A Scientific Approach to Innovation Management: Theory and Evidence from Four Field Experiments

Abstract

This paper studies the implications of an approach in which managers and entrepreneurs make decisions under uncertainty by formulating and testing theories such as scientists do. By combining the results of four Randomized Control Trials (RCTs) involving 754 start-ups and small-medium enterprises and 10,730 data points over time, we find that managers and entrepreneurs who adopt this approach terminate more projects, do not experiment with many new ideas, and perform better. We develop a model that explains these results.

JEL Classification: L21, L26, M13, M21

Keywords: Scientific entrepreneurship, Management Practices, field experiments, strategic decision-making, uncertainty

Arnaldo Camuffo - arnaldo.camuffo@unibocconi.it
Department of Management and Technology & ICRIOS, Bocconi University, Milan

Alfonso Gambardella - alfonso.gambardella@unibocconi.it
Department of Management and Technology & ICRIOS, Bocconi University, Milan and CEPR

Danilo Messinese - danilo.messinese@unibocconi.it
Department of Management & Technology & ICRIOS, Bocconi University, Milan

Elena Novelli - elena.novelli.1@city.ac.uk
Cass Business School, City, University of London

Emilio Paolucci - emilio.paolucci@polito.it
Department of Management Engineering and Production, Politechnique of Turin, Turin, Italy

Chiara Spina - chiara.spina@insead.edu
INSEAD, Singapore

Acknowledgements
We gratefully acknowledge funding support from the ICRIOS Center at Bocconi University, the Innovation Growth Lab, the Strategy Research Foundation, the Department for Business, Energy and Industrial Strategy of the UK Government, City, University of London, and the Polytechnic of Turin. This research was approved by Bocconi Research Ethics Committee with protocol references 2016-III/13.696, 77133-2, and by City Ethics Committee with reference ETH1819-0351. The four studies presented in this manuscript are included in the AEA RCT Registry (AEARCTR-0006578, AEARCTR-0002205, AEARCTR-0006579, AEARCTR-0003875.)

1 Introduction

Managers and entrepreneurs do not have solid routines or methods to make decisions under uncertainty, such as decisions regarding the launch of new products, services or new businesses. This observation is supported by ample evidence from the world of practice and academic research. For instance, a Harvard Business Review report surveyed 646 managers and showed that many managers rely on gut feeling rather than systematic and well-organized judgments (Harvard Business Review Analytic Services, 2016). McKinsey analyzed the decisions of 2207 executives, and found that in 28% of the cases they make good decisions, in 60% they make bad decisions, and in 12% they make infrequent good decisions (Lovallo and Sibony, 2010). Moreover, CEOs and entrepreneurs tend to be overconfident (Camerer and Lovallo, 1999; Astebro, 2003; Malmendier and Tate, 2008; Galasso and Simcoe, 2011; Astebro et al., 2014) and 84.8% of the US start-up do not report successful results at least within their first 7 years (Fairlie and Miranda, 2016).

Against this background, this paper studies an approach to managerial decisions under uncertainty that starts with a rigorous framing of the problem, develops theories that predict the outcomes of actions from logical connections between antecedents and consequences, and tests these theories using existing data or data drawn from well-defined experimental designs. We call this approach "scientific" because it overlaps to a good extent with the approach that scientists use to develop new knowledge.

Extant studies have documented the poor logic of managerial decisions under uncertainty, and some scholars and practitioners advocate greater rigor and logic in defining frameworks, formulating models, and testing them (e.g. Martin 2009; Felin and Zenger, 2009; Csazar and Levinthal, 2016; Eisenhardt and Bingham, 2017). Our study is also part of a broader effort to highlight the importance of managerial practices (Bertrand and Schoar, 2003; Bloom and Van Reenen, 2007; Bloom et al., 2013; Gosnell, List and Metcalfe, 2020).

We provide evidence of the implications of a scientific approach to decision-

making through four randomized control trials (RCTs) involving start-ups and small-medium enterprises (SMEs). We obtain three main results. First, scientific decision-makers are more likely to terminate projects, and to do so in early stages. Second, they apply changes to their ideas fewer times before they commit to one or terminate the project. Third, they perform better.

We develop a model that provides a guide to interpret the empirical results. Recent models of entrepreneurial behaviour (Akcigit and Kerr, 2018; Jones and Pratap, 2020) focus on exploration for innovation decisions, but neglect the heterogeneity in how entrepreneurs decide to explore the environment and act based on feedback gathered through exploration. Our model focuses on some key features of the scientific approach and their implications for decision-making.

Our research design is based on four randomized control trials (RCTs) in which we conduct an evaluation of the same intervention across contexts (with the exception of minor changes due to operational constraints). We followed recent studies that have implemented a similar design (see, for instance, Allcott, 2015; Banerjee et al. 2015, Bowers et al. 2017). In so doing, we address an important shortcoming of research based on field experiments, whose results may be sensitive to scale or differences in implementation or economic environments (Levitt and List, 2009; Angrist and Pischke, 2010; List, Maniadis and Tufano, 2017; Milkman, Gromet, Ho et al., 2021).

We conducted our four RCTs in Milan (in 2016 and 2017), Turin (2018), and London (2019), involving start ups and small and medium enterprises (SMEs). Taken together, the four RCTs involve 754 [1] firms and 10,730 data points.

[1]For the sake of transparency, we clarify that these numbers are slightly different compared to the numbers we indicated in the registration. For RCT2 and RCT3: our registration indicated, respectively, 265 firms and 132 firms as the target. We eventually retained in our sample only 250 for RCT2 and 127 firms for RCT 3 as (1) some of the recruited firms eventually decided not to take part in the program and (2) upon close investigation it emerged that some were not meeting the target characteristics. For both RCT 2 and 3 we observed firms for 18 observation points as planned. For RCT4: our registration indicated that our target was 240 firms observed for 10 data points. We eventually involved 274 firms due to the high demand to take part in the program, but we retained only 261 in the sample as some of them decided that they did not want to provide data despite having taken part

We decided, deliberately, to adopt this research design after observing the preliminary results of our first RCT that uses data on 116 firms reported in Camuffo et al. (2020). However, not only does the larger scale of the present study provide more robust evidence about our phenomeon, but it also enables us to produce new analyses and results. Overall we believe that this is a sizable contribution in that it provides a far more complete and detailed picture of our phenomeon, including results or effects that could not see or did not study in Camuffo et al. (2020). In the Conclusions we highlight the rich a novel conclusions of this study compared to Camuffo et al. (2020). More generally, we believe that this has methodological value as well. It shows how much we can lever in terms of both robustness, novelty, and depth when we scale up initial research finidngs.

In all four RCTs, the experimental design involved the same training sessions over a period of about sixteen weeks, and the collection of data for about one year. All firms received training on how to manage innovation decisions using standard content on innovation management and entrepreneurship. About 80% of the content was the same for both treatment and control groups. Specifically, both treated and control groups were exposed to frameworks that they could use to support decision making, such as for instance the business model canvas and the balance scorecard; both groups were also exposed to the importance of using data and evidence to make decisions and where therefore exposed to multiple techniques for collecting data from customers, suppliers and key partners such as qualitative interviews, surveys and A/B testing. However, the treatment group was exposed to a scientific approach to decision-making: they were taught to use the frameworks learnt to develop a theory of the problem and draw hypotheses that flew logically from it; they were then taught to use the evidence-gathering techniques learnt in class to test those hypotheses and to evaluate the results in a disciplined way.

Our results highlight the importance of systematic decision-making in innova-

in the program or provided implausible information. We observed for 8 data points (as opposed to 10) because the RCT started later then expected and the funding had a time limit that could not be extended.

tion and entrepreneurial contexts characterized by uncertainty. Prior work has emphasized the importance, in these contexts, of flexibility and experimentation (Kerr, Nanda and Rhodes-Kropf, 2014, Ewens, Nanda and Rhodes-Kropf, 2018), but this paper suggests that when entrepreneurs conduct experiments like scientists — which is, complementing experiments with theories and hypotheses — they benefit greatly. In other words, a theoretical framework helps entrepreneurs to envision scenarios more precisely, design more informative tests, and more successfully update priors. As noted by Kerr Nanda and Rhodes-Kropf, (2014, page.38) "Successful experimentation requires being able to capitalize on experiments that reveal positive outcomes, and these upside scenarios can be as tricky as termination decisions". The scientific approach helps to capitalize on experiments with positive or not so positive outcomes (when decision-makers need to pivot). As the results show, this has powerful consequences for the number and type of innovative projects that are commercialized, and for the performance of these projects. This explains an important mechanism related to the 'up or out' pattern (i.e. start-ups that fail to grow will abandon an economy) documented by Haltiwanger, Jarmin and Miranda, (2013) and Decker et al. (2014), among others.

Section 2 presents our model and its main predictions. Section 3 describes the experimental design, sample, and data collection. Section 4 presents our results. Section 5 concludes.

2 Model

2.1 Set-up

Decision-makers explore ideas in stages. The goal of exploration is to identify a strategy or an activity that they can pursue and that produces an outcome. For short, we call this strategy or activity an "idea." The outside option of decision makers is $\pi_0 > 0$. We assume for simplicity that in each stage of exploration decision-makers run an experiment that tests one idea. The focus on one idea per experiment helps to streamline our discussion. The relevant

twist of this assumption is that we rule out that decision-makers have the resources to run all their ideas once and in parallel.

We assume that decision-makers enter the exploration process with a new idea different from the outside option. Thus, in our model the entry into the process is given, and the question is how the process unfolds after entry. At the beginning of the first stage of exploration, decision-makers receive a signal on whether it is worth running a costly experiment on this new idea to obtain more information on whether its value is higher than the outside option. Based on this signal they decide whether to run this experiment. If they do, at the end of the experiment they observe the value of the new idea and compare it to the outside option. During the experiment, they also obtain information about new ideas that they could test with a new experiment. If they decide to run the new experiment, they enter a new stage of exploration. Thus, at the end of the experiment in each stage, decision-makers decide whether to *terminate* the project and earn the outside option, *pivot* to a new idea, which means running a new experiment to test a new idea, or *commit* to the idea they have just experimented with.

They terminate if the idea they have just experimented with is less valuable than the outside option and they do not see any signal about new ideas such that the expected value of a new experiment is higher than the outside option. They pivot if the value of this new experiment is higher than the current idea and the outside option. They commit to the current idea if it is more valuable than both the outside option and the expected value of a new experiment.

We assume for simplicity that all the ideas that decision-makers can experiment with in the different stages of exploration can take values $\pi > \pi_0$ or $0 < \pi_0$. Before the experiment they do not observe these realizations. However, the probability that they infer π, the value of the idea, is $p \in (0, 1)$ and depends on the characteristics of decision-makers, such as the fact that they are cautious in evaluating their idea, or other characteristics such as their education, experience or knowledge of the phenomenon they deal with. We assume that $p \sim F(p)$, where F is a cumulative probability distribution of

p across decision-makers. The experiment reveals whether the value of the idea is π or 0. Our simplification implies that if decision-makers observe π they commit to the idea they have just experimented with and do not pivot because in a new costly experiment the best they can do is to earn π again. If decision-makers observe 0, either they pivot, if the expected value of the new experiment exceeds the outside option, or they terminate. In the Appendix we develop an extended model in which ideas can take a continuum of values, which allows for pivoting even when the experiment is successful. This extended model does not change our main results.

At the beginning of each stage, decision-makers know that if at the end of the experiment they observe 0, they earn the outside option. Thus, before the experiment, the expected value of running the experiment is $\pi p + \pi_0(1-p) - k > \pi_0$, or

$$(\pi - \pi_0)\, p - k > 0 \tag{1}$$

where k is the cost of running the experiment. Solving for p, it is easy to see that the share of decision-makers who satisfy this inequality is $1 - F(x)$ where $x \equiv \left(\frac{k}{\pi - \pi_0}\right)$. Again heterogeneity across decision-makers, such as in terms of education or experience, implies that some decision-makers have lower k or F, for given p. They run experiments more efficiently, and are more likely to satisfy inequality (1).

2.2 Decision process

The first step of the exploration process is determined by whether decision-makers run the first experiment. The share of decision-makers who run this experiment is $1 - F(x)$. Since p and all the parameters of (1) depend on given characteristics of decision-makers, if this inequality is satisfied in stage 1, it will be satisfied in all future stages of exploration. Moreover, as noted, if at the end of the experiment in stage 1 decision-makers observe π they commit to the current idea they have just experimented with because they do better than the outside option and it is not worth paying k again for an experiment that cannot produce better outcomes than π.

Let Φ be the probability that decision-makers observe that the value of the idea is 0. Note that Φ is different from $1-p$. The latter is the probability inferred by decision-makers that the idea is worth 0 before running the experiment, while Φ is the actual share of decision-makers who observe 0 after the experiment. In this respect, Φ is the "true" probability of observing 0. Decision-makers pivot, instead of terminating, if they observe 0 and a valuable potential new idea. We make two assumptions that describe the discovery process in our model.

Assumption 1. *Decision-makers search for new ideas in a closed space made of A ideas*

Assumption 2. *If at the end of stage of exploration $s \geq 1$ decision-makers observe 0, they:*

- *discover with probability $\theta_s \in [0, 1]$ a new idea that they can test in a new experiment*
- *learn that the value of other $\lambda_s A$ unexplored ideas is also 0, with $0 < \lambda_s \leq 1$, where $\lambda > 0$ because at least the idea actually tested in the experiment will be discarded.*

Assumption 1 says that decision-makers search in a space bounded by their experience, knowledge or by their preferences to focus on one activity (e.g. they do not want to operate in some industries or markets).

The first part of Assumption 2 says that at the end of a failed experiment, decision-makers have the chance to see new ideas. The second part of Assumption 2 says that the decision-makers apply logical frameworks that enable them to link the outcome of the experiment focused on a specific idea to other ideas. Specifically, if the experiment indicates that the value of the focal idea is 0, they realize, prior or without making a new experiment, that $\lambda_s A$ ideas not yet assessed in previous experiments will also have value 0. Therefore, they do not need to test them. We call θ_s the discovery rate of ideas and λ_s the exclusion rate.

Assumption 2 is crucial in our analysis. It says that while decision-makers see new ideas, they can use their logic to compare them to ideas they have already tested with negative results, and thus realize that it is not worth paying the cost of testing them in a new experiment. As we will see in the next section, and illustrate with examples, this ability to logically identify ideas that would also be associated with a negative outcome without testing them is an important characteristic of scientific decision-makers.

Thus, at the end of stage 1, if the experiment yields a negative outcome, decision-makers see with probability θ_1 a new idea they can experiment with. However, the failure of the idea they have just tested implies that they rule out $\lambda_1 A$ ideas from their search space of A ideas. Thus, the probability that the new idea is a genuinely new idea worth testing is $\theta_1 (1 - \lambda_1)$, where $(1 - \lambda_1)$ is the probability that the new idea is not one of the ideas that decision-makers exclude *a priori*.

To summarize, at the end of stage 1, $F(x)$ decision-makers have terminated before running the first experiment; $[1 - F(x)]\, \Phi\theta_1(1-\lambda_1)$ run the first experiment, observe 0 and pivot to a new idea they have identified when running the first experiment; $[1 - F(x)]\, \Phi\, [1 - \theta_1(1 - \lambda_1)]$ run the first experiment, observe 0 and terminate; and $[1 - F(x)]\, (1-\Phi)$ run the first experiment, observe π and therefore commit.

The process continues across stages in the same fashion. Thus, if decision-makers run the experiment in stage 2, at the end of it, either they observe π and commit, or they observe 0 with probability Φ. In this case, they see a new idea they can pivot to with probability θ_2. However, the failed experiment in stage 2 suggests that a share λ_2 of the $(1 - \lambda_1)A$ ideas available at the end of stage 1 are also unfeasible. Thus, the probability that the decision-maker pivots is $\theta_2\, [1 - \lambda_1 - \lambda_2(1 - \lambda_1)] = \theta_2(1 - \lambda_1)(1 - \lambda_2)$, where you now rule out both the ideas discarded at the end of stage 1 and those discarded at the end of stage 2.

