Understanding Probabilistic Reasoning in Innovation

Andrea Coali[1], Alfonso Gambardella[1], and Elena Novelli[2]

[1]Bocconi University
[2]Bayes Business School

Preliminary Draft
Please do not circulate without authors' permission

January 2022

Abstract

Decision-making processes in the context of innovation are characterized by high uncertainty and prone to decision-making biases. In this paper we explore the implications of adopting what we call a *scientific approach to decision making*, based on probabilistic reasoning. We focus in a context in which the innovation underlying the decision making process is a new entrepreneurial venture. We develop a structural model to disentangle and identify two separate but complementary effects of this approach. The estimation of our structural model, based on data from two randomized control trials (RCTs) involving early stage start-ups, shows that scientific entrepreneurs tend to be more conservative in assessing the value over their ideas, an effect that we call *debiasing effect*. It also shows that, conditional on their decision to remain operational, scientific entrepreneurs tend to perform better, an effect that we call *learning effect*. We finally show how the selection induced by the debiasing effect does not significantly increase the rate of false-negatives for treated firms compared to the control group. We discuss the implications for future research and practice.

1 Introduction

An astonishing 90% of newly-born start-ups fail within ten years, with around 21% of them failing already in their first year (NationalBusinessCapital, 2020). Part of the reasons behind this pattern relates to the fact that entrepreneurs, and innovators developing new projects more in general, face a decision-making process that is characterized by high uncertainty along multiple dimensions (McGrath, 1997, Folta, 1998). In the presence of uncertainty, the assessment on the value of novel ideas, and therefore, entrepreneurial decision-making becomes difficult.

One way to deal with uncertainty would be to use a probabilistic decision-making process, making decisions on uncertain outcomes based on probabilistic information. However, research shows that decision makers often deviate from "rational" or probabilistic reasoning. Prior research has shown that entrepreneurs often do not follow a systematic decision-making process (Bloom and Van Reenen, 2007) and even ignore important information (Bennett and Chatterji, 2019, Tversky and Kahneman, 1974). Alternatively, they rely on heuristic principles to reduce the complex tasks of probability assessment and value prediction to a simpler task (Tversky and Kahneman, 1974). Whereas this is certainly a useful process and can lead to good results (Bingham and Eisenhardt, 2011), research shows that it can also lead to a plethora of severe and systematic biases (Tversky and Kahneman, 1974).

Differently from prior research that has largely explored deviations from a probabilistic approach to decision making, in this paper we aim to understand more about what happens when entrepreneurs are induced to employ probabilistic decision-making processes more systematically. Specifically, we address the following research question: *What are the implications of probabilistic reasoning on entrepreneurial decision making?*

This question has been so far under investigated, but this is perhaps unsurprising given the research-design challenges that addressing this question involves. First, answering this question requires an exogenous shock that induced entrepreneurs to reason and make decisions in probabilistic terms. Second, it requires observing the decision-making process of entrepreneurs in detail as well as the outcomes originating from that process, such as the specific decisions made and their performance. Third, it requires comparing entrepreneurs using a probabilistic approach to a proper counterfactual.

To respond to these challenges, we employ a randomized control trial (RCT) design, where we teach a sample of entrepreneurs to reason in probabilistic terms, developing a theory of their business idea and the problems it would likely solve, developing hypotheses flowing logically from it, designing tests that can provide them with signals regarding the probabilities of those hypotheses being supported with data, and evaluating those results in a disciplined way against their prior theoretical expectations. Following related work (Ashraf, Banerjee, and Nourani, 2021; Camuffo et al., 2020), we call this "a scientific approach to innovation management" as this process resembles the one followed by scientists in developing new knowledge. We maintain the other half of the sample in a control condition, where they are delivered equivalent management content but without a scientific approach. We then monitor these entrepreneurs for a variety of months, collecting detailed data on their decision-making process, choices, and performance.

We dig one step deeper compared to prior research (e.g. Camuffo et al., 2020) and develop a structural model that allows us to disentangle and identify precisely two different effects that the exposure to a scientific approach has on entrepreneurial decision making. The estimation of our structural model, based on data from two randomized control trials

(RCTs) conducted in Milan (2017) and Turin (2018), involving 377 early stage start-ups, shows that entrepreneurs following a scientific approach to decision making perform on average better, an effect that we call the *learning effect*; but they also make an earlier and faster downward adjustment of their business' expected values, ultimately showing a higher rate of project termination, an effect that we call the *debiasing effect*. This in line with the intuition that entrepreneurs tend to sometimes pursue “falsely positive ideas” and that probabilistic reasoning can help them to reach a more conservative evaluation of their ideas. But that it can also help them understand the problem faced better and identify a better solution, achieving superior performance.

The co-existence of these two effects leads to a natural follow up question. Given that a scientific approach to decision making leads both to a more conservative assessment of ideas but also to superior learning and performance, are scientific entrepreneurs excessively cautious when terminating their projects, effectively discarding ideas that could, instead, be successful? In other words, is it possible that while this method leads entrepreneurs to reduce the number of ideas that others would have falsely seen as positive, it also leads entrepreneurs to discard too many (falsely negative) ideas?

Answering this question is no easy task, as it would require the determination of the value of the terminated ideas, were they not terminated, which is clearly not possible. However, in the attempt to nurture this important debate, we provide some suggestive evidence on the matter by combining three main pieces of evidence. First, we asked a team of professional to provide an evaluation of the ideas developed by entrepreneurs in the two RCTs, based on the pitches that entrepreneurs provided at the baseline. Second, we analyze the pattern of revenue growth over time comparing firms that terminated with those that remained active in the market, conditional on the treatment received. Third, we analyze data on the share of firms receiving external finance. These data suggest that the selection of firms induced by the *scientific training* has been a positive one. In other words, we do not find evidence in favor of the idea that treated entrepreneurs who decided to terminate were too conservative and that their rate of false negative was higher that their rate of reduction in false positives. As an additional exercise, leveraging on the results of the structural estimation, we identify different assumptions and use them to calculate two different scenarios at different end of the spectrum of possibilities. In what we call the *lower-bound* condition, we assume that the value model for firms that stayed in the market is exactly the same as for those firms that terminate, not considering the positive learning effect of ”scientific” entrepreneurs. Results show that, under this condition, the reduction in false positives is compensated by the increase in false negatives. In what we call the *upper-bound* condition we consider, instead, the positive learning effect of the scientific approach. Results show again that the selection has been a ”positive” one, meaning that the reduction in false positives more than compensates the increase in false negatives. Finally, we also replicate results on termination and on the selection trade-off using a business simulation game.

The remainder of the paper is structured as follows. Section II elaborates on the scientific approach and its implications. Section III details the structural framework we develop. Section IV describes our methodology and data. Section V reports the estimation results. In Section VI we discuss the trade-off between the false-positive and false-negative rate, while Section VII offer concluding reflections.

2 The Scientific Approach and Probabilistic Reasoning

Consider an entrepreneur with the goal of developing an innovative product or service, or willing to launch a new business. Typically, she starts with an intuition coming from observation of real-world phenomena, spotting a problem that would need an innovative solution to be solved. Before deciding to embark in a new project, our decision-maker will evaluate whether her idea is worth the development efforts and this assessment will be made at regular intervals throughout the life of the project. At every round of assessment, her decision can be represented as a choice between three mutually exclusive alternatives (Kirtley and O'Mahony, 2020, Lieberman, Lee, and Folta, 2017, Eisenhardt and Bingham, 2017, Gans, Stern, and Wu, 2019): 1) *terminate* the project, if the entrepreneur believes it won't generate sufficient value; 2) change substantial elements of the idea or project to improve its value (what we refer to as *pivoting*); 3) *continue* the development of the project along its current trajectory.

Along the way her assessment will be based on considerations regarding the multiple potential scenarios she could face in the environment in which she operates, over which there is uncertainty. This uncertainty could originate, for instance, from the fact that she is not yet familiar with customers' preferences in the environment she targets; or from the fact that these preferences might be subject to change. She will also consider the actions that she can take to deal with the multiple scenarios she might be facing. At the very early stage of her process, actions could concern the development of the idea. At later stages, actions could be linked to the idea commercialization and could include, for example, the development of different versions of the same product, service, or business model, or the implementation of alternative marketing strategies.

Of course, every action she envisions might have a different value under different scenarios. Suppose for instance that our entrepreneur's idea is about developing an innovative service for car-sharing, but there is uncertainty regarding the extent to which cars are going to be relevant in the medium term in the context in which she is operating. If the context in which she operates is going through a massive drop in the use of cars and an increase in the use of bikes, the action of pursuing such car-sharing project could have a negative value. Instead, if renting cars is a valuable option in the context in which she operates, the action of pursuing such car-sharing project idea could have a high value. Depending on what her assessment on the scenario more likely to manifest itself and what value she envisions her actions to have, she could decide to terminate the project, or to pivot to a new version of the project, or to simply continue the development of the project along its natural trajectory.

Prior literature in management, entrepreneurship and innovation has shown that very often entrepreneurs make these decisions simply following their gut feelings as opposed to trying and predicting the likelihood of scenarios and the value of actions (Bennett and Chatterji, 2019). Other streams of research have instead documented the use of structured approaches that support entrepreneurial decision making, such as the use of structured practices (Bloom and Van Reenen, 2007; Yang et al., 2020), trial and error (von Hippel and Tyre, 1995), effectuation (Sarasvathy, 2001), experimentation and lean methods (Ries, 2011; Thomke, 1998), heuristics (Bingham and Eisenhardt, 2011; Bingham and Haleblian, 2012). Whereas these approaches can be beneficial and lead to superior performance (e.g. Bingham and Eisenhardt, 2011), existing research is also replete with examples that show that the use of these practices can also lead to important biases (Tversky and Kahneman, 1974; Felin et al., 2020)

The key question we, thus, address in this paper is: to what extent can entrepreneurs, in the face of uncertainty, use probabilistic reasoning to discover relevant scenarios and assess the value of their entrepreneurial ideas under those scenarios? And what would the implications for entrepreneurial decision making of using such an approach be?

To this end, we explore a decision making approach that is based on "probabilistic" reasoning. Following related work (Camuffo et al., 2020, Ashraf, Banerjee, and Nourani, 2021), we refer to this approach as to *a scientific approach to decision making* due to its resemblance to the process followed by scientists when they approach a problem. The key tenet of this methods is that "scientific" entrepreneurs follow a five-step process in making decisions. They start from thinking about the problem in a broad way, effectively developing a theory of the problem and identifying the key elements on which they should focus when developing their projects, for instance the relevant scenarios they should take into account and the relevant actions they could consider in those scenarios (step 1). Scientific entrepreneurs then formulate testable and falsifiable hypotheses based on their theory (step 2) and they test them via carefully designed tests (step 3). The outcomes of these tests can be used as "signals" by the entrepreneur to assess the value of the idea. Signals will then be rigorously evaluated against their theory and prior beliefs. Such evaluation (step 4) will ultimately lead to a decision on the future of the idea (step 5).

For instance, if our entrepreneur approached the problem in a scientific way, she would start by developing a theory about the problem that her such car-sharing service addresses and the way in which it addresses it, and how the value of those actions would change under different relevant scenarios. She would then develop some core hypotheses regarding the scenarios she is facing and the value of actions in those scenarios, such as that car transportation in large cities is highly valued by certain categories of individuals, that owning a car in a large city is not practical due to the high fixed costs and the limited use per person, and that individuals consider sharing cars a viable option. She will test such hypotheses by collecting relevant information from a representative group of target customers. She will then evaluate the results obtained from the test against her theory, to ultimately reach a decision about whether to continue with the development of her idea, terminate the project, or pivot.

