ISSN 0265-8003

A SCIENTIFIC APPROACH TO ENTREPRENEURIAL DECISION MAKING: LARGE SCALE REPLICATION AND EXTENSION

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

Discussion Paper DP15972
First Published 20 December 2021
This Revision 29 September 2023

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

A SCIENTIFIC APPROACH TO ENTREPRENEURIAL DECISION MAKING: LARGE SCALE REPLICATION AND EXTENSION

Abstract

This large-scale replication of Camuffo et al. (2020) - 759 firms in 4 randomized control trials - confirms that a scientific approach to entrepreneurial decisions can be taught and leads to superior results. The paper yields novel contributions. First, the adoption of a scientific approach generates fewer pivots, which is associated with higher performance. Second, it develops a theoretical framework that sheds light on the underlying mechanisms: methodic doubt (greater caution) and efficient search (better information). We show that fewer pivots imply that, in our sample, the former mechanism dominates. Third, results are robust to the use of a measure of the adoption of the scientific approach instrumented by the treatment, and to models that account for the joint determination of variables.

JEL Classification: L21, L26, M13, M21

Keywords: entrepreneurship

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

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

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

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

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

Chiara Spina - chiara.spina@insead.edu
Institut Européen d'Administration des Affaires, Singapore

Acknowledgements
We gratefully acknowledge funding support from the ICRIOS Center at Bocconi University, the Innovation Growth Lab, the Strategy Research Foundation, the Department for Business, Energy and Industrial Strategy of the UK Government, City, University of

London, and the Polytechnic of Turin. This research was approved by Bocconi Research Ethics Committee with protocol references 2016-III/13.696, 77133-2, and by City Ethics Committee with reference ETH1819-0351. The four studies presented in this manuscript are included in the AEA RCT Registry (AEARCTR-0006578, AEARCTR-0002205, AEARCTR-0006579, AEARCTR-0003875.).

INTRODUCTION

How should entrepreneurs make decisions under uncertainty? Entrepreneurship research and practice point to the fact that entrepreneurs do not have solid routines or methods to make these decisions. For instance, a comprehensive large-scale survey by Bennett & Chatterji (2019) shows that entrepreneurs often fail to take even the lowest cost steps to discover the true value of their ideas. This calls for the importance of studying structured methods that can lead entrepreneurs to improve the quality of their strategic decisions.

Against this background, a recent stream of research advances the idea of a scientific approach to decision making (Camuffo et al. 2020; Felin & Zenger, 2017; Zellweger & Zenger, 2021). This approach claims that, once founders have identified a problem that they believe is worth investigating (Nickerson and Zenger, 2004; Baer et al, 2013; Nickerson et al. 2018), they benefit from developing a theory of the problem and from articulating sets of hypotheses that logically flow from the theory. These predictions should then be tested rigorously. This approach is named "scientific" because it resembles the exploratory approach used by scientists in their research. It leverages on multiple streams of strategy research, including those that emphasize the importance of problem framing, discovery and formulation (Nickerson & Zenger, 2004), the role of mental representations in navigating uncertain environments (Csazar & Levinthal, 2016; Eisenhardt & Bingham, 2011; Felin et al., 2019; Ott & Eisenhardt, 2020), as well as the importance of experimentation and testing to mitigate uncertainty (Ghosh et al., 2020; Kohavi & Thomke, 2017; Koning, Hasan & Chatterji, 2022; Murray & Tripsas, 2004; Ott, Eisenhardt and Bingham, 2017). The main thrust of this approach is that it improves decisions because it ensures that they are based on a more accurate understanding of the contingencies that will lead to higher performance.

Despite the promise of this approach, empirical evidence is limited. A recent lively debate in the entrepreneurship literature (Sergeeva, Bhardwaj, & Dimov, 2021; Zellweger & Zenger, 2022) highlights two relevant questions. First, how valid is the scientific approach to entrepreneurial decision making, i.e. whether entrepreneurs using such an approach "*act differently, and whether the outcomes that result are superior*" (Zellweger & Zenger, 2022, p. 698). Second, whether a scientific approach can be effectively taught to entrepreneurs.

STUDY MOTIVATION

To date, only one small scale study has been published in this area (Camuffo et al. 2020). It reported that entrepreneurs taught to use a scientific approach to decision making were more likely to terminate their projects and to pivot radically. But the study's small scale and specific institutional setting outs natural boundaries on the credibility of these results. A large-scale replication in other settings is therefore critical to establish the validity of this approach.

Our research design combines four randomized control trials (RCTs), including Camuffo et al.'s (2020) original one carried out in Milan in 2016. The three additional RCTs apply the same intervention (except for minor changes due to operational constraints) across different contexts and time windows (Milan in 2017, Turin, 2018, London, 2019). Overall, we analyze data for 759 firms over time (11,463 data points). Notable studies implemented similar large-scale RCTs across contexts (Allcott, 2015; Banerjee et al. 2015; Bowers et al. 2017; Davis et al., 2023).

We follow Bettis, Helfat & Shaver's (2016) recommendation that replication studies should move "in stages" altering the original design and context incrementally in order to understand how these changes affect the original detected impact. In line with this guidance, RCT2 was conducted in the same context as the original study (Milan, Italy) but in a different year (2017). RCT3 was conducted in Turin, Italy (2018), in order to capture a potentially different population of "tech" entrepreneurs (Turin is a technological hub). RCT4 was conducted in London, United Kingdom (2019) and extended to a broader population that did not include only early-stage startups.

These contexts present similarities and differences. The similarities allow to appreciate the robustness of the effects of the scientific approach across contexts, highlighting its more fundamental, context-independent characteristics and impact. The differences allow to capture unique effects due to the heterogeneity across contexts. In the former case, larger scale generates clearer evidence about effects non-detectable or negligible in smaller samples. In the latter case, it generates clearer evidence on if and to what extent differences in context matters.

PREVIEW OF RESULTS AND CONTRIBUTION

Our results are generally consistent with Camuffo et al. (2020). However, the larger scale of this study generated new results.

First, differently from the original study we find a positive and precisely estimated effect of the intervention on termination across all the three new RCTs. The limited sample size of the Camuffo et al.'s (2020) study did not allow to detect this effect consistently.

Second, while in Camuffo et al. (2020) the effect of the intervention on the number of radical pivots is positive and statistically significant, we instead show that the intervention makes firms less likely to pivot radically more than once or twice. This again depends on the fact that in the original study's smaller sample only a few firms pivot radically more than once.

Third, we show that treated firms perform better when they pivot once as opposed to many times. Entrepreneurs who do not have a conceptual structure (a theory) that guide their experiments and activities are more likely to pivot without necessarily benefiting from it.

Fourth, compared to Camuffo et al (2020), our larger sample improves the precision of our estimates of the intervention on performance.

Fifth, the larger statistical power of this study enables us to go beyond the intention-to-treat estimates in Camuffo et al. (2020). We run an instrumental variable analysis that estimates the effect of the actual adoption of the scientific approach (measured by an index of scientific intensity) instrumented by the intervention. Albeit this estimation rests on the exclusion restriction that the instrument affects the dependent variables only through the index, it increases the confidence that the effects that we detect reflect treatment compliance (adoption of the scientific approach) rather than a generic effect of the treatment. We also provide evidence of the robustness of our results using models that account for the joint determination of the key variables.

Sixth, we show that the estimates of the effect of the intervention have the same sign and comparable magnitudes across all the RCTs. Of course, their degree of precision varies, given the randomness across contexts and the smaller scale of each RCT. However, the similarity of the estimates across RCTs and the stronger precision of the aggregated results in the combined sample, point to consistent effects of the scientific approach across contexts. We believe that this is an interesting result in and of itself. There is something systematic and invariant across contexts in how this approach affects entrepreneurial decision making.

Building on these novel findings, we extend Camuffo et al. (2020) by providing a new and more precise theoretical interpretation of the mechanisms of the scientific approach. Through abductive reasoning, (King et al, 2021; Pillai et al, 2020; Behfar & Okhuysen, 2018), we conceptualize two plausible mechanisms: *methodic doubt* and *efficient search*. We then discuss why our results are consistent with a framework in which methodic doubt dominates the effect of efficient search. This highlights the importance of replications not only for obtaining more robust *empirical* results, but also because more robust empirical results provide the opportunity to improve *theorizing.*

Overall, this paper offers two contributions.

First, it responds to the call for systematic replication to consolidate the credibility of preliminary findings in entrepreneurship and strategy research (Bettis, Helfat & Shaver, 2016) and increase their generalizability. A large-scale replication enables us to: (1) address issues of sensitivity to scale (Angrist & Pischke, 2010; King, Goldfarb & Simcoe, 2021; Levitt & List, 2009; List, Maniadis & Tufano, 2017); (2) understand variation in treatment effects across different geographical, industrial and institutional contexts (Hsu, Simmons & Wieland, 2017; Milkman, et al. 2021; Patel & Fiet, 2010); and (3) reveal novel insights otherwise undetectable with smaller experiments (Camerer et al., 2016; Crawford et al., 2022; Tsang & Kwan, 1999).

Second, our results contribute to the emerging research program on the scientific approach to entrepreneurial decision making (Camuffo et al. 2020; Zellweger and Zenger, 2021, 2022). The evidence we provide is consistent with the thesis that entrepreneurs can be taught to reason in a scientific way and that this type of training positively affects their performance in terms of termination of non-promising project, selective pivoting and higher performance. These results resonate with extant entrepreneurship research advocating the superiority of decision-making approaches that combine cognition with action -thinking and doing- to learn under uncertainty (McDonald & Eisenhardt, 2020; Ott et al. 2017).

In the next section we provide the theoretical background of the scientific approach to entrepreneurial decision making and recall the key results of Camuffo et al.'s (2020) study. We then describe the research design, data and method used in our empirical investigation and reports the analyses and findings of the experimental studies. We then offer a novel theoretical interpretation of the findings based on a stylized description of the entrepreneurial decision-making process which. We abduct from the findings our two plausible mechanisms through which the adoption of the scientific approach affects entrepreneurial decision-making outcomes: methodic doubt and search efficiency. We show that our empirical results are compatible with a condition in which the effect of methodic doubt dominates that of efficient search. The final section discusses the theoretical and practical implication of our replication study.

THEORETICAL BACKGROUND

A scientific approach to decision making

In order to navigate uncertainty throughout the entrepreneurial process, entrepreneurship research has recently developed a systematic approach that broadly refers to the idea that entrepreneurs should behave like scientists (Camuffo et al., 2020; Zellweger & Zenger, 2021; Agarwal et al., 2023). This approach encourages entrepreneurs to first envision future states through a theory, and then use such theory as the guide to select which experiments to conduct and use the associated evidence to update their beliefs (Felin & Zenger, 2009, 2017; Ehrig & Schmidt, 2022).[1] Theories boil down to envisioning a future state space -- comprising selected variables and causal relationships among them -- to which founders assign a belief. Beliefs on a theory can be stronger or weaker and -- for simplicity – they can be operationalized as subjective probabilities. Belief formation through theorizing is thus the first step for the entrepreneur-as-scientist approach, and it involves engaging in deliberate cognitive efforts to frame a problem (Park & Baer, 2022) and carefully formulate it as a conceptual structure (Felin & Zenger, 2009; Zellweger & Zenger, 2022;

[1] We define theories as systems of ideas or concepts intended to explain, predict or hypothesize the existence of something, which are based on general principles independent of the thing whose existence is explained, predicted or hypothesized. They comprise of two components: causal structures and believes. Theory-as-causal structure captures the number, the type and the relationships among the conceptual elements of the theory. Theory-as-belief captures the "plausibility" of the theory, i.e., its subjective evaluation of to what extent it is likely to envisage possible and valuable future states (Camuffo et al., 2023).

Baer, Dirks & Nickerson, 2013). Theories are then translated into testable hypotheses, which are tested through experiments. These experiments elicit signals that provide evidence about the hypotheses and their assumptions (Camuffo et al., 2020). This information is then incorporated -- following a Bayesian updating process -- into updated beliefs (Zellweger & Zenger, 2021), leading entrepreneurs to make more informed decisions about whether to continue with the current plan, pivot to a different business model, or terminate the venture (Camuffo et al., 2020; Ries, 2011).

Once entrepreneurs have chosen their domain of interest, they frame and formulate the problem (the idea) they want to investigate as a theory, i.e., a conceptual structure of the problem that they are facing (Ehrig & Schmidt, 2022; Felin & Zenger, 2017) – and a belief on it (Zellweger & Zenger, 2021). A theory clarifies the key dimensions of the problem entrepreneurs consider and on which they focus their attention (Camuffo et al., 2020; Felin et al., 2020a). The next step of developing hypotheses that flow logically from the theory helps to modularize the problem into manageable blocks (Leatherbee & Katila, 2020), making the decision process more efficient. Next, testing hypotheses using rigorous methods helps decision makers to elicit less noisy signals thus exploring its components without overcommitting (Kerr, Nanda & Rhodes-Kropf, 2014; Koning, Hasan & Chatterji, 2022).