If at the end of stage 3 the experiment fails again, decision-makers see a new

idea with probability θ_3, and rule out $1-(1-\lambda_1)(1-\lambda_2)-\lambda_3(1-\lambda_1)(1-\lambda_2) = (1-\lambda_1)(1-\lambda_2)(1-\lambda_3)$ ideas. More generally, at the end of stage s the probability of pivoting conditional on reaching this stage and observing a negative outcome in the experiment in stage s is $\theta_s \prod_{j=1}^{s}(1-\lambda_j)$. Overall, the share of decision-makers who terminate at the end of this stage is $[1-F(x)]\,\Phi\left[1-\theta_s \prod_{j=1}^{s}(1-\lambda_j)\right]$, the share of decision-makers who pivot is $[1-F(x)]\,\Phi\theta_s \prod_{j=1}^{s}(1-\lambda_j)$, and those who commit to the idea experimented in stage s is $[1-F(x)]\,(1-\Phi)$

2.3 Scientific Decision-Makers

We focus on two characteristics of scientific decision-makers.

First, scientific decision-makers analyze problems using logical frameworks. They develop models of the problem identifying its components and the relationship between them. This helps them to identify logical gaps in their ideas, and more generally in their reasoning about the potential success of their ideas. Also, high quality data and rigorous tests are more likely to deliver signals on the value of the ideas that are more closely associated with their actual value. Both factors push decision-makers to assess problems more objectively, reducing the natural tendency of managers and entrepreneurs to overconfidence (Astebro, 2003; Malmendier and Tate, 2008; Galasso and Simcoe, 2011; Astebro et al., 2014).

Second, they better understand problems because of better frameworks that help them to better understand the space of solutions. This has two implications. The first one is that a better assessment of the solutions makes them more efficient in discovering new ideas. The second implications is that the logical links in their frameworks help them to associate, through logic, analogies, or both, the outcomes of their experiments to other potential ideas.

Our examples in the next section illustrate these two characteristics of scientific decision-makers. Here we summarize them in the following two assumptions.

Assumption 3. *Given x and p, scientific decision-makers are more likely to*

show a higher value of the distribution function F

Assumption 4. *Scientific decision-makers are more likely to exhibit higher θ_s and λ_s, $s \geq 1$*

The higher F implies that, other things being equal, scientific decision-makers are more likely to exhibit a lower probability p because they are more cautious in evaluating the idea. This makes it less likely that they satisfy (1). The higher θ_s implies that they are more likely to discover new ideas, while the higher λ_s implies that, if they see a negative outcome from an experiment, they identify a higher number of untested ideas with negative outcomes.

The following two propositions capture the predictions that we test with our RCT. First, in early stages, a scientific approach is more likely to lead to terminate projects. Second, scientific decision-makers are less likely to pivot many times.

The intuition of the first proposition is that the reduction of overconfidence makes scientific decision-makers more cautious, leading to a distribution F with greater weight on lower values of p. Assumption 3 then implies that they are less likely to satisfy condition (1), and therefore they are more likely to terminate before running the experiment in stage 1.

However, for ideas that have not been terminated in the early stages, scientific decision makers are likely to show a lower level of termination due to the fact that, following Assumption 4, scientific decision makers are also characterized by a higher probability of finding new ideas θ_1, which may offset the effect of F. It may take a few stages to outweigh the initial higher share of termination of scientific decision-makers. Thus, in early stages scientific decision-makers are more likely to terminate.

The intuition of the second proposition is that, because the explorable space of search is bounded, the higher discovery rate together with the higher exclusion rate implies that scientific decision-makers cover the search space more quickly. Thus, scientific decision-makers commit or terminate earlier. In earlier stages, we cannot say if the probability of pivoting increases or decreases. The initial

effect produced by the higher discovery rate θ_1 may increase the probability of pivoting. However, the combined effect of covering the space more quickly, and the higher exclusion rate of ideas, make it unambiguously more likely that, eventually, scientific decision-makers are less likely to pivot.

Proposition 1. *If scientific decision-makers exhibit a higher F, in early stages they are more likely to terminate projects*

Proof. The share of decision-makers who terminate up to the end of stage s is

$$F(x) + [1 - F(x)]\,\Phi \sum_{i=1}^{s} \left[1 - \theta_i \prod_{j=1}^{i} (1 - \lambda_j) \right]$$

Assumption 3 implies that scientific decision-makers exhibit higher F. Therefore, the first term of this expression indicates that they are more likely to terminate before launching the first experiment. Assumption 4 implies that we cannot sign the subsequent terms unambiguously. However, the probability of termination after the first experiment is $[1 - F(x)]\,\Phi\cdot\,[1 - \theta_1(1 - \lambda_1)]$. Using $dF > 0$ and $d\,[\theta_1(1 - \lambda_1)]$ to denote differences between scientific and non-scientific decision-makers, the overall difference in the probability of termination at the end of the first stage is

$$\{1 - \Phi\,[1 - \theta_1(1 - \lambda_1)]\}\,dF - [1 - F(x)]\,\Phi d\,[\theta_1(1 - \lambda_1)]$$

The first term of this expression is positive, suggesting that, other things being equal, the overall effect of the higher F still implies that scientific decision-makers exhibit a higher probability of termination at the end of the first stage. We cannot sign unambiguously $d\,[\theta_1(1 - \lambda_1)]$, or the other terms of the probability of termination at the end of stage s. However, we cannot rule out that the initial boost will be absorbed gradually, and at least up to some initial stages of exploration scientific decision-makers are more likely to terminate. ■

Proposition 2. *Scientific decision-makers do not pivot many times*

Proof. In the generic stage s, the probability that decision-makers pivot more

than s times is $[1 - F(x)]\,\Phi\theta_s \prod_{j=1}^{s}(1-\lambda_j)$. Approximate $\theta_s \prod_{j=1}^{s}(1-\lambda_j)$ with a geometric Brownian motion with expected value $\theta e^{-\lambda s}$, where $-\lambda < 0$ is the negative drift. Assumption 4 implies that scientific decision-makers exhibit higher θ and λ. Using $d\theta$, $d\lambda > 0$ to denote differences for scientific decision-makers this implies $d\theta e^{-\lambda s} = e^{-\lambda s} d\theta - \theta s e^{-\lambda s} d\lambda > 0$ for $s < \frac{d\theta}{\theta d\lambda}$. Since $F(x)$ is higher for scientific decision-makers, for s sufficiently large scientific decision-makers are unambiguously less likely to pivot. ■

2.4 Examples

We provide two examples of companies from our RCTs consistent with the assumptions about scientific decision-making in the previous section.

Inkdome is a start-up that decided to launch an online search engine to find the right artist for getting tattoos. They developed a clear theory. They expect the service to be viable if four hypotheses are corroborated: 1) consumers use different artists for new tattoos; 2) they search online; 3) the search takes time; 4) they can find online all the information they need to find the ideal artist for the service they seek. Before testing these hypotheses with data from a survey interview, they set the following rule for launching the project: a) all hypotheses have to be corroborated; b) one hypothesis is corroborated if more than 60% of the interviewees respond positively. Using this 60% rule, the experiment corroborated the first three hypotheses, but not the fourth one. Inkdome abandoned the project.

Under similar conditions, many counterfactual entrepreneurs in our RCT did not set such clear hypotheses and tests. They based their assessments on generic discussions and interviews, and unclear rules. The lack of good frameworks implied that they did not set clear rules or consider that their assessment might be affected by over- (or under-) confidence. More generally, at the end of the first RCT in Milan in 2015 we asked the 88 surviving start-ups to score between 1-7 how likely they would be to close a potential new start-up that they found. The median score of the treatment group was 4.4 vs 3.2 of the control group, with a p-value smaller than 1%. Of course, we cannot exclude

that the additional level of caution triggered by the treatment might lead entrepreneurs to even become underconfident. While we cannot object to this statement, we rely on the evidence from past literature that suggests that managers and particularly entrepreneurs are typically overconfident.

A good example of our assumption about the implications of good frameworks is represented by Mimoto. Mimoto is a start-up that planned to offer an electric moto-sharing service in Milan. The founders started with a theory that focused on target customers. The logic is that analogous services such sa car- or bike-sharing are likely to meet a wide demand by different types of customers. Motos, instead, are likely to be used by special groups of people. They then theorized that the ideal target customers are young people, with mobility needs, and ability to pay. This prompted them to focus on college students.

However, when they tested their theory in practice, the use of the service was disappointing, both by college students and others. They went then back to their theory. They still thought that the focus was young people with ability to pay, but they had to look for customers with *unpredictable* mobility needs. College students have predictable schedules dictated by the regularity of the timing of their classes. This makes public or private transportation competitive because they can plan mobility needs in advance either by following public transportation schedules or by planning when and where to park depending on the schedule. A moto service is most useful when you suddenly have to grab some means of transportation to reach quickly a relatively far location in the city. Mimoto then saw clearly where to turn for a new target customer. It focused on young professionals, who are also young and need mobility in large cities, but have unpredictable mobility needs.

The general counterfactual non-scientific entrepreneurs in our RCTs did not have a clear theory and did not run rigorous experiments. As a result, many of them did not have clear directions about the actions to take after negative signals. On some occasions, they run experiments one after the other, and on other occasions they did not run any experiment but just used the evidence

as a basis for internal discussion. Also, during the experiment Mimoto noted that women had a hard time using the sturdy motorcycles that they employed in the experiment. They realized that while this was a problem for women, it was in fact a general problem that could also be faced by other people, while they observed that in Milan many different people employed scooters. Thus, they switched from motos to scooters, ruling out all combinations of target customers and motos. To summarize, thanks to general frameworks, scientific decision-makers know more promptly where to go when they receive feedback information from experiments. This helps them not to wander by trying several ideas without a solid logic.

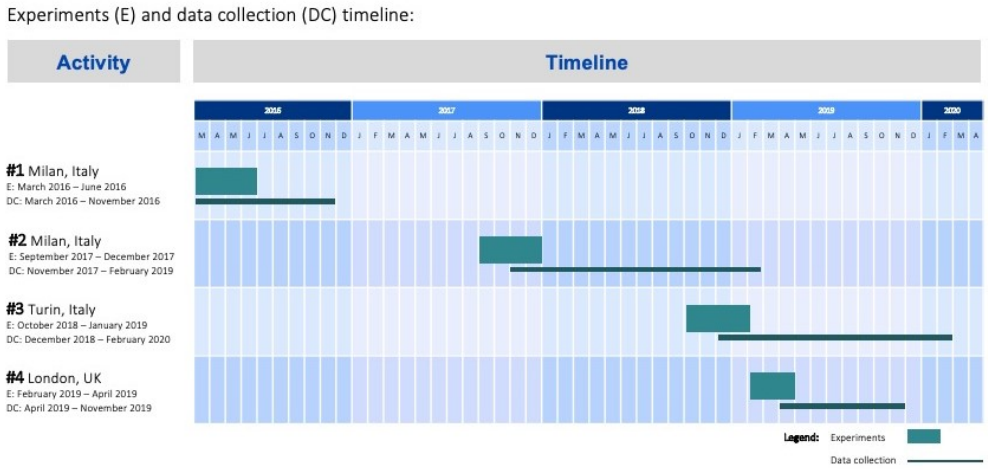
3 Context and Research Design

3.1 Experimental Design

We offered four training programs free of charge to entrepreneurs in Milan (Italy), Turin (Italy), and London (UK). We randomly assigned 50% of the participants to a training condition (training using the scientific approach) and the other 50% to a control condition (training without the scientific approach). Entrepreneurs provided data on their decision-making processes and performance before and for several months after the end of the training programs. The structure, type of intervention, and the data collection process of the four experiments were the same, with a view to conduct an internal replication of the same study across samples and regions. Figure 1 provides more detail on the timeline of each RCT and of the data collection.

For each study, the research team created new training programs for entrepreneurs. Training programs are effective ways to treat entrepreneurs (Field, Jayachandran and Pande, 2010; Campos et al., 2017; Anderson et al., 2018). The four programs were asynchronous with a gap of at least a few months between each other, given the heavy set-up required for each experiment. In every location, the research team recruited large teams of research assistants that were trained on how to advertise the program, interact with entrepreneurs,

Figure 1: Timeline of the Four RCTS



and collect data. In addition, we recruited qualified instructors and taught them how to deliver the training material, which was designed by the research team.

3.2 Step 1: Recruitment of Participants

In each study, we advertised the program at a national level over multiple online and offline channels, including social media posts, newsletters, magazines and events. One of the reasons why the program was appealing to entrepreneurs is that it was advertised and delivered under the brand of the countries' top business schools. We conducted the advertisement campaign over several weeks to ensure the recruitment of at least 100 entrepreneurs. Regardless of the media used, the campaign promoted the training program generically as a business support program (to avoid self-selection based on interest in a specific topic) offered free of charge to entrepreneurs operating in any industry. To apply to the program, we required entrepreneurs to provide information about their business, team, and decision-making practices via an online survey and a brief telephone interview. We did not admit to the program entrepreneurs who failed to complete the survey or the interview.

Since our study investigated the decisions that entrepreneurs make about their business, our empirical design required that the subjects receiving the treatment were the key decision-makers in the firm. We were more likely to meet this condition in micro-enterprises (with less then 10 employees), where owners are highly involved in the management of the firm. In the field experiment in London we admitted in the program only firms that met this condition. In the three field experiments conducted in Italy entrepreneurs met this condition naturally because all firms were at a very initial stage and had yet to start offering their service or product to the market.

The recruitment campaigns attracted entrepreneurs from different Italian regions (in the first three experiments) and from different parts of England (in the fourth experiment). The final sample included 754 entrepreneurial firms.

3.3 Step 2: Intervention Details

We assigned entrepreneurs in each experiment to either a treatment or a control group through simple randomization. We also broke down the treatment and control groups into smaller groups, and randomly assigned each subgroup to an experienced instructor. We administered a baseline survey to all entrepreneurs prior to the intervention, and we used the information in the survey to check that the observable characteristics were balanced across the treatment and the control groups using balance tests. The Online Appendix reports the results of the balance tests across groups for each of the four RCTs and for the variables that were common across all RCTs. As the four RCTs were conducted asynchronously, the research team had the opportunity over time to introduce additional relevant dimensions to the baseline survey. As a result, the list of observables is larger for later RCTs.

Treatment and control groups attended the same number of sessions, covering the same topics related to strategy and entrepreneurship. The sessions were highly experiential and smaller groups ensured that instructors provided feedback to each participant. About 80% of the content in the two classes was the same in terms of topics delivered and teaching material. Specifically,

both treated and control groups were taught frameworks that they could use to support decision making, such as the business model canvas or the balance scorecard; both groups were also exposed to evidence gathering techniques such as qualitative interviews, surveys and A/B testing. Both groups were taught to apply these frameworks and techniques to their specific contexts and were given feedback from their peers and instructor. However, the treatment group was taught to apply the frameworks and techniques used to make decisions using a scientific approach, which is by developing a theory of the problem and hypotheses that flew logically from it, by testing those hypotheses and eventually by evaluating the results of the test in comparison with the theory originally developed. The control group, instead, was free to apply these frameworks and techniques in the way they found more appropriate.

An example can help clarify the difference between the two groups. One of the first sessions of the training program focused on the Business Model Canvas (henceforth, BMC), a tool widely used in entrepreneurial education that concisely and visually represents a company's business model. It is composed of nine elements that describe a firm's customer value proposition, customer segments, channels, customer relationships, revenue streams, key resources, key partners, key activities, and cost structure. The control group was exposed to the basic content of the BMC and was taught to use this tool to provide a general overview of their business and discuss its implications. Entrepreneurs in the control group were then encouraged to apply the framework to their own business and given a dedicated time slot to discuss about it with their peers. They were then encouraged to present the application in front of the classroom and received general feedback from the instructor and peers. This is the typical way in which BMC is taught in MBAs and Executive programs.

The training offered to the treatment group shared some similarities with that offered to the control group: the treatment group too was exposed to the basic content of the BMC, asked to apply it to their business and discuss it with their peers. But differently from the control group, the treatment group was also explicitly invited to develop a theory that emerged from the application

of the BMC to their business and develop explicit hypotheses. For example, imagine an entrepreneur that, when filling in their BMC, indicated that they were running an electronics retail business using an online distribution channel. If this entrepreneur was part of the control group, they would be invited to generally discuss about the motivation behind this choice, its alignment with other choices made by the company, and would be given feedback on its suitability. The same entrepreneur in the treatment group, instead, would be explicitly asked to formulate the hypotheses underlying this choice, which, if supported, would make this choice a valuable one. For example, one such hypothesis might be "the majority of my target customers in the city where I am located buys electronics online".