What is the value of a "scientific approach" compared to other approaches? The thrust of our work is that the scientific approach has two main effects on entrepreneurial decision making. First, it improves entrepreneurs' ability to develop a more objective and conservative assessment of the value of the business, reducing the impact of decision making biases such as, for instance, overconfidence. We call this the *debiasing effect* of the scientific method. The development of a general theory of the problem and its articulation into hypotheses, helps scientific entrepreneurs focusing on the relevant assumptions behind the business idea that need to hold for the value proposition to generate value, effectively leading to the formulation of more structured prior probabilities. This is complemented with the design of high quality tests that can provide them with signals about the extent to which their theory and hypotheses, and their priors more in general, are actually supported by data from the environment. Relating signals received back to the broad theory leads to a validation of the theory or to a rejection of it. This results in an update of their priors toward something more objective and to a more conservative expectation on the value of the idea. If this is too low, entrepreneurs may be thus choose to terminate the project. For instance, if our entrepreneur collected a negative signal on people's willingness to share cars due to hygiene concerns in a pandemic world, she would be more likely to form a negative value expectation and terminate the project. This effect is likely to lead scientific entrepreneurs to terminate their projects more often than non-scientific entrepreneurs.

The second effect that the scientific approach has on entrepreneurial decision making is that it likely improves the ability of scientific entrepreneurs to identify the changes to the business proposition would lead it to develop more or less value more easily or more rapidly. We call this the *learning effect* of the scientific method. The development of a theory and its articulation into hypotheses leads to a clear identification of the core elements or the problem and the relationships between them. This facilitates a quicker and more efficient search of the solution space, as it leads actors to identify ex-ante the characteristics of the solution (e.g., Camuffo et al., 2020, Felin, Kauffman, and Zenger, 2020). For example, if our entrepreneur obtained a positive signal on her hypotheses that car transportation in large cities is highly valued by households with young children who cannot use other types of transportation such as bikes easily, she would immediately be able to understand that this will also make the service appealing to households that include the elderly and could pivot in this direction. This effect is likely to lead scientific entrepreneurs to perform better, conditional on the fact that they do not terminate their project.

3 Structural Framework

Our theory suggests that there are two effects associated with the application of a "scientific approach": a *debiasing effect* and a *learning effect*. The goal of this paper is that of disentangling and identifying both effects. Research on the impact of decision making on performance has been limited by the fact that studying this issue requires facing many challenges. One of this is that firm performance is observable only for firms that have not terminated their activities, creating a source of selection bias to deal with. To address this challenge, we develop a structural decision-making framework and estimate it with a multi-equation simultaneous maximum-likelihood model. Another challenge is that the choice to use a specific decision-making process is endogenous to firm performance, and to other firm characteristics. In our setting, as explained in detail in Section 4, this challenge is mitigated by recurring to a randomized setting and to an exogenous treatment offered to entrepreneurs, thus leading to an ideally unbiased parameter estimation.

We start with a *value equation*. We consider the realized value (v) of the business idea and model it as:

$$v = a + \theta T + \sigma \epsilon \tag{1}$$

Where $a = \gamma X$, with X being a set of controls recorded at the baseline period. We assume that the realized value of the business is a function of a set of controls X and of whether the entrepreneur has been trained with the scientific approach. Purposefully, the dummy T separates entrepreneurs trained with the "scientific approach" with entrepreneurs in the control group, hence θ identifies what we have labeled to be the *learning effect*.

Our model postulates that entrepreneurs explore their business ideas and form expectations of their potential value and probability of success over time. Let us denote these expectations by $\hat{v}$. We assume that entrepreneurs decide to keep developing their business if such expectations are higher than their outside option w. Therefore, in our framework, such estimations are crucial as the decision between continuing with the development of the business or terminating the project is based on the evaluation of that expectation with respect to an individual outside option.

In our model, we represent the entrepreneur's decision making process as characterized by four crucial points in time: (i) the baseline, before the training (0 - the *Baseline Evaluation*), (ii) after the training (E - the *Early Evaluation*), (iii) later in time after the training (L - the *Late Evaluation*), and (iv) at the time of the decision whether to remain active or terminate the business (F - the *Final Evaluation*). To clarify what we mean with the *Late Evaluation* period, it is worth explaining briefly the structure of our data collection process. We follow entrepreneurs for multiple data points, recording whether they are still operational in the market at each of them. Once an entrepreneur decides to terminate his/her project, our data collection reaches a natural end. Hence, we consider as *Late Evaluation* the last available data point before such decision. If an entrepreneur never terminates the project within the observation window, we consider as *Late Evaluation* the end of our observation window. Hence, we develop four equations:

$$\hat{v_0} = c + c_0 + \sigma_0 \epsilon \tag{2}$$

$$\hat{v_E} = c + c_0' + c_E + (c_{ET} + \theta)T + \sigma_E \epsilon \tag{3}$$

$$\hat{v_L} = c + c_0' + c_L + (c_{LT} + \theta)T + \sigma_L \epsilon \tag{4}$$

$$\hat{v_F} = c + c_0' + c_F + (c_{FT} + \theta)T + \sigma_F \epsilon \tag{5}$$

The baseline evaluation (Eq. 2) happens before the training and therefore it depends on a series of factors independent of the training, such as education levels, age or previous startup experience, which we include in the vector c_0, whereas c represents a constant term.

Once the intervention starts, we assume it to have an effect on the evaluation. Equations relating the early (Eq. 3), late (Eq. 4) and final (Eq. 5) evaluations include c, which is the constant term, c_0' that identifies constant idiosyncratic factors as above but that we assume could vary in terms of magnitude (as represented by the apostrophe), and c_j (where $j = E, L, F$), which identify contemporaneous factors affecting the value estimation.

In addition to this, we assume that the intervention has two different effects on the evaluation made by entrepreneurs, i.e.the *learning effect* θ and the *debiasing effect* c_{jT} (which is not restricted to be constant over time) cannot be identified empirically. This is because we assume that the scientific approach helps entrepreneurs understand the opportunity for positive performance since the beginning of its application, but that its debiasing effect might vary overtime. To achieve the goal of this paper of identifying these two effects, additional structure in our model is needed.

We first build on the previous steps, and generalize the decision-making process as:

$$\hat{v_j} = c + c_0' + c_j + (c_{jT} + \theta)T + \sigma_j \epsilon \geq w_j \tag{6}$$

With j representing the different time periods, and w_j representing the entrepreneur's outside option (which we assume vary over time and that we represent as w_j=w_{j-1}+b_j). This condition is verified if and only if:

$$\epsilon \geq \frac{w_j - c - c_0' - c_j - (c_jT + \theta)T}{\sigma_j} = z_j \tag{7}$$

We relabel the right hand side of equation (Eq. 7) z_j. When the decision of staying in the market is made (which we labeled with F), for entrepreneurs who choose to terminate their project, we cannot observe the values above. Rather, we only observe the final outcome. Hence, for F, we consider the following equation based on a latent model for the probability of remaining active in the market:

$$Pr(Stay) = \Phi(\frac{-w_F + c + c_0' + c_L + (c_{FT} + \theta)T}{\sigma_F}) \tag{8}$$

We can now re-arrange some equations to retrieve the structural parameters of interest. Let us rewrite Eq. 7 for the first three data points (0, E and L).

$$z_0 = \frac{w_0 - c - c_0}{\sigma_0} \tag{9}$$

$$z_E = \frac{w_E - c - c_0' - c_E - (c_{ET} + \theta)T}{\sigma_E} \tag{10}$$

$$z_L = \frac{w_L - c - c_0' - c_L - (c_{LT} + \theta)T}{\sigma_L} \tag{11}$$

Plugging Eq. 11 into Eq. 8, Eq. 10 into Eq. 11 and Eq. 9 into Eq. 10, we obtain:

$$z_E = \frac{\sigma_0 z_0 + b_E - (c_0' - c_0) - c_E - (c_{ET} + \theta)T}{\sigma_E} \tag{12}$$

$$z_L = \frac{\sigma_E z_E + b_L - (c_L - c_E) - (c_{LT} - c_{ET})T}{\sigma_L} \tag{13}$$

$$Pr(Stay) = \Phi(\frac{-\sigma_L z_L - b_F + (c_F - c_L) + (c_{FT} - c_{LT})T)}{\sigma_F}) \tag{14}$$

We now turn to the description of the data and methodology used to estimate the whole model and the structural coefficients of interest.

4 Methodology and Data

4.1 Experimental Design

To estimate the structural framework we leverage on data from two field experiments, delivered in the context of a business support program that was offered to entrepreneurs in Milan and Turin (Italy). Both RCTs shared the same structure, type of intervention, and data collection process. The two RCTs were held asynchronously[1].

Both programs were advertised nationally over multiple offline and online channels. The advertisement campaign lasted for several weeks to ensure recruitment of at least 100 entrepreneurs per batch. The campaign promoted the program as a cutting edge business support program, offered free of charge to early stage entrepreneurs operating in any industry. The focus on early stage startups ensured that participants into the programs were highly involved in the decision making process. To apply, entrepreneurs were required to fill in an online survey and complete a telephone interview. In total, the first RCT (Milan) recruited 250 entrepreneurs, and the second (Turin) recruited 127.

Entrepreneurs were assigned to either a treatment or a control group through simple randomization. We checked that the randomization was successful with a set of balance checks across groups (Tables A1 and A2 in the Appendix). Then, each group was broken down into smaller groups and assigned to an instructor, thus creating different classes of entrepreneurs. To avoid potential biases due to instructors' teaching style, each instructor was in charge of teaching to both one treated and one control classroom.

Entrepreneurs in both groups attended the same number of sessions. All the sessions were highly experiential and the division in small classes ensured that instructors provided feedback to each participant. Both groups of entrepreneurs were exposed to general managerial frameworks (such as the balance scorecard or the business model canvas) and to data gathering techniques (such as interview techniques, surveys and A/B testing). However, the treatment group was taught to apply this content using a scientific approach. Treated entrepreneurs learnt to develop a theory of the problem faced, to develop hypotheses that flow logically from it, and to use the evidence gathering techniques to test those hypotheses and relate the results back to the theory. For instance, both groups were exposed to the Business Model Canvas (BMC), a widely used tool in entrepreneurship, which helps entrepreneurs graphically schematize a company's business model. Entrepreneurs in the control group were exposed to this method and taught to apply it to their business. Instead, treated entrepreneurs were taught to use the BMC as a starting point for their theoretical reasoning. Each component of the BMC was translated in an hypotheses to be tested. In later sessions entrepreneurs were exposed to different testing designs. Entrepreneurs in the control group were generally encouraged to

[1]Both experiments have been pre-registered on the AEA RCT Registry. Codes: AEARCTR-0002205 (Milan); AEARCTR-0006579 (Turin)

apply these techniques to the problems they were facing in their business, whereas treated entrepreneurs were explicitly encouraged to use those techniques to test the hypotheses developed in the previous sessions.

Contamination between treatment groups was prevented by scheduling classes in different days or times of the week, according to the offered training. Moreover, all the communication was separated by treatment group and the research team checked whether applicants to the program knew other applicants, allocating them to the same experimental group.

4.2 Data Collection Process

We asked entrepreneurs to provide data on their decision-making processes and business performance throughout the training program for up to 66 weeks after the beginning of the training programs to a team of research assistants (RAs) via a set of phone interviews. RAs were purposefully trained by the research team and were responsible of conducting monthly telephone interviews with entrepreneurs. Overall, for each entrepreneur we collected the baseline and up to 18 data points.