While rigorous and extensive experimentation has also been the core of the burgeoning, practitioner-oriented, entrepreneurship literature on the lean startup method (Blank, 2013; Contigiani & Levinthal, 2019; Furr & Dyer, 2014; Leatherbee & Katila, 2020; Ries, 2011), the idea of a scientific approach to decision making goes beyond merely testing sequences of hypotheses (Felin et. al. 2020; Blank & Eckhardt, 2023). It emphasizes that the collection of systematic evidence benefits when it is directed by a theory, which operates like a guide and helps focusing the collection of evidence on the most promising regions of the space of possibilities

(Gavetti & Levinthal, 2000; Gavetti & Rivkin, 2007; Levinthal, 2017), acknowledging the holistic and systemic nature of problems, their complexity and the interdependence among their constitutive components. Finally, the disciplined interpretation of evidence leads to an update of beliefs about the theory (Zellweger & Zenger, 2022) which might improve decisions for example avoiding overcommitment to non-promising paths (Keil & Mahring, 2010; Sleesman et al. 2018).

Scientific approach and problem-based view

The scientific approach to entrepreneurial decision making is tightly coupled with the scholarly work on problem framing, discovery and formulation (Nutt, 1998; Baer et. al, 2013; Nickerson, Silverman and Zenger, 2007; Nickerson and Argyres, 2018; Cummings and Nickerson, 2021; Park & Baer, 2022). Problem formulation's antecedents are the discovery and identification of a problem, i.e. the observation of "a deviation from a desired set of specific -- or a range of acceptable -- conditions resulting in a symptom or a web of symptoms recognized as needing to be addressed" (Baer et al. 2013 p.199). Problem discovery and identification, in turn, derive from attention processes (Shepherd, Mcmullen & Ocasio, 2017), search images (Felin & Koenderink, 2022) and mental representations (Csaszar & Levinthal, 2016), which "frame" the problem contributing to identify which "deviation from desired conditions" is salient for the decision maker.

The theory component of the scientific approach intervenes at the problem formulation stage. The more a problem is novel and ill-structured, the higher the uncertainty, as multiple, competing views of the problem are possible and plausible. These views correspond to alternative, future state spaces. The scientific approach posits that founders should ground problem formulation (i.e. future state space definition) on theories (Karni, 2022), i.e. on conceptual structures and beliefs that are causal, plausible and, hence, potentially valuable.

While problem formulation determines what problem is solved and, hence, conditions its solution, it does not refer to the process of searching for a solution to the problem. The key question for problem formulation is how decision makers can "organize a search to identify and select a problem whose resolution can be expected to generate significant value" (Nickerson, Silverman and Zenger, 2007, p. 213). Instead, the key question for problem solving is how decision makers can "organize an efficient search for high-value solutions to an already formulated problem".

Consider, for example, Elon Musk's acknowledgement back in 2002 that, in order to profitably grow SpaceX in the emerging space travel industry, rockets and spaceships should have been re-usable. This is a matter of problem formulation which is grounded on the "theory" that, if rockets and spaceships are re-usable, the cost of space missions (launch cost plus unit loading cost) will become low enough to make space trips affordable to a large enough market. Problem solving, instead, refers to the generation, evaluation, and selection of alternative solutions (Baer et al. 2013). In Elon Musk's example, such alternatives may be whether to buy rockets and spaceships from Russian manufacturers or design and build them internally.

All in all, the scientific approach posits that entrepreneurs, once they have picked a problem in their domain of interest, should formulate it as a causal conceptual structure and a belief that explains why it is plausible and potentially valuable. This implies identifying which elements of the problem are deemed to be relevant, logically connect them through causal links and assign a subjective probability to this future state space (Camuffo et al., 2023). Once entrepreneurs believe the problem has been correctly formulated, they can solve it choosing among alternative solutions. Experiments (like, for example split tests on different versions of a new product) allow to collect information to test and update the beliefs about which solutions work best.

Mechanisms

The positive effects of the adoption of the scientific approach on entrepreneurial decision making are driven by two mechanisms: *efficient search* (better information) and *methodic doubt* (greater caution).

The former mechanism refers to the fact that decision makers are *more efficient in their search* for solutions to the identified problem. The use of theories improves problem formulation effectiveness (Felin & Zenger, 2017; Felin et al., 2019) because once the key dimensions of the problem are clear, also the characteristics of the potential solutions become less ambiguous. Thus, theory-based problem formulation changes the process through which solutions are sought, from a "searching through" to a "searching for" process (Felin, Kauffman & Zenger, 2020, p. 9). It enables decision-makers to specify, *ex-ante*, what they are looking for, and streamline the costly process of cycling through all options (Rivkin, 2000; Lippman & Rumelt, 2003). Solutions to a well-formulated problem can be more efficiently tested as experiments conducted under this condition deliver less noisy signals about the validity of the hypothesized solution or about variations to the solution that could deliver more value.

The latter mechanism refers to the fact that the scientific approach "instills" in entrepreneurs a *methodic doubt* about the problems they formulate and the solutions they evaluate. As highlighted by Camuffo et al. (2020) and Zellweger & Zenger (2021), the scientific method is the universal and systematic approach through which humans pose and resolve doubt. The scientific approach, due to its investigative nature, fosters critical thinking, contrasting decision makers' tendencies to unawareness (Karni & Vierø, 2017), overconfidence (Astebro et al., 2014) and the neglect of opportunity costs (Bennett & Chatterji, 2023; Chen et al., 2018). Entrepreneurs-as-scientists question and thoroughly articulate why their ideas should work and collect evidence (tests) to validate or disprove such beliefs. By choosing to make the causal structure underlying their reasoning explicit and by gathering information that might conform or go against it, entrepreneurs-as-scientists essentially raise the bar for accepting something as true or valuable and do so only in the presence of clear arguments or consistent evidence (Kerr, Nanda & Rhodes-Kropf, 2014; Koning, Hasan & Chatterji, 2022).

Existing evidence: The Camuffo et al. (2020) study

Despite the logical validity of these conceptual arguments, the empirical evidence about the effects of a scientific approach to entrepreneurial decision-making is limited to the work of Camuffo et al. (2020). The study exposes a small sample of 116 founding teams to a training program that teaches founders to “behave as scientists” when making decisions.

Camuffo et al.’s (2020) main proposition is that entrepreneurs who adopt a scientific approach are more likely to terminate their projects (“exit”, in their terminology) and to pivot radically. Termination is interpreted as a positive outcome, because it prevents entrepreneurs from pursuing projects that are not valuable, saving crucial resources that could be used in better endeavors. Investing in entrepreneurial projects with no potential – a type I decision error or false positive – entails significant opportunity costs at the individual and collective level.

The logic behind why adopting the scientific approach leads to termination is based on the idea that entrepreneurs using a scientific approach are more likely to make a fair assessment of the value of their projects and recognize those with a low probability of success, avoiding biases of overconfidence and judgment variability that often plague their decisions (Astebro, Herz, Nanda & Weber, 2014; Camerer & Lovallo, 1999; Kahneman, Lovallo & Sibony, 2011). In addition to termination, the Camuffo et al.’s (2020) study focuses also on radical pivoting -- defined as strategic changes to the original business model (Kirtley & O’Mahony, 2023). The study shows that entrepreneurs who use a scientific approach radically pivot more because they are more efficient in learning about and re-directing to what portions of space of possibilities is more promising and valuable. Finally, Camuffo et al. (2020) estimates the impact of the scientific approach on performance, showing that it increases revenues.

Despite the uniqueness of the Camuffo et al.'s (2020) study and its value from a research and practice perspective, its small scale limits the generalizability of its findings. Furthermore, the small sample constrained the authors' choice of empirical specification and estimation methods. Moreover Camuffo et al. (2020) does not provide detailed insights about the mechanisms with which a scientific approach to decision making generates higher termination, pivots, and superior performance.

RESEARCH DESIGN, DATA AND MEASURES

Experimental design

We focus on entrepreneurs who have identified a problem and investigate how exposure to a scientific approach helps them to better formulate and address this problem. The dataset we analyze includes the data from the original Camuffo et al.'s (2020) study (henceforth, RCT1) as well as three additional datasets deriving from three randomized control trials that replicate the original study's design (henceforth, RCT2, RCT3 and RCT4). RCT1 was conducted in Milan, Italy (2016). As discussed earlier, we selected the contexts of the additional studies following Bettis, Helfat and Shaver (2016) who posit that "quasi-replications" that test the impact of specific research design to different contexts and populations "hold especially strong promise for the field of strategic management, because (…) inform us about how well results hold up in multiple settings" (p. 2196). In line with this guidance, we conducted RCT2 in the same context of the original study (Milan, Italy) but in a different year (2017). We conducted RCT3 in Turin, Italy (2018), to cater the intervention to a different population of entrepreneurs. While Milan offers entrepreneurs a relevant economic ecosystem, Turin is a technological hub, which also extends the study to a more varied pool of industries. We conducted RCT4 in London, United Kingdom (2019) with the goal of targeting a broader pool of potential applicants not restricted to early-stage startups. Thus, overall, the four RCTs cover a set of varied entrepreneurial types, ranging from entrepreneurs in non-high tech sectors (RCT1 and RCT2), to high-tech entrepreneurs (RCT3), and more established small firms (RCT4).

Since we wanted to replicate RCT1, we organized the three new RCTs as two-arm field experiments in which participant firms were randomly assigned to a "treatment" and a "control" group. Both groups underwent an intervention (a training program), but only firms in the treatment group were trained to adopt the scientific approach. We observed and compared over time the pre/post-intervention difference in firm behaviors, decisions and performance between firms in the treatment and control groups. Taken together, the four RCTs involved 759 firms and 11,463 data points.[2] Sections 1 and 2 in the Appendix report results of the randomization checks and the treatment of attrition.

Recruitment of participants

As in Camuffo et al. (2020), we advertised each program at a national level over multiple online and offline channels, including social media posts, newsletters, magazines, and events. One of the reasons why these programs were appealing to potential applicants is that it was advertised and delivered under the brand of some of the Countries' top business schools. In addition, the instructors were experienced mentors – which reinforced the perception of the value of the program. We conducted the advertisement campaign over several weeks. Regardless of the media used, the campaign promoted the program as a management training program (to avoid self-

[2] All RCTs are pre-registered at the AEA registry and subjected to the relevant IRB/Ethics Committees for approval. Detailed information about each RCT is available upon request.

selection based on interest in a specific topic) offered free of charge to firms operating in any industry. Potential application was not restricted in any way, in order to comply with standard, anti-discrimination ethical guidelines for RCTs. In the application process, participants were required to provide information about their business, team, and decision-making practices via an online survey and a brief telephone interview. We did not admit to the program applicants who failed to complete the survey or the interview.

Sample Composition and participants' characteristics

The previous sections already provided some information about the characteristics of the firms involved in the four analyzed RCTs in terms of industry and stage of development. We provide more details about the participants in Tables 1a and 1b, which provide descriptive statistics and pairwise correlations about meaningful covariates for the samples of the three new RCTs. On average, entrepreneurs involved in these new RCTs hold at least a Bachelor Degree, are in their 30s and 65 percent of them are male. Thirty-six percent have a STEM background (positively correlated with male gender). Their average entrepreneurial experience is 2.46 years. They have 4.51 years of industry experience and 3.66 years of managerial experience. Table A1 of the Appendix provides an industry breakdown of the participants. We used the NACE Rev. 2 - the official statistical classification of economic activities in the European Community - to classify firms into sectors. More than 70% of participants' firms are in the service sector with a strong presence of ICT services (20.82%), professional, scientific and technical activities (15.15%), food services (18.05%) and manufacturing (11.59%). The three new RCTs cover a broader range of sectors than Camuffo et al. (2020). For instance, there are no firms in the art sector in the three RCTs conducted in Italy, while there are 23 firms in this sector in the RCT run in UK. The new RCTs cover also other sectors under-represented in RCT1 - such as agricolture, forestry and fishing, education, human health and social work activities, professional, scientific and technical activities – improving the cross-sector representativeness of the full sample of firms.

INSERT Table 1a and 1b about here

Intervention Details

In each RCT, we assigned firms to either a treatment or a control group through simple randomization. We also broke down the treatment and control groups into smaller classes/learning groups, and randomly assigned each subgroup to an experienced instructor. The baseline survey administered to participants prior to the intervention provided a wide array of observable characteristics which we used to test whether the composition of the treatment and the control groups were balanced. Tables A2 to A6 in the Appendix report the randomization checks for the full sample and each RCT separately. As the four RCTs were conducted asynchronously, the research team had the opportunity over time to introduce additional relevant dimensions to the baseline survey. As a result, the list of observables is larger for later RCTs.

Treatment and control groups attended the same number of training sessions, covering the same topics related to strategy and entrepreneurship. The sessions were highly experiential and smaller groups ensured that instructors provided feedback to each participant.

About 80% of the content in the two classes was the same in terms of topics delivered and teaching materials. Specifically, both treated and control groups were taught frameworks that they could use to support decision making, such as the business model canvas or the balances scorecard; both groups were also exposed to evidence gathering techniques such as qualitative interviews, surveys and A/B testing. Both groups were taught to apply these frameworks and techniques to their specific contexts and were given feedback from their peers and instructor.

The key difference between the two groups was that the treatment group was taught to apply the frameworks and techniques in accordance with a scientific approach to decision making, i.e., by developing a theory of the problem and hypotheses that flew logically from it, by testing those hypotheses through rigorous experiment or equivalent method and by revising their beliefs based on the gathered evidence. The control group, instead, was free to apply these frameworks and techniques in the way they found more appropriate.