In subsequent sessions, entrepreneurs in both groups were taught techniques to collect data in support of their decisions. For instance, they were taught about qualitative interviews, surveys, and experiments, and the strengths and weaknesses of each of these methodologies. Entrepreneurs in both groups were then invited to think about which techniques they could use in their businesses and discuss with their peers and instructor about one specific implementation in their context. The control group was let free to choose the context or problem to which they applied those techniques and was given general feedback on the way in which the technique was applied. The treatment group, instead, was explicitly invited to use these techniques to test the hypotheses formulated in the previous sessions and was given specific feedback on whether the proposed design was consistent with the hypothesis that they wanted to test. Of course, entrepreneurs in both groups were given genuine and valuable feedback. For example, if an entrepreneur proposed to administer a survey to a very small sample of target customers, the instructor would recommend them to increase their sample size irrespective of whether they belonged to the treatment or the control group; or if an entrepreneur formulated a survey question in a way that could be improved in terms of clarity, they would be offered suggestions regarding how to improve it, irrespective of whether they were in the treatment or in the control group.

It is important to note that for every RCT, each instructor was teaching both a treatment and a control group at different times of the day or different days of the week. This choice allowed to include instructor-fixed effects in our analyses and control for the different teaching styles of instructors and ultimately affect the absorption of the content taught to participants. Although instructors were not blind to the treatment, we directly supervised the delivery of each session to ensure high teaching standards and that the content was in line with the experimental design described above.

To prevent participants from meeting and potentially discussing key elements of the treatment, we offered training sessions on different days of the week or on the same day of the week to both groups, but at different times of the day. To further prevent contamination, the research team kept all communication to the two groups of entrepreneurs attending the program discrete and separate. For the same reason, the research team checked if applicants to the program had any acquaintance with other applicants and allocated them to the same experimental group.

3.4 Step 3: Data Collection

We systematically collected data on all participants through telephone interviews conducted by a team of research assistants over the span of several months. We hired research assistants for the purpose of these experiments and the research team trained them extensively. Research assistants were undergraduate or graduate students that were selected on the basis of their academic performance, basic knowledge of the entrepreneurial process, communication and analytical skills. The research team interviewed research assistants, and tested their communication and analytical skills through various activities (analysis of a business case, interviewing an entrepreneur and coding responses according to a simple, predefined coding scheme), to ensure they would be able to perform the tasks required by the project.

Research assistants performed regular phone calls that followed a predefined script that included open and closed-ended questions focusing on changes in

the business model, decision-making, and performance outcomes. In all RCTs but the first one, we recorded telephone interviews and subjected them to random checks to ensure that research assistants were conducting calls in accordance with the guidelines provided by the research team. The main variables used in this study refer to outcomes such as termination, pivot and amount of revenue and were therefore collected through closed-ended questions. Following an approach similar to the one used by Bloom et al. (2012), we also included a number of open-ended questions that elicited — without asking leading questions — what type of approach to decision-making entrepreneurs were using. Specifically, we instructed research assistants to code for the occurrence and the extent to which entrepreneurs employed themes related to theory, hypotheses, tests and evaluation. We use these data in supplementary analyses and provide more detail about this in Section 6 of the Online Appendix.

The data collection process continued for up to 14 months after the training program ended. In one of the RCTs (London), we could only collect observations for 7 months after the training program due to funding constraints. We take into account the duration of the data collection process in discussing our results.

4 Results

4.1 A Glimpse at the Data

Tables 1 and 2 report the descriptive statistics for the variables collected after the intervention through telephone interviews and refer, respectively, to the cross-sectional and longitudinal samples.

Firms in our sample pivot at most six times during the observation window. Fifty nine percent of the sample never pivots, and six percent of the firms pivot more than two times. To measure pivoting, we referred to the BMC taught to entrepreneurs during the training program. During each of the interviews, we asked entrepreneurs to report and describe any changes made to any of the

nine dimensions of the BMC (value proposition, customers, channels, customer relationships, key activities, key partners, key resources, revenue streams, cost structure). We classified a firm as having pivoted at time t if they reported a major change to their value proposition or customer segment, two key dimensions of their business.

Thirty four percent of the firms terminate their projects within the observation window. The average amount of revenue is EUR 15,538, with large variation in the sample since a substantial number of firms has zero revenue within the observation window. The number of entrepreneurs participating to each RCT changes mostly in light of budget constraints and venue capacity of each study. In all RCTs, half of the sample is assigned to the treatment condition. In all of our analyses we will report the results obtained on the full sample of 754 firms. However, in the Online Appendix, we report the results obtained for each RCT separately, for each of the models we present in this paper.

Figure 2a and 2b provide a visual representation of our data. The evidence is consistent with our predictions. In Figure 2a the number of treated firms that terminate the project within the observation window is higher than the number of control firms. Moreover, as we will show with our regressions, they are more likely to terminate earlier, as predicted by Proposition 1. Also, treated firms are more likely to pivot once, while control firms are more likely to pivot more times. This result, which we confirm below with our regressions, is interesting in light of our model. Our model predicts that a higher share of pivoting in early stages depends on the fact that the ability of scientific decision-makers to learn how to run better experiments from initial observations outweigh the more conservative effect of precision that encourages termination. Figure 2b shows that, on average, the revenue of treated firms grows faster than that of control firms.

4.2 Termination

We start our analysis by examining the impact of the intervention on termination. Table 3 reports the results of our analyses. Column (1) reports the

Table 1: Variables and Descriptive Statistics - Cross Section Sample

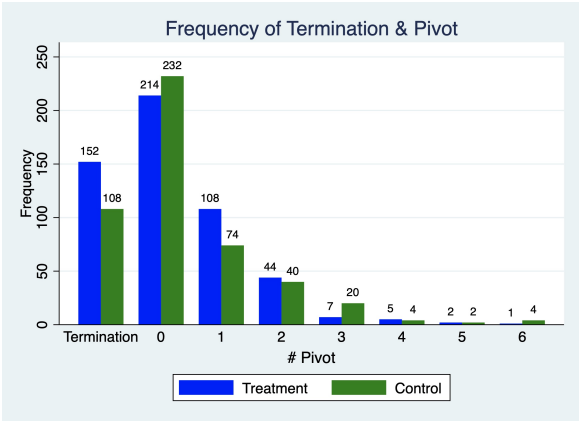
Variable	Description	Obs	Mean	SD	Min	Max
Number of Pivots	Number of times the firm pivoted within the observation period	754	0.68	1.05	0	6
Pivot=0	Dummy equal to 1 if the firm did not pivot within the observation window; 0 otherwise	754	0.59	0.49	0	1
Pivot=1	Dummy equal to 1 if the firm pivoted once within the observation window; 0 otherwise	754	0.24	0.43	0	1
Pivot=1-2	Dummy equal to 1 if the firm pivoted once or twice within the observation window; 0 otherwise	754	0.35	0.48	0	1
Pivot=2	Dummy equal to 1 if the firm pivoted twice within the observation window; 0 otherwise	754	0.11	0.31	0	1
Pivot=2+	Dummy equal to 1 if the firm pivoted more than twice within the observation window; 0 otherwise	754	0.06	0.24	0	1
Termination	Dummy equal to 1 if the firm terminated the project within the observation window; 0 otherwise	754	0.34	0.48	0	1
Revenue	Firm's cumulative revenue in EURO	754	15538	83240.95	0	1489026
RCT1	Milan 1	754	0.15	0.36	0	1
RCT2	Milan 2	754	0.33	0.47	0	1
RCT3	Turin	754	0.17	0.37	0	1
RCT4	London	754	0.35	0.48	0	1
Intervention	Dummy equal to 1 if the firm was treated; 0 otherwise	754	0.5	0.5	0	1
Average Scientific Intensity	Score reflecting the extent to which the firm's decision making process follows a scientific approach	754	2.23	1.21	0	5
Postgraduate	Dummy equal to 1 if team average on a score that reflected the highest level of education of each member of the team reported by the entrepreneur was higher than 3, where the educational level attained by team members is coded as follow: 5= PhD, 4=MBA, 3=MSc, 2=BA, 1=high school, 0=otherwise	754	0.17	0.38	0	1
Experience: Industry	Dummy equal to 1 if the team average experience in the focal industry is higher than 5 years; 0 otherwise	754	0.28	0.45	0	1
Experience: Managerial	Dummy equal to 1 if the team average managerial experience is higher than 5 years	754	0.23	0.42	0	1
Mature	Dummy equal to 1 if the team average age is higher than 30 years. NA for RCT1	638	0.56	0.5	0	1

Table 2: Variables and Descriptive Statistics - Longitudinal Sample

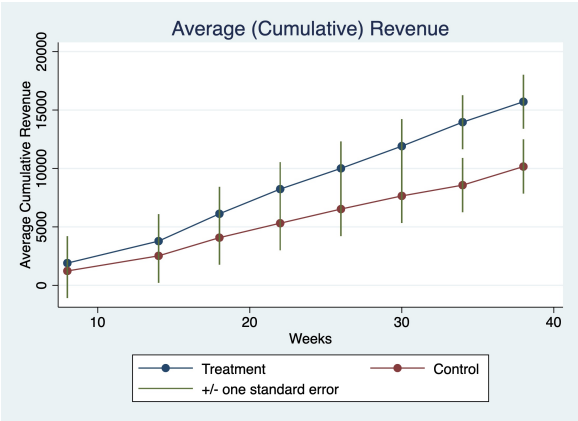
Variable	Description	Obs	Mean	SD	Min	Max
Termination	Dummy equal to 1 if the firm terminated the project within the observation window; 0 otherwise	8508	0.03	0.17	0	1
Week of Termination	Week at which the firm terminated the project	754	41.72	16.89	6	66
Revenue (Flow)	Firm's revenue (flow) in EURO	10730	1098.48	8581.3	0	231000
Intervention	Dummy equal to 1 if the firm was treated; 0 otherwise	10730	0.5	0.5	0	1
Scientific Intensity	Score reflecting the extent to which the firm's decision making process follows a scientific approach	10730	2.24	1.26	0	5

Figure 2: Termination, Pivot and Performance

(a) Termination and Pivot



(b) Performance (euros)



...

results of a cross section linear probability model that shows that the intervention raises the probability of termination by 10.4 percentage points (p=0.001). As a robustness check we also run a probit regression and report the results in Column (2). The marginal effect calculated at the observed values suggests that treated firms report a probability to terminate that is 10.4 percentage points higher than that of control firms(p=0.000).

Table 4 reports the results of a Cox proportional hazard model in Column (1). We corroborated the proportionality assumption using the Schoenfeld residuals. We find that the hazard rate of termination is higher for treated firms than for control firms. In Column (2) we replicate this analysis using a regression estimated with OLS to predict the week of termination. We find that, on average, treated firms terminate their project about 2.3 weeks earlier than control firms (p=0.012). Overall, we find that scientific decision-makers are more likely to terminate their projects earlier.

In the Online Appendix (Section 3, Tables from A6 to A9) we compare these results with the results of obtained for each individual RCT.

Table 3: Termination

VARIABLES	(1) Termination OLS Cross-Section Full Sample	(2) Termination Probit Cross-Section Full Sample
Intervention	0.104***	0.299***
	(0.001)	(0.000)
Constant	0.283***	-5.038***
	(0.000)	(0.000)
Observations	754	754
R-squared	0.078	
Dummies for mentors and RCT	Yes	Yes
Clustered Errors	Intervention Mentor RCT	Intervention Mentor RCT
Number of id	754	754

Robust pval in parentheses, *** p<0.01, ** p<0.05, * p<0.1. Conservatively, all specifications control for the variables that were unbalanced between the treatment and control group despite randomization. Specifically, Model (3) controls for "Background: Economics" and "Background: STEM", Model (4) controls for "Self-regulation", Model (1) controls for the interaction between each RCT dummies and variable that was unbalanced in that specific RCT. Section 2 of the Online Appendix provides definitions of all variables. However, we clarify that results are similar when not controlling for these variables.

Table 4: Termination Time

VARIABLES	(1) Hazard of termination Survival Full Sample	(2) Week of termination OLS Full Sample
Intervention	0.375***	-2.322**
	(0.000)	(0.012)
Constant		32.446***
		(0.000)
Observations	754	754
R-squared		0.242
Dummies for mentors and RCT	Yes	Yes
Clustered Errors	Intervention Mentor RCT	Intervention Mentor RCT

Robust pval in parentheses, *** p<0.01, ** p<0.05, * p<0.1. All specifications control for the variables that were unbalanced between the treatment and control group despite randomization.

4.3 Pivot

Table 5 reports the results of our analysis on the number of pivots. Column (1) reports the results of a cross section regression, estimated with OLS where the dependent variable is the number of pivots made by the firms within the observation window. The regression includes dummies for mentors and RCTs and we clustered standard errors at the intervention-mentor-RCT level and reports a not significant effect of the intervention. In Column (2) we report the results of a linear probability regression in which the dependent variable is a dummy equal to 1 if the firm has pivoted only once within the observation window and equal to 0 otherwise (i.e., if it has not pivoted or has pivoted more than once). By distinguishing between the choice to pivot once vs. not pivoting or pivoting many times, the results in Column (2) show that the intervention raises the probability of pivoting only once by 8.7 percentage points (p-value = 0.001) vis-à-vis no pivot or more than one pivot. As noted earlier, this is an interesting result. Our model suggests that this effect depends on the fact that scientific decision-makers' learning from observing the results of the experiments enable them to outweigh the conservative effect of their greater precision.

Table 5: Number of Pivots

	(1)	(2)
	# Pivots	Pivoting once
	OLS	OLS
	Cross-Section	Cross-Section
VARIABLES	Full Sample	Full Sample
Intervention	-0.032	0.087***
	(0.654)	(0.001)
Constant	0.432***	0.083***
	(0.000)	(0.007)
Observations	754	754
R-squared	0.120	0.082
Dummies for mentors and RCT	Yes	Yes
Clustered Errors	Intervention Mentor RCT	Intervention Mentor RCT

Robust pval in parentheses, *** p<0.01, ** p<0.05, * p<0.1. All specifications control for the variables that were unbalanced between the treatment and control group despite randomization.

Results from a multinomial probit specification, reported in Table 6, confirm this finding. Columns (1), (2) and (3) refer, respectively, to the probability that a firm pivots once, twice, or more than twice vis-à-vis the no-pivot baseline. In Figure 3 we show the marginal effects of intervention calculated at the observed values for the entire sample. The intervention raises the probability of pivoting once or twice and lowers the probability of not pivoting or pivoting more than twice. Specifically, when we look at the full sample, the intervention decreases the probability of not pivoting by 5.7 percentage points (although this result is not significant at the conventional level, with p=0.120), increases the probability of pivoting once by 8.6 percentage points (p=0.002), increases the probability of pivoting twice by 0.9 percentage points (although this result is not significant at the conventional level, with p=0.672), and decreases the probability of pivoting more than twice by 3.7 percentage points (p=0.005).

We report in the Online Appendix (Section 4) a comparison between these results and those obtained for each individual RCT (Tables A10-A15).

Table 6: Pivot Multinomial Probit

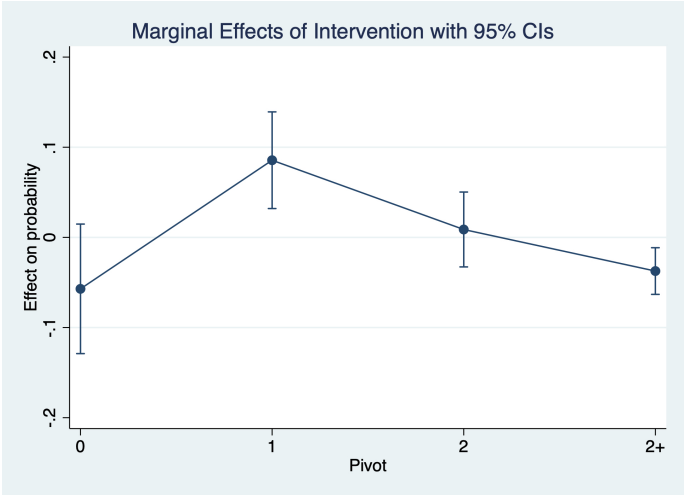
VARIABLES	(1) Pivoting only once Multinomial Probit Cross-Section Full Sample	(2) Pivoting twice Multinomial Probit Cross-Section Full Sample	(3) Pivoting more than twice Multinomial Probit Cross-Section Full Sample
Intervention	0.370***	0.148	-0.287
	(0.010)	(0.397)	(0.117)
Constant	-1.374***	-2.104***	-2.438***
	(0.000)	(0.000)	(0.000)
Observations	754	754	754
Dummies for mentors and RCT	Yes	Yes	Yes
Clustered Errors	Intervention Mentor RCT	Intervention Mentor RCT	Intervention Mentor RCT

Robust pval in parentheses, *** p<0.01, ** p<0.05, * p<0.1. All specifications control for the variables that were unbalanced between the treatment and control group despite randomization.