Each phone interview was based on a standardized semi-structured interview script, including both open and closed-ended questions. Inquired topics included business performance, decision-making practices and any change introduced to the business idea. Each interview was recorded and stored in an encrypted storage, while RAs were also instructed to encode qualitative answers into quantitative information.

Each entrepreneur was interviewed up until the end of the project or up until the time they declared to have terminated the development of their business idea; thus, for firms that exited the market we have information only up to such exit decision. For firms that did not terminate before the end of our observation window, we have information up to 66 weeks after the beginning of the study.

4.3 Data and Estimation technique

We turn now to the description of the data employed in the empirical estimation of the structural model. To allow for a consistent estimation, we collapse our panel dataset into a cross-sectional form, creating distinct variables corresponding to the three mentioned data points.

To measure selection, i.e. entrepreneurs whose projects are still operational at the end of our investigation period, we create a dummy variable that takes value 1 for entrepreneurs that are still operating in the market and 0 for those that instead terminate their project at any point in time. For the former, we measure overall performance (or value) by computing the revenue growth between the first (baseline) and last available data point. We also check the robustness of results by computing the average of the revenue growths between each collected data point. To measure entrepreneurs' perceived value or estimation of future value, we rely on survey and interview data recording two main components. First, we asked entrepreneurs to provide a predicted probability of termination at the baseline, early and late data points on a scale from 0 to 100. Second, we asked entrepreneurs to directly estimate the minimum and maximum potential future value of their business ideas, on a scale from 0 to 100. To compute the estimated value, we take the logged average of the two. Also in this case, we record entrepreneurs' own estimations at three main data points: before the start of the training (baseline), after eight weeks from the first class and in the last available data point. The latter means that, for entrepreneurs that remained active in the market, we have the full set of information. Instead, for entrepreneurs that terminated, we have information up to the data point prior to which they declared having terminated, which we treat as our "last available" data point. Finally, as to capture idiosyncratic factors that could affect both the project value and entrepreneurs' estimations, we employ pre-training data on team size (number of people in the founding team), team average age, weekly hours worked, years of experience with startups and the team-average education levels.

Table 1 includes some descriptive statistics about these variables by treatment group.

Table 1: Descriptive Statistics

	Scientific		Control		Total	
	Mean	SD	Mean	SD	Mean	SD
Revenue Growth (Stay = 1)	1.70	3.486	1.05	2.687	1.33	3.069
Average Revenue Growth (Stay = 1)	0.09	0.19	0.06	0.15	0.07	0.17
Stay (Dummy)	0.54	0.499	0.69	0.465	0.61	0.488
Probability of Termination (Baseline)	0.17	0.196	0.21	0.209	0.19	0,204
Probability of Termination (Week 8)	0.17	0.228	0.16	0.192	0.16	0.210
Probability of Termination (Last)	0.26	0.288	0.29	0.287	0.27	0.288
Estimated Value (Baseline - log)	4.16	0.285	4.12	0.299	4.14	0.292
Estimated Value (Week 8 - log)	4.06	0.403	4.11	0.302	4.09	0.356
Estimated Value (Last - log)	3.99	0.478	3.99	0.400	3.99	0.440
Startup Experience (Years)	1.28	3.082	1.21	2.323	1.24	2.721
Team Size (Baseline)	2.31	1.441	2.23	1.365	2.27	1.401
Education	1.91	0.794	1.99	0.906	1.950	0,852
Age	31.19	8.541	31.10	7.635	31.145	8.084
Hours Worked (Baseline)	12.88	18.812	12.92	19.334	12.89	19.053

Descriptive statistics on both baseline characteristics and outcomes, by treatment group. For balance checks and related tests, please refer to Tables A1 and A2 in the Appendix.

By assuming a cumulative normal distribution, we can estimate the value of z_j by simply calculating the inverse of the latter given the predicted probabilities of termination (p_j). Mathematically, since $p_j = \Phi(\frac{w_j - c - c_0' - (c_{jT} + \theta)T}{\sigma_j}) = \Phi(z_j)$, we can retrieve z_j as:

$$z_j = \Phi^{-1}(p_j) \tag{15}$$

We cannot know the z estimate at the exact time in which the decision has been taken (what we labelled with F). We, therefore, employ a selection model, where we include as our selection variable the estimate z_L, and we rely on a latent estimation for such probability.

If we were to only estimate the first two equations of the structural model described above, this could be done with a standard Heckman model where the exclusion restriction would be satisfied by the inclusion in the selection equation of the estimate z_L from Eq. 15 evaluated in the *late* period. This would allow us to estimate the *learning effect* θ conditional on the decision to stay in the market. However, relying solely on such two equations does not allow us to estimate the *debiasing effect*.

The opportunity to leverage data on the entrepreneurs' estimation of the potential value of their ideas enables us to retrieve all the structural parameters of interest and be able to separate the *debiasing* effect from the *learning effect*. Particularly, we leverage on the first two post-training data points (E and L) and consider such predicted values for two additional equations, that we label with *. Indeed, it is the availability of the own estimations by entrepreneurs that allow to estimate empirically both Eq. 3* and Eq. 4* and ultimately retrieve the two variances σ_E and σ_L that allow us to estimate the variance σ_F from Eq. 8. This additional step is what allows us to identify the *debiasing effect* in the three different data points we are considering. Indeed, by estimating the three variances, we are able to subtract θ from the estimated coefficients on T in Eq. 14 and Eq. 12 and finally compute the debiasing effect for Eq. 13.

We thus end up with the following structural model to be estimated, made up of six equations:

$$v = a + \theta T + \sigma\epsilon \tag{1}$$

$$Pr(Stay) = \Phi(\frac{-\sigma_L z_L - b_F + (c_F - c_L) + (c_{FT} - c_{LT})T}{\sigma_F}) \tag{14}$$

$$z_L = \frac{\sigma_E z_E + b_L - (c_L - c_E) - (c_{LT} - c_{ET})T}{\sigma_L} \tag{13}$$

$$z_E = \frac{\sigma_0 z_0 + b_E - (c_0' - c_0) - c_E - (c_{ET} + \theta)T}{\sigma_E} \tag{12}$$

$$\hat{v_L^*} = c + c_0' + c_L + (c_{LT} + \theta)T + \sigma_L\epsilon \tag{4*}$$

$$\hat{v_E^*} = c + c_0' + c_E + (c_{ET} + \theta)T + \sigma_E\epsilon \tag{3*}$$

We estimate these equations through a multi-equation conditional mixed-process estimator using the `cmp` user-written command in STATA 16 (Roodman, 2011). The fitted algorithm is a modified version of a seemingly unrelated regressions estimator. In other words, it employs a maximum likelihood (ML) estimator with the assumption that the errors from the different, independent, equations are distributed according to a joint normal distribution. The `cmp` estimator allows us to model a simultaneous equation framework where endogenous variables in a multi-staged process appear both on the right and left end sides of six empirical equations representing the structural model described in the previous subsection. We estimate the following set of empirical equations, linked to the structural equations above:

Eq. 1 : $v^* = \alpha_v + \beta_v X + \theta T + D + \epsilon_v$

Eq. 14 : $\Phi(\alpha_F + \gamma_F z_L + \beta_F T + D)$

Eq. 13 : $z_L = \alpha_L + \gamma_L z_E + \beta_L T + +\epsilon_L$

Eq. 12 : $z_E = \alpha_E + \gamma_E z_0 + \beta_E T + \lambda_E X + D + \epsilon_E$

Eq. 4* : $v_L^* = \alpha_{v_L} + \beta_{v_L} T + \lambda_{v_L} X + D + \epsilon_{v_L}$

Eq. 3* : $v_E^* = \alpha_{v_E} + \beta_{v_E} T + \lambda_{v_E} X + D + \epsilon_{v_E}$

Where D is a set of dummies for RCT and class instructors, X is a set of controls recorded at the baseline period as described above and the α represent constant terms of each equation. All the equations are linearly estimated, but the selection one (Eq. 14) which follows a probit model. Again, equations are estimated simultaneously assuming a joint normal distribution of the error terms. Standard errors are clustered by classroom.

From the estimated coefficients of the above regressions, we can thus retrieve all the parameters of interest that belong to our theoretical structural model. Specifically, the *learning effect* is straightforwardly estimated from the first equation, and it is the coefficient θ on the treatment dummy computed from the first model. All the other structural coefficient have instead to be computed leveraging on the estimated variances and coefficients from the econometric models. Particularly, the computation entails a non-linear combination of different estimated parameters. We conduct such computation using the `nlcom` routine on STATA.

Retrieving the OLS variances from Eq. 3* and Eq. 4*, we can estimate the variance of the model related to the decision (selection equation) from Eq. 14 and we compute all the structural coefficient related to the *debiasing effect* at different points in time from the other equations. Recall that in all equations but the value one, the estimated coefficient on the treatment dummy captures both the hypothesized effects. Thanks to the estimation of variances, we can subtract the estimated *learning effect* (θ) from such coefficients and finally retrieve the correct estimation for the *debiasing effect*. Table 2 details the calculations.

Table 2: Structural Parameter Computation

Parameter	Computation	Equations
θ	θ	1
σ_E	*OLS variance*	3*
σ_L	*OLS variance*	4*
σ_F	$-\frac{\sigma_L}{\lambda F}$	14, 4*
c_{ET}	$\beta_{v_E} - \theta$	3*, 1
c_{LT}	$\beta_{v_L} - \theta$	4*, 1
c_{FT}	$\beta_F \sigma_F + c_{LT}$	14, 4*

This computation strategy leverages on the straightforward calculations from Eq, 3* and Eq. 4*. An alternative computation strategy is shown in the Appendix.

Before computing the full-fledged structural estimation, we also estimate a three-step extended selection model where we only leverage on the entrepreneurs' prediction of idea value rather than on both the latter and the predicted probability of termination. First, we account for selection by running a simple Heckman selection model using the predictions at the last available data point to identify the selection equation, always controlling for baseline characteristics. This step is a different way of modelling the first two equations in the full structural model. Differently from that, we directly employ the prediction on the idea value rather than relying on the perceived probability of termination. However, since our theory explains how entrepreneurs' value estimations should be a function of the treatment, such identification might be subject to endogeneity. We thus add a third equation, instrumenting the late predictions with the baseline, exogenous, prediction of idea value. In the final step, to fully disentangle the *debiasing effect* of the treatment, we add a fourth equation introducing entrepreneurs' value estimations in the *early* period, thus setting up a recursive instrumentation structure. To run this stepwise estimation, we also rely on the `cmp` command in STATA.

5 Results

5.1 Extended Selection Model

Table 3 reports the results of the four-equations extended selection model. In the Appendix we also report the results of the first two steps described in Section 4.

Table 3: Extended Selection Model

	Value Equation	Selection Equation	$\hat{v}_L$	$\hat{v}_E$
Intervention	0.764**	-0.287**	0.0383	-0.0599**
	(0.371)	(0.123)	(0.0416)	(0.0269)
$\hat{v}_L$		1.905***		
		(0.304)		
$\hat{v}_E$			1.086***	
			(0.351)	
$\hat{v}_0$				0.259***
				(0.0687)
Startup Experience	0.190***	-0.032	0.004	0.003
	(0.0653)	(0.0244)	(0.0083)	(0.0051)
Team Size (Baseline)	0.261	-0.049	0.008	0.0018
	(0.181)	(0.0392)	(0.0132)	(0.0140)
Education	0.293	0.001	-0.007	-0.037*
	(0.230)	(0.0727)	(0.0228)	(0.0203)
Age	-0.074***	0.014	0.003	0.003
	(0.0221)	(0.00865)	(0.00288)	(0.00242)
Hours Worked (Baseline)	0.009	0.005	-0.002	0.002**
	(0.0128)	(0.0054)	(0.0016)	(0.0001)
Constant	1.692**	-7.603***	-0.576	3.088***
	(0.705)	(1.350)	(1.463)	(0.317)
Correlation	-0.187***			
	(0.0687)			
RCT Dummies	Yes	Yes	Yes	Yes
Mentor Dummies	Yes	Yes	Yes	Yes
N	377			

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

The value equation is estimated through an OLS conditioned on the selection equation, estimated through a *probit* model. The last two equations are estimated through OLS.