An example clarifies the difference between the two groups. One of the first sessions of the training program focused on the Business Model Canvas (henceforth, BMC), a tool widely used in business education that concisely and visually represents a company's business model. It is composed of nine elements that describe a firm's customer value proposition, customer segments, channels, customer relationships, revenue streams, key resources, key partners, key activities, and cost structure.

The control group was exposed to the basic content of the BMC and was taught to use this tool to develop a general overview of their business and discuss its implications with peers. This is the typical way in which BMC is taught in MBAs and Executive programs.

The treatment group, too, was exposed to the basic content of the BMC, asked to apply it to their business and requested to discuss it with their peers. However, differently from the control group, it was explicitly nudged to articulate a theory about why the elements included in the BMC would contribute to value creation, how they were logically connected and why they, as a whole, would generate value for customers. Besides, based on the BMC, treated participants were asked to explicitly formulate testable hypotheses.

For example, one of the entrepreneurs in the program when filling in their BMC, indicated that they were running an electronics retail business through an online distribution channel. If this entrepreneur were part of the control group, she would be invited to generally discuss about the motivation behind this choice and would be given feedback on its suitability. The same entrepreneur in the treatment group, instead, would be explicitly asked to clarify how this element logically connects to the other elements of the BMC, what were the cause-effect relationships, and how they, together, generate value for the customer. This entrepreneur would then be asked to explicitly formulate testable hypotheses derived from the BMC.

In subsequent sessions, participants in both groups were taught techniques to collect data to better inform their decisions. For instance, they were taught about qualitative interviews, surveys, and experiments, the strengths and weaknesses of each of these methodologies and the conditions under which they can be effectively applied. Participants in both groups were then invited to think about which techniques they could use in their businesses and discuss with their peers and instructor their implementation. The control group was instead free to choose how to apply these techniques and was given general feedback about how the technique was applied. The treatment group was explicitly invited to use these techniques to test the hypotheses formulated in the previous sessions and was given specific feedback on whether the proposed design was consistent with the hypotheses they set out to test.

The following example illustrates how treated entrepreneurs articulated the theories underlying their ideas. An entrepreneur participated to the UK RCT and joined the training program with an idea centered around selling consumer electronics in a customer-friendly way. As a result of his participation to the program, he formulated the problem articulating a theory grounded on the belief that *"(…) customers are not technologically savvy and all they care about is solving their specific problem (….)"* and that, consequently, *"(….) in order to make them happy and be on top of competitors, we have to make things very simple, paradoxically "hiding"* *technical information that are redundant to them (…)".* As a further implication, he also believes that "*(….) in order to increase sales we must not advertise products, but solutions. (…)"* This entrepreneur set out to test the hypotheses derived from this theory to *"(…) see whether the theory is going to work (….)".* Among others, he ran an experiment on a specific product line: "(…) *We started advertising some memory cards as solutions for specific devices (…) we found out that these cards started to sell much better than the others which were instead advertised as products (…) right now memory cards represent a third of our business, and we are selling 5,000 of them a month (…)."*

This simple example illustrates how the treatment made participants articulate a theory underlying their idea, i.e. a conceptual structure (attributes and causal links) and beliefs upon which to base development efforts.

We gave all participants genuine and valuable feedback. For example, if a participant proposed to administer a survey to a very small sample of target customers, the instructor would recommend them to increase the sample size irrespective of whether they belonged to the treatment or the control group; or if a participant formulated a survey question in a way that could be improved in terms of clarity, they would be offered suggestions regarding how to improve it, irrespective of whether they were in the treatment or in the control group.

However, participants in the treatment group were nudged to think about information gathering as "scientific" experimentation. For example, they were prompted to clarify what was the counterfactual they were considering, the effects size they were looking for and the belief threshold they were setting to consider the new information as consistent or inconsistent with their hypotheses.

In every RCT, each instructor taught both a treatment and a control group at different times of the day or different days of the week. This choice allowed us to control for cross-instructor differences (e.g. teaching styles) in our analyses (instructor fixed effects) that might affect the absorption of the content taught to participants. Although instructors were not blind to the treatment, we directly supervised the delivery of each session to ensure high teaching standards and that the content was in line with the experimental design described above. We prevented cross-condition contamination ensuring that participants in the experimental and control group did not meet and potentially share key elements of the treatment. For example, we offered the training sessions of the two groups on different days of the week or on the same day of the week, but at different times of the day. To further prevent contamination, the research team kept all communication to the two groups of decision makers attending the program discrete and separate. For the same reason, the research team checked if applicants to the program had any acquaintance with other applicants and allocated them to the same experimental group.

Data Collection

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

Research assistants performed regular phone calls that followed a predefined script that included open and closed-ended questions focusing on changes in the business model, decision-making, and performance outcomes. We targeted an interval of about four weeks between each interview, with some variations due to the entrepreneurs' schedule availability. In all three replication RCTs, we recorded telephone interviews and subjected them to random checks to ensure that research assistants were conducting calls in accordance with the guidelines provided by the research team.

The main variables used in this study refer to outcomes such as project termination, radical pivot and amount of revenue and were therefore collected through closed-ended questions. Following an approach similar to that illustrated by Bloom & Van Reenen (2010), we included a number of open-ended questions that elicited — without asking leading questions — what type of approach to decision-making they were using. Specifically, we instructed research assistants to code entrepreneurs' interviews for the occurrence and intensity of themes related to theory, hypotheses, tests and evaluation. We use these data in supplementary analyses and provide more detail about this in Section 3 of the Appendix. The data collection process continued for up to 18 observation points after the beginning of the program, corresponding to up to 16 months. In one of the RCTs (London), we could only collect observations for 9 months after the beginning of the program due to funding constraints. We consider the duration of the data collection process in discussing our results.

Measures

Following Camuffo et al. (2020) we focus on the following dependent variables.

Termination (referred to as "exit" in Camuffo 2020) is a dummy equal to 1 if the firm terminated the project within the observation window; 0 otherwise.

Number of Radical Pivots is the number of times the firm pivoted within the observation period. We calculated this variable from information collected during the interviews. During each interview, we referred to the Business Model Canvas (BMC) taught to decision makers during the training program and asked decision makers to describe any changes made to any of the nine dimensions of the BMC (value proposition, customers, channels, customer relationships, key activities, key partners, key resources, revenue streams, cost structure). We detected the presence of a pivot at time t if the firm reported a major change to their value proposition or customer segment, two key dimensions of their business. Table 2 provides examples of radical pivots.

INSERT Table 2 about here

As per our definition, these are examples of radical pivots, since they entail changes in core components of the business model – either its customer base, value proposition or both. In general, entrepreneurs pivoted in the most literal sense: stand on some of their past knowledge and focus and turn towards new factors that change their overall product or business (Hampel et al., 2020).

Performance is measured as the firm's cumulative revenue in EURO at the date of the last observation available for each trial since the beginning of the training. This represents a different time period for each RCT. However, we include in our analyses a set of RCT dummies that on average control for these differences.

The key independent variable is *Intervention*, a dummy equal to 1 if the firm was in the treatment group; 0 otherwise. We include in our analyses a set of RCT dummies, to control for any difference across RCTs. We also employ instructor dummies.

Methodology

We structure the analyses by outcome of interest (termination, radical pivot, performance). In line with Camuffo et al. (2020) we found our main analyses on linear regressions (OLS) in which we estimate our dependent variables as a function of the intervention and controls. We present the results obtained in the large-scale analysis and those obtained in the individual RCTs. We can then compare results across RCTs and highlight similarities and differences. In all the specifications, we cluster the standard errors at the intervention-instructor-RCT level and include RCT dummies to control for differences across RCTs. Our results do not change if we run all the analysis with standard errors clustered only at intervention-RCT level.

Table 3 reports descriptive statistics and pairwise correlation between variables of the combined large sample (759 firms). Thirty-four percent of the firms in our sample terminate their projects within the observation window. Firms in our sample radically pivot at most nine times during the observation window. Fifty-nine percent of the sample never radically pivots, and six percent of the firms radically pivot more than two times. The average amount of revenue is EUR 15,753, with large variation in the sample since a substantial number of firms has zero revenue within the observation window. The number of firms participating to each RCT varies across RCT contingent on financial constraints and resource availability for each study.

Figures 1a and 1b provide a visual representation of our data. Figure 1a shows that the number of treated firms that terminate the project within the observation window is higher than the number of control firms. It also shows that treated firms are more likely to radically pivot once, while control firms are more likely to radically pivot more times. Finally, Figure 1b shows that, on average, the revenue of treated firms grows faster than that of control firms.

INSERT Table 3 about here

INSERT Figure 1a and 1b about here

RESULTS

Termination

Table 4 shows our estimates of the impact of the intervention on termination, the first dependent variable investigated by Camuffo et al. (2020). Column (1) reports the results of a cross-section linear probability model that shows that the intervention raises the probability of termination by 9.8 percentage points (p=0.001). In Columns (2-5) we report the results by RCT. The effect of the intervention on termination, which is weak and not very precise in the original study (RCT1), is stronger and more precise in the large-scale sample.

Table 5 reports the results of a Cox proportional hazard model in Column (1). We corroborated the proportionality assumption using the Schoenfeld residuals. We find that the hazard rate of termination is higher for treated than control firms. In Column (2) we replicate this analysis using an OLS regression that predicts the week of termination. We find that, on average, treated firms terminate their project about 2.7 weeks earlier than control firms (p=0.009). Overall, treated entrepreneurs are more likely to terminate their projects and they do it earlier.

Radical Pivoting

Table 6 shows our estimates of the impact of the intervention on radical pivoting, the second dependent variable investigated by Camuffo et al. (2020). Column (1) reports the results of a cross-section regression, estimated by OLS where the dependent variable is the number of radical pivots made by the firms within the observation window. Overall, the intervention does not have a precise linear impact on the number of radical pivots (p=0.486). In Columns (2-5) we report the results by

RCT. Interestingly, only RCT1's result revealed a linear effect of the intervention on the number of radical pivots. This result does not replicate in the subsequent RCTs.

Results from a multinomial probit, reported in Table 7, shed light on this finding. Columns (1), (2) and (3) refer, respectively, to the probability that a firm pivots radically once, twice, or more than twice *vis-á-vis* the no-radical pivot baseline. In Figure 2 we show the marginal effects of the intervention calculated at the observed values for the entire sample. The intervention raises the probability of pivoting radically once or twice and lowers the probability of not pivoting radically or pivoting radically more than twice. When we look at the full sample, the intervention does not have a precise impact on the probability of not pivoting radically. It increases the probability of pivoting radically once by 8.3 p.p. (p=0.003), does not have a precise impact on the probability of pivoting radically twice and decreases the probability of pivoting radically more than twice by 3.7 p.p. (p=0.001). Results from the individual RCTs (not reported for brevity, but available upon request) are consistent with these patterns.[3]

Treated firms are more likely to pivot radically (as opposed to not pivoting radically), but pivot radically fewer times relative to control firms. This highlights the importance of large-scale replication. In our larger study, most control firms do not pivot radically, and a few of them pivot radically two or more times. The treated firms, instead, pivot radically once or twice.

Camuffo et al. (2020) instead showed that treated firms pivot radically more often. But this was driven by the fact that, in their small sample, control firms did not pivot radically and treated firms pivoted mostly once. The sample was not large enough to include a sufficient number of control firms that pivoted two or more times to detect the curvilinear effect. Our larger sample shows that a scientific approach does not induce more pivots but more focused pivots. Treated

[3] For RCT 1, the Multinomial Probit model does not converge because only a few firms pivoted more than once.

entrepreneurs, through theories and experiments, pivot more selectively. Compared to the control group, they are stuck with the initial idea (no radical pivot) and do not waste time and resources by changing their ideas too often.

INSERT Tables 6 and 7 about here

INSERT Figure 2 about here

Performance

Although Camuffo et al. (2020) did not predict the effect of the intervention on performance, it explored the question empirically. We follow the original study and present in Table 8 the results of OLS regressions about the impact of the intervention on the cumulative revenue of firms in our sample (in EUR) at the date of the last observation in each trial.

Results show that, on average, treated firms earn EUR 6999.327 more than control firms (p=0.030). The small effect size reflects the fact that many firms in the total sample earn no revenue as they are early-stage start-ups that started their activities with our training program. Within the observation period, some of the firms started earning revenues in the order of dozen thousand EUR, very much in line with the average amount earned by start-ups in their first few months of operation. The increase in revenue between the times of the first and last interview ranges from 0 to EUR 1,489,026, with revenues increasing of EUR 29,568 at the 90th percentile.

INSERT Table 8 about here

When comparing intervention coefficients across the four RCTs, the *suest* (STATA package) test does not reject the null hypothesis that there are no clear differences across the four RCTs. The same result holds for the termination and radical pivoting analysis reported previously in the paper. The only exception is the RCT 1 intervention coefficient in Table 6 which differs from the coefficients in RCT 3 and 4, as implied by the discussion in the previous paragraph.