4.4 Performance

Our model predicts that in early stages treated firms are more likely to terminate their projects and they do not pivot many times. However, it does not

Figure 3: Marginal Effects of Intervention on Pivot



predict whether this yields higher performance because performance depends on whether the conjectures entrepreneurs make are correct. It may be that control firms linger on projects that should not be terminated, or it might be that by not pivoting, or pivoting many times, they enjoy better outcomes. We therefore explore this question empirically.

In Table 7 we reports the results of our analysis of the impact of the intervention on the cumulative revenue of firms in our sample (in EUR) at the date of our last observation in each trial, estimated by OLS. This represents a different time period for each one of our trials. However, our RCT dummies control for these differences, on average. We also employ mentor dummies and cluster the standard errors at the intervention-mentor-RCT level. Results show that, on average, treated firms earn EUR 6,504.108 more than control firms (p=0.046). The small effect size reflects the fact that many firms in our first three RCTs earn no revenue as they are start-ups that started their activities at the time of our training program. Within the observation period, some of the firms started earning revenues of the order of dozen thousand EUR, very much in line with the average earned by start-ups in their first few months of operation. The increase in revenue between the time of the first interview and the last ranges from 0 to EUR 1,320,396, with revenues increasing of EUR 23,100 at

the 90th percentile.

In Table A16 of the Online Appendix (Section 5), we report these results in comparison with the results of obtained for each individual RCT.

Table 7: Performance OLS

VARIABLES	(1) Revenue OLS Cross-section Full Sample
Intervention	6,504.108**
	(0.046)
Constant	9,039.968***
	(0.006)
Observations	754
Dummies for mentors and RCT	Yes
Clustered Errors	Intervention Mentor RCT

Robust pval in parentheses, *** p<0.01, ** p<0.05, * p<0.1. All specifications control for the variables that were unbalanced between the treatment and control group despite randomization.

4.5 Instrumenting Scientific Intensity

Our analyses so far provided estimates of the intent-to-treat effect. This is interesting from a policy perspective as it provides an estimate of the effect of the treatment that takes into account that not all individuals targeted by an intervention are necessarily compliers (Gelman, Hill and Vehtari, 2020). However, to maximize what we can learn from the intervention, we asked our research assistants who were conducting regular phone interviews with the entrepreneurs to use a predefined interview protocol (based on 16 items) to assess the level of scientific intensity used by entrepreneurs in making decisions. This protocol led to the determination of a score (on a scale from 0 to 5) that measured the level of scientific intensity of each entrepreneur at each

observation point. We use this score to conduct an additional set of analyses using the intervention as an instrument for the level of scientific intensity exhibited by decision makers. This enables us to provide a more precise estimate of the complier average casual effect. In the Online Appendix, in Section 6, we provide more details on how this score was created. The analyses in Table A18 of the Online Appendix, show that treated firms demonstrated higher levels of scientific intensity than firms in the control group.

Table 8 presents the results of a cross-section specification estimated using two-stage least squares. Results on termination, reported in Column (1), show that the increase of one unit in the average scientific intensity increases the probability of terminating of 29.9 percentage points (0.001). In Column (2), we report the results of our analysis on pivot, which show that the increase of one unit in the average scientific intensity increases the probability of pivoting once (versus 0 or more than once) of 25 percentage points (p=0.000). Looking at the effect on performance, results in Column (3) show that an increase of one unit in the average scientific intensity is associated with an increase of EUR 18,703.974 (p=0.056). As a robustness check, we replicate our analyses on termination and pivot using a IV probit specification and using a two-stage least square approach on the longitudinal sample. We report results in the Online Appendix (Tables A19-A23).

Table 8: Instrumenting Scientific Intensity

VARIABLES	(1) Termination 2SLS Cross-Section Full Sample	(2) Pivoting once 2SLS Cross-Section Full Sample	(3) Revenue 2SLS Cross-Section Full Sample
Average Scientific Intensity	0.299***	0.251***	18,703.974*
	(0.001)	(0.000)	(0.056)
Constant	-0.283	-0.391**	-26,351.771
	(0.125)	(0.016)	(0.208)
Observations	754	754	754
R-squared	-0.485	-0.225	0.064
Dummies for mentors and RCT	Yes	Yes	Yes
Clustered Errors	Intervention Mentor RCT	Intervention Mentor RCT	Intervention Mentor RCT

Robust pval in parentheses, *** p<0.01, ** p<0.05, * p<0.1. All specifications control for the variables that were unbalanced between the treatment and control group despite randomization.

4.6 Attrition

Unfortunately, not all firms continued to answer our interviews until the end of the program. Notoriously attrition is more the norm than the exception in field experiments (Gerber and Green, 2012). To contrast this tendency, we designed the program so that the core training was followed by a series of monthly events focused on relevant themes for entrepreneurs. The events included no treatment and were delivered in separate days but in the same way for treated and control firms. However, participation in these events was allowed only to firms that showed their continued engagement with the data collection. Nevertheless, some of these firms did not reach the last interview round. The motivation provided by entrepreneurs in expressing their unavailability to be interviewed was that, since the main training was over, their incentive to answer the interviews was lower. Overall, 22% of firms in our sample withdrew at different points of the program. To verify that attrition did not affect our results, we check there was not any significant difference between treated and control in their early withdrawal from the program. In Table 9, we estimate early withdrawal from the program as a function of the intervention, which we show has no significant impact. In our main analyses, we addressed attrition by inputting the missing values of those who left the study, making the conservative assumption that the performance of firms that left the program did not change after they left the program.

4.7 Heterogeneous Treatment Effects

In addition to testing whether our treatment has a main effect on the outcome variables of interest, we are interested in exploring the presence of heterogeneous treatment effects, that is, whether certain groups react differently to the treatment. This exploration was not part of the pre-analysis plan, but to connect the results of this paper to the ongoing debate in the literature we are particularly interested in three dimensions that the literature suggests support entrepreneurial decision making: education, work experience (industry and managerial), and age. We measure these constructs at the level of the

Table 9: Early Withdrawal from Program

VARIABLES	Early Withdrawn OLS Cross-section Full Sample
Intervention	-0.020
	(0.433)
Constant	0.405***
	(0.000)
Observations	754
R-squared	0.159
Dummies for mentors and RCT	Yes
Clustered Errors	Intervention Mentor RCT

Robust pval in parentheses, *** p<0.01, ** p<0.05, * p<0.1. All specifications control for the variables that were unbalanced between the treatment and control group despite randomization.

entrepreneurial team, since our focus on micro-businesses implies that each member of the team is likely to have played a relevant role in decision making. Table 10 reports the results of a regression analysis, estimated with OLS, where we regress the key outcome variables in this study (termination, pivot and performance) against Postgraduate (a dummy variable equal to 1 for firms where the average level of highest education degree attained by the member of the team of the entrepreneur was at least the Master degree), the interaction between Intervention and Postgraduate, and the interaction between Intervention and its complement (Non Postgraduate). For entrepreneurs with an average education lower than the postgraduate level, the intervention increases the probability of termination of 9.8 percentage points (p=0.001) and the probability of pivoting once of 10.1 percentage points (p=0.000); it increases revenue of EUR 7,645.550 (p=0.091). For entrepreneurs with a postgraduate degree the intervention does not have a significant impact on the dependent variables. We interpret this result to suggest that whereas the effect of the intervention is clearer for entrepreneurs with a lower education degree, for entrepreneurs with a postgraduate degree the intervention has an impact that shows a higher degree of variation and might depend on other factors. This is interesting because it reveals how the effect of the intervention varies for different types of entrepreneurs, not only in terms of point estimate but also

in terms of variability.

Table 10: Education OLS Cross-Section

VARIABLES	(1) Termination OLS Cross-Section Full Sample	(2) Pivoting once OLS Cross-Section Full Sample	(3) Revenue OLS Cross-section Full Sample
Postgraduate	0.019	0.040	547.530
	(0.809)	(0.482)	(0.958)
Intervention X Postgraduate	0.141	-0.036	-2,986.407
	(0.197)	(0.679)	(0.796)
Intervention X Non Postgraduate	0.098***	0.101***	7,645.550*
	(0.001)	(0.000)	(0.091)
Constant	0.275***	0.087***	9,906.638***
	(0.000)	(0.000)	(0.007)
Observations	754	754	754
R-squared	0.080	0.085	0.087
Dummies for mentors and RCT	Yes	Yes	Yes
Clustered Errors	Intervention Mentor RCT	Intervention Mentor RCT	Intervention Mentor RCT

Robust pval in parentheses, *** p<0.01, ** p<0.05, * p<0.1. All specifications control for the variables that were unbalanced between the treatment and control group despite randomization.

Table 11 reports the results of a regression analysis, estimated with OLS, where we regress the key outcome variables in this study against High Industry Experience (a dummy variable equal to 1 for firms where the team of the entrepreneur had on average more than five years of industry experience, corresponding to the 75 percentile of the distribution of industry experience), the interaction between Intervention and High Industry Experience, and with its complement. For entrepreneurs with low industry experience, the intervention increases the probability of termination by 12.3 percentage points (p=0.002) and the probability of pivoting once by 10.8 percentage points (p=0.001). For highly experienced teams the intervention has a non significant effect on the dependent variables.

In Table 12 we perform a similar analysis but we look at the level of managerial experience, using four years as the threshold to distinguish between higher and lower experience (corresponding to the 75 percentile of the distribution of managerial experience). Results show that the treatment increases the probability of termination by 12.5 percentage points (p=0.001) for firms with

lower experience, while it does not have a statistically significant effect on termination for firms with high managerial experience. Instead, the treatment increases the probability of pivoting once by 6.4 percentage points (p=0.040) for firms with lower experience, while it increases the same probability by 14.3 percentage points (p=0.003) for firms with high managerial experience. The treatment also increases revenue by EUR 8,070.775 (p=0.038) for firms with low managerial experience, but it does not have a statistically significant effect for more experienced firms, showing a pattern that is similar to that observed in the case of education presented above.

Table 11: Industry Experience OLS Cross-Section

	(1)	(2)	(3)
	Termination	Pivoting once	Revenue
	OLS	OLS	OLS
	Cross-Section	Cross-Section	Cross-section
VARIABLES	Full Sample	Full Sample	Full Sample
High Industry Experience	-0.011	0.034	12,559.028*
	(0.835)	(0.512)	(0.066)
Intervention X High Industry Experience	0.053	0.039	21,955.047
	(0.382)	(0.430)	(0.302)
Intervention X Low Industry Experience	0.123***	0.108***	647.409
	(0.002)	(0.001)	(0.906)
Constant	0.299***	0.075**	-204.030
	(0.000)	(0.029)	(0.980)
Observations	754	754	754
R-squared	0.080	0.083	0.103
Dummies for mentors and RCT	Yes	Yes	Yes
Clustered Errors	Intervention Mentor RCT	Intervention Mentor RCT	Intervention Mentor RCT

Robust pval in parentheses, *** p<0.01, ** p<0.05, * p<0.1. All specifications control for the variables that were unbalanced between the treatment and control group despite randomization.

Table 12: Managerial Experience OLS Cross-Section

VARIABLES	(1) Termination OLS Cross-Section Full Sample	(2) Pivoting once OLS Cross-Section Full Sample	(3) Revenue OLS Cross-section Full Sample
High Managerial Experience	-0.017	-0.031	5,462.768
	(0.704)	(0.516)	(0.357)
Intervention x High Managerial Experience	0.058	0.143***	3,046.304
	(0.368)	(0.003)	(0.620)
Intervention x Low Managerial Experience	0.125***	0.064**	8,070.775**
	(0.001)	(0.040)	(0.038)
Constant	0.294***	0.090**	8,004.405*
	(0.000)	(0.018)	(0.089)
Observations	754	754	754
R-squared	0.081	0.084	0.091
Dummies for mentors and RCT	Yes	Yes	Yes
Clustered Errors	Intervention Mentor RCT	Intervention Mentor RCT	Intervention Mentor RCT

Robust pval in parentheses, *** p<0.01, ** p<0.05, * p<0.1. All specifications control for the variables that were unbalanced between the treatment and control group despite randomization.

Finally, in Table 13, we report the results of a regression analysis, estimated with OLS, where we regress the key outcome variables in this study against Mature (a dummy variable equal to 1 for firms where the team of the entrepreneur was on average older than 30 years old), the interaction between intervention and Mature and with complement (Younger). We conduct this analysis using only the data of RCT2, RCT3 and RCT4 because we did not collect the variable age in RCT1. For younger entrepreneurs, the treatment increases the probability of termination by 14.2 percentage points (p=0.000) and of pivoting once by 8.4 percentage points (p=0.015). It increases revenue by 12,579.958 (p=0.094). For more mature entrepreneurs, the intervention increases the probability of pivoting once by 14.6 percentage points (p=0.006), but it does not have a statistically significant effect on termination and revenue.

Table 13: Age OLS Cross-Section

VARIABLES	(1) Termination OLS Cross-Section Full Sample	(2) Pivoting once OLS Cross-Section Full Sample	(3) Revenue OLS Cross-section Full Sample
Mature	0.006	-0.003	10,459.473
	(0.927)	(0.955)	(0.321)
Intervention X Mature	0.025	0.146***	-11,760.632
	(0.726)	(0.006)	(0.110)
Intervention X Younger	0.142***	0.084**	12,579.958*
	(0.000)	(0.015)	(0.094)
Constant	0.283***	0.075*	7,134.708
	(0.000)	(0.052)	(0.119)
Observations	638	638	638
R-squared	0.077	0.075	0.083
Dummies for mentors and RCT	Yes	Yes	Yes
Clustered Errors	Intervention Mentor RCT	Intervention Mentor RCT	Intervention Mentor RCT

Robust pval in parentheses, *** p<0.01, ** p<0.05, * p<0.1. All specifications control for the variables that were unbalanced between the treatment and control group despite randomization.

Overall, these results suggest that the impact of a scientific approach on termination, pivoting and revenue is consistently positive and statistically significant for relatively less educated, experienced and mature entrepreneurs. The effect of the intervention for relatively more educated, experienced and mature entrepreneurs is often not statistically significant. As mentioned, we interpret this result to suggest that whereas the effect of the training is clearer for the former type of entrepreneurs, for the latter type of entrepreneurs the intervention has an impact that shows a higher degree of variation and might depend on other factors. This is an important result that also provides insights for the potential welfare effects of our intervention. We believe that this result does not only speak about the potential of our training, but of the scientific approach as a rigorous tool for decision-making under uncertainty in the contexts that we studied. Given the cost of experience, this suggests that training managers and entrepreneurs to adopt the scientific approach that scientists use in their research, can reach high benefit/cost ratios.

5 Conclusion

Entrepreneurs and innovators make choices under conditions of uncertainty. We examine the implications of a scientific approach to decision-making in these cases. Our study underscores the importance of teaching managers and entrepreneurs more than basic business skills (such as accounting or marketing) or soft skills. Our empirical results and model emphasize the importance of teaching them to develop frameworks about what they do and the decisions they have to make, and to test the implications of these frameworks through experiments.

In addition to contributing to research on decision-making, this paper contributes to a larger debate on the design of field experiments. In particular, it outlines the limitations of conducting interventions that focus on a small scale and a limited set of contexts. Compared to Camuffo et al (2020), which focused on the first trial, we observe relevant differences.

First, in Camuffo et al. (2020) we do not find a statistically significant effect of the intervention on termination in most of the regressions, while we find a positive and significant effect across all the other three RCTs, as shown in the Online Appendix (Section 3). The limited size of the sample in Camuffo et al. (2020) did not produce a sufficient number of terminations to detect this effect.