Standard errors clustered at the classroom level.

The coefficient on the intervention dummy from the value equation signals that entrepreneurs following a scientific approach have a revenue growth with respect to the baseline 76.4 pp higher than those of entrepreneurs in the control group (the *learning effect*). The negative coefficient in the selection equation on the same intervention dummy instead signals that treated entrepreneurs are more likely to terminate their projects, a result consistent with Camuffo et al., 2020. The significant correlation between these first two equations signals the necessity of controlling for selection when analyzing performance. Particularly, the negative sign of such correlation is in line with our theoretical expectations. Indeed, starting from the assumption that entrepreneurs tend to overestimate the value of their ideas, a negative correlation signals that on average entrepreneurs with higher perceived evaluations (and thus higher likelihood of remaining active on the market) ultimately have a lower realized value.

The last two equations instead describe the *debiasing* effects at the *Late* and *Early* evaluation data points. Particularly, we find that the *debiasing* effect is mostly found at the early stage of evaluation, rather than at later stages. These results are not surprising if we think that in the *Late* period, all entrepreneurs regardless of their treatment status should make better estimations of the value of their ideas. At the baseline, the value of the average idea is the same between the two groups, thus in principle we should see no differences in the average self-evaluation if the treatment had no effect. A positive, even if not significant coefficient, could signal that entrepreneurs in the treated group recognized that their ideas, given the early downward *debiasing* effect, should be better than those in the control group given the intervention (the *learning effect* is taken into account in their predictions). However, these results can only be confirmed with the structural model, that fully disentangles the two effects.

To provide additional evidence in support of this mechanism, we directly look at the estimates made by entrepreneurs on the potential value of their ideas at different points in time. We compare the averages of entrepreneurs estimation across four groups defined by two dimensions: whether the entrepreneur belonged to the treatment versus control group and whether they terminated the project within the observation window. We report these metrics in Figure 1 below.

Figure 1: Entrepreneurs' Evaluations

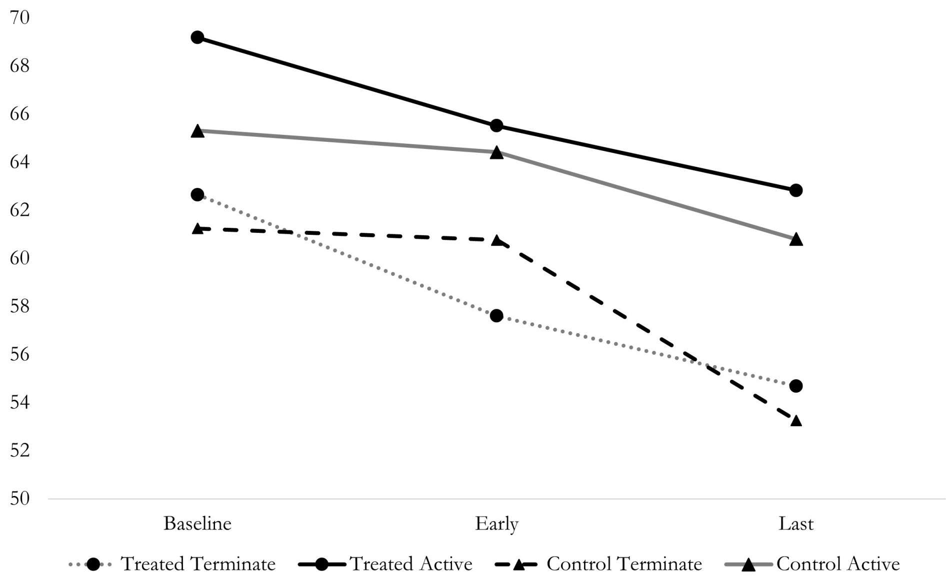


Figure 1 shows how projects that were not terminated show higher estimation value than those that were terminated, in line with the idea that entrepreneurs, on average, terminate the projects that they assess to have a lower value. Second, for all groups, estimates are progressively lower over time. Third, the path of reduction is different for treated and control entrepreneurs. As found in the econometric results, treated entrepreneurs reduced their own estimation already at the very early stages, regardless on whether their final decision was to stay in or exit from the market, while the estimates from control entrepreneurs are constant between the baseline and the early period. There is also a further reduction when looking at the last period available, which changes from firm to firm according to their period of exit. When looking at control entrepreneurs instead the bigger drop happens at a later stage.

We now turn to the full-fledged structural estimation, to clearly disentangle the *debiasing* and *learning* effects.

5.2 Structural Estimation Results

Table 4 reports the results of the structural estimation.

Table 4: Estimated Structural Parameters

	Parameter	Std. Err	z-score
θ	0.80	0.390	2.06
σ_E	0.35	0.032	10.66
σ_L	0.43	0.040	10.83
σ_F	1.75	1.375	1.27
c_{ET}	-0.85	0.387	-2.20
c_{LT}	-0.82	0.374	-2.19
c_{FT}	-1.65	0.501	-3.29

Structural parameters retrieved after ML estimation of the six equations described in Section 4.3. Parameters and their standard errors are retrieved using the `nlcom` routine in Stata 16. All estimated equations include dummies for RCT and instructor, with standard errors clustered at the classroom level. Full estimation results of the six equations are reported in the Appendix.

Our estimation results show a positive and significant θ coefficient, that represents what we called the *learning effect*. On average, "scientific entrepreneurs" experience an increase in revenues of around 80 percentage points compared to traditional ones, conditional on the decision to stay on the market. This result is in line with the one found by the extended selection model described in section 5.1.

The three variance estimates, σ_E, σ_L, σ_F grow in magnitude as the final decision is taken. Particularly, the variance related to the decision equation is around five times the variance experienced in earlier stages (1.75 vs 0.35).

Parameters c_{ET}, c_{LT} and c_{FT} are those identifying the *debiasing effect* in the three decision periods that we consider in our structural framework. Differently from the extended selection model estimation, we leverage on the structural model to clearly separate them from the *learning effect*. Results show that treated entrepreneurs are more likely to perceive a lower potential future value of their business ideas at all the three stages. Particularly, and coherently with the results of the extended selection model, the *debiasing effect* materializes already at an early stage, i.e. eight weeks after the beginning of the training. It seems to be persistent over time, signalling a more conservative approach of treated entrepreneurs when estimating the future value of their ideas.

In the Appendix, we also report the results of the six equations estimated via `cmp` and an alternative computation of structural parameters. To test the robustness of the following results, we also run a number of checks. We first include additional controls in all empirical equations to take into account of some imbalances between groups prior to the training. Then, we employ an alternative measurement for z, where - instead of considering the last available data point - we considered the previous one. Finally, we also conduct the same analyses using the average revenue growth over time rather than the revenue growth with respect to the baseline. Results, reported in the Appendix, are consistent with our main analyses.

Overall, these results suggest that treated entrepreneurs following a scientific approach perform better, on average, compared to traditional entrepreneurs, even when taking into account the effect of selection (the decision to not terminate the project). They also show that they tend to make a downward adjustment to their estimation of their business' ideas values. This downward adjustment on the potential future value of their business idea, is what we believe is the mechanism behind the higher rate of projects' termination by "scientific entrepreneurs" shown in Camuffo et al., 2020. Thus, despite the positive *learning effect* that would allow treated entrepreneurs to perform better on average, this mechanism of reduction in the value of their expectations is what drives more firms towards market exit. The fact that treated entrepreneurs tend to be more cautious can have positive implications, since we can expect that many terminated ideas would not have been successful if still active. This positive selection leads to efficient resource savings, both in terms of time and money.

However, it might also be that following a scientific approach, while ruling out bad ideas and consequently reducing the rate of false positives, leads also to the termination of potentially good projects. The following section discusses this topic in depth.

6 The Trade-Off between Retaining and Discarding Ideas for Scientific Entrepreneurs

In the previous section, we have shown that treated entrepreneurs are more conservative when evaluating their ideas' value and that they reach a more cautious evaluation more quickly than control. This result is in line with the idea that entrepreneurs following a scientific approach reach more quickly a more realistic evaluation of the idea, and more quickly identify "falsely positive ideas".

However, we cannot exclude the possibility that the treatment rather reduces the confidence of entrepreneurs, leading them to discard truly good ideas, increasing - in other words - the number of falsely negative ideas that they terminate. This is a very important question, but one that is, nevertheless, not trivial to answer. Answering this question would require knowing what could have been the "true" realized value of terminated projects, were they not terminated; this is clearly not possible.

To try to mitigate this issue, we provide some suggestive evidence that might at least partially address this concern. First, we gather an external evaluation of entrepreneurs' ideas from a leading firm, leveraging on idea pitches collected at the baseline period. Second, we leverage on additional data collected in the two RCTs. Third, we gather we use the result from the structural estimation to derive additional insights. Fourth, we run a novel experiment using a business simulation game played with MSc students.

6.1 External Evaluation

We partnered with a leading Italian company operating in the innovation and entrepreneurship landscape to obtain an external evaluation on the value of the entrepreneurs' ideas by experienced professionals. We asked the company to evaluate the pitches submitted by entrepreneurs at the baseline, that is, before the start of the training [2]. We assume that the baseline idea is a good proxy of the potential success and future value of the business idea.

The evaluation has been made on three main elements, on a scale from 1 to 100 for each of them: 1) *Profitability*; whether the idea can turn out to be a huge commercial success; 2) *Innovativeness*: whether the idea contains significant innovations; 3) *Feasibility*: whether the idea is realistic and possible to be realized. We then average these three scores to create an *expert evaluation score* ranging from 0 to 100. In the following analyses, we also look at the *profitability* score alone, as it is the one which is more directly comparable to the potential monetary value of the idea.

Again, our empirical results have shown how treated entrepreneurs that are still active in the market, despite being less than those in the control group, perform better. Thus, it seems that the selection induced by applying the *scientific approach* reduces the rate of false positives. To suggest that this beneficial reduction in the false positive rate is not hindered by a stronger increase in the false negative rate, we would like to see that the *expert evaluation* for treated ideas that terminated is not different from those that terminated in the control group. Table 5 reports the averages for both the *expert evaluation* and *profitability* scores, by treatment group and termination decision.

Table 5: Expert Evaluation Scores

	Expert Evaluation		Profitability Evaluation	
	Terminate	Active	Terminate	Active
Control	40.43 (*22.58*)	36.87 (*21.53*)	41.11 (*25.67*)	35.49 (*25.11*)
Scientific	37.21 (*20.73*)	36.99 (*20.41*)	36.71 (*24.67*)	38.74 (*24.77*)

Standard deviations in parentheses. *N = 327; Control = 167; Treated = 160.* There are no significant differences at conventional levels between groups.