Instrumenting scientific intensity

Our analyses so far provided estimates of the intention-to-treat effect, which neglects that participants may not comply with the intervention (Gelman, Hill & Vehtari, 2020). In this paper, we take advantage of the large scale of our sample to address this issue. We asked the research assistants who conducted the regular phone interviews with the decision makers to use a predefined coding scheme (based on 16 items) to assess the level of scientific intensity of decision makers (on a scale from 0 to 5). During the phone interviews, research assistants asked open-ended questions whose content was coded to measure whether participants used theories and experiments to form, test and update their beliefs. Interviewees did not know and were unaware of the coding scheme. We measured scientific intensity in each observation points (about once every four week) and in our cross-section regressions we used the average scientific intensity across all observation points. Table A10 in section 5 of the Appendix provides more details about the measure of scientific intensity. Table 9 in the paper reports a higher level of scientific intensity of the treated vs control group, confirming that the intervention was successful. As an important side result of our study, this finding responds to Zellweger and Zenger (2022)'s question about whether entrepreneurs can be taught to use a scientific approach to decision making.

Table 10 presents the results of a cross-section specification estimated using two-stage least squares where the intervention is used as an instrument for scientific intensity. Results on termination, reported in Column (1), show that the increase of one unit in the average scientific intensity increases the probability of terminating by 29.6 p.p. (p=0.003). In Column (2), we report the results of our analysis on radical pivot, which shows that the increase of one unit in the average scientific intensity increases the probability of radically pivoting once (versus 0 or more than once) of 25.2 p.p. (p=0.001). Looking at the effect on performance, results in Column (3) show that an increase of one unit in the average scientific intensity is associated with an increase of EUR 21,047.762 (p=0.039).

INSERT Tables 9 and 10 about here

Additional analyses

Our three dependent variables – termination, radical pivot and revenue – depend on correlated entrepreneurial choices. Estimating the three equations separately might then not account for the correlation across the errors of these equations.

Table A7a of the Appendix presents Seemingly Unrelated Regression (SUR) results, which control for any covariance across the three regressions. The regression coefficients obtained simultaneously are largely consistent with those presented in the previous sections. The intervention increases the probability of termination and of pivoting once. The estimates remain precise with very low p-value thresholds. The effect on performance remains positive and precise, albeit with lower effect sizes. While this puts a cautionary note on the magnitude of the effect, we know that there is heterogeneity in these effects, and thus we can only expect some variation in this estimate. However, the precision of the estimate (p = 0.001) implies that on average this variation is largely within a positive range.

We also run, as robustness check, a copula model for the joint determination of the binary variables termination and pivoting once, using the STATA package *rbicopula*. We report the results in Table A7b of the Appendix. The estimated dependence between error terms of the two regressions is clearly different from zero (Wald test of theta=0: Prob > chi2 = 0.054), and the total average marginal effects of Intervention on the joint probability *Pr*(termination=1, pivoting only once=1) is positive (B=0.056, p= 0.000). Overall, the treatment plays two roles: first, it increases the two separate unconditional probabilities, second it increases the joint probability as well, consistently with our framework.

Our result about pivoting prompted us to dig to a greater extent into the relationships between treatment and pivot. In particular, we studied whether the decision of treated entrepreneurs to pivot once (as opposed to zero or many times) is associated with greater performance. Unfortunately, we cannot conduct a proper causality test because we would need an independent instrument for pivoting, in addition to the treatment, which we do not have. However, we explored the correlations to check whether we find contradicting evidence. Figure A1 in the Appendix shows that in our cross-section of 759 entrepreneurs, treated entrepreneurs achieve their best performance for 0 pivots, and this is the best performance across all groups of treated and control entrepreneurs with by number of pivots. Treated entrepreneurs perform slightly worse when they pivot once, and much worse when they pivot more than once. Control entrepreneurs who pivot 2, 4 or 5 times perform like treated entrepreneurs who pivot once, and worse than these treated entrepreneurs when control entrepreneurs pivot 0, 1, or 3 times. In line with this outcome, Table A8 in the Appendix provides a cross-section specification using two-stage least squares, where the dummy variable for pivoting once is instrumented by the intervention. Column (1) reports results on termination. Firms that pivot only once are clearly more likely to terminate (p=0.019). Pivoting once increases revenues by EUR 83,489.59 (p=0.090) as well, as reported in Column (2).

Overall these results corroborate our conjecture that treated entrepreneurs work with better models. On the one hand, our pivot variable represents radical pivot, and not incremental changes. Thus, a consistent interpretation of our finding about the best performing group, treated entrepreneurs with 0 pivots, is that these individuals understand whether they are working with a good model, and thus do not change, or they make incremental changes to achieve higher performance. When they expect, instead, that the model is not working, they know where to pivot, and in one pivot they achieve their next best result. Control entrepreneurs instead, perform much worse at 0 and 1 pivots. They perform better at 2 or more pivots. Interestingly, in these cases they reach performance similar to treated entrepreneurs who pivot once. When entrepreneurs understand that they have to change their model, the scientific approach enables entrepreneurs to achieve the same performance of other entrepreneurs who take instead more steps to achieve the same result.

THEORETICAL INTERPRETATION BASED ON FINDINGS

Decision making process

In this section we develop a framework that provides a plausible explanation of our empirical results. To streamline this discussion, we use a stylized representation of the entrepreneurial decision-making process in which entrepreneurs evaluate ideas in exploration stages. In each stage they evaluate one idea. At the end of each stage they decide whether to commit to develop the idea, pivot to a new idea, or terminate the process. This representation is consistent with the one illustrated by Gans, Stern, and Wu (2019). Figure 3 provides a graphical representation of this process. When entrepreneurs commit or terminate, the process ends. When they pivot, the process starts again with a new idea, and the entrepreneurs can pick one of the three options again.

INSERT Figure 3 about here

At the end of each stage entrepreneurs have a subjective belief (a probability) about whether the idea will be successful. We assume that if this probability is higher than a threshold they commit to the idea. This is natural in that the entrepreneurial process is riddled with uncertainty, and the fact that most new ventures fail (Nobel, 2011; Fairlie, Miranda & Zolas, 2019) suggests that entrepreneurs know that the likelihood of success of their ideas is not high. Thus, as soon as they see an idea higher than the threshold, they know that it most unlikely that they will be able to do better in the future, and commit to it. If instead the probability is lower than the threshold, they could pivot or terminate.

To understand the process, consider that if entrepreneurs reach a given stage, they must have discarded all the ideas explored in in the previous stages, and they have decided to pivot instead of terminating. Thus, if entrepreneurs reach stage *i*, they have explored $i - 1$ ideas (starting with idea 1), all these previous ideas were below the threshold, the idea they are currently evaluating is the i^{th} idea, and they have pivoted $i - 1$ times.

Figure 4 provides a snapshot of the entrepreneurs' options at the generic stage *i*. We represent the probabilities of this process in a simple way. We denote by p_i the subjective probability of success of the i^{th} idea at the end of stage *i* and by $\tilde{p}$ the subjective probability threshold. At the end of stage *i*, if the subjective probability of success of the i^{th} idea is higher than the threshold, that is $p_i > \tilde{p}$, entrepreneurs commit to this idea, and the process ends. If it is smaller, that is $p_i < \tilde{p}$, they either terminate or pivot to another idea.

INSERT Figure 4 about here

To choose between termination and pivot, entrepreneurs think of new ideas that they could explore in the next stage $i + 1$. Suppose that $\hat{p}_{i+1}$ is the expected probability of success of an idea that they can explore in this new stage. This is the expected probability, at the end of stage *i*, of the subjective probability of success p_{i+1} that they will observe at the end of the new stage. If this expected probability is lower than the threshold, that is $\hat{p}_{i+1} < \tilde{p}$, entrepreneurs terminate because both the probability of success of the current idea and the expected probability of success of the new idea that they can come up with is smaller than the threshold. As shown in Figure 4, this corresponds to the event $p_i < \tilde{p}$ and $\hat{p}_{i+1} < \tilde{p}$, together with the event that all subjective probabilities of the ideas in previous stages are below the threshold – that is, $p_1 < \tilde{p}$, $p_2 < \tilde{p}$, …, $p_{i-1} < \tilde{p}$.

If instead, they expect a higher probability of success for this new $(i + 1)^{\text{th}}$ idea, they pivot to the new stage. In this case, like for termination, $p_i < \tilde{p}$ and all the probabilities in previous stages are smaller than the threshold. However, the expected probability of success of the next-stage idea is now higher than the threshold, that is $\hat{p}_{i+1} > \tilde{p}$.

If entrepreneurs move to the new stage $i + 1$, at the end of this new stage (i.e., after exploring the new idea), they update $\hat{p}_{i+1}$ to p_{i+1}. If the update is still higher than the threshold, they commit to the new idea; if it is smaller, they terminate or pivot following the same logic of stage i at the end of the new stage $i + 1$.[4]

Decision rules for termination, pivoting, and commitment

From this discussion, we derive the decision rules for termination, pivot and commitment at any exploration stage i, which we summarize in Figure 4.

If entrepreneurs have reached the end of stage i, all the probabilities of the $i – 1$ ideas up to the beginning of stage i are smaller than the threshold. Then, at the end of stage i:

- entrepreneurs terminate their project if the probability of success p_i of the i^{th} idea and the expected probability of success of the next stage idea, $\hat{p}_{i+1}$, are both smaller than the threshold, that is $p_i < \tilde{p}$ and $\hat{p}_{i+1} < \tilde{p}$;

[4] Clearly, because the new information acquired during the exploration stage $i + 1$ update the expected probability of success of the $(i + 1)^{\text{th}}$ idea it can very well be that at end of stage i this expected probability is higher than the threshold, that is $\hat{p}_{i+1} > \tilde{p}$, and therefore entrepreneurs pivot. However, at the end of stage $i + 1$ it falls below the threshold, that is $p_{i+1} < \tilde{p}$.

- entrepreneurs pivot if the probability of success p_i of the i^{th} idea is smaller than the threshold and the expected probability of success of the next stage idea, $\hat{p}_{i+1}$, is higher than the threshold, that is $p_i < \tilde{p}$ and $\hat{p}_{i+1} > \tilde{p}$;
- entrepreneurs commit if the probability of success p_i of the i^{th} idea is higher than the threshold, that is $p_i > \tilde{p}$.

The effects of the scientific approach: efficient search vs "methodic doubt"

The scientific approach can affect decision-making in two ways. On the one hand, theories act as maps that help entrepreneurs to select what to explore, the information to collect and how to incorporate them in the decision-making process. This process of belief formation, testing and updating enables entrepreneurs to focus on better opportunities (*efficient search*). In our framework, this corresponds to higher values of p_i, $\hat{p}_{i+1}$ and p_{i+1}. On the other hand, it nudges entrepreneurs to investigate the logic behind their intuitions, articulate the reason why their idea is plausible, question their assumptions and seek evidence that corroborates or disproves each aspect of it. Consequently, the adoption of the scientific approach makes entrepreneurs skeptical. They *"(….) acknowledge that opportunities are hard to uncover as entrepreneurs have legitimate doubts that opportunities they see are in fact valuable (….)* (Zellweger & Zenger, 2021, p.7). In our framework, this corresponds to higher values of $\tilde{p}$. This *methodic doubt* is pervasive and might regard the goodness of their theories, the validity of their tests and how they incorporate new information in their reasoning.

These two effects are countervailing, and it is not clear which one prevails. We approach this issue empirically. First, we conceptually develop the implications that we expect of the two effects separately for *the probabilities of terminating, pivoting, and commitment discussed above.* Then, we compare these expectations with the empirical results from our analyses to abduct which of the two effects dominate in our data (King et al, 2021; Pillai et al, 2020; Behfar & Okhuysen, 2018). As we clarify below, in our data the effect of the methodic doubt dominates the effect of efficient search.

Efficient search

"Scientific" entrepreneurs search more efficiently, because theories and rigorous tests enable them to learn more and to recognize better opportunities. Because they are guided by a theory, they explore the domain of ideas in rank order, beginning with those that they believe have a higher probability of success. Hence, other things being equal, probabilities p_i and $\hat{p}_{i+1}$ are larger than for non-scientific entrepreneurs at any given stage, increasing the probability to commit and reducing the probability to terminate or pivoting.

However, an open issue is the extent to which scientific entrepreneurs learn from ideas that they discard and use them as the basis for conceiving new ideas. We envision two possibilities and Figure 5 helps us to illustrate our logic.

INSERT Figure 5 about here

The first possibility is that scientific entrepreneurs have a superior ability to conceive new ideas building upon ideas that they ruled out in previous exploration stages. This could happen because their better theories and experiments enable them to learn from the information they collect about the idea through the process, even when the tested idea holds less promise than expected. Good theories enable entrepreneurs to understand what to do even when they gather negative evidence: if the outcome of their test is not in line with the idea A, their theory helps them to understand that the outcome might be in line with another idea, B. Thus, not only their test helps them falsify idea A, but it also leads them to acquire signals about the potential of idea B.

Compared to the case of non-scientific entrepreneurs in which falsification of idea A offers no clues about other ideas, in this case the scientific approach generates a relative expansion of the entrepreneur's domain of search because each exploration stage, and the signals from this exploration, engender new feasible ideas. In Figure 5 we represent this possibility with the curve marked as "efficient search with domain expansion".

The second possibility is that scientific entrepreneurs do not have a superior ability to learn from failures. In this case, when an idea reveals lack of promise, the entrepreneur's incentive to look further is reduced. More generally, they will interpret the lack of support for the original idea as a signal that the overall domain that they are investigating is not promising, and thus new ideas that they can generate out of this domain are even less promising. In Figure 5 this case is represented by the curve marked as "efficient search without domain expansion", which declines more sharply with increasing exploration stages.