Second, Camuffo et al. (2020) showed a positive effect of the intervention on the number of pivots (Online Appendix, Section 4), which again depends on the fact that in the smaller sample only a few firms pivot more than once. In the larger sample of this paper we discover that the intervention makes pivoting more focused because treated firms are less likely to pivot more than one or two times. As shown by our model this underlies an important characteristic of the impact of a scientific approach to decision-making that we did not discover in the previous paper: to the extent that decision-makers search in a limited space, this approach enables them to see more quickly where are the opportunites rather than lingering on continuous experimentation.

Third, the Online Appendix shows, more generally, that in all RCTs the estimates of the intervention have the same sign and comparable magnitudes, while of course the statistical significance varies. This suggests that, taken individually, the problem with each trial is statistical significance and precision, not model specification.

Fourth, the larger sample size of this paper enables us to study heterogenous effects that we could not study with the limited sample size of Camuffo et al. (2020). The interesting insight is that overall the effect of our treatment is statistically significant for entrepreneurs with lower education, younger, and less experienced. The point estimate is the same for more educated, older and experienced, but the standard error is larger. This suggests that for the former category the absorption of the approach does not depend on other elements, while for the latter category there are other factors that affect this absorption and that represent an interesting line of future research.

Finally, the limited sample size of Camuffo et al. (2020) prevented us to test the impact of a specific measure of our intervention. Bloom et al. (2013) tested a 2SLS model of the impact of a variable "Management" that measures the adoption of managerial practices instrumented by the intervention. In this paper, and not in Camuffo et al. (2020), we run all our regressions about termination, pivot and performance using a measure of "scientific intensity" instrumented by our intervention. As in Bloom et al. (2013) this comes with the exclusion restriction that the treatment only has an indirect effect through the instrumented variable. However, the 2SLS results produce results in line with the intention-to-treat regressions.

Our study also has practical implications. Research in economics and management has generated many theories and models that prescribe concrete managerial actions. While we teach these theories and models in academic programs, managers and entrepreneurs rarely use them to make decisions, and prefer to rely on their intuitions, experience, gut feelings, or their own logic. This is a serious gap that makes academic research in economics and management less relevant than it could be. Our prediction is that if we nurture a culture

of scientific decision-making in firms, the value and the use of theories from academic research in economics and management will also increase.

This also helps us to address the question whether the benefits of our training program are worth the costs. This was a complex RCT with many elements and dimensions and we could not collect data to address this point. This is a limitation of this study and an open question for future research. However, we believe that the message of our findings is broader. We focused on a sample of entrepreneurs very much like Bloom et al. (2013) focused on simple Indian companies to make the more general point that management matters. Our broader point is about the value of business training that teaches students to use in practice the logic of many theories in business or economics, and to test the implications of their decisions. In this respect, our findings speak to opportunity to understand better the value of a more "scientific" business education, as well possibly of scientific decision-making in other realms where decisions under uncertainty matter, such as for instance politics.

Our study is not without limitations. There are limitations in our RCT due to natural difficulties in producing a perfect design in a complex project. For example, research assistants followed the businesses for months, asking them about their approach to entrepreneurship. We cannot exclude that these follow-up interviews did not affect the behaviour of entrepreneurs. Some data depend on subjective coding. We did our best to ensure that this did not affect our findings. In particular, in each RCT a member of the team was in charge of listening to a random sample of the recordings of the interviews conducted by each research assistant and check their coding. If we observed major discrepancies, we would provide feedback to the relevant research assistant. In general we felt reassured by the fact that our findings tend to go in the same directions across RCT, or when we use different estimation methods or variables, such as the intention-to-treat and 2SLS regressions that use a measure of scientific-intensity. Also, in the case of pivots, we provide a detailed account of how we coded it in the Online Appendix. Another limitation of our study is that it only covers a relatively short time period. It would be

interesting to explore the effects of our treatment in the medium or long term. Finally, our model provide plausible explanations for our results: in particular, closed search space and faster search due to a better understanding of the opportunity space, produces the non-linear effect on pivots that we document empirically. However, we do no test for this or other exact mechanisms.

More in general, we need more research to better understand the implications of a scientific approach to decision-making and how it can, in detail, create opportunities for better innovation and entrepreneurial decisions, and how different types of firms or individuals can take advantage of these opportunities. We hope that future research can shed light on these important micro-foundations of economic performance.

References

Akcigit, Ufuk, and William R. Kerr. 2008. "Growth Through Heterogeneous Innovations." *Journal of Political Economy* 126 (4): 1374-1443.

Allcott, Hunt. "Site Selection Bias in Program Evaluation." 2015. *The Quarterly Journal of Economics* 130 (3): 1117-1165.

Anderson, David R., Dennis J. Sweeney, Thomas A. Williams, Jeffrey D. Camm, and James J. Cochran, 2018. *An Introduction to Management Science: Quantitative Approach*. Cengage learning.

Angrist, Joshua D., and Jörn-Steffen Pischke. 2010. "The Credibility Revolution in Empirical Economics: How Better Research Design Is Taking the Con Out of Econometrics." *Journal of Economic Perspectives* 24 (2): 3-30.

Astebro, Thomas. 2003. "The Return to Independent Invention: Evidence of Risk Seeking, Extreme Optimism or Skewness-Loving?" *Economic Journal* 113 (484): 226–39.

Astebro, Thomas, Herz, Holger, Nanda, Ramana, and Weber Robert A. 2014. "Seeking the Roots of Entrepreneurship: Insights from Behavioral Economics." *Journal of Economic Perspectives* 28 (3): 49–70

Banerjee, Abhijit, Esther Duflo, Nathanael Goldberg, Dean Karlan, Robert Osei, William Parienté, Jeremy Shapiro, Bram Thuysbaert, and Christopher Udry. 2015. "A Multifaceted Program Causes Lasting Progress for the Very Poor: Evidence from Six Countries." *Science* 348 (6236).

Bertrand, Marianne, and Antoinette Schoar. 2003. "Managing with Style: The Effect of Managers on Firm Policies." *The Quarterly Journal of Economics* 118 (4): 1169-1208.

Bloom, Nicholas, and John Van Reenen. 2007. "Measuring and Explaining Management Practices across Firms and Countries." *The Quarterly Journal of Economics* 122 (4): 1351-1408.

Bloom, Nicholas, Benn Eifert, Aprajit Mahajan, David McKenzie, and John Roberts. 2013. "Does Management Matter? Evidence from India." *The Quarterly Journal of Economics* 128 (1): 1-51.

Bowers, Jake, Nathaniel Higgins, Dean Karlan, Sarah Tulman, and Jonathan Zinman. 2017. "Challenges to Replication and Iteration in Field Experiments: Evidence from Two Direct Mail Shots." *American Economic Review* 107 (5): 462-65.

Camerer, Colin, and Dan Lovallo. 1999. "Overconfidence and excess entry: An experimental approach." *American Economic Review* 89 (1): 306-318.

Campos, Francisco, Michael Frese, Markus Goldstein, Leonardo Iacovone, Hillary C. Johnson, David McKenzie, and Mona Mensmann. 2017. "Teaching Personal Initiative Beats Traditional Training in Boosting Small Business in West Africa." *Science* 357 (6357): 1287-1290.

Camuffo, Arnaldo, Alessandro Cordova, Alfonso Gambardella, and Chiara Spin. 2020. "A Scientific Approach to Entrepreneurial Decision-Making: Evidence from a Randomized Control Trial." *Management Science* 66 (2), 564-586.

Csaszar, Felipe A., and Daniel A. Levinthal. 2016. "Mental Representation and the Discovery of New Strategies." *Strategic Management Journal* 37 (10):

2031-2049.

Decker, Ryan, John Haltiwanger, Ron Jarmin, and Javier Miranda. 2014. "The Role of Entrepreneurship in US Job Creation and Economic Dynamism." *Journal of Economic Perspectives* 28 (3): 3-24.

Eisenhardt, Kathleen M., and Christopher B. Bingham. 2017. "Superior Strategy in Entrepreneurial Settings: Thinking, Doing, and the Logic of Opportunity." *Strategy Science* 2 (4): 246-257.

Ewens, Michael, Ramana Nanda, and Matthew Rhodes-Kropf. 2018. "Cost of Experimentation and the Evolution of Venture Capital." *Journal of Financial Economics* 128 (3): 422-442.

Fairlie, Robert W., and Javier Miranda. 2017. "Taking the Leap: The Determinants of Entrepreneurs Hiring their First Employee." *Journal of Economics and Management Strategy* 26 (1): 3-34.

Felin, Teppo, and Todd R. Zenger. 2009. "Entrepreneurs as Theorists: On the Origins of Collective Beliefs and Novel Strategies." *Strategic Entrepreneurship Journal* 3 (2): 127-146.

Field, Erica, Seema Jayachandran, and Rohini Pande. 2010. "Do Traditional Institutions Constrain Female Entrepreneurship? A Field Experiment on Business Training in India." *American Economic Review* 100 (2): 125-29.

Galasso, Alberto, and Timothy S. Simcoe. 2011. "CEO Overconfidence and Innovation." *Management Science* 57 (8): 1469-1484.

Gelman, Andrew, Jennifer Hill, and Aki Vehtari. 2020. *Regression and Other Stories*. Cambridge University Press.

Gerber, A.S. and Donald P. Green. 2012. *Field Experiments: Design. Analysis, and Interpretation*. WW Norton.

Gosnell, Greer K., John A. List, and Robert Metcalfe. 2020. The Impact of Management Practices on Employee Productivity: A Field Experiment with Airline Captains. *Journal of Political Economy* 128(4), 1195-1233.

Haltiwanger, John, Ron S. Jarmin, and Javier Miranda. 2013. "Who Creates Jobs? Small versus Large versus Young." *Review of Economics and Statistics* 95 (2): 347-361.

Harvard Business Review Analytic Services. 2016. *The Evolution of Decision Making: How Leading Organizations Are Adopting a Data-Driven Culture.*

Jones, John Bailey, and Sangeeta Pratap. 2020. "An Estimated Structural Model of Entrepreneurial Behavior." *American Economic Review* 110 (9): 2859-98.

Kerr, William R., Ramana Nanda, and Matthew Rhodes-Kropf. 2014. "Entrepreneurship as Experimentation." *Journal of Economic Perspectives* 28 (3): 25-48.

Levitt, Steven D., and John A. List. 2009. "Field Experiments in Economics: The Past, the Present, and the Future." *European Economic Review* 53 (1): 1-18.

List, John, Zacharias Maniadis, and Fabio Tufano. 2017. To Replicate or Not To Replicate? Exploring Reproducibility in Economics through the Lens of a Model and a Pilot Study. No. 00646. *The Field Experiments Website.*

Lovallo, Dan, and Olivier Sibony. 2010. "The Case for Behavioral Strategy," *McKinsey Quarterly*, March.

Malmendier, Ulrike, and Geoffrey Tate. 2008. "Who Makes Acquisitions? CEO Overconfidence and the Market's Reaction." *Journal of Financial Economics* 89 (1): 20-43.

Maniadis, Zacharias, Fabio Tufano, and John A. List. 2015. "How to Make Experimental Economics Research More Reproducible: Lessons from Other Disciplines and a New Proposal." In *Replication in Experimental Economics.* Emerald Group Publishing Limited.

Martin, Roger, and Roger L. Martin. 2009. "The Design of Business: Why Design Thinking Is the Next Competitive Advantage." *Harvard Business Press*,

2009.

McCullough, Bruce D. 2009. "Open Access Economics Journals and the Market for Reproducible Economic Research." *Economic Analysis and Policy* 39 (1): 117-126.

Milkman, Katherine, Gromet, Dena, Ho, Hung et al. 2021. "Megastudies Improve the Impact of Applied Behavioural Science." *Nature* 600, 478–483 (2021).

Zimmermann, Christian. 2015. "On the Need for a Replication Journal." FRB St. Louis Working Paper 2015-16.

A Scientific Approach to Innovation Management: Online Appendix

1 Extension of the model

In this Appendix we extend our model by assuming that after the experiment decision-makers observe outcome $\pi \sim F$, with $\pi \in [0, B]$, where B is a finite upper bound such that $B > \pi_0$, and π_0 is the outside option of the decision-maker at the outset of the process. This also implies that we extend our model by allowing for the possibility that decision-makers can pivot even if, at the end of the experiment that they run in the generic stage s, they observe an outcome higher than the outside option.

Before the experiment in stage $s \geq 1$, decision-makers know that if at the end of the experiment they observe $\pi \leq \pi_{s-1}$, where π_{s-1} is the best option they realized up to this stage, they earn π_{s-1}. Therefore, they run the experiment if $\int_{\pi_{s-1}}^{B} \pi dF(\pi) + \pi_{s-1}F(\pi_{s-1}) - k > \pi_{s-1}$, or after integration by parts

$$B - \int_{\pi_{s-1}}^{B} F(\pi)d\pi - \pi_{s-1} - k > 0 \tag{A1}$$

The decision-makers who satisfy this inequality run the experiment. In order to define the share of these decision-makers we can focus on two sources of randomness across decision-makers: the cost of the experiment k or any parameter of the distribution F. We are agnostic about the exact source of randomness and, for $s \geq 1$, we define $G_{s-1} \equiv G_{s-1}(z_{s-1})$ to be the distribution that represents the share of decision-makers that satisfy this inequality, where z_{s-1} is the left-hand side of (A1). Other things being equal, the distribution function G_{s-1} decreases with k and increases with parameters that lower F over its entire support. For example, suppose that ξ is a parameter that affects F in the sense of first order stochastic dominance. We can think of $G_{s-1} \equiv \Gamma_{s-1}(z_{s-1} \mid \xi)h(\xi)$, where Γ_{s-1} is the distribution function of z_{s-1} conditional on ξ, and $h(\cdot)$ is the marginal probability of ξ. It is also easy to see that G_{s-1} decreases with π_{s-1}.

Thus, at the beginning of stage 1, G_0 decision-makers satisfy (A1). In particular, $\pi_{s-1} = \pi_0$ is the outside option. After the experiment, decision-makers observe π and we set $\Phi_0 \equiv \Phi(\pi_0)$ to be the probability that the experiment yields $\pi \leq \pi_0$. We now assume that, like in the text, decision-makers discover new ideas and they can rule out some ideas in the space A using logic and analogies based on their frameworks. Specifically, we assume that, θ_1 is the probability that they discover a new idea. However, we assume that if they observe $\pi \leq \pi_0$, they rule out $\lambda_{10}A$ ideas, while if they observe $\pi > \pi_0$, they rule out $\lambda_{11}A$ ideas, with $\lambda_{10} < \lambda_{11}$. The rate of discovery of new ideas is the same, θ_1, but if they observe $\pi > \pi_0$, it is natural to assume that they rule out more ideas because any new idea viable for experimentation has to overcome a higher threshold.

Let $q_{10} \equiv \theta_1(1 - \lambda_{10})$ and $q_{11} \equiv \theta_1(1 - \lambda_{11})$. Following the same logic of the model in the text, $1-G_0$ decision-makers terminate before the experiment in stage 1; after the experiment in stage 1, $G_0\left[\Phi_0 G_0 q_{10} + (1 - \Phi_0)G_1 q_{11}\right]$ decision-makers pivot, $G_0\Phi_0 G_0(1 - q_{10})$ terminate, and $(1 - \Phi_0)G_1(1 - q_{11})$ commit to the new idea.