The first aspect worth to be analyzed is how the expert evaluations are significantly lower than the self-evaluations made by entrepreneurs on the same 1-100 scale, as displayed in Figure 1. If we trust the judgment of external evaluators, this reinforce our initial idea that the *debiasing effect* induced by the *scientific approach* leads to a more careful

[2]We used 220 pitches for the RCT conducted in Milan, and 107 pitches for the RCT conducted in Turin. The missing pitches were not available due to corrupted data in our storage facilities. We checked whether the firms for which we do not have the pitch were systematically different from the others, finding no significant differences at the baseline on the variables used in the main analyses. Our final sample included 327 pitches, of which 167 in the control group and 160 in the treatment group. Balance checks still hold for this subsample of firms, meaning that the absence of the pitch is likely a random occurrence.

and realistic evaluation, since entrepreneurs tend to overestimate their ideas' potential at the baseline. Coming back to the trade-off between the *false positive* and *false negative* rates, this data shows how treated firms that decided to terminate do not have an expert evaluation that is significantly higher than those in the control group. Instead, from a qualitative point of view, it even seems that ideas in the control group that terminated have the highest scores.

Combined with the evidence coming from the regressions, these results points towards the idea that the selection induced by the *scientific approach* is indeed a "positive" one. In other words, the reduction in the *false positive* rate is not hindered by a significant increase in the *false negative* rate with respect to the control group.

6.2 Additional Evidence

An alternative but objective way to infer the potential value of ideas is to look at at whether firms have received financing from external investors over time. In our data collection efforts, we asked at any data point whether entrepreneurs received external financing (for instance, from venture capitalists or business angels). We thus create a dummy taking value 1 if the firm has received financing within the observation period, and 0 otherwise. In Table 6 we report, for each cell, the share of firms having received external financing, separated by intervention (treatment vs. control) and final decision (termination vs. being active in the market).

Table 6: Share of Firms Having Received External Financing

	Terminate	Active	Difference
Control	2%	10%	-8%
Scientific	1%	19%	-18%
Difference	1%	-9%	10%

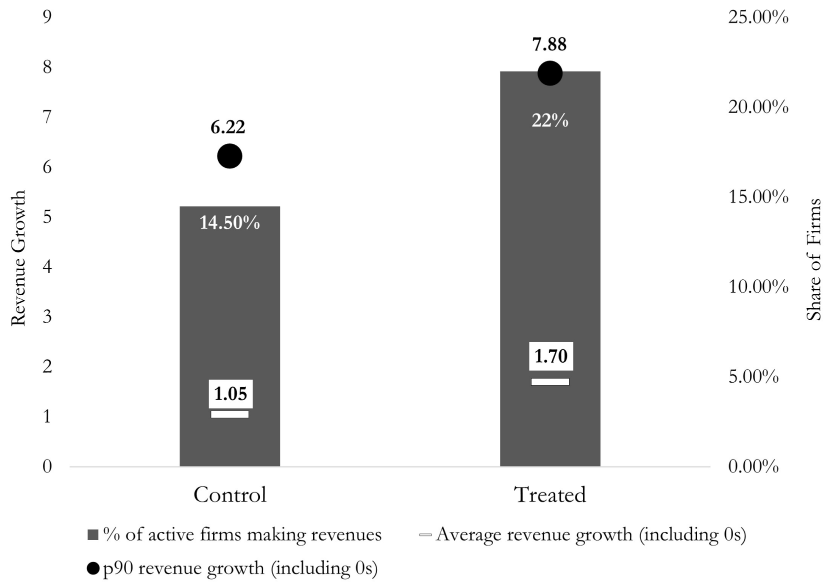
Looking at results for the treatment group, only 1% (1 firm out of 85) of firms that terminated the project collected external finance before their decision to terminate. This share corresponds, instead, to 2% (1 firm out of 59) for those in the control group. This goes in the direction of suggesting that ideas terminated by treated entrepreneurs are not better than those terminated by control entrepreneurs. Conversely, looking at entrepreneurs that decided to stay in the market, we see that 19% (19 firms out of 100) of treated entrepreneurs received external financing, which corresponds to almost double the 10% (13 firms out of 131) recorded in the "control" group[3].

[3] We also run a simple linear probability model, regressing the financing dummy on the interaction between the dummy for staying in the market and the intervention. We add as controls RCT and mentor dummies, clustering the standard errors by classroom. The coefficient on the interaction (i.e. the Difference-in-Differences coefficient) is significant at the 10% level.

These numbers are also consistent with the intuition that projects retained by scientific entrepreneurs tend to be of higher quality, with a selection of false positive projects taking place. Projects that were retained (i.e. stayed in the market) were on average more appreciated by external investors, who we can safely assume were blind with respect to the decision-making approach adopted by entrepreneurs. This is a strong signal towards our theory that the "scientific approach" helps selecting the best projects ex-ante.

Second, we provide some additional evidence on the distribution of revenues across treatment groups and termination decision, since it is also at the backbone of our empirical estimation.

Figure 2: Additional Evidence on Revenue Growth



The columns indicate the share of firms with positive revenue growth, conditional on their decision to stay operational (right axis). The white bar and the black dot indicate, respectively, the 90th percentile and the average of the distribution of revenue growth (including 0s), conditional on the decision to stay operational (left axis).

It is worth noting that in both RCTs all firms started with no revenues, thus explaining the sizeable magnitude of the *learning effect*. Again, we compute the growth in revenues as the difference in logged revenues between the last available data point and the baseline, adding 1 to the latter value as to compute logs for the 0s.
The distribution of revenue growth at the end of the observation window is indeed very skewed, with few firms having positive values. To further explore this phenomenon, we created a dummy variable for firms still operational on the market at the end of the observation period, taking value 1 whether a firm shows a positive revenue growth. Figure 2 summarizes the share of firms making revenues together with two key moments of the revenue growth's distribution.

More precisely, only 14.5% of those in the control group (namely, 19/131) made revenues, versus the 22% (22/100) of those in the scientific group[4]. Looking at the average revenue growth of all operational firms, including those with no revenue growth, Figure 2 shows an higher average revenue growth for scientific entrepreneurs. Whereas the medians for both groups are set to 0, the Figure shows that the value of 90th percentile is higher for scientific entrepreneurs.

This evidence brings further support to the results of our econometric and structural estimations, reinforcing the idea that scientific entrepreneurs make less false positives. Ideas that have been selected by scientific entrepreneurs, despite being fewer, have not only average higher revenues but also a higher chance of reaching the revenue stage

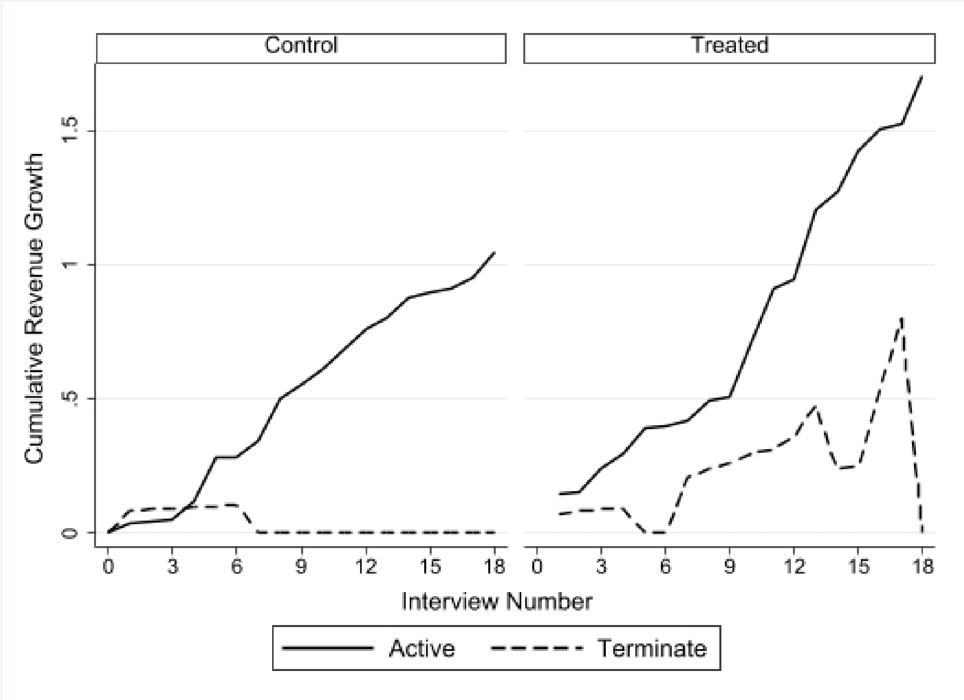
However, since we are also interested in what happens on the false-negative side, we follow the logic from Elfenbein and Knott, 2015 and classify firms into two types (*good*, or *bad*), based on whether they make revenues or not.

First, we leverage on the panel structure of our database and look at data on the revenue growth over time. Specifically, for each firm in the sample, we computed its cumulative revenue growth from the baseline to each observation in our panel. For firms that decide to remain active in the market, we expect a growing trend. For firms that terminate, we expect a more noisy pattern, as their revenue growth naturally goes to zero after their decision to terminate the project (and we conservatively set them as missing values in our database). We then compute the average by treatment group and the final termination decision. Figure 3 shows the results of these computations. The figure shows that, looking at firms that remain active, firms in the scientific group perform better, in line with previous findings.

Instead, what is more interesting is that firms in the scientific group that terminated their projects did made some revenues, although these revenues were lower than the ones of firms that stayed at the very same point in time. This is a first signal that, on average, ideas that were discarded performed less well than those that were not discarded, at least up to their termination decision.

[4]We also run a Probit with a Heckman selection model (`heckprobit`) using as a dependent variable the dummy recording positive revenues. The fitted model mimics the one run in the last two steps of the full structural model, using z_L as the selection variable. Results are in line with the intuition that the probability of making revenues conditional on the decision to stay in the market is significantly higher for scientific entrepreneurs. We also run simpler tests (probit, t-test and chi2 test) on the subsample of operational firms, thus not accounting for selection. While the t-test shows a significant difference in the expected direction (one-tailed, $p = 0.07$), the other two tests do not show significant results.

Figure 3: Panel Data on Revenue Growth



The graph show the pattern of average revenue growth by treatment and final decision of staying or not in the market. For firms that exited, the value of revenue growth is set as missing after their decision to terminate, explaining the noisier pattern.).

When looking at the control group, our model suggests that the share of firms remaining operational in the market is likely to include false positives, but it could also includes projects that the treatment group would have discarded as false negatives. However, the facts that 1) the revenue growth of scientific firms that are still active in the market is higher than that of control firms, and 2) that the revenue growth for scientific firms that terminated is on average always lower than the one of control firms that stayed, suggests that overall the reduction in false positives compensates the potential increase in false negatives experienced by the treatment group. To sum up, this suggestive evidence is in line with the results coming from the analysis of the *expert evaluations*, reinforcing the idea that the scientific approach leads to a reduction in the rate of *false positives*, but to a less than proportional increase in the rate of *false negatives*.

6.3 Structural Estimation: Additional Insights

To further support the insights drawn from descriptive data, we go back to our structural model and focus on the first two equations, estimated with linear regression for the *value equation* (Eq. 1) and with a probit model for the selection equation (Eq. 14). We retrieve the correlation coefficient ρ between the two equations and the Mills' ratios from the selection equation for firms that stayed and terminate their projects.