Figure 5 compares these two cases against the case of non-scientific entrepreneurs who perform a "less efficient search". This curve does not fall across stages as much as in the case of scientific entrepreneurs because non-scientific entrepreneurs do not explore ideas following a clear rank order, from higher to lower potential. Therefore, the probability of pivoting tends to be steadier across stages because the value of ideas tends to be less correlated across stages.

Ex-ante we cannot predict which of the two "efficient search" curves is a better representation of the actual behavior of scientific entrepreneurs, but we can predict the pattern that we should observe in the data under these scenarios. Under the first scenario ("efficient search with domain expansion") we should observe that scientific entrepreneurs pivot more than non-scientific entrepreneurs, because they keep seeing better opportunities they can pivot to. Under the second scenario ("efficient search without domain expansion") we should observe that scientific entrepreneurs are more likely to pivot than non-scientific entrepreneurs in earlier stages and less likely to pivot in later stages because in earlier stages their ideas are more likely to be better than non-scientific entrepreneurs, and vice versa in later stages. Our results, which show that treated entrepreneurs are more likely to pivot a few times than no-pivot or pivoting several times, implies that the behavior of scientific entrepreneurs is consistent with the idea of efficient search without domain expansion.

Methodic doubt

The *methodic doubt* makes entrepreneurs who adopt the scientific approach more cautious about their ideas because these entrepreneurs understand that theory theories are mental representations of reality, and thus could be wrong. We represent this attitude with a higher decision threshold $\tilde{p}$.

Net effect: method doubt dominates efficient search

The effects of efficient search and methodic doubt on the probability of termination go in opposite directions. Efficient search raises the likelihood that entrepreneurs adopting the scientific approach develop better ideas. This makes them more likely to explore ideas with higher probabilities of success, which raises, at each exploration stage, probabilities p_i and $\hat{p}_{i+1}$. Other things being equal, this lowers the overall probability of termination throughout the entrepreneurial process.

Methodic doubt instead raises the threshold $\tilde{p}$. Other things being equal, this raises the probability of termination.

Once again, we cannot predict unambiguously the direction of the effect of the scientific approach on termination. Comparing the logic described in our framework with our empirical findings we can however establish which effect dominates in our data. Our results show that treated entrepreneurs are more likely to terminate, particularly in earlier stages. This observed effect implies that methodic doubt dominates efficient search.

Our results on pivot and termination are consistent with each other. A higher threshold increases the probability of termination and, but to a lesser extent, the probability of pivoting. This probability depends on the difference between two probabilities, which are not affected unambiguously by the higher threshold. The smaller probability of pivoting at later stages is instead explained by the lower ability to envision new ideas from the negative information collected throughout the entrepreneurial process.

We believe that these are interesting results that open avenues for future research. Our framework claims that there are two mechanisms underlying the scientific approach. The first one is questioning and has to do with the fact that entrepreneurs adopting the scientific approach are aware that they deal with unknowns and, hence, they are more demanding about the decision rules that they adopt. The second one is growth-oriented. It has to do with the ability to see general frameworks that generate growing opportunities from the exploration process. As we discussed, this ability is different from the ability to recognize wrong theories. Also, it is cognitively more demanding because it involves the development of solid alternative theories. We find that in our sample the former mechanism dominates. However, this does not imply that there are no samples in which the latter mechanism might be more important. Our framework in this section helps to see these two implications of the scientific approach, and to understand when, under what conditions, and why one effect can dominate the other.

Performance implications

Our framework cannot predict whether our empirical results that scientific entrepreneurs terminate to a greater extent, especially in earlier stages, and pivot only a few times, are associated with higher performance. This is an empirical question. However, Tables 8 and 10 (column 3) and Figure 1b provide reasonable evidence that scientific entrepreneurs exhibit higher performance than their non-scientific counterfactual. Even if we do not find that they expand their domain of search, their more questioning behavior about termination and pivoting still generates better performance than the control group.

Comparison with Camuffo et al. (2020)

Compared with Camuffo et al. (2020), our study provides similar (albeit stronger) evidence about termination. However, it highlights that the scientific approach does not generate more pivots but, rather, implies more selective pivoting. Our multiple RCTs and larger sample reveal that scientific decision makers navigate more efficiently and quickly their search space across stages and have weaker beliefs about the benefits of multiple additional pivots.

Of course, our replication study does not put the final word on which mechanism (i.e., combination of efficient search and methodic doubt) underlies the performance effects of the scientific approach. Multiple patterns may be equifinal in this respect. However, the results of Camuffo et al. (2020) depend on the small scale of the experiment that does not capture firms that pivot more than once.

Our findings showcase an important problem, recently summarized in the applied economics literature as the “voltage effect” (List, 2022). Experimental research based on small samples, not designed to “scale-up” and build external validity, can lead to incorrect inference and an incomplete theoretical understanding of a phenomenon even holding constant the context and research design. If research about the scientific approach to entrepreneurial decision-making – or any other approach – aims at having a widespread impact, it must achieve 'high voltage', i.e., it must replicate at scale. This study, and its results, represent a step in this direction.

Our results *vis-á-vis* Camuffo et al. (2020) are even more important if we relate them to the paucity of systematic replication in entrepreneurship and strategy research, where most studies

– especially experimental ones – are small scale and offer hardly actionable and generalizable preliminary findings. This exacerbates and reiterates the importance of Bettis et al (2016)'s call for replication studies for the advancement of the field.

DISCUSSION AND CONCLUSIONS

This paper provides a large-scale empirical investigation of the implications of a scientific approach to decision making (Camuffo et al. 2020; Zellweger & Zenger, 2021). It responds to Bettis et al (2016)'s call for systematic replication to consolidate the credibility of preliminary findings in strategic management research by replicating the results of Camuffo et al. (2020) – the first and only empirical test of the theory-based approach to decision making – on a large-scale sample.

The paper contributes to existing research on decision making in multiple ways. First, it answers a recent lively debate in research on decision making by showing that decision makers can be taught to make decisions using a scientific approach and that the use of this approach is associated with superior performance (Sergeeva, Bhardwaj, & Dimov, 2021; Zellweger & Zenger, 2022). Across contexts, time, and institutional settings, we find that a relatively short treatment embedded in a training program can lead decision makers to adopt a scientific approach when making decisions.

Second, our large-scale investigation reveals novel aspects of the way in which a scientific approach affects outcomes, showing that decision makers using this approach (1) are more likely to terminate their project and do so earlier; (2) are more likely to pivot, but do so in a more focused way (few times as opposed to not pivoting or pivoting many times); and (3) perform better.

Third, it contributes to research in this area with a framework that provides insights on the mechanisms underlying these results, offering a conceptual basis for future research. Our results are consistent with a scenario in which the dominating effect of a scientific approach is that decision makers are more doubtful and critical than efficient in their search. This still leads to higher performance that, according to our framework, depends mainly on the selection effect of removing false positives (projects that would turn out to be unsuccessful) than on the ability to build on negative feedback from testing ideas to develop better ideas.

In addition to contributing to research on decision-making, this paper contributes to a larger debate on the design of field experiments. It outlines the limitations of conducting interventions that focus on small samples and a limited set of contexts. Compared to Camuffo et al (2020), our larger scale study surface relevant differences. First, Camuffo et al. (2020) do not find a statistically significant effect of the intervention on termination in most of the regressions, while we find a positive and clear effect across all the other three RCTs. The limited size of the sample in Camuffo et al. (2020) did not produce enough terminations to fully detect this effect. Second, Camuffo et al. (2020) showed a positive effect of the intervention on the number of pivots, which again depends on the fact that in the smaller sample only a few firms pivot more than once. In the larger sample of this paper, we find that the intervention makes pivoting more focused because treated firms are less likely to pivot more than one or two times. Third, the larger sample size made the results about performance statistically more robust, enabled us to provide a more precise estimate of the complier average casual effect using an index of the scientific-intensity of the decisions instrumented by the intervention, and allowed to study heterogeneous treatment effects. Finally, we show that in all RCTs the estimates of the intervention have the same sign and comparable magnitudes, while of course the effect precision varies. This implies that, taken individually, the problem with each trial is the precision of the effect, not model specification.

Our study has practical implications. Research in economics and management has generated many theories and models that prescribe concrete managerial actions. While we teach these theories and models in academic programs, entrepreneurs rarely use them to make decisions, and prefer to rely on their intuitions, experience, gut feelings, or their own logic. This is a serious gap that makes academic research in economics and management less relevant than it could be. While academics may not make their best effort to make their research relevant, the lack of "demand for theory" by decision makers is also likely to contribute to this gap. Our prediction is that if we nurture a culture of scientific decision-making in firms, the value and the use of theories from academic research in economics and management will also increase.

This paper has shown that the impact of a scientific approach on termination, pivot and performance is robust across different contexts, despite some variations in the effect size and precision of the estimates. The investigation of the possible factors contributing to such variations would be very important and insightful, although it would require a specifically-designed research project. We believe that future research would derive valuable insights from pursuing this line of investigation.

An important limitation of this study is that it only covers a short time period. It would be interesting to explore the treatment effects in the medium or long term. Also, it would be important to understand to what extent the effect of scientific intensity is long-lived and whether it dissipates more quickly for some people or businesses rather than for others. More in general, we need more research to better understand the implications of a scientific approach to decision-making and how it can, in detail, create opportunities for better strategy, innovation and entrepreneurial decisions, and how different types of firms or individuals can take advantage of these opportunities. Future work might also investigate to what extent and under which conditions other decision-making heuristics might compare with the scientific approach in making search more efficient and instilling methodic doubt. We hope that future research can shed light on these important micro-foundations of business performance.

REFERENCES

Agarwal, R., Bacco, F., Camuffo, A., Coali, A., Gambardella, A., Msangi, H., Sonka, S.T., Temu, A., Waized, B. & Wormald, A. (2023), “Does a Theory-of-Value Add Value? Evidence from A Randomized Control Trial with Tanzanian Entrepreneurs”, working paper available at SSRN: https://ssrn.com/abstract=4412041 or http://dx.doi.org/10.2139/ssrn.4412041.

Allcott H. 2015. Site selection bias in program evaluation. *The quarterly journal of economics* **130** (3): 1117-1165.

Angrist JD, Pischke J. 2010. The Credibility Revolution in Empirical Economics: How Better Research Design is Taking the Con out of Econometrics. *Journal of Economic Perspectives* **24** (2): 3-30.

Åstebro T, Herz H, Nanda R, Weber RA. 2014. Seeking the Roots of Entrepreneurship: Insights from Behavioral Economics. *The Journal of economic perspectives* **28** (3): 49-69.

Baer M, Dirks KT, Neckerson JA. 2013. Microfoundations of strategic problem formulation. *Strategic management journal* **34** (2): 197-214.

Banerjee A, Duflo E, Goldberg N, Karlan D, Osei R, Pariente W, Shapiro J, Thuysbaert B, Udry C. 2015. A multifaceted program causes lasting progress for the very poor: Evidence from six countries. *Science (American Association for the Advancement of Science)* **348** (6236): 1260799.

Behfar K, Okhuysen GA. 2018. Discovery Within Validation Logic: Deliberately Surfacing, Complementing, and Substituting Abductive Reasoning in Hypothetico-Deductive Inquiry. *Organization science (Providence, R.I.)* **29** (2): 323.

Bennett VM, Chatterji AK. 2023. The entrepreneurial process: Evidence from a nationally representative survey. *Strategic management journal* **44** (1): 86-116.

Bettis RA, Helfat CE, Shaver JM. 2016. The necessity, logic, and forms of replication. *Strategic management journal* **37** (11): 2193-2203.

Blank S. 2013. Why the lean start-up changes everything. *Harvard business review* **91** (5): 64.

Blank S, Eckhardt JT. 2023. The Lean Startup as an Actionable Theory of Entrepreneurship. *Journal of management* : 14920632311680.

Bloom N, Van Reenen J. 2010. New Approaches to Surveying Organizations. *American Economic Review* **100** (2): 105-109.

Bowers J, Higgins N, Karlan D, Tulman S, Zinman J. 2017. Challenges to Replication and Iteration in Field Experiments: Evidence from Two Direct Mail Shots. *The American economic review* **107** (5): 462-465.

Camerer CF, Dreber A, Forsell E, Ho T, Huber J, Johannesson M, Kirchler M, Almenberg J, Altmejd A, Chan T, Heikensten E, Holzmeister F, Imai T, Isaksson S, Nave G, Pfeiffer T, Razen M, Wu H. 2016. Evaluating replicability of laboratory experiments in economics. *Science (American Association for the Advancement of Science)* **351** (6280): 1433-1436.

Camerer C, Lovallo D. 1999. Overconfidence and Excess Entry: An Experimental Approach. *American Economic Review* **89** (1): 306-318.

Camuffo A, Cordova A, Gambardella A, Spina C. 2020. A Scientific Approach to Entrepreneurial Decision Making: Evidence from a Randomized Control Trial. *Management science* **66** (2): 564-586.

Camuffo A, Gambardella A, Pignataro A. 2023. Theory-Driven Strategic Management Decisions, CEPR Dicussion Paper 17664-2.

Chen JS, Croson DC, Elfenbein DW, Posen HE. 2018. The Impact of Learning and Overconfidence on Entrepreneurial Entry and Exit. *Organization science (Providence, R.I.)* **29** (6): 989-1009.