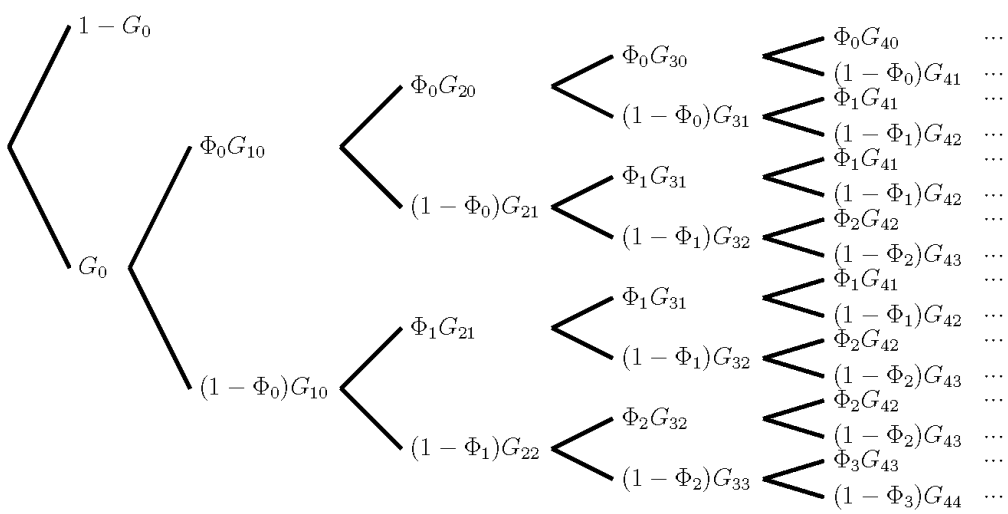
At the end of the experiment in stage 2, we have 4 cases: if in stage 1 decision-makers observed $\pi \leq \pi_0$, at the end of stage 2 they could observe $\pi \leq \pi_0$ or $\pi > \pi_0$; otherwise, they could observe $\pi \leq \pi_1$ or $\pi > \pi_1$, where we now take into account that they have to overcome the higher threshold $\pi_1 > \pi_0$. The probabilities of these 4 cases are, respectively, Φ_0^2, $\Phi_0(1 - \Phi_0)$, $(1 - \Phi_0)\Phi_1$, $(1 - \Phi_0)(1 - \Phi_1)$, where $\Phi_1 \equiv \Phi(\pi_1)$. Followng the same logic above and in the text, the probabilities that in these 4 cases they find a new idea they can pivot to are, respectively, $\theta_2(1 - \lambda_{10})(1 - \lambda_{20})$; $\theta_2(1 - \lambda_{10})(1 - \lambda_{21})$; $\theta_2(1 - \lambda_{11})(1 - \lambda_{21})$; $\theta_2(1 - \lambda_{11})(1 - \lambda_{22})$.

More generally, these pattern repeat themselves following the same logic at the end of each stage. In general, the new threshold is the highest observed outcome π in previous experiments. To streamline this representation, consider the end of the generic stage s in which there have been $j \geq s$ updates of π, such that the best available option after the sth experiment is π_j. The probability that decision-makers observe a new idea worth pivoting to in a new experiment in the following stage can be approximated by a geometric Brownian motion with negative drift $-\gamma_j$. We can then write the expected value of this probability as $q_{sj} = \theta e^{-\gamma_j s}$ such that γ_j increases with j. Intuitively, other things being equal, the higher the number of updates the lower the probability that decision-makers find a new idea. We can then define $G_{sj} \equiv G_s q_{sj}$ as the share of decision-makers who reached stage s, have a history of $j \leq s$ updates, and pivot.

Figure A1 shows how the pivoting decisions unfold. Let $\Phi_j \equiv \Phi(\pi_j)$, $j = 0, 1, 2, \ldots$. In

each branch of the Figure we have the probabilities of pivoting conditional on the history of observations up to that point. In the initial branch, $1 - G_0$ is the probability that decision-makers terminate without running the experiment in stage 1. If they run it, which happens with probability G_0, in the next branch they pivot with probability $\Phi_0 G_{10}$ or $(1 - \Phi_0)G_{11}$ depending on whether they observe $\pi_1 \leq \pi_0$ or $\pi_1 > \pi_0$. At the end of stage 2, which is the next branching in Figure A1, the upper branching reports the probability of pivoting if decision-makers observed $\pi_1 \leq \pi_0$ at the end of stage 1 and they observe $\pi_2 \leq \pi_0$ or $\pi_2 > \pi_0$ at the end of stage 2. In the lower branching, we have the equivalent probabilities but we now take into account that the best option they carry on from the previous stage is $\pi_1 > \pi_0$ rather than π_0. The logic of the following steps is the same.

Figure A1: Share of Pivots



Assumptions 3 and 4 extend naturally to this more general case. In particular, scientific decision-makers exhibit a lower G_0 produced by a higher F, and higher θ and γ_j $\forall j \leq s$.

The assumptions relaxed by this extended model do not affect the logic of our model in the text regarding termination. The reason is that termination implies that decision-makers never observe $\pi > \pi_0$. Thus, the probability of termination replicates the one showed in the proof of Proposition 1, and the logic discussed in this proof applies here as well. Thus, in this extended model scientific decision-makers are also more likely to terminate in early stages.

We then focus on the propensity to pivot. The share of decision-makers who pivot at any given stage, conditional upon reaching the stage, is the sum of all the vertical terms in each column corresponding to a given stage in Figure A1. These probabilities change between

scientific and non-scientific decision-makers because of G_{sj}. In G_{sj}, G_s only differs between scientific and non-scientific decision-makers because of F. However, differences in F implied by Assumption 3 disappear if decision-makers run the experiment of stage 1 because after this experiment all decision-makers satisfy condition (A1). This leaves differences in q_{sj}. As shown in the proof of Proposition 2 in the text, these expression eventually decrease for s large enough. Therefore, for s large enough, a sufficient number of vertical terms for a given stage in the columns of Figure A1 will be smaller for scientific decision-makers such that the sum of these vertical terms will also be smaller.

In order to establish that they are less likely to pivot after a given stage, it is easy to see from Figure A1 that if in stage s decision-makers reached a node where we observe G_{sj}, the share of these decision-makers who pivot and run the $s+1$ experiment is $\Phi_j G_{s+1j} + (1-\Phi_j) G_{s+1j+1}$. This implies that if in stage s, the share G_{sj} of scientific decision-makers, $\forall j \leq s$, is smaller than that of the non-scientific decision-makers, their probability of pivoting to stage $s+1$ is smaller than that of non-scientific decision-makers because both $G_{s+1j} < G_{sj}$ and $G_{s+1j+1} < G_{sj}$, and both G_{s+1j} and G_{s+1j+1} are smaller for scientific decision-makers than non-scientific decision-makers. As a result, even in this extended model, conditional upon reaching a given stage, scientific decision-makers are eventually less likely to pivot.

The question is whether at the outset of the process scientific decision-makers are more likely to pivot. But this question is not different from what we have discussed in the text, as it involves an assessment of the conditions in the very first two branches of Figure A1 that are the same as the ones in the text. Thus, if scientific decision-makers are more likely to pivot initially, they will eventually be less likely to pivot. If they are not more likely to pivot initially, they will be less likely to pivot from the outset of the process.

2 Balance Checks

Table A1: Balance Checks RCT1

Variable Name	Description	Treatment		Control		Difference	
		Mean	SD	Mean	SD	b	p
Currently Employed	Proportion of team members employed at the time of the training	0.68	0.39	0.72	0.42	0.04	(0.567)
Currently Studying	Proportion of team members enrolled in an education program at the time of training	0.19	0.37	0.28	0.42	0.09	(0.250)
Education Level	Highest educational level attained by team members (5=PhD, 4=MBA, 3=MSc, 2=BA, 1=high school, 0=otherwise; Team Average)	2.34	0.86	2.12	0.91	-0.22	(0.192)
Experience: Entrepreneurial Founder or Employee	Number of years of experience working with companies other than focal as founder or employee (Team Average)	0.92	2.51	0.32	1.18	-0.59	(0.107)
Experience: Entrepreneurial Mentor	Number of years of experience working with companies other than focal as mentor or consultant (Team Average)	0.02	0.13	0.02	0.13	0.00	(0.981)
Experience: Industry	Number of years of experience in industry (Team Average)	2.55	4.64	2.56	4.78	0.01	(0.991)
Experience: Managerial	Number of years of managerial experience (Team Average)	2.03	3.34	1.22	3.32	-0.81	(0.192)
Idea Stage	Dummy variable assuming value of 1 when the company has one business idea and 0 when the company has started working on the project but has not launched it on the market yet	0.63	0.49	0.65	0.48	0.02	(0.807)
Lombardy	Dummy variable assuming value of 1 when the majority of team members comes from the Italian region of Lombardy and 0 otherwise	0.32	0.47	0.4	0.49	0.08	(0.366)
Sector: Furniture	Dummy variable assuming value of 1 when the company operates in the furniture sector and 0 otherwise	0.25	0.44	0.25	0.43	-0.01	(0.916)
Sector: Internet	Dummy variable assuming value of 1 when the company operates in the internet sector and 0 otherwise	0.44	0.5	0.51	0.5	0.07	(0.467)
Sector: Retail	Dummy variable assuming value of 1 when the company operates in the retail sector and 0 otherwise	0.1	0.3	0.07	0.26	-0.03	(0.548)
Team Size	Number of team members	2.85	1.36	2.72	1.31	-0.13	(0.606)
Observations		59		57		116	

Table A2: Balance Checks RCT2

Variable Name	Description	Treatment		Control		Difference	
		Mean	SD	Mean	SD	b	p
Age	Age (Team Average)	31.47	8.18	31.41	7.90	-0.06	(0.950)
Analytic Thinking	Agreement on a 1-10 scale with the following statements (Team Average): "Analyzing the situation and looking at the evidence is critical to our company's decision-making", "We carefully assess all the possible alternatives before making a choice for our company", "We prefer to gather all the relevant information before making a decision for our company", "Multiple elements are taken into account when making a decision for our company, pros and cons are carefully evaluated in every situation"	8.38	3.68	8.07	3.28	-0.32	(0.475)
Background: Economics	Team members with an economics background (%)	0.41	0.42	0.31	0.37	-0.10**	(0.046)
Background: Other	Team members with no economics backgrounds (%)	0.22	0.36	0.20	0.33	-0.02	(0.696)
Background: STEM	Team members with a STEM (Science Technology Engineering Mathematics) backgrounds (%)	0.38	0.40	0.49	0.41	0.11**	(0.032)
Certainty	Agreement on a 1-10 scale with the following statements (Team Average): "We are sure about our business model", "We are sure about our strategy"	5.93	1.94	5.61	1.91	-0.32	(0.191)
Consensus	Answer on a 1-10 scale to the following questions (Team Average): "To what extent do you and your team members have consensus on the long term objectives of the firm?", "To what extent do you and your team members have consensus on the short term objectives of the firm?", "To what extent do you and your team members have consensus on the survival strategy of the firm?"	8.85	1.67	8.86	1.66	0.00	(0.990)
Education	Highest educational level attained by team members (5=PhD, 4=MBA, 3=MSc, 2=BA, 1=high school, 0=otherwise; Team Average)	1.94	0.74	1.95	0.80	0.00	(0.969)
Experience: Entrepreneurial	Number of years of entrepreneurial experience (Team Average)	1.09	2.19	0.93	1.44	-0.17	(0.480)
Experience: Industry	Number of years of experience in industry (Team Average)	2.84	3.82	2.33	3.62	-0.51	(0.280)
Experience: Managerial	Number of years of managerial experience (Team Average)	2.29	3.69	2.27	4.18	-0.02	(0.971)
Experience: Work	Number of years of work experience (Team Average)	8.73	7.75	9.02	8.85	0.28	(0.788)
Full Time	Percentage of team members working full-time	0.57	0.43	0.62	0.42	0.05	(0.390)
Gender (Female)	Proportion of women in the team	0.27	0.37	0.25	0.36	-0.03	(0.541)
Hours: Total Weekly	Weekly hours dedicated to the company (Team Average)	10.17	9.65	10.96	11.45	0.78	(0.560)
Idea Potential	Independent assessment of the value of the idea	47.22	21.22	47.31	23.25	0.09	(0.975)
Idea Value: Max	Maximum estimated value of the project (0 to 100)	85.08	16.29	85.67	16.16	0.59	(0.773)
Idea Value: Mean	Estimated value of the project (mean, 0 to 100)	65.40	15.53	64.52	16.69	-0.88	(0.668)
Idea Value: Min	Minimum estimated value of the project (0 to 100)	45.71	19.86	43.21	22.93	-2.50	(0.357)
Idea Value: Range	Estimated value of the project (range, 0 to 100)	39.37	18.85	42.46	20.99	3.10	(0.221)
Intuitive Thinking	Agreement on a 1-10 scale with the following statements (Team Average): "We are prone to following our intuitions when making company-related decisions", "We consider feelings and intuitions rather than analysis in our startup decisions", "First impressions are important when making decisions", "It is important to rely on gut feelings and intuition when making decisions"	4.09	1.70	3.83	1.74	-0.25	(0.244)
Lombardy	Dummy variable taking value of 1 when the majority of team members comes from the Italian region of Lombardy, 0 otherwise	0.56	0.47	0.57	0.46	0.01	(0.883)
Months to Revenue	Number of months to revenue	11.52	5.80	11.51	5.85	-0.01	(0.987)
Part Time	Percentage of team members working part-time	0.08	0.18	0.08	0.17	0.00	(0.941)
Probability Termination	Probability of terminating the project	31.64	32.53	32.35	31.60	0.70	(0.863)
Team Size	Number of team members	2.25	1.46	2.28	1.37	0.03	(0.858)
Observations		125		125		250	