Using the previously computed variances of the *value equation* (σ) and of the selection equation (σ_F), we can thus compute, for each entrepreneur in our sample, the expected value of the correction in the value equation by treatment condition and by decision to terminate the project or not. Mathematically, for firms that stayed in the market, this corresponds to:

$$correction = \rho \times \frac{\sigma}{\sigma_F} \times \frac{\phi(x\beta)}{\Phi(x\beta)} \tag{17}$$

Instead, for entrepreneurs that terminated, this corresponds to:

$$correction = -\rho \times \frac{\sigma}{\sigma_F} \times \frac{\phi(x\beta)}{(1-\Phi(x\beta))} \tag{18}$$

The intuition behind this analysis is that the correction provides us with a measure of the extent to which the value of ideas needs to be adjusted due to the selection. A positive value of the correction suggests that entrepreneurs using a scientific approach underestimated the value of the project; a negative sign suggests that scientific entrepreneurs overestimate the value of the project. A positive difference in the correction between those who terminate and those who stay suggests that the underestimation of those who terminate is higher than the overestimation of those who stay. We are interested in the difference between control and treated entrepreneurs.

We compute these differences in Table 7, were we make the conservative assumption that the value model for entrepreneurs who remained active and those who terminate their projects is identical for entrepreneurs who terminate and those who stay. We call this the *lower bound* condition.

Table 7: Same Value Model for Terminate and Stay

	Terminate	Stay	Difference
Control	0.41 (*0.09*)	-0.19 (*0.07*)	0.595
Scientific	0.31 (*0.09*)	-0.27 (*0.08*)	0.581
Difference	0.098	0.083	0.015

Standard Deviation in Parentheses

The negative ρ coefficient estimated through the structural model leads to a negative correction for firms that stayed in the market. While it can be challenging to interpret such coefficient in the light of the Heckman selection model, the negative direction signals that entrepreneurs on average tend to overestimate the value of their ideas resulting in a negative correlation when looking at realized performance. Such effect moreover could be mostly driven by the weakest bias reduction provided by the control group, given the results from our structural estimation for the treated entrepreneurs. Importantly, the difference-in-difference calculation leads to a number close to zero and not statistically significant (0.015). This suggests that there is not a significant difference between treated and control entrepreneurs when it comes to the balance between overestimated and underestimated projects.

Our theory and empirics also suggest that "scientific" entrepreneurs perform better on average due to what we called "the *learning effect*". But under the stated assumption that the value model for entrepreneurs that terminated vs. did not terminate their project remains the same, the difference-in-difference estimation does not change.

We next relax the assumption that the value model does not change depending on whether projects were discarded or not and rather assume that the value model is different according to the decision taken. This assumption will lead us to what we call our *upper-bound* condition. Under this assumption, we subtract the *learning effect* $\theta = 0.80$ to value of the correction for the projects of scientific entrepreneurs who terminate their projects, which now becomes -0.49. We subtract the estimated learning effect since the value model we estimated already considers the treatment effect for scientific entrepreneurs. The negative correction signals the existence of a bias reduction also for entrepreneurs that terminated their projects. We report these results in Table 8. The difference-in-difference estimation becomes 0.82, suggesting that the selection results in a lower reduction of value for treated (vs. control) entrepreneurs. Bad ideas are effectively ruled out, without a substantial increase in the *false negative* rate.

Table 8: Different Value for Terminate and Stay

	Terminate	Stay	Difference
Control	0.41 (*0,09*)	-0.19 (*0.07*)	0.595
Scientific	-0.49 (*0.09*)	-0.27 (*0.08*)	-0.221
Difference	0.900	0.083	0.819

Standard Deviation in Parentheses

These cases represent two extremes, one where the selection induced by the "scientific approach" is particularly positive (the *upper-bound*) and one where the approach leads to some adverse selection processes (the *lower-bound*), but close to zero. Despite these results should be interpreted with caution as they are based on assumptions, we believe they provide encouraging suggestive evidence of a well-balanced trade-off between the extent to which scientific entrepreneurs discard bad projects at the expense of good projects: in the worst case scenario (*lower bound*) these two effects essentially compensate each other; in the best case scenario (*upper bound*) the positive effect dominates the negative one.

6.4 Business Simulation Game

Our evidence so far suggests that the strong decrease in the false-positive rate caused by the application of the "scientific approach" more than compensates the increase in the false-negative rate. To further corroborate this interpretation, we run an additional experiment using a business simulation game with Master of Science (MSc) students. Business games are widely used in entrepreneurship and strategy education, and are claimed to provide a high value in the whole education process (e.g. Fox, Pittaway, and Uzuegbunam, 2018). We leverage on the potential of a real-life computer simulation for our research purposes, trying to replicate the results found in a real-life setting.

While it is out of the scope of this paper to describe the mechanics of the game, it is worth briefly explaining how it works. The game simulates the activities of an early-stage startup in its idea validation phase. The player, being in the co-founder role, has to infer the potential value of such startup and ultimately decide whether to launch it on the market or to terminate the project. To understand the idea type with which she is playing, the player can conduct several activities that mimic the real-life experience of an early-stage entrepreneur in the phase of idea validation. For instance, she can brainstorm with the virtual co-founders and create a business model canvas of her idea, or can validate her assumptions by running virtual interviews or questionnaires. The game includes a time dimension, with the market changing conditions over the game days: the player also receives market information in the form of short virtual newspaper articles. Once the player makes her decision, the game ends. The performance are evaluated on a different set of metrics, including market performance and scientific performance (in line with our definition of the scientific approach for decision-making).

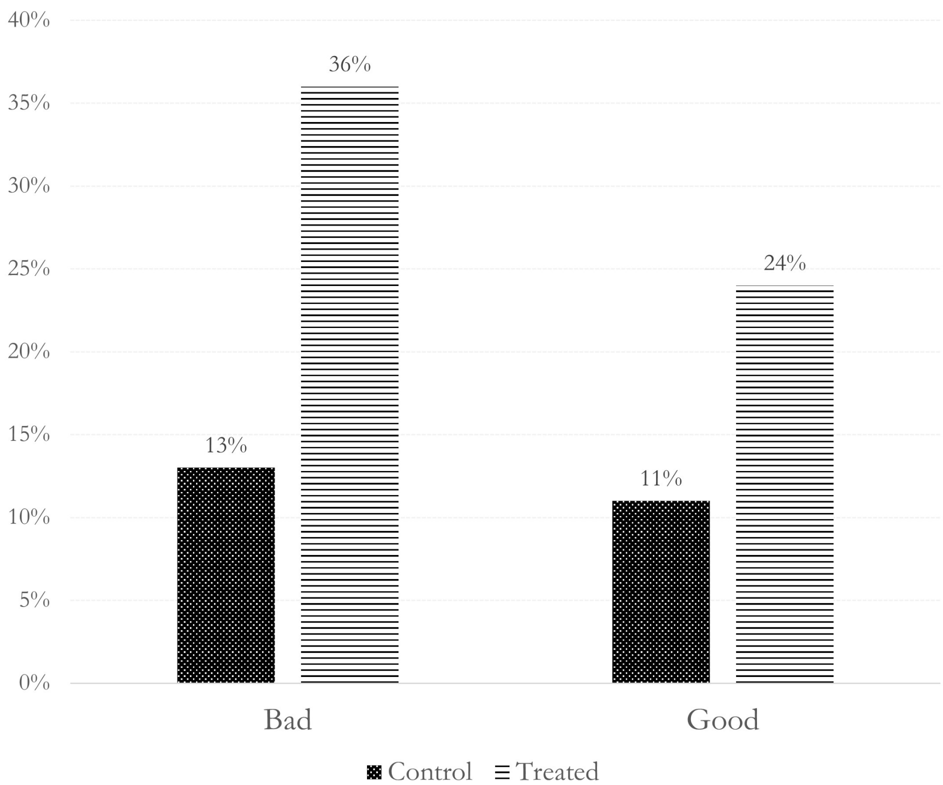
Following again the experiment run in Elfenbein and Knott (2015), we force the simulation game to only have two type of scenarios: a *good* and a *bad* one. In the *good scenario*, the underlying idea to be evaluated by the player would be profitable. In this case, the best decision that the player could make would be to *launch* the project. In the *bad scenario* instead, the underlying idea to be evaluated can never be profitable. In this case, the best decision by the player would be to *terminate* the project. Given the nature of the game, the final profitability of the idea also depend on the specific choices that the player makes during the game. However, what is important for our testing is that such choices cannot change the fundamental value of the idea, that can only be low in the *bad scenario* and high in the *good scenario*.

In the context of a common lecture involving different Master of Science (MSc) programs at our home university, we asked students to play this game on a voluntary basis. Students were all enrolled in their first year of studies and came from three different programs. In one program only, students attended a course on scientific decision-making and thus we consider them as our treated group ($N = 28$). Students in the other two MSc programs did not attend such course and thus become our control group ($N = 50$). For game-related technical reasons, we randomized the in-game scenario (*good* vs. *bad*) before students came to class, stratifying by MSc program. The distribution of conditions turned out to be quite balanced, with 54% and 60% of students in the control and treated group respectively being assigned to the *good scenario*.

Students played for a maximum of 60 minutes, meaning that before the end of the lecturing hour they had to make their final decision. Again participation to the common lecture was voluntary. To incentivize players and also reproduce the monetary incentives of an entrepreneurial activity, we offered three 20€ gift cards (one per each MSc program) to the top players according to their in-game performance. To avoid biases, we introduced the game to students without any explicit reference to the "scientific approach". We mentioned that the main goal was to make the best decision given the players' own evaluations of the in-game idea, not suggesting any path or methodology to follow when running such evaulation.

Our outcome variable is a dummy taking value 1 if the player decides to not launch the idea. We have one main hypotheses based on the evidence from our RCTs: treated students following a scientific approach will terminate more often regardless of the underlying idea. Second, we would like to have further confirmation about the positive trade-off between the rate of false-positive and false-negatives. By comparing the two groups within each scenario, we expect that the treated students terminate projects relatively more often in the *bad scenario* than in the *good scenario*. Figure 4 shows the results by treatment group and in-game scenario.

Figure 4: Share of Terminated Projects



The graph show share of players deciding to terminate the idea development in the business simulation game. *Bad* and *Good* refer to the in-game scenarios, identifying the true potential of the underlying business idea to be evaluated. In the *good scenario* the best decision would be not to terminate, while the converse is true for the *bad scenario*.

Our results suggest how treated players are more likely to decide to terminate their ideas. Despite the small sample at our disposal a two-tailed t-test on the termination dummy between the two groups shows a significant difference (two-tailed, $t = -1.84; M1 = 0.12; M2 = .29$). When looking within scenarios, we also find an interesting pattern. In the *bad scenario*, where the best decision would be to terminate, treated players are three-times more likely to decide to do so. In the *good condition*, where the best decision would be to launch the startup, treated players are only two-times more likely to choose termination as their final decision. Thus, qualitatively, it seems that the reduction in the false-positive rate more than compensates the increase in the false-negative one. Moreover, the better ability of treated players in discriminating between ideas when compared to the control group is also signalled by the constant share of players belonging to the latter that decide to terminate.

Given the small sample size, these differences are not statistically significant[5]. However, putting together all the evidence coming from expert evaluators, the RCTs, the structural modeling and the simulation game provides a signal favorable towards the idea that applying a scientific approach to decision-making in entrepreneurial contexts could lead to positive selection outcomes, being the reduction in false-positives more than compensating the increase in the false-negative rate.

7 Discussion and Conclusion

In this paper we have explored the implications of encouraging entrepreneurs to employ a "scientific-approach to decision making". This approach, based on developing a theory of the problem faced, a set of hypotheses logically flowing from it, a series of tests to validate those hypotheses and a disciplined evaluation of results, is expected to induce entrepreneurs to reason in more probabilistic terms. Our empirical estimations and structural model predict that entrepreneurs following this approach are more likely to terminate their projects, as a result of a *debiasing effect* that leads entrepreneurs to develop a more conservative estimation of the value of their ideas. They also predict that treated entrepreneurs perform better because the "scientific approach" leads them to a better understanding of the problem and the solution space, an effect that we have called *learning effect*.