Contigiani A, Levinthal DA. 2019. Situating the construct of lean start-up: adjacent conversations and possible future directions. *Industrial and corporate change* **28** (3): 551-564.

Crawford GC, Skorodziyevskiy V, Frid CJ, Nelson TE, Booyavi Z, Hechavarria DM, Li X, Reynolds PD, Teymourian E. 2022. Advancing Entrepreneurship Theory Through Replication: A Case Study on Contemporary Methodological

Challenges, Future Best Practices, and an Entreaty for Communality. *Entrepreneurship theory and practice* **46** (3): 779-799.
Csaszar FA, Levinthal DA. 2016. Mental representation and the discovery of new strategies. *Strategic management journal* **37** (10): 2031-2049.
Cummings T, Nickerson J. 2021. A Protocol Mechanism for Solving the "Right" Strategic Problem. *Strategic Management Review*
Davis AM, Flicker B, Hyndman K, Katok E, Keppler S, Leider S, Long X, Tong JD. 2023. A Replication Study of Operations Management Experiments in Management Science. *Management science.*
Dyer J, Christensen CM, Furr N. 2014. *The innovator's method: bringing the lean start-up into your organization.* Harvard Business Review Press: Boston.
Ehrig T, Schmidt J. 2022. Theory-based learning and experimentation: How strategists can systematically generate knowledge at the edge between the known and the unknown. *Strategic management journal* **43** (7): 1287-1318.
Eisenhardt KM, Bingham CB. 2017. Superior Strategy in Entrepreneurial Settings: Thinking, Doing, and the Logic of Opportunity. *Strategy science* **2** (4): 246-257.
Fairlie RW, Miranda J, Zolas N. 2019. Measuring Job Creation, Growth, and Survival among the Universe of Start-ups in the United States Using a Combined Start-up Panel Data Set.
Felin T, Gambardella A, Stern S, Zenger T. 2020. Lean startup and the business model: Experimentation revisited. *Long Range Planning* **53** (4): 101889.
Felin T, Koenderink J. 2022. A Generative View of Rationality and Growing Awareness. *Frontiers in psychology* **13** : 807261.
Felin T, Zenger TR. 2017. The Theory-Based View: Economic Actors as Theorists. *Strategy science* **2** (4): 258-271.
Felin T, Zenger TR. 2009. Entrepreneurs as theorists: on the origins of collective beliefs and novel strategies. *Strategic entrepreneurship journal* **3** (2): 127-146.
Gans JS, Stern S, Wu J. 2019. Foundations of entrepreneurial strategy. *Strategic management journal* **40** (5): 736-756.
Gavetti G, Levinthal D. 2000. Looking Forward and Looking Backward: Cognitive and Experiential Search. *Administrative science quarterly* **45** (1): 113-137.
Gavetti G, Rivkin JW. 2007. On the Origin of Strategy: Action and Cognition over Time. *Organization science (Providence, R.I.)* **18** (3): 420-439.
Gelman A, Hill J, Vehtari A. 2020. *Regression and Other Stories.* Cambridge University Press:
Ghosh S, Thomke S, Pourkhalkhali H. Aug 2020. The Effects of Hierarchy on Learning and Performance in Business Experimentation. *Academy of Management Annual Meeting Proceedings* **2020** (1): 20500.
Hsu DK, Simmons SA, Wieland AM. 2017. Designing Entrepreneurship Experiments. *Organizational research methods* **20** (3): 379-412.
Kahneman D, Lovallo D, Sibony O. 2011. Before you make that big decision. *Harvard business review* **89** (6): 50-137.
Karni E. 2022. A theory-based decision model. *Journal of economic theory* **201** : 105444.
Karni E, Vierø M. 2017. Awareness of unawareness: A theory of decision making in the face of ignorance. *Journal of economic theory* **168** : 301-328.
Keil M, Mähring M. 2010. Is Your Project Turning into a Black Hole? *California management review* **53** (1): 6-31.
Kerr WR, Nanda R, Rhodes-Kropf M. 2014. Entrepreneurship as Experimentation. *Journal of Economic Perspectives* **28** (3): 25-48.
King A, Goldfarb B, Simcoe T. 2021. Learning from Testimony on Quantitative Research in Management. *The Academy of Management review* **46** (3): 465-488.
Kirtley J, O'Mahony S. 2023. What is a pivot? Explaining when and how entrepreneurial firms decide to make strategic change and pivot. *Strategic management journal* **44** (1): 197-230.
Kohavi R, Thomke S. 2017. The Surprising Power of Online Experiments. *Harvard business review* : 74.
Koning R, Hasan S, Chatterji A. 2022. Experimentation and Start-up Performance: Evidence from A/B Testing. *Management science* **68** (9): 6434-6453.
Leatherbee M, Katila R. 2020. The lean startup method: Early-stage teams and hypothesis-based probing of business ideas. *Strategic entrepreneurship journal* **14** (4): 570-593.
Levinthal DA. 2017. Mendel in the C-Suite: Design and the Evolution of Strategies. *Strategy science* **2** (4): 282-287.
Levitt SD, List JA. 2009. Field experiments in economics: The past, the present, and the future. *European economic review* **53** (1): 1-18.
Lippman SA, Rumelt RP. 2003. A bargaining perspective on resource advantage. *Strategic management journal* **24** (11): 1069-1086.

List, J.A., 2022. *The voltage effect: How to make good ideas great and great ideas scale*. Currency.
List J, Zacharias Maniadis, Tufano F. 2017. To Replicate or Not To Replicate? Exploring Reproducibility in Economics through the Lens of a Model and a Pilot Study. *IDEAS Working Paper Series from RePEc*
McDonald RM, Eisenhardt KM. 2020. Parallel Play: Startups, Nascent Markets, and Effective Business-model Design. *Administrative science quarterly* **65** (2): 483-523.
Milkman KL, Gromet D, Ho H, Kay JS, Lee TW, Pandiloski P, Park Y, Rai A, Bazerman M, Beshears J, Bonacorsi L, Camerer C, Chang E, Chapman G, Cialdini R, Dai H, Eskreis-Winkler L, Fishbach A, Gross JJ, Horn S, Hubbard A, Jones SJ, Karlan D, Kautz T, Kirgios E, Klusowski J, Kristal A, Ladhania R, Loewenstein G, Ludwig J, Mellers B, Mullainathan S, Saccardo S, Spiess J, Suri G, Talloen JH, Taxer J, Trope Y, Ungar L, Volpp KG, Whillans A, Zinman J, Duckworth AL. 2021. Megastudies improve the impact of applied behavioural science. *Nature (London)* **600** (7889): 478-483.
Murray F, Tripsas M. 2004. The Exploratory Processes Of Entrepreneurial Firms: The Role Of Purposeful Experimentation. In *Business Strategy over the Industry Lifecycle,* Anonymous Emerald Group Publishing Limited; 45-75
Nickerson JA, Zenger TR. 2004. A Knowledge-Based Theory of the Firm--The Problem-Solving Perspective. *Organization science (Providence, R.I.)* **15** (6): 617-632.
Nickerson JA, Silverman BS, Zenger TR. 2007. The 'problem' of creating and capturing value. *Strategic organization* **5** (3): 211-225.
Nickerson J, Argyres N. 2018. Strategizing Before Strategic Decision Making. *Strategy science* **3** (4): 592-605.
Nobel C. 2011. Why companies fail--and how their founders can bounce back. *Harvard Business Review*
Nutt PC. 1998. Framing Strategic Decisions. *Organization science (Providence, R.I.)* **9** (2): 195-216.
Ott TE, Eisenhardt KM. 2020. Decision weaving: Forming novel, complex strategy in entrepreneurial settings. *Strategic management journal* **41** (12): 2275-2314.
Ott TE, Eisenhardt KM, Bingham CB. 2017. Strategy Formation in Entrepreneurial Settings: Past Insights and Future Directions. *Strategic entrepreneurship journal* **11** (3): 306-325.
Park CH, Baer M. 2022. Getting to the Root of Things: The Role of Epistemic Motivation and Construal Levels in Strategic Problem Formulation. *Strategy science* **7** (4): 284-299.
Patel PC, Fiet JO. 2010. Enhancing the internal validity of entrepreneurship experiments by assessing treatment effects at multiple levels across multiple trials. *Journal of economic behavior & organization* **76** (1): 127-140.
Pillai SD, Goldfarb B, Kirsch DA. 2020. The origins of firm strategy: Learning by economic experimentation and strategic pivots in the early automobile industry. *Strategic management journal* **41** (3): 369-399.
Ries E. 2011. *The Lean Startup: How Today's Entrepreneurs Use Continuous Innovation to Create Radically Successful Businesses.* Books on Tape:
Rivkin JW. 2000. Imitation of Complex Strategies. *Management Science* **46** (6): 824-844.
Sergeeva A, Bhardwaj A, Dimov D. 2021. In the heat of the game: Analogical abduction in a pragmatist account of entrepreneurial reasoning. *Journal of business venturing* **36** (6)
Shepherd DA, McMullen JS, Ocasio W. 2017. Is That An Opportunity? An Attention Model Of Top Managers' Opportunity Beliefs For Strategic Action. *Strategic management journal* **38** (3): 626-644.
Sleesman DJ, Lennard AC, McNamara G, Conlon DE. 2018. Putting Escalation of Commitment in Context: A Multilevel Review and Analysis. *The Academy of Management annals* **12** (1): 178-207.
Tsang EWK, Kwan K. 1999. Replication and theory development in organizational science: A critical realist perspective. *The Academy of Management review* **24** (4): 759-780.
Zellweger TM, Zenger TR. 2021. Entrepreneurs as scientists: A pragmatist approach to producing value out of uncertainty. *The Academy of Management review*
Zellweger T, Zenger T. 2022. Entrepreneurs as Scientists: A Pragmatist Alternative to the Creation-Discovery Debate. *The Academy of Management review* **47** (4): 696-699.

TABLES

Table 1a: Descriptive statistics at the baseline – Focus RCT 2-3-4

Variable	Description	Obs	Mean	Std. Dev.	Min	Max
Hours: Total Weekly	Weekly hours dedicated to the company (Team Average)	643	18.88	17.02	0	100

Education	Highest educational level attained by team members (5=PhD, 4=MBA, 3=MSc, 2=BA, 1=high school, 0=otherwise; Team Average)	643	2.23	0.89	0	5
Experience: Entrepreneurial	Number of years of entrepreneurial experience (Team Average)	643	2.46	3.97	0	30
Experience: Industry	Number of years of experience in industry (Team Average)	643	4.51	5.95	0	35
Experience: Managerial	Number of years of managerial experience (Team Average)	643	3.66	4.96	0	30
Age	Age (Team Average)	643	33.13	8.70	18	68
Gender (Male)	Proportion of women in the team	643	0.65	0.40	0	1
Background: Economics	Team members with Economics background (%)	643	0.24	0.36	0	1
Background: Other	Team members with no economics background (%)	643	0.22	0.34	0	1
Background: STEM	Team members with a STEM (Science Technology Engineering Mathematics) background (%)	643	0.36	0.41	0	1
Idea Value: Mean	Estimated value of the project (mean, 0 to 100))	643	65.83	17.25	1	100

Note: some values for some of the variables were missing (about 2% of the data). We filled in these missing values with the median value of the variable

Table 1b: Pairwise correlations pre-intervention – Focus RCT 2-3-4

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
(1) Hours: Total Weekly	1.00										
(2) Education	0.23	1.00									
	(0.000)										
(3) Experience: Entrepreneurial	0.30	0.17	1.00								
	(0.000)	(0.000)									
(4) Experience: Industry	0.27	0.20	0.45	1.00							
	(0.000)	(0.000)	(0.000)								
(5) Experience: Managerial	0.30	0.20	0.60	0.53	1.00						
	(0.000)	(0.000)	(0.000)	(0.000)							
(6) Age	0.20	0.28	0.47	0.58	0.66	1.00					
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)						
(7) Gender (Male)	-0.04	-0.14	-0.01	-0.07	-0.07	-0.12	1.00				
	(0.295)	(0.000)	(0.895)	(0.094)	(0.090)	(0.002)					
(8) Background: Economics	-0.08	0.01	-0.14	-0.16	-0.14	-0.16	0.06	1.00			
	(0.041)	(0.838)	(0.000)	(0.000)	(0.000)	(0.000)	(0.103)				
(9) Background: Other	-0.24	-0.21	-0.06	-0.11	-0.11	-0.04	-0.12	-0.30	1.00		
	(0.000)	(0.000)	(0.143)	(0.007)	(0.006)	(0.284)	(0.003)	(0.000)			
(10) Background: STEM	-0.01	0.04	-0.06	0.03	-0.05	0.01	0.25	-0.38	-0.42	1.00	
	(0.696)	(0.259)	(0.121)	(0.422)	(0.224)	(0.845)	(0.000)	(0.000)	(0.000)		
(11) Idea Value: Mean	0.13	-0.02	0.10	-0.01	0.05	-0.01	0.10	0.03	-0.04	-0.02	1.00
	(0.001)	(0.555)	(0.012)	(0.841)	(0.209)	(0.695)	(0.009)	(0.465)	(0.336)	(0.680)	

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

Table 2: Examples of radical pivots

Company	Initial business idea	Process that led to pivoting	Revised business idea	Element(s) changed
A	Website to connect travellers with special mobility needs with B&B and guesthouses that provide food-related experiences	During the program the team discovered that travellers with special mobility needs are not interested in this offering through extensive interviews and surveys and that B&B and guesthouses would welcome the opportunity to attract more customers through a dedicated website	Website where travellers can book accommodations and food-related experiences at B&B and guesthouses.	Target customers (from travellers with special mobility needs to all travellers)
B	E-commerce that sells craft food and delicacies that are hard to find in supermarkets, particularly targeted at university students that move to another city to study and may miss their local food	During the program the team realized that the willingness to spend of university students was not very high, and that shipping fees were preventing them from ordering from the website. After revisiting their vision and interviewing food producers, they decided to switch to tourists and create vending machines in touristy areas.	Vending machines selling craft food and delicacies that are hard to find in supermarkets, with a focus on tourists	Target customers (from university students to tourists), value proposition, key resources.
C	Umbrella-sharing services for individuals, with umbrella renting stations located close to metro stations or in central areas	During the program the team realized that the demand from individuals was quite low. They thought that a similar service could, however, provide value for organizations such as hotels, convention centres, corporates, and universities – that would bring higher demand and offer the service as a benefit to their customers/employees.	Umbrella-sharing services sponsored by hotels, convention centres, corporates and universities.	Target (from individuals to corporates).
D	Investment-consulting firm specializing in advising clients on wine investments (i.e., purchasing bottles of fine wines that are likely to increase in value with age and resell them for a premium after they age).	During the program the team realized that customers wanted a full-range of services including a dedicated cellar managed by TYI so that the start-up would oversee the ageing process.	Investment-consulting firm specializing in advising and managing wine investments for clients.	Value proposition (from advising on investments to advising and managing investments for customers).