Table A3: Balance Checks RCT3

Variable Name	Description	Treatment		Control		Difference	
		Mean	SD	Mean	SD	b	p
Age	Age (Team Average)	30.60	9.29	30.53	7.14	-0.07	(0.963)
Analytic Thinking	Agreement on a 1-5 scale with the following statements (Team Average): "Analyzing the situation and looking at the evidence is critical to our companyÕs decision-making", "We carefully assess all the possible alternatives before making a choice for our company", "We prefer to gather all the relevant information before making a decision for our company" and "Multiple elements are taken into account when making a decision for our company, pros and cons are carefully evaluated in every situation"	4.30	0.63	4.40	0.56	0.11	(0.318)
Background: Economics	Team members with Economics backgrounds (%)	0.18	0.31	0.20	0.36	0.02	(0.701)
Background: Other	Team members with no Economics/STEM backgrounds (%)	0.56	0.43	0.44	0.46	-0.11	(0.152)
Background: STEM	Team members with a STEM (Science Technology Engineering Mathematics) backgrounds (%)	0.26	0.38	0.36	0.45	0.09	(0.223)
Confidence	Agreement on a 1-5 scale with the following statements (Team Average): "We are confident in our entrepreneurial skills", "We are sure we are deploying the best strategy for our business", "We are confident in our ability to manage our business", "We master the competences necessary for our venture" and "We are sure there is no better business model for our idea"	3.41	0.52	3.32	0.65	-0.09	(0.397)
Currently Studying	Number of team members enrolled in an education program at the time of training	0.26	0.30	0.21	0.30	-0.04	(0.426)
Education	Highest educational level attained by team members (5=PhD, 4=MBA, 3=MSc, 2=BA, 1=high school, 0=otherwise; Team Average)	1.85	0.89	2.06	1.09	0.21	(0.240)
Experience: Business Plan	Dummy taking value of 1 if the team had years of experience in business plan design, 0 otherwise	0.26	0.36	0.35	0.43	0.09	(0.228)
Experience: Entrepreneurial	Number of years of entrepreneurial experience (Team Average)	1.65	4.38	1.73	3.37	0.08	(0.908)
Experience: Industry	Number of years of experience in industry (Team Average)	2.77	5.72	3.03	5.04	0.25	(0.792)
Experience: Managerial	Number of years of managerial experience (Team Average)	1.54	2.78	1.76	3.76	0.22	(0.705)
Gender (Female)	Proportion of women in the team	0.31	0.38	0.25	0.36	-0.06	(0.356)
Hours: Total Weekly	Weekly hours dedicated to the company (Team Average)	11.39	10.06	11.76	12.36	0.37	(0.853)
Idea Maturity	Maturity of the idea (in months)	9.32	9.43	11.98	11.63	2.66	(0.158)
Idea Potential	Independent assessment of the value of the idea (two evaluators, average) based on five criteria: innovation, feasibility, sustainability, team competence, market size	49.22	11.99	49.16	12.86	-0.06	(0.978)
Idea Value: Mean	Estimated value of the project (mean)	65.82	18.53	63.30	16.05	-2.52	(0.415)
Intuitive Thinking	Agreement on a 1-5 scale with the following statements (Team Average): "We are prone to following our intuitions when making company-related decisions" and "We consider feelings and intuitions rather than analysis in our startup decisions"	2.74	0.83	2.70	0.99	-0.03	(0.838)
Later Stage	Dummy variable taking value of 1 if the firm is at a more advanced stage than others, 0 otherwise	0.13	0.34	0.11	0.31	-0.03	(0.666)
Locus of Control	Agreement on a 1-7 scale with the following statements (Team Average): "In most jobs you need a lot of luck to excel", "One typically earns what they are worth", "To make money you just need to know the right people", "To get a good position you need luck", "Income is mainly the result of hard work", "There is a direct relationship between a personÕs abilities and the position he/she holds", "Many of the difficulties encountered at work concerns senior colleagues", "Generally, people who work well get rewarded", "Promotions are awarded to people who work well", "To find a good job, having a good network is more important than actual skills", "A well-trained person always finds a satisfycing job" and "To get a really good job you have to have high-level acquaintances"	3.84	0.67	3.79	0.70	-0.05	(0.707)
Months to Revenue	Number of months to revenue	12.69	11.37	14.68	10.58	1.99	(0.310)
Piedmont	Dummy variable taking value of 1 when the majority of team members comes from the Italian region of Piedmont and 0 otherwise	0.55	0.45	0.52	0.48	-0.03	(0.748)
Probability Pivot Idea	Probability of changing the business idea	31.89	22.96	32.53	26.75	0.65	(0.884)
Probability Pivot Other	Probability of changing other components of the business model	52.20	22.97	52.92	26.17	0.73	(0.868)
Probability Pivot Problem	Probability of changing the problem and customer segment	34.57	22.49	34.48	25.20	-0.09	(0.983)
Probability Termination	Probability of terminating the project	13.64	16.53	17.42	21.66	3.78	(0.268)
Risk-averse	Agreement on a 1-7 scale with the following statements (Team Average): "In important matters I never take unnecessary risks, which can be avoided", "In important situations I never deliberately chose to take risks I could have avoided", "I always try to avoid situations that put me at risk of getting into trouble with other people", "I am always very careful and I put safety first" and "I prefer to avoid doing things that expose me to criticism and liability"	4.23	1.03	3.96	1.04	-0.27	(0.151)
Risk-taker	Agreement on a 1-7 scale with the following statements (Team Average): "I can be pretty reckless and take some big risks", "I think I often act boldly and courageously", "I am a brave and daring person and I like to tempt fate in various situations", "There is a direct relationship between a personÕs abilities and the position he/she holds" and "I think I am often less cautious than other people"	4.04	1.13	3.98	0.91	-0.05	(0.766)
Scientific intensity: 1 Theory	Theory development score	2.92	1.32	3.05	1.20	0.13	(0.559)
Scientific intensity: 2 Hypothesis	Hypothesis development score	2.14	1.63	1.98	1.51	-0.16	(0.571)
Scientific intensity: 3 Test	Test score	1.32	1.73	1.29	1.69	-0.03	(0.919)
Scientific intensity: 4 Valuation	Valuation score	0.84	1.49	0.94	1.63	0.09	(0.742)
Self-efficacy	Agreement on a 1-7 scale with the following statements (Team Average): "I think I will always be able to achieve a goal even if I have to perform a difficult task", "Faced with new tasks and challenges, I am always confident that I will be able to complete them", "I am sure I will succeed", "When I have a goal, I almost always get better results than others", "When I take a test or an exam I am sure I can pass it successfully", "I am confident that my results will be recognized and appreciated by others", "I am not worried about difficult situations, because so far I have always managed to get by with my skills", "I never had any problem understanding and facing even the most complicated situations" and "I think I get the crux of the matter first"	5.46	1.07	5.57	0.96	0.11	(0.557)
Self-regulation	Agreement on a 1-7 scale with the following statements (Team Average): "People can count on me to meet the set and planned deadlines", "I can hardly say no", "I change my mind quite often", "Others would describe me as an impulsive person", "I wish I had more self-discipline", "I get carried away by my feelings", "I am not easily discouraged", "Sometimes I canÕt stop but do something, even though I know it is wrong", "I often act without thinking about all the alternatives", "I often do things that seem right in the present, even at the expense of future goals" and "When I pursue a goal I follow the original plan, even when I realize that it is not the best"	4.99	0.82	5.25	0.85	0.25*	(0.090)
Startcup	Dummy variable taking value of 1 if the firm takes part to a local competition, 0 otherwise	0.11	0.32	0.18	0.39	0.07	(0.290)
Team Size	Number of team members	2.51	1.48	2.14	1.36	-0.37	(0.144)
Observations		61		66		127	

Table A4: Balance Checks RCT4

Variable Name	Description	Treatment		Control		Difference	
		Mean	SD	Mean	SD	b	p
Age	Age (Team Average)	35.77	8.56	36.37	9.20	0.60	(0.590)
Background: Economics	Team members with Economics backgrounds (%)	0.15	0.29	0.15	0.29	0.00	(0.940)
Background: Other	Team members with no economics backgrounds (%)	0.08	0.11	0.09	0.16	0.01	(0.410)
Background: STEM	Team members with a STEM (Science Technology Engineering Mathematics) backgrounds (%)	0.30	0.39	0.36	0.43	0.06	(0.260)
Business Age	Age of the business (years)	2.48	3.22	3.28	5.17	0.80	(0.140)
Confidence	Agreement on a 1-5 scale with the following statements (Team Average): "We are confident in our entrepreneurial skills", "We are sure we are deploying the best strategy for our business", "We are confident in our ability to manage our business", "We master the competences necessary for our venture", "We are sure there is no better business model for our idea"	3.41	0.70	3.34	0.76	-0.07	(0.440)
Education	Highest educational level attained by team members (5=PhD, 4=MBA, 3=MSc, 2=BA, 1=high school, 0=otherwise; Team Average)	2.67	0.81	2.58	0.79	-0.10	(0.340)
Experience: Enterpreneurial	Number of years of entrepreneurial experience (Team Average)	3.85	3.49	4.64	5.95	0.79	(0.200)
Experience: Industry	Number of years of experience in industry (Team Average)	6.75	6.47	7.70	7.56	0.95	(0.280)
Experience: Managerial	Number of years of managerial experience (Team Average)	5.96	5.29	6.22	6.16	0.26	(0.730)
Experience: Work	Number of years of work experience (Team Average)	13.02	7.98	13.53	8.59	0.51	(0.620)
Gender (Female)	Proportion of women in the team	0.42	0.42	0.50	0.44	0.08	(0.150)
Hours: % Innovation monthly	Working hours dedicated to the design of new products or services in the last month (January 2019, %)	39.46	34.16	36.84	34.59	-2.62	(0.540)
Hours: % Innovation yearly	Working hours dedicated to the design of new products or services in the last year (2018, %)	46.05	33.35	40.02	32.68	-6.04	(0.140)
Hours: Total Weekly	Weekly hours dedicated to the company (Team Average)	31.55	18.57	29.61	17.18	-1.94	(0.390)
Idea Value: Mean	Estimated value of the project (mean, 0 to 100)	66.73	17.05	66.62	20.22	-0.11	(0.960)
Idea Value: Range	Estimated value of the project (range, 0 to 100)	39.26	22.03	38.00	21.94	-1.26	(0.650)
Probability Expansion	Probability of expanding the business outside of the current industry or market	68.25	27.40	66.59	28.12	-1.67	(0.630)
Probability Pivot Idea	Probability of making a radical change to the business	45.85	28.18	42.12	26.99	-3.72	(0.280)
Probability Pivot Problem	Probability of changing the problem and customer segment	38.18	26.16	40.55	26.26	2.38	(0.470)
Scientific Intensity	Scientific intensity	2.61	1.18	2.41	1.25	-0.20	(0.180)
Team Size	Number of team members	2.14	1.95	2.31	2.14	0.18	(0.490)
Turnover Annual	Annual turnover (2018) £	50616.11	145448.79	71977.35	195899.81	21361.24	(0.320)
Turnover Monthly	Monthly turnover (January 2019) £	5113.83	17734.76	6099.50	24490.47	985.67	(0.710)
Observations		133		128		261	

Table A5: Balance Checks Full Sample

Variable Name	Description	Treatment		Control		Difference	
		Mean	SD	Mean	SD	b	p
Business Age	Age of the business (years)	0.87	2.24	1.12	3.39	0.24	(0.244)
Education	Highest educational level attained by team members (5=PhD, 4=MBA, 3=MSc, 2=BA, 1=high school, 0=otherwise; Team Average)	2.24	0.88	2.21	0.91	-0.04	(0.585)
Experience: Managerial	Number of years of managerial experience (Team Average)	3.36	4.52	3.32	5.17	-0.04	(0.915)
Experience: Entrepreneurial	Number of years of entrepreneurial experience (Team Average)	2.12	3.43	2.21	4.21	0.09	(0.749)
Experience: Industry	Number of years of experience in industry (Team Average)	4.11	5.60	4.30	6.11	0.19	(0.663)
Team Size	Number of team members	2.34	1.65	2.33	1.67	-0.01	(0.924)
Turnover: Annual	Annual turnover EUR	20266.69	102520.24	28300.88	137456.09	8034.19	(0.364)
Observations		378		376		754	

3 Termination

Table A6: Termination OLS Cross-Section

VARIABLES	(1) Termination OLS Cross-Section Full Sample	(2) Termination OLS Cross-Section RCT1	(3) Termination OLS Cross-Section RCT2	(4) Termination OLS Cross-Section RCT3	(5) Termination OLS Cross-Section RCT4
Intervention	0.104***	0.035	0.096**	0.194**	0.097**
	(0.001)	(0.647)	(0.044)	(0.037)	(0.035)
Constant	0.283***	0.316	0.364***	0.761**	0.287***
	(0.000)	(0.219)	(0.001)	(0.011)	(0.002)
Observations	754	116	250	127	261
R-squared	0.078	0.183	0.034	0.158	0.026
Dummies for mentors	Yes	Yes	Yes	Yes	Yes
Dummies for RCTs	Yes	-	-	-	-
Clustered Errors	Intervention Mentor RCT	Intervention Mentor	Intervention Mentor	Intervention Mentor	Intervention Mentor

Robust pval in parentheses, *** p<0.01, ** p<0.05, * p<0.1. All specifications control for the variables that were unbalanced between the treatment and control group despite randomization.

Table A7: Termination Probit Cross-Section

VARIABLES	(1) Termination Probit Cross-Section Full Sample	(2) Termination Probit Cross-Section RCT1	(3) Termination Probit Cross-Section RCT2	(4) Termination Probit Cross-Section RCT3	(5) Termination Probit Cross-Section RCT4
Intervention	0.299***	0.105	0.249**	0.613***	0.295**
	(0.000)	(0.635)	(0.023)	(0.008)	(0.015)
Constant	-5.038***	-5.154***	-0.279	1.342**	-0.581***
	(0.000)	(0.000)	(0.114)	(0.050)	(0.002)
Observations	754	111	250	127	261
Dummies for mentors	Yes	Yes	Yes	Yes	Yes
Dummies for RCTs	Yes	-	-	-	-
Clustered Errors	Intervention Mentor RCT	Intervention Mentor	Intervention Mentor	Intervention Mentor	Intervention Mentor

Robust pval in parentheses, *** p<0.01, ** p<0.05, * p<0.1. All specifications control for the variables that were unbalanced between the treatment and control group despite randomization.

Table A8: Hazard of Termination

VARIABLES	(1) Hazard of termination Survival Full Sample	(2) Hazard of termination Survival RCT1	(3) Hazard of termination Survival RCT2	(4) Hazard of termination Survival RCT3	(5) Hazard of termination Survival RCT4
1.intervention	0.375***	0.101	0.334**	0.512	0.416**
	(0.000)	(0.664)	(0.012)	(0.158)	(0.014)
Observations	754	116	250	127	261
Dummies for mentors	Yes	Yes	Yes	Yes	Yes
Dummies for RCTs	Yes	-	-	-	-
Clustered Errors	Intervention Mentor RCT	Intervention Mentor	Intervention Mentor	Intervention Mentor	Intervention Mentor

Robust pval in parentheses, *** p<0.01, ** p<0.05, * p<0.1. All specifications control for the variables that were unbalanced between the treatment and control group despite randomization.

Table A9: Week of Termination

VARIABLES	(1) Week of termination OLS Full Sample	(2) Week of termination OLS RCT1	(3) Week of termination OLS RCT2	(4) Week of termination OLS RCT3	(5) Week of termination OLS RCT4
Intervention	-2.322**	-1.114	-4.137*	-1.407	-1.606*
	(0.012)	(0.551)	(0.062)	(0.452)	(0.090)
Constant	32.446***	42.557***	51.376***	31.697***	32.103***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Observations	754	116	250	127	261
R-squared	0.242	0.176	0.042	0.101	0.041
Dummies for mentors	Yes	Yes	Yes	Yes	Yes
Dummies for RCTs	Yes	-	-	-	-
Clustered Errors	Intervention Mentor RCT	Intervention Mentor	Intervention Mentor	Intervention Mentor	Intervention Mentor

Robust pval in parentheses, *** p<0.01, ** p<0.05, * p<0.1. All specifications control for the variables that were unbalanced between the treatment and control group despite randomization.

4 Pivot

Table A10: Number of Pivots

VARIABLES	(1) # Pivots OLS Cross-Section Full Sample	(2) # Pivots OLS Cross-Section RCT1	(3) # Pivots OLS Cross-Section RCT2	(4) # Pivots OLS Cross-Section RCT3	(5) # Pivots OLS Cross-Section RCT4
Intervention	-0.032	0.261**	0.012	-0.370	-0.038
	(0.654)	(0.021)	(0.841)	(0.311)	(0.588)
Constant	0.432***	0.536	1.238***	1.107	0.435***
	(0.000)	(0.217)	(0.000)	(0.448)	(0.000)
Observations	754	116	250	127	261
R-squared	0.120	0.105	0.070	0.068	0.019
Dummies for mentors	Yes	Yes	Yes	Yes	Yes
Dummies for RCTs	Yes	-	-	-	-
Clustered Errors	Intervention Mentor RCT	Intervention Mentor	Intervention Mentor	Intervention Mentor	Intervention Mentor

Robust pval in parentheses, *** p<0.01, ** p<0.05, * p<0.1. All specifications control for the variables that were unbalanced between the treatment and control group despite randomization.

Table A11: Pivot OLS

VARIABLES	(1) Pivoting once OLS Cross-Section Full Sample	(2) Pivoting once OLS Cross-Section RCT1	(3) Pivoting once OLS Cross-Section RCT2	(4) Pivoting once OLS Cross-Section RCT3	(5) Pivoting once OLS Cross-Section RCT4
Intervention	0.087***	0.027	0.123***	0.080	0.083*
	(0.001)	(0.726)	(0.001)	(0.288)	(0.079)
Constant	0.083***	-0.013	0.336***	0.807**	0.085**
	(0.007)	(0.736)	(0.004)	(0.023)	(0.030)
Observations	754	116	250	127	261
R-squared	0.082	0.147	0.064	0.059	0.040
Dummies for mentors	Yes	Yes	Yes	Yes	Yes
Dummies for RCTs	Yes	-	-	-	-
Clustered Errors	Intervention Mentor RCT	Intervention Mentor	Intervention Mentor	Intervention Mentor	Intervention Mentor

Robust pval in parentheses, *** p<0.01, ** p<0.05, * p<0.1. All specifications control for the variables that were unbalanced between the treatment and control group despite randomization.

Table A12: Pivot Multinomial Probit Full Sample

VARIABLES	(1) Pivoting once Multinomial Probit Cross-Section Full Sample	(2) Pivoting twice Multinomial Probit Cross-Section Full Sample	(3) Pivoting more than twice Multinomial Probit Cross-Section Full Sample
Intervention	0.370***	0.148	-0.287
	(0.010)	(0.397)	(0.117)
Constant	-1.374***	-2.104***	-2.438***
	(0.000)	(0.000)	(0.000)
Observations	754	754	754
Dummies for mentors and RCTs	Yes	Yes	Yes
Clustered Errors	Intervention Mentor RCT	Intervention Mentor RCT	Intervention Mentor RCT

Robust pval in parentheses, *** p<0.01, ** p<0.05, * p<0.1. All specifications control for the variables that were unbalanced between the treatment and control group despite randomization.

For RCT 1, a Multinomial Probit model does not converge due to the fact that only a few firms have pivoted more than once.