We estimated our models using data from two randomized control trial that involved 377 startups. The results validate the models and support the intuition that the method leads entrepreneurs to a being more conservative in selecting project, reducing the rate of "false positive", but also to enhance the value of any project they focus on.

[5]Alongside acknowledging the fact that this experiment is underpowered, we also acknowledge the limitation of using a business simulation game as a testing tool. Indeed, a simulation game cannot reproduce the affection mechanisms and emotional dynamics that real-life entrepreneurs might display when it comes to the development of their own business idea.

To better understand the potential value of this finding for scholars and practitioners, we reflect upon the extent to which the conservative attitude of scientific entrepreneurs might actually lead them to increase their rate of "false negative", that is, of good projects that they discard. We provide suggestive evidence coming from expert evaluations on the entrepreneurs' ideas, the RCTs, the structural estimation and a business simulation game. All these pieces of evidence support the idea that this possible effect is more than compensated by the beneficial effect of the reduction in false positives.

Overall, we believe that these findings might inform existing research on innovation and entrepreneurship as well as policy and practice. Educating entrepreneurs to follow a scientific approach to decision-making can indeed lead to a better selection process, effectively discarding projects that ultimately would perform poorly. This can lead to large tangible and intangible resource savings. Moreover, teaching entrepreneurs and students to think in "scientific" terms helps them in devising better strategies and development trajectories, resulting in higher performance. From a scientific point of view, our development of a structural model, estimated with data from two randomized control trials, give us the opportunity to overcome some of the intrinsic limitations faced by studies in the area and enables us to disentangle and identify two separate effects that the use of probabilistic reasoning might have for entrepreneurs.

References

Ashraf, Nava, Abhijit Banerjee, and Vesall Nourani (2021). "Learning to Teach by Learning to Learn". en. In: *Working paper*, p. 115.

Bennett, Victor M. and Aaron K. Chatterji (2019). "The Entrepreneurial Process: Evidence from a Nationally Representative Survey". en. In: *Strategic Management Journal*. ISSN: 1097-0266. DOI: 10.1002/smj.3077.

Bingham, Christopher B. and Kathleen M. Eisenhardt (2011). "Rational Heuristics: The 'Simple Rules' That Strategists Learn from Process Experience". en. In: *Strategic Management Journal* 32.13, pp. 1437–1464. ISSN: 1097-0266. DOI: 10.1002/smj.965.

Bingham, Christopher B. and Jerayr Haleblian (2012). "How Firms Learn Heuristics : Uncovering Missing Components of Organizational Learning". eng. In: *Strategic entrepreneurship journal : SEJ*. Strategic Entrepreneurship Journal : SEJ. - Chichester : Wiley & Sons, ISSN 1932-4391, ZDB-ID 2393106-1. - Vol. 6.2012, 2, p. 152-177 6.2.

Bloom, Nicholas and John Van Reenen (Nov. 2007). "Measuring and Explaining Management Practices Across Firms and Countries*". In: *The Quarterly Journal of Economics* 122.4, pp. 1351–1408. ISSN: 0033-5533. DOI: 10.1162/qjec.2007.122.4.1351.

Camuffo, Arnaldo et al. (Aug. 2020). "A Scientific Approach to Entrepreneurial Decision Making: Evidence from a Randomized Control Trial". In: *Management Science* 66.2, pp. 564–586. ISSN: 0025-1909. DOI: 10.1287/mnsc.2018.3249.

Eisenhardt, Kathleen M. and Christopher B. Bingham (Dec. 2017). "Superior Strategy in Entrepreneurial Settings: Thinking, Doing, and the Logic of Opportunity". en. In: *Strategy Science* 2.4, pp. 246–257. ISSN: 2333-2050, 2333-2077. DOI: 10.1287/stsc.2017.0045.

Elfenbein, Daniel W. and Anne Marie Knott (2015). "Time to exit: Rational, behavioral, and organizational delays". In: *Strategic Management Journal* 27952.7, pp. 957–975.

Felin, Teppo, Stuart Kauffman, and Todd Zenger (Feb. 2020). *Microfoundations of Resources: A Theory*. en. SSRN Scholarly Paper ID 3549865. Rochester, NY: Social Science Research Network. DOI: 10.2139/ssrn.3549865.

Felin, Teppo et al. (Aug. 2020). "Lean Startup and the Business Model: Experimentation Revisited". en. In: *Long Range Planning* 53.4, p. 101889. ISSN: 0024-6301. DOI: 10.1016/j.lrp.2019.06.002.

Folta, Timothy B. (1998). "Governance and Uncertainty: The Trade-off between Administrative Control and Commitment". In: *Strategic Management Journal* 19.11, pp. 1007–1028. ISSN: 0143-2095.

Fox, Joe, Luke Pittaway, and Ikenna Uzuegbunam (2018). "Simulations in Entrepreneurship Education: Serious Games and Learning Through Play". In: *Entrepreneurship Education and Pedagogy* 1.1.

Gans, Joshua S., Scott Stern, and Jane Wu (2019). "Foundations of Entrepreneurial Strategy". en. In: *Strategic Management Journal* 40.5, pp. 736–756. ISSN: 1097-0266. DOI: 10.1002/smj.3010.

Kirtley, Jacqueline and Siobhan O'Mahony (2020). "What Is a Pivot? Explaining When and How Entrepreneurial Firms Decide to Make Strategic Change and Pivot". en. In: *Strategic Management Journal* n/a.n/a. ISSN: 1097-0266. DOI: 10.1002/smj.3131.

Lieberman, Marvin B., Gwendolyn K. Lee, and Timothy B. Folta (2017). "Entry, Exit, and the Potential for Resource Redeployment". en. In: *Strategic Management Journal* 38.3, pp. 526–544. ISSN: 1097-0266. DOI: 10.1002/smj.2501.

McGrath, Rita Gunther (1997). "A Real Options Logic for Initiating Technology Positioning Investments". In: *The Academy of Management Review* 22.4, pp. 974–996. ISSN: 0363-7425. DOI: 10.2307/259251.

NationalBusinessCapital (Jan. 2020). *2019 Small Business Failure Rate: Startup Statistics by Industry.*

Ries, Eric (Sept. 2011). *The Lean Startup: How Today's Entrepreneurs Use Continuous Innovation to Create Radically Successful Businesses.* en. Crown. ISBN: 978-0-307-88791-7.

Roodman, D (2011). "Fitting fully observed recursive mixed-process models with cmp". In: *Stata Journal* 1, pp. 159–206.

Sarasvathy, Saras D. (2001). "Causation and Effectuation: Toward a Theoretical Shift from Economic Inevitability to Entrepreneurial Contingency". In: *The Academy of Management Review* 26.2, pp. 243–263. ISSN: 0363-7425. DOI: 10.2307/259121.

Thomke, Stefan H. (June 1998). "Managing Experimentation in the Design of New Products". In: *Management Science* 44.6, pp. 743–762. ISSN: 0025-1909. DOI: 10.1287/mnsc.44.6.743.

Tversky, Amos and Daniel Kahneman (Sept. 1974). "Judgment under Uncertainty: Heuristics and Biases". In: *Science* 185.4157, pp. 1124–1131. DOI: 10.1126/science.185.4157.1124.

von Hippel, Eric and Marcie J. Tyre (Jan. 1995). "How Learning by Doing Is Done: Problem Identification in Novel Process Equipment". en. In: *Research Policy* 24.1, pp. 1–12. ISSN: 0048-7333. DOI: 10.1016/0048-7333(93)00747-H.

Yang, Mu-Jeung et al. (2020). "How Do CEOs Make Strategy?" In: *NBER Working Paper* 36.

Appendix

Balance Checks

Table A1: Balance Checks Milan RCT

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 A2: Balance Checks Turin RCT

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	

Extended Selection Model: Full Estimation

We report here the results of the first two steps of the four-equations extended selection model, estimated separately. Table A3 shows the results of the Heckman selection model with entrepreneurs' own predictions of idea value at the Late stage ($\hat{v}_L$) used to identify the selection equation. Table A4 adds the intermediate equation that instruments $\hat{v}_L$ with the pre-training (baseline) evaluations $\hat{v}_0$.

Table A3: Heckman Selection Model

	Value Equation	Selection Equation
Intervention	0.803**	-0.409***
	(0.350)	(0.111)
$\hat{v}_L$		0.573***
		(0.148)
Startup Experience	0.187***	-0.021
	(0.0660)	(0.0242)
Team Size (Baseline)	0.261	-0.033
	(0.182)	(0.0508)
Education	0.291	-0.084
	(0.224)	(0.0861)
Age	-0.075***	0.028***
	(0.0225)	(0.00827)
Hours Worked (Baseline)	0.008	0.008
	(0.0127)	(0.0051)
Constant	1.758**	-2.415***
	(0.748)	(0.752)
Correlation	-0.201***	
	(0.0642)	
RCT Dummies	Yes	Yes
Mentor Dummies	Yes	Yes
N	377	

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

Standard errors clustered at the classroom level

Table A4: Extended Heckman Selection Model

	Value Equation	Selection Equation	$\hat{v}_L$
Intervention	0.755**	-0.287**	-0.0268
	(0.368)	(0.124)	(0.0499)
$\hat{v}_L$		1.904***	
		(0.305)	
$\hat{v}_0$			0.282***
			(0.0692)
Startup Experience	0.190***	-0.0320	0.0078
	(0.0647)	(0.024)	(0.0069)
Team Size (Baseline)	0.259	-0.0489	0.0100
	(0.181)	(0.0392)	(0.0144)
Education	0.290	0.002	-0.047**
	(0.231)	(0.0725)	(0.0225)
Age	-0.073***	0.0141	0.0058**
	(0.0219)	(0.0087)	(0.0026)
Hours Worked (Baseline)	0.009	0.005	0.0004
	(0.0127)	(0.0054)	(0.0015)
Constant	1.666**	-7.598***	2.777***
	(0.706)	(1.352)	(0.323)
Correlation	-0.186***		
	(0.0670)		
RCT Dummies	Yes	Yes	Yes
Mentor Dummies	Yes	Yes	Yes
N		377	

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

Standard errors clustered at the classroom level

Full Estimation Results

We report in Table A5 the results of the full `cmp` estimation, used to retrieve the structural coefficients shown in Table 4 the main text.