Table 3: Descriptive statistics and pairwise correlations

	Variable	Obs	Mean	SD	Min	Max	1	2	3	4	5	6	7	8	9	10	11	12	13
1	Termination	759	0.34	0.48	0	1	1.00												
2	Number of Pivots	759	0.71	1.12	0	9	-0.10	1.00											
3	Not Pivoting	759	0.59	0.49	0	1	0.07	-0.76	1.00										
4	Pivoting once	759	0.24	0.43	0	1	-0.01	0.15	-0.67	1.00									
5	Pivoting twice	759	0.11	0.31	0	1	-0.03	0.41	-0.42	-0.20	1.00								
6	Pivoting more than twice	759	0.06	0.25	0	1	-0.08	0.74	-0.31	-0.15	-0.09	1.00							
7	Performance (Revenues)	759	15753.89	83051.78	0	1489026	-0.03	-0.04	0.03	-0.00	-0.02	-0.04	1.00						
8	Intervention	759	0.50	0.50	0	1	0.12	-0.03	-0.05	0.10	0.01	-0.08	0.04	1.00					
9	Average scientific intensity	759	2.14	1.17	0	5	-0.02	0.27	-0.31	0.15	0.17	0.15	0.08	0.14	1.00				
10	RCT1	759	0.15	0.36	0	1	0.03	-0.13	0.13	-0.05	-0.08	-0.08	-0.05	0.01	-0.21	1.00			
11	RCT2	759	0.33	0.47	0	1	0.11	0.07	-0.12	0.08	0.06	0.02	-0.12	-0.01	0.15	-0.30	1.00		
12	RCT3	759	0.17	0.38	0	1	-0.03	0.25	-0.22	0.08	0.11	0.16	-0.07	-0.01	0.07	-0.20	-0.32	1.00	
13	RCT4	759	0.34	0.48	0	1	-0.12	-0.17	0.19	-0.11	-0.09	-0.09	0.21	0.01	-0.05	-0.31	-0.51	-0.33	1.00

Table 4: Termination OLS Cross-Section

VARIABLES	(1) Termination OLS Cross-Section Full Sample	(2) Termination OLS Cross-Section RCT1	(3) Termination OLS Cross-Section RCT2	(4) Termination OLS Cross-Section RCT3	(5) Termination OLS Cross-Section RCT4
Intervention	0.098	0.035	0.096	0.158	0.097
	(0.001)	(0.647)	(0.044)	(0.084)	(0.035)
Constant	0.284	0.316	0.374	0.730	0.287
	(0.175)	(0.219)	(0.001)	(0.012)	(0.002)
Observations	759	116	250	132	261
R-squared	0.076	0.183	0.034	0.138	0.026
Dummies for Instructors	Yes	Yes	Yes	Yes	Yes
Dummies for RCTs	Yes	-	-	-	-
Clustered Errors	Intervention Instructor RCT	Intervention Instructor	Intervention Instructor	Intervention Instructor	Intervention Instructor

Robust pval in parentheses. Our balance checks have shown that three variables in two RCTs were unbalanced between the treatment and control group despite randomization. To address this issue, we conservatively included these variables as controls in all specifications. Specifically, Model (3) controls for “Background: Economics” and ”Background: STEM”, Model (4) controls for ”Self-regulation”, Model(1) controls for the interaction between each RCT dummies and variable that was un-balanced in that specific RCT. However, results are similar when not controlling for these variables.

Table 5: Termination Time

VARIABLES	(1) Hazard of termination Survival - Full Sample	(2) Week of termination OLS - Full Sample
Intervention	0.365	-2.699
	(0.000)	(0.009)
Constant		43.350
		(0.000)
Observations	759	759
R-squared		0.280
Dummies for instructors and RCTs	Yes	Yes
Clustered Errors	Intervention Instructor RCT	Intervention Instructor RCT

Robust pval in parentheses. Our balance checks have shown that three variables in two RCTs were unbalanced between the treatment and control group despite randomization. To address this issue, we conservatively included these variables as controls in all specifications. Specifically, Model (3) controls for “Background: Economics” and ”Background: STEM”, Model (4) controls for ”Self-regulation”, Model(1) controls for the interaction between each RCT dummies and variable that was un-balanced in that specific RCT. However, results are similar when not controlling for these variables.

Table 6: Number of Pivots

VARIABLES	(1) # Pivots OLS – Cross-Section Full Sample	(2) # Pivots OLS – Cross-Section RCT1	(3) # Pivots OLS – Cross-Section RCT2	(4) # Pivots OLS – Cross-Section RCT3	(5) # Pivots OLS – Cross-Section RCT4
Intervention	-0.053	0.261	0.011	-0.474	-0.038
	(0.486)	(0.021)	(0.860)	(0.194)	(0.588)
Constant	0.693	0.536	1.245	1.174	0.435
	(0.175)	(0.217)	(0.000)	(0.451)	(0.000)
Observations	759	116	250	132	261
R-squared	0.128	0.105	0.071	0.073	0.019
Dummies for Instructors	Yes	Yes	Yes	Yes	Yes
Dummies for RCTs	Yes	-	-	-	-
Clustered Errors	Intervention Instructor RCT	Intervention Instructor	Intervention Instructor	Intervention Instructor	Intervention Instructor

Robust pval in parentheses. Our balance checks have shown that three variables in two RCTs were unbalanced between the treatment and control group despite randomization. To address this issue, we conservatively included these variables as controls in all specifications. Specifically, Model (3) controls for “Background: Economics” and ”Background: STEM”, Model (4) controls for ”Self-regulation”, Model(1) controls for the interaction between each RCT dummies and variable that was un-balanced in that specific RCT. However, results are similar when not controlling for these variables.

Table 7: Pivot Multinomial Probit

VARIABLES	(1) Pivoting only once Multinomial Probit Cross-Section Full Sample	(2) Pivoting twice Multinomial Probit Cross-Section Full Sample	(3) Pivoting more than twice Multinomial Probit Cross-Section Full Sample
Intervention	0.350	0.112	-0.282
	(0.014)	(0.515)	(0.077)
Constant	-1.363	-2.062	-2.429
	(0.000)	(0.000)	(0.000)
Observations	759	759	759
Dummies for instructors and RCTs	Yes	Yes	Yes
Clustered Errors	Intervention Instructor RCT	Intervention Instructor RCT	Intervention Instructor RCT

Robust pval in parentheses. Our balance checks have shown that three variables in two RCTs were unbalanced between the treatment and control group despite randomization. To address this issue, we conservatively included these variables as controls in all specifications. Specifically, Model (3) controls for “Background: Economics” and ”Background: STEM”, Model (4) controls for ”Self-regulation”, Model(1) controls for the interaction between each RCT dummies and variable that was un-balanced in that specific RCT. However, results are similar when not controlling for these variables.

Table 8: Performance OLS

VARIABLES	(1) Revenue OLS Cross-section Full Sample	(2) Revenue OLS Cross-section RCT1	(3) Revenue OLS Cross-section RCT2	(4) Revenue OLS Cross-section RCT3	(5) Revenue OLS Cross-section RCT4
Intervention	6999.327	10,799.493	1,517.117	3253.979	12,227.935
	(0.030)	(0.125)	(0.136)	(0.027)	(0.164)
Constant	-2,999.664	-4,899.746	-466.455	5808.953	6,297.301
	(0.369)	(0.403)	(0.855)	(0.257)	(0.344)
Observations	759	116	250	132	261
R-squared	0.085	0.220	0.023	0.096	0.036
Dummies for Instructors	Yes	Yes	Yes	Yes	Yes
Dummies for RCTs	Yes	-	-	-	-
Clustered Errors	Intervention Instructor RCT	Intervention Instructor	Intervention Instructor	Intervention Instructor	Intervention Instructor

Robust pval in parentheses. Our balance checks have shown that three variables in two RCTs were unbalanced between the treatment and control group despite randomization. To address this issue, we conservatively included these variables as controls in all specifications. Specifically, Model (3) controls for “Background: Economics” and ”Background: STEM”, Model (4) controls for ”Self-regulation”, Model(1) controls for the interaction between each RCT dummies and variable that was un-balanced in that specific RCT. However, results are similar when not controlling for these variables.

Table 9: Scientific Intensity

VARIABLES	(1) Scientific Intensity OLS Cross-section Full Sample	(2) Scientific Intensity OLS Cross-section RCT1	(3) Scientific Intensity OLS Cross-section RCT2	(4) Scientific Intensity OLS Cross-section RCT3	(5) Scientific Intensity OLS Cross-section RCT4
Intervention	0.319	0.581	0.196	0.296	0.321
	(0.000)	(0.002)	(0.072)	(0.281)	(0.015)
Constant	1.285	1.155	2.493	1.082	2.086
	(0.006)	(0.006)	(0.000)	(0.172)	(0.000)
Observations	759	116	250	132	261
R-squared	0.173	0.178	0.090	0.064	0.028
Dummies for Instructors	Yes	Yes	Yes	Yes	Yes
Dummies for RCTs	Yes	-	-	-	-
Clustered Errors	Intervention Instructor RCT	Intervention Instructor	Intervention Instructor	Intervention Instructor	Intervention Instructor

Robust pval in parentheses. Our balance checks have shown that three variables in two RCTs were unbalanced between the treatment and control group despite randomization. To address this issue, we conservatively included these variables as controls in all specifications. Specifically, Model (3) controls for “Background: Economics” and ”Background: STEM”, Model (4) controls for ”Self-regulation”, Model(1) controls for the interaction between each RCT dummies and variable that was un-balanced in that specific RCT. However, results are similar when not controlling for these variables.

Table 10: Instrumenting Scientific Intensity

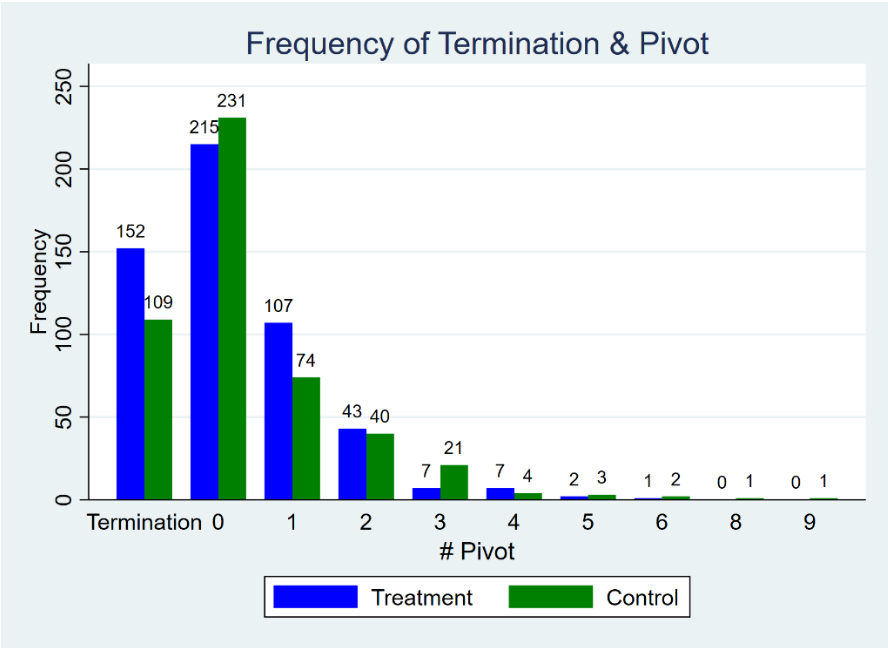
VARIABLES	(1) Termination 2SLS Cross-Section Full Sample	(2) Pivoting once 2SLS Cross-Section Full Sample	(3) Revenue 2SLS Cross-Section Full Sample
Average scientific intensity	0.296	0.252	21,047.762
	(0.003)	(0.001)	(0.039)
Constant	-0.108	-0.377	-30,936.525
	(0.592)	(0.008)	(0.068)
Observations	759	759	759
Dummies for instructors and RCTs	Yes	Yes	Yes
Clustered Errors	Intervention Instructor RCT	Intervention Instructor RCT	Intervention Instructor RCT

Robust pval in parentheses. Our balance checks have shown that three variables in two RCTs were unbalanced between the treatment and control group despite randomization. To address this issue, we conservatively included these variables as controls in all specifications. Specifically, Model (3) controls for “Background: Economics” and ”Background: STEM”, Model (4) controls for ”Self-regulation”, Model(1) controls for the interaction between each RCT dummies and variable that was un-balanced in that specific RCT. However, results are similar when not controlling for these variables.