Table A13: Pivot Multinomial Probit RCT2

VARIABLES	(1) Pivoting once Multinomial Probit Cross-Section RCT2	(2) Pivoting twice Multinomial Probit Cross-Section RCT2	(3) Pivoting more than twice Multinomial Probit Cross-Section RCT2
Intervention	0.465**	0.107	-0.229
	(0.014)	(0.744)	(0.156)
Constant	-0.371	-1.192*	-0.936**
	(0.447)	(0.052)	(0.016)
Observations	250	250	250
Dummies for mentors and RCTs	Yes	Yes	Yes
Clustered Errors	Intervention Mentor RCT	Intervention Mentor RCT	Intervention Mentor RCT

Robust pval in parentheses, *** p<0.01, ** p<0.05, * p<0.1. All specifications control for the variables that were unbalanced between the treatment and control group despite randomization.

Table A14: Pivot Multinomial Probit RCT3

VARIABLES	(1) Pivoting once Multinomial Probit Cross-Section RCT3	(2) Pivoting twice Multinomial Probit Cross-Section RCT3	(3) Pivoting more than twice Multinomial Probit Cross-Section RCT3
Intervention	0.117	0.027	-0.892
	(0.795)	(0.941)	(0.190)
Constant	1.062	-1.680	-1.757
	(0.485)	(0.254)	(0.405)
Observations	127	127	127
Dummies for mentors and RCTs	Yes	Yes	Yes
Clustered Errors	Intervention Mentor RCT	Intervention Mentor RCT	Intervention Mentor RCT

Robust pval in parentheses, *** p<0.01, ** p<0.05, * p<0.1. All specifications control for the variables that were unbalanced between the treatment and control group despite randomization.

Table A15: Pivot Multinomial Probit RCT4

VARIABLES	(1) Pivoting once Multinomial Probit Cross-Section RCT4	(2) Pivoting twice Multinomial Probit Cross-Section RCT4	(3) Pivoting more than twice Multinomial Probit Cross-Section RCT4
Intervention	0.422*	-0.070	-0.314
	(0.100)	(0.819)	(0.269)
Constant	-1.259***	-1.415***	-1.916***
	(0.000)	(0.000)	(0.000)
Observations	261	261	261
Dummies for mentors and RCTs	Yes	Yes	Yes
Clustered Errors	Intervention Mentor RCT	Intervention Mentor RCT	Intervention Mentor RCT

Robust pval in parentheses, *** p<0.01, ** p<0.05, * p<0.1. All specifications control for the variables that were unbalanced between the treatment and control group despite randomization.

5 Performance

Table A16: Performance OLS Cross-Section

VARIABLES	(1) Revenue OLS Cross-section Full Sample	(2) Revenue OLS Cross-section RCT1	(3) Revenue OLS Cross-section RCT2	(4) Revenue OLS Cross-section RCT3	(5) Revenue OLS Cross-section RCT4
Intervention	6,504.108**	10,799.493	1,514.605	263.431	12,227.935
	(0.046)	(0.125)	(0.136)	(0.269)	(0.164)
Constant	9,039.968***	-4,899.747	-445.998	-594.302	6,297.301
	(0.006)	(0.403)	(0.859)	(0.484)	(0.344)
Observations	754	116	250	127	261
R-squared	0.086	0.220	0.023	0.052	0.036
Dummies for mentors	Yes	Yes	Yes	Yes	Yes
Dummies for RCTs	Yes	-	-	-	-
Clustered Errors	Intervention Mentor RCT	Intervention Mentor	Intervention Mentor	Intervention Mentor	Intervention Mentor

Robust pval in parentheses, *** p<0.01, ** p<0.05, * p<0.1. All specifications control for the variables that were unbalanced between the treatment and control group despite randomization.

6 Scientific Intensity

In this section, we provide additional details about the coding scheme used to assess the extent to which entrepreneurs use a scientific approach to decision-making. A team of research assistants conducted regular phone calls with all entrepreneurs taking part to our programs. Calls followed a detailed protocol with a script including a number of open-ended questions which are used to measure scientific decision-making. In using open-ended questions, we follow an approach similar to the one employed by Bloom and Van Reenen (2007, 2010) for their World Management Survey. Like Bloom and Van Reenen, we asked open-ended ques-

tions until an accurate assessment of the decision-making practices could be made by the research assistants. This allowed them to gather detailed information about decision-making approaches rather than ask directly about respondents' perception and aspirations. In addition, respondents were not aware that their responses were being scored against a predefined coding scheme, which helped ensure the collection of unbiased information. Decision-making practices were scored from 0 (lowest score) to 5 (highest score), across four key areas: 1) Theory, 2) Hypotheses, 3) Tests, 4) Evaluation.

When entrepreneurs formulate a theory, they elaborate a set of core ideas (and the key relationship between them) that explains why their business proposition should be viable. A theory generates a firm-specific point of view (Felin and Zenger, 2017) that flashes out what key assumptions decision-makers hold. These assumptions are then articulated as hypotheses or predictions that flow logically from the theory (Popper, 1972). Hypotheses provide the basis for a data gathering process that provides evidence in support or against of such hypotheses. Data can be gathered through tests of various nature, including qualitative (interviews, observations, etc.) and quantitative (data collection through surveys, A/B testing, etc.) data gathering techniques. Following the data collection process, entrepreneurs carefully analyze the results of their tests and re-evaluate their theory in light of these results.

In line with key literature in this area, we consider each component of the scientific decision-making approach as a multifaceted construct. For instance, the articulation of a theory rests on a wide variety of aspects, such as its clarity, level of detail, the extent to which it is based on evidence and the extent to which it considers alternative explanations. To adequately capture the multiple dimensions of each component, we identified some sub-components that measure the key aspects that define theory, hypotheses, tests, and evaluation. In addition, each of these sub-elements can greatly vary in quality across entrepreneurs. One entrepreneur might have an extremely clear theory related to how his/her firm generates value for customers, while another might have a very murky explanation for his/her value creation process. All research assistants received extensive training prior to performing calls. The multiple training and practice sessions organized by the research team clarified how to score each sub-component. These sessions also provided clear examples with related scores to create an objective standard research assistants could refer to when coding. We provide an overview of the sub-components of the scientific approach and their related scores in Table A20 below.

Table A17: Scientific Intensity Components

Component	Sub-component	Definition	Score
Theory	Clarity of theory	The extent to which the theory is understandable	0 (no theory), 1 (not clear at all) to 5 (extremely clear)
Theory	Articulation of theory	The extent to which the theory is detailed	0 (no theory), 1 (not detailed at all) to 5 (extremely detailed)
Theory	Consideration of alternatives	The extent to which the theory includes alternative possible options	0 (no theory), 1 (no consideration of alternatives at all) to 5 (careful consideration of many alternatives)
Theory	Theory based on evidence	The extent to which the theory is based on objective evidence	0 (no theory), 1 (not based on objective evidence at all) to 5 (extremely based on objective evidence)
Hypotheses	Explicitness of hypotheses	The extent to which the respondent can articulate the fundamental assumptions that make his/her business viable	0 (no hypotheses), 1 (not explicit at all) to 5 (extremely explicit)
Hypotheses	Coherence of hypotheses	The extent to which hypotheses are coherent with the theory	0 (no hypotheses), 1 (not coherent at all) to 5 (extremely coherent)
Hypotheses	Level of details of hypotheses	The extent to which hypotheses clearly indicate the details of what the entrepreneur wishes to learn and how to measure it	0 (no hypotheses), 1 (not detailed at all) to 5 (extremely detailed)
Hypotheses	Falsifiability of hypotheses	The extent to which it is possible to clearly determine (after tests) whether the hypotheses are supported or not	0 (no hypotheses), 1 (not falsifiable at all) to 5 (extremely falsifiable)
Tests	Coherence of tests	The extent to which the test is coherent with the hypotheses	0 (no tests), 1 (not coherent at all) to 5 (extremely coherent)
Tests	Validity of tests	The extent to which the test has been conducted in a context similar to which the business operates	0 (no tests), 1 (not valid at all) to 5 (extremely valid)
Tests	Representativeness of tests	The extent to which the test has been conducted with a sample that is representative of the broad group the firm targets	0 (no tests), 1 (not representative at all) to 5 (extremely representative)
Tests	Rigorousness of tests	The extent to which the appropriate test and procedure for that type of test have been chosen for hypotheses-testing	0 (no tests), 1 (not rigorous at all) to 5 (extremely rigorous)
Evaluation	Data-based assessment	The extent to which the evaluation is based on data	0 (no evaluation), 1 (not based on data at all) to 5 (extremely based on data)
Evaluation	Coherence of measures	The extent to which the measure used are consistent with the learning objective the entrepreneur has in mind	0 (no evaluation), 1 (not coherent at all) to 5 (extremely coherent)
Evaluation	Systematic evaluation	The extent to which the evaluation is based on systematically collected and analysed data	0 (no evaluation), 1 (not systematic at all) to 5 (extremely systematic)
Evaluation	Explanatory power of evaluation	The extent to which the evaluation results in clarity on the main findings from the test and their implications for the business	0 (no evaluation),1 (not explanatory at all) to 5 (extremely explanatory)

Table A18: Scientific Intensity OLS Cross-Section

	(1)	(2)	(3)	(4)	(5)
	Scientific intensity	Scientific intensity	Scientific intensity	Scientific intensity	Scientific intensity
	OLS Cross-section	OLS Cross-section	OLS Cross-section	OLS Cross-section	OLS Cross-section
VARIABLES	Full Sample	RCT1	RCT2	RCT3	RCT4
Intervention	0.331***	0.581***	0.206*	0.341	0.321**
	(0.000)	(0.002)	(0.060)	(0.164)	(0.015)
Constant	2.081***	1.155***	2.488***	1.298**	2.086***
	(0.000)	(0.006)	(0.000)	(0.018)	(0.000)
Observations	754	116	250	127	261
R-squared	0.120	0.178	0.092	0.057	0.028
Dummies for mentors	Yes	Yes	Yes	Yes	Yes
Dummies for RCTs	Yes	-	-	-	-
Clustered Errors	Intervention Mentor RCT	Intervention Mentor	Intervention Mentor	Intervention Mentor	Intervention Mentor

Robust pval in parentheses, *** p<0.01, ** p<0.05, * p<0.1. All specifications control for the variables that were unbalanced between the treatment and control group despite randomization.

7 Instrumenting Scientific Intensity

Table A19: Termination 2SLS

VARIABLES	(1) Termination 2SLS Cross-Section Full Sample	(2) Termination 2SLS Cross-Section RCT1	(3) Termination 2SLS Cross-Section RCT2	(4) Termination 2SLS Cross-Section RCT3	(5) Termination 2SLS Cross-Section RCT4
Average Scientific Intensity	0.299***	0.082	0.344*	0.597	0.253***
	(0.001)	(0.615)	(0.060)	(0.177)	(0.003)
Constant	-0.283	0.345	-0.213	-0.191	-0.188
	(0.125)	(0.289)	(0.559)	(0.853)	(0.290)
Observations	754	116	250	127	261
R-squared	-0.485	0.101	-0.832	-1.737	-0.309
Dummies for mentors	Yes	Yes	Yes	Yes	Yes
Dummies for RCTs	Yes	-	-	-	-
Clustered Errors	Intervention Mentor RCT	Intervention Mentor	Intervention Mentor	Intervention Mentor	Intervention Mentor

Robust pval in parentheses, *** p<0.01, ** p<0.05, * p<0.1. All specifications control for the variables that were unbalanced between the treatment and control group despite randomization.

Table A20: Termination IV Probit

VARIABLES	(1) Termination IV Probit Cross-Section Full Sample	(2) Termination IV Probit Cross-Section RCT1	(3) Termination IV Probit Cross-Section RCT2	(4) Termination IV Probit Cross-Section RCT3	(5) Termination IV Probit Cross-Section RCT4
Average Scientific Intensity	0.603***	0.257	0.576***	0.833***	0.611***
	(0.000)	(0.547)	(0.000)	(0.000)	(0.000)
Constant	-5.456	-5.737***	-1.657***	-0.998	-1.896***
	(0.000)	(0.000)	(0.000)	(0.275)	(0.000)
Observations	754	116	250	127	261
Dummies for mentors	Yes	Yes	Yes	Yes	Yes
Dummies for RCTs	Yes	-	-	-	-
Clustered Errors	Intervention Mentor RCT	Intervention Mentor	Intervention Mentor	Intervention Mentor	Intervention Mentor

Robust pval in parentheses, *** p<0.01, ** p<0.05, * p<0.1. All specifications control for the variables that were unbalanced between the treatment and control group despite randomization.

Table A21: Pivot 2SLS

VARIABLES	(1) Pivoting once 2SLS Cross-Section Full Sample	(2) Pivoting once 2SLS Cross-Section RCT1	(3) Pivoting once 2SLS Cross-Section RCT2	(4) Pivoting once 2SLS Cross-Section RCT3	(5) Pivoting once 2SLS Cross-Section RCT4
Average Scientific Intensity	0.251***	0.063	0.441***	0.248	0.219*
	(0.000)	(0.693)	(0.007)	(0.134)	(0.057)
Constant	-0.391**	-0.119	-0.477	0.334	-0.325
	(0.016)	(0.693)	(0.136)	(0.553)	(0.180)
Observations	754	116	250	127	261
R-squared	-0.225	0.130	-0.846	-0.191	-0.306
Dummies for mentors	Yes	Yes	Yes	Yes	Yes
Dummies for RCTs	Yes	-	-	-	-
Clustered Errors	Intervention Mentor RCT	Intervention Mentor	Intervention Mentor	Intervention Mentor	Intervention Mentor

Robust pval in parentheses, *** p<0.01, ** p<0.05, * p<0.1. All specifications control for the variables that were unbalanced between the treatment and control group despite randomization.

Table A22: Pivot IV Probit

VARIABLES	Pivoting once IV Probit Cross-Section Full Sample	Pivoting once IV Probit RCT1	Pivoting once IV Probit Cross-Section RCT2	Pivoting once IV Probit Cross-Section RCT3	Pivoting once IV Probit Cross-Section RCT4
Average Scientific Intensity	0.684***	0.283	0.816***	0.605**	0.660***
	(0.000)	(0.658)	(0.000)	(0.021)	(0.000)
Constant	-5.687***	-6.558***	-2.302***	-0.290	-2.236***
	(0.000)	(0.000)	(0.000)	(0.816)	(0.000)
Observations	754	116	250	127	261
Dummies for mentors	Yes	Yes	Yes	Yes	Yes
Dummies for RCTs	Yes	-	-	-	-
Clustered Errors	Intervention Mentor RCT	Intervention Mentor	Intervention Mentor	Intervention Mentor	Intervention Mentor

Robust pval in parentheses, *** p<0.01, ** p<0.05, * p<0.1. All specifications control for the variables that were unbalanced between the treatment and control group despite randomization.

Table A23: Performance 2SLS

VARIABLES	(1) Revenue 2SLS Cross-Section Full Sample	(2) Revenue 2SLS Cross-Section RCT1	(3) Revenue 2SLS Cross-Section RCT2	(4) Revenue 2SLS Cross-Section RCT3	(5) Revenue 2SLS Cross-Section RCT4
Average Scientific Intensity	18,703.974*	25,182.411*	5,437.553	812.963*	32,031.798
	(0.056)	(0.087)	(0.155)	(0.056)	(0.172)
Constant	-26,351.771	-47,474.210*	-10,420.817	-1,979.193*	-53,791.488
	(0.208)	(0.090)	(0.141)	(0.061)	(0.273)
Observations	754	116	250	127	261
R-squared	0.064	0.106	-0.134	0.043	0.009
Dummies for mentors	Yes	Yes	Yes	Yes	Yes
Dummies for RCTs	Yes	-	-	-	-
Clustered Errors	Intervention Mentor RCT	Intervention Mentor	Intervention Mentor	Intervention Mentor	Intervention Mentor

Robust pval in parentheses, *** p<0.01, ** p<0.05, * p<0.1. All specifications control for the variables that were unbalanced between the treatment and control group despite randomization.

8 References

Bloom, Nicholas, and John Van Reenen. 2007. "Measuring and Explaining Management Practices across Firms and Countries." *The Quarterly Journal of Economics* 122 (4): 1351-1408.

Bloom, Nicholas, and John Van Reenen. 2010. "New approaches to surveying organizations." *American Economic Review*, 100(2), pp.105-09.

Felin, Teppo, and Todd R. Zenger. 2017 "The theory-based view: Economic actors as theorists." *Strategy Science* 2.4: 258-271.

Popper, Karl. 1972. "Objective Knowledge, an Evolutionary Approach". *Clarendon Press.*