Table A5: Structural Model: Full Estimation

Model	Value (Eq 1) *OLS*	Selection (Eq 14) *Probit*	z_L (Eq 13) *OLS*	z_E (Eq 12) *OLS*	v_L^* (Eq 4*) *OLS*	v_E^* (Eq 3*) *OLS*
Intervention	0.802**	-0.473***	-0.162**	0.0534	-0.0169	-0.0507*
	(0.390)	(0.128)	(0.0770)	(0.0862)	(0.0498)	(0.0268)
z_L		-0.245				
		(0.186)				
z_E			1.012***			
			(0.178)			
z_0				0.317***		
				(0.0419)		
Startup Experience	0.165**			0.00313	0.0103	0.007
	(0.0666)			(0.0180)	(0.00715)	(0.00512)
Team Size (Baseline)	0.259			-0.139***	0.019	0.009
	(0.159)			(0.0439)	(0.0149)	(0.0140)
Education	0.259			0.114*	-0.0447*	-0.0326
	(0.221)			(0.0636)	(0.0238)	(0.0208)
Age	-0.068***			-0.024***	0.005	0.002
	(0.0220)			(0.0071)	(0.0031)	(0.0023)
Hours Worked (Baseline)	0.0103			-0.0003	0.0005	0.002***
	(0.0125)			(0.0034)	(0.0014)	(0.0001)
Constant	1.562**	0.304	0.413	-0.334	3.941***	4.155***
	(0.787)	(0.306)	(0.352)	(0.289)	(0.143)	(0.0601)
Correlation	-0.236*					
	(0.128)					
RCT Dummies	Yes	Yes	Yes	Yes	Yes	Yes
Mentor Dummies	Yes	Yes	Yes	Yes	Yes	Yes
Equation $ln(\sigma)$ (OLS only)	1.079***		0.171**	-0.008	-0.846***	-1.061***
	(0.0778)		(0.0710)	(0.0370)	(0.0923)	(0.0938)
N	377					

All equations contain dummies for RCT and instructor, with standard errors clustered at the classroom level (in parentheses). *** p<0.01, ** p<0.05, * p<0.1

Alternative Computation of Structural Coefficients

Table A6: Structural Parameter Computation

Alternative computation from Z equations

Parameter	Computation	Equations
θ	θ	1
σ_E	$OLS\ variance$	3*
σ_L	$OLS\ variance$	4*
σ_F	$-\frac{\sigma_L}{\lambda F}$	14, 4*
c_{ET}	$beta_E \sigma_E - \theta$	3*, 12
c_{LT}	$-\beta_L \sigma_L + c_{ET}$	4*, 13, 12
c_{FT}	$\beta_F \sigma_F + c_{LT}$	14, 4*, 13, 12

Table A7: Estimated Structural Parameters

Alternative estimation from Z equations

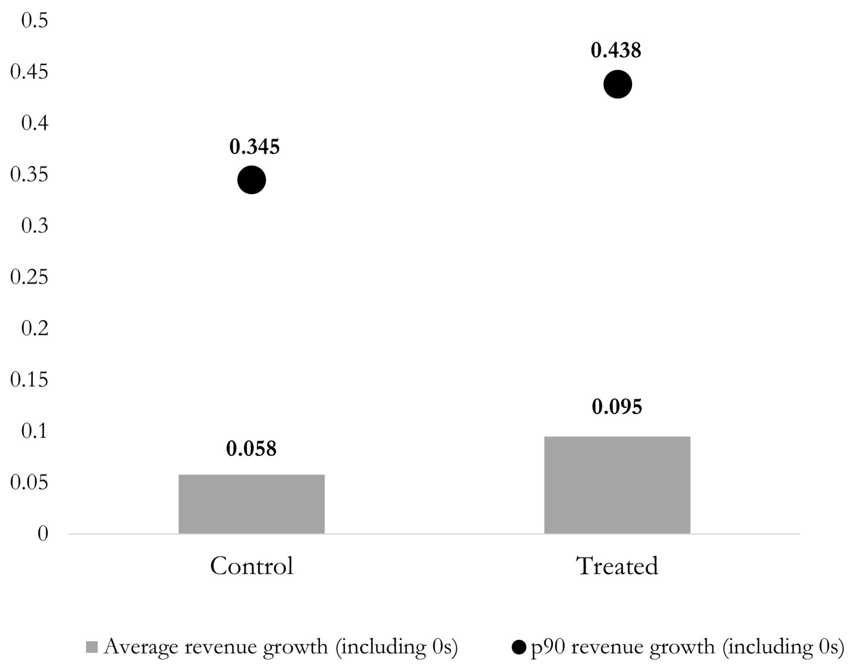
	Parameter	Std. Err	z-score
θ	0.80	0.390	2.06
σ_F	0.35	0.032	10.66
σ_L	0.43	0.040	10.83
σ_F	1.75	1.375	1.27
c_{ET}	-0.82	0.371	-2.21
c_{LT}	-0.75	0.381	-1.97
c_{FT}	-1.58	0.502	-3.15

Structural parameters retrieved after ML estimation of the six equations described in Section 4.3. This alternative computation retrieves the parameters c_{ET} and c_{LT} from Eq. 12 and 13 rather than from Eq. 3* and Eq. 4*. Parameters and their standard errors are retrieved using the `nlcom` routine in Stata 16. All estimated equations include dummies for RCT and instructor, with standard errors clustered at the classroom level.

Robustness Checks: Average Revenue Growth results

In this subsection we report the results for the extended selection model and structural models when using the average revenue growth over time as dependent variable. First, Figure 5 shows the statistics pertaining to average revenue growth. The latter has been computed as the average of the growth of revenues between each data point in the sample, only for firms remaining active for the whole observation window.

Figure 5: Average Revenue Growth Over Time



The graph show the mean and 90th percentile values of the average revenue growth over time for firms active in the market, by treatment condition.

Consistently with the figures on revenue growth with respect to the baseline discussed in the main text, treated firms have a higher growth rate when compared to control firms, both on average and at the 90th percentile.

Table A8 reports the results of the four-equations extended selection model, using as dependent variable the average revenue growth over time.

Table A8: Extended Selection Model

	Value Equation	Selection Equation	$\hat{v}_L$	$\hat{v}_E$
Intervention	0.042**	-0.287**	0.0383	-0.0599**
	(0.0206)	(0.123)	(0.0416)	(0.0269)
$\hat{v}_L$		1.905***		
		(0.304)		
$\hat{v}_E$			1.086***	
			(0.351)	
$\hat{v}_0$				0.259***
				(0.0687)
Startup Experience	0.011***	-0.032	0.004	0.003
	(0.0036)	(0.0244)	(0.0083)	(0.0051)
Team Size (Baseline)	0.015	-0.049	0.008	0.0018
	(0.010)	(0.0392)	(0.0132)	(0.0140)
Education	0.016	0.001	-0.007	-0.037*
	(0.0128)	(0.0727)	(0.0228)	(0.0203)
Age	-0.004***	0.014	0.003	0.003
	(0.001)	(0.00865)	(0.00288)	(0.00242)
Hours Worked (Baseline)	0.0004	0.005	-0.002	0.002**
	(0.0007)	(0.0054)	(0.0016)	(0.0001)
Constant	0.094**	-7.603***	-0.576	3.088***
	(0.0392)	(1.350)	(1.463)	(0.317)
Correlation	-0.187***			
	(0.0687)			
RCT Dummies	Yes	Yes	Yes	Yes
Mentor Dummies	Yes	Yes	Yes	Yes
N	377			

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

The value equation is estimated through an OLS conditioned on the selection equation, estimated through a *probit* model. The last two equations are estimated through OLS.

Standard errors clustered at the classroom level.

The *learning* effect is still significant and estimated to lead to an additional average growth of 4 pp for treated entrepreneurs. The effect is quite sizeable given the averages shown in Figure 5.

Table A9 reports the results of the structural estimation using the average revenue growth over time as dependent variable. The *learning effect* is consistent with previous estimations and the same is valid for the different *debiasing effects*, despite the substantial change in the nature of the dependent variable.

Table A9: Estimated Structural Parameters

	Parameter	Std. Err	z-score
θ	0.04	0.022	2.06
σ_E	0.35	0.032	10.66
σ_L	0.43	0.040	10.83
σ_F	1.75	1.375	1.27
c_{ET}	-0.10	0.032	-2.96
c_{LT}	-0.06	0.046	-1.33
c_{FT}	-0.89	0.602	-1.48

Structural parameters retrieved after ML estimation of the six equations described in Section 4.3, but using the average revenue growth over time as dependent variable. Parameters and their standard errors are retrieved using the `nlcom` routine in Stata 16. All estimated equations include dummies for RCT and instructor, with standard errors clustered at the classroom level.

Robustness Checks: Additional Controls

In this robustness check, we include two additional controls to the main estimation that resulted to be statistically different between the two treatment groups at the baseline. The dependent variable for the *value equation* is the revenue growth with respect to the baseline, as in the analyses reported in the main text.

We add the variable *Self-regulation*, which accounts for the team-level discipline in organization and decision-making activities measured through a 11-item Likert scale. This variable is only available for the RCT conducted in Turin, with a statistical difference between treatment groups significant at 10% (see Table A2).

Second, we add the variable *Background: Economics*, which records the percentage of team members with a degree in economics. This variable is only available for the RCT conducted in Milan, with a statistical difference between treatment groups significant at 5% (see Table A1). We set at 0 the value of these two variables for the RCTs where they were not recorded.

Table A10 reports the results of the four-equation extended selection model with these additional controls. Estimates show a consistent *learning effect*, albeit the significance is slightly reduced. The likelihood of being active on the market at the end of the observation window is still lower for treated entreperneurs. Consistently with the main results, treated entrepreneurs tend to estimate lower values when asked to estimate the value of their ideas in the early data point.

Table A11 shows the structural parameters from the full fledged estimation. All the parameters estimated are consistent with what we find in the main analyses, despite a non-significant parameter for the *debiasing effect* in the *Late* data point.

Table A10: Extended Selection Model

	Value Equation	Selection Equation	$\hat{v}_L$	$\hat{v}_E$
Intervention	0.678*	-0.290**	0.0556	-0.0583**
	(0.360)	(0.116)	(0.0443)	(0.0279)
$\hat{v}_L$		1.884***		
		(0.318)		
$\hat{v}_E$			1.121***	
			(0.363)	
$\hat{v}_0$				0.248***
				(0.0701)
Startup Experience	0.185***	-0.0311	0.00429	0.00338
	(0.0660)	(0.0251)	(0.00834)	(0.00507)
Team Size (Baseline)	0.260	-0.0506	0.00394	0.00238
	(0.172)	(0.0394)	(0.0142)	(0.0140)
Education	0.228	-0.00486	-0.00289	-0.0389*
	(0.244)	(0.0783)	(0.0243)	(0.0206)
Age	-0.0669***	0.0142*	0.00167	0.00270
	(0.0220)	(0.00768)	(0.00282)	(0.00237)
Hours Worked (Baseline)	0.0102	0.00481	-0.00198	0.00202**
	(0.0127)	(0.00514)	(0.00163)	(0.000847)
Background: Economics	1.129*	-0.0147	-0.141**	0.0219
(RCT Milan only)	(0.584)	(0.210)	(0.0709)	(0.0347)
Self-regulation	-0.197	0.0908	0.0550	0.0356
(RCT Turin only)	(0.249)	(0.158)	(0.0488)	(0.0276)
Constant	1.209*	-7.496***	-0.649	3.135***
	(0.711)	(1.450)	(1.501)	(0.322)
Correlation	-0.162**			
	(0.0794)			
RCT Dummies	Yes	Yes	Yes	Yes
Mentor Dummies	Yes	Yes	Yes	Yes
N	377			

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

The value equation is estimated through an OLS conditioned on the selection equation, estimated through a *probit* model. The last two equations are estimated through OLS.

Standard errors clustered at the classroom level.

Table A11: Estimated Structural Parameters

	Parameter	Std. Err	z-score
θ	0.72	0.382	1.88
σ_E	0.34	0.033	10.57
σ_L	0.42	0.040	10.72
σ_F	3.42	5.883	0.58
c_{ET}	-0.77	0.380	-2.02
c_{LT}	-0.72	0.368	-1.95
c_{FT}	-2.12	2.063	-1.03

Structural parameters retrieved after ML estimation of the six equations described in Section 4.3, adding additional controls to the equations. Parameters and their standard errors are retrieved using the `nlcom` routine in Stata 16. All estimated equations include dummies for RCT and instructor, with standard errors clustered at the classroom level.