FIGURES

Figure 1: Termination, Pivot and Performance

(a) Termination and Pivot



(b) Performance (Revenues in EURO)

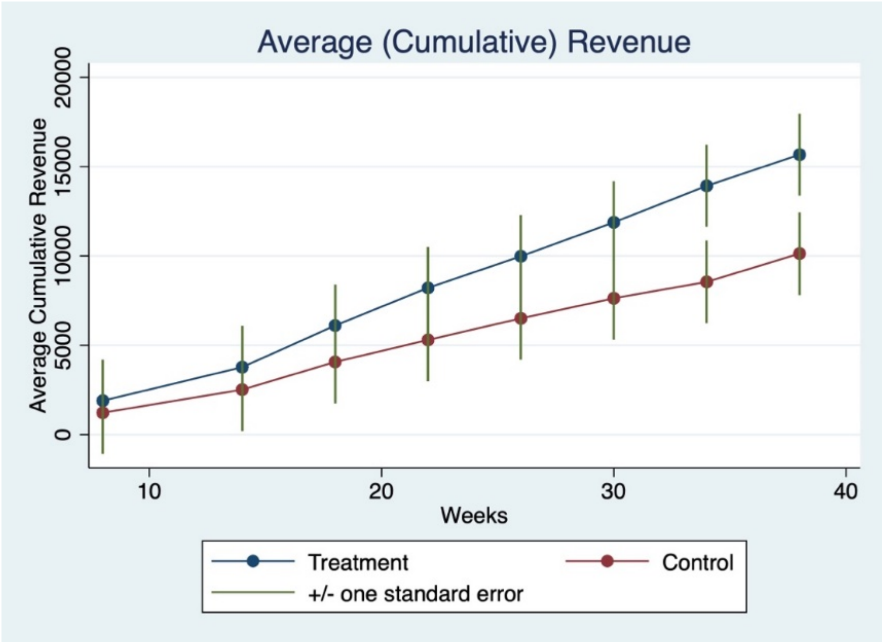


Figure 2: Marginal Effects of Intervention on Pivot

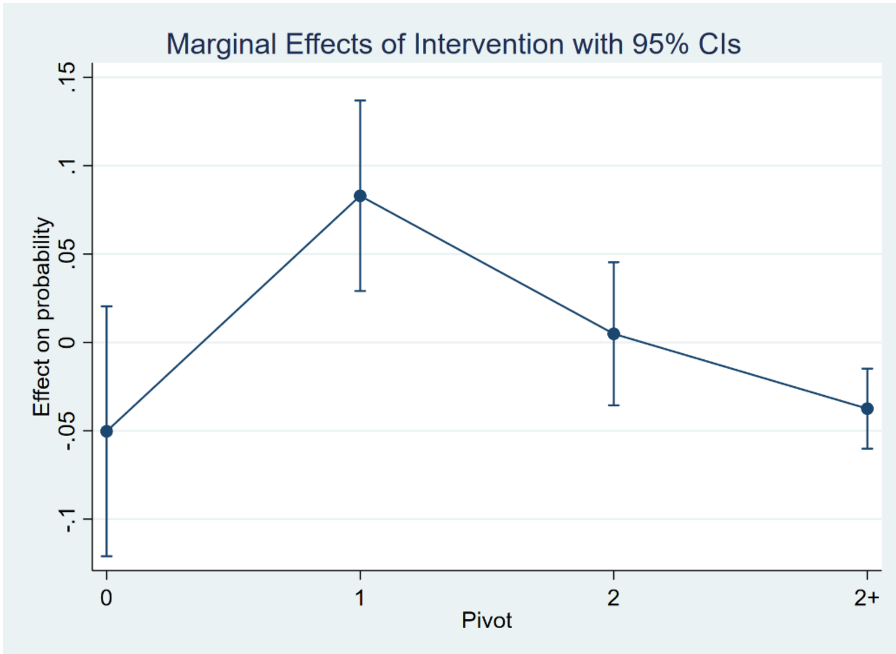


Figure 3: Decision-Making Process, Exploration Stages

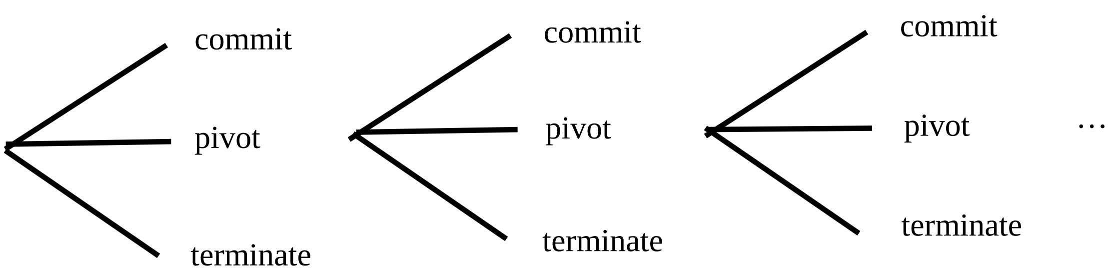
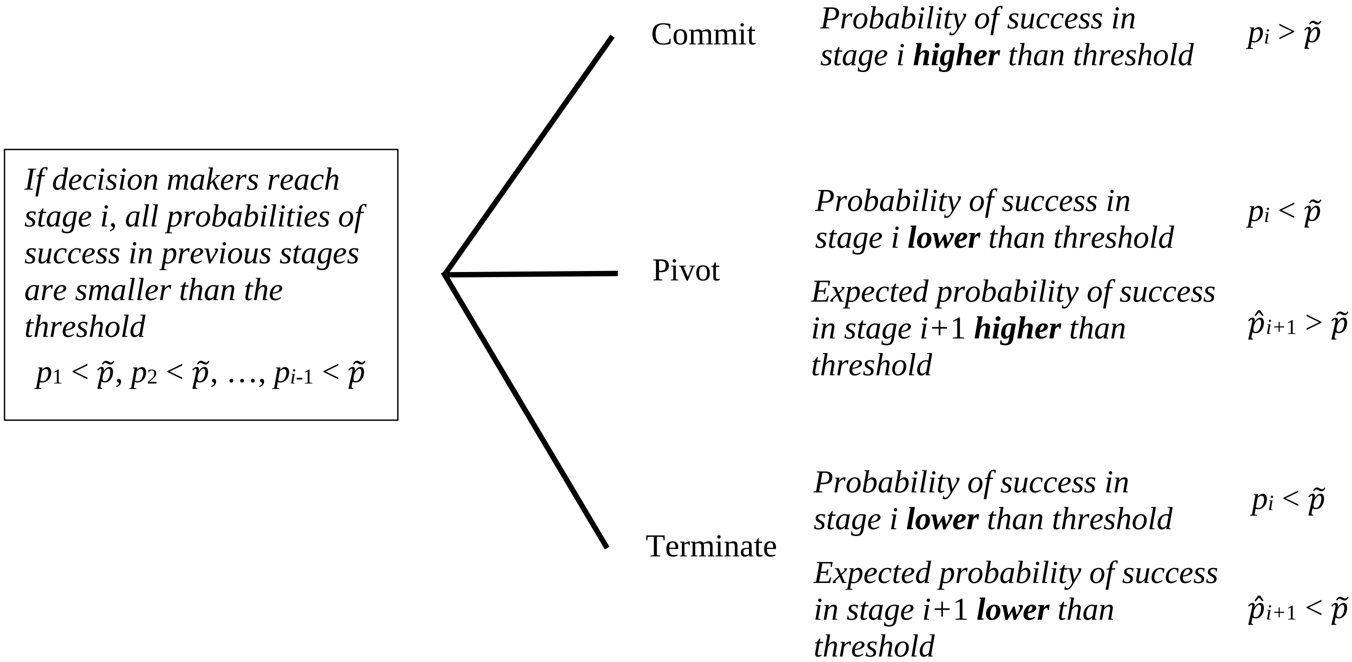
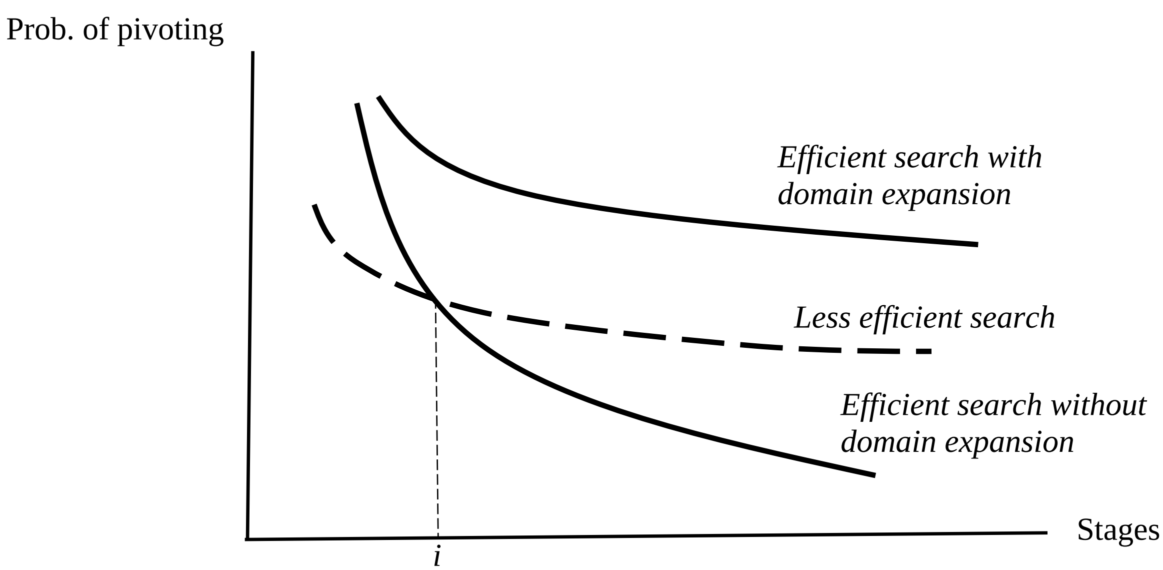


Figure 4: Decision-Makers' Options at Stage *i*



p_i = probability of success of i^{th} idea at the end of stage i
$\hat{p}_{i+1}$ = expected probability of success of $(i + 1)^{th}$ idea at the end of stage i
$\tilde{p}$ = threshold probability beyond which decision makers commit to the idea

Figure 5: Efficient Search and Probability of Pivoting Across Stages



If negative information about ideas in previous stages reveal new ideas that expand the domain of search, efficient search raises the odds of exploring new high-quality ideas in later stages. Otherwise, the probability of pivoting falls more rapidly across stages because, other things being equal, better ideas are explored first. With less efficient search, the sequence of exploration of the ideas is less correlated with their quality.

APPENDIX

1. Sector details

Table A1: Participants breakdown by RCT and industry (number and % frequency)

	RCT				
	1	2	3	4	Total
Accommodation and food service activities	7	77	40	13	137
	5.11%	56.20%	29.20%	9.49%	100.00%
	6.03%	30.80%	30.30%	4.98%	18.05%
Administrative and support service activities	3	19	11	5	38
	7.89%	50.00%	28.95%	13.16%	100.00%
	2.59%	7.60%	8.33%	1.92%	5.01%
Agriculture, forestry and fishing	1	13	2	2	18
	5.56%	72.22%	11.11%	11.11%	100.00%
	0.86%	5.20%	1.52%	0.77%	2.37%
Arts, entertainment and recreation	0	0	0	23	23
	0.00%	0.00%	0.00%	100.00%	100.00%
	0.00%	0.00%	0.00%	8.81%	3.03%
Construction	0	0	0	2	2
	0.00%	0.00%	0.00%	100.00%	100.00%
	0.00%	0.00%	0.00%	0.77%	0.26%
Education	0	11	12	14	37
	0.00%	29.73%	32.43%	37.84%	100.00%
	0.00%	4.40%	9.09%	5.36%	4.87%
Electricity, gas, steam and air conditioning supply	0	3	0	2	5
	0.00%	60.00%	0.00%	40.00%	100.00%
	0.00%	1.20%	0.00%	0.77%	0.66%
Financial and insurance activities	0	27	0	0	27
	0.00%	100.00%	0.00%	0.00%	100.00%
	0.00%	10.80%	0.00%	0.00%	3.56%
Human health and social work activities	0	0	0	20	20
	0.00%	0.00%	0.00%	100.00%	100.00%
	0.00%	0.00%	0.00%	7.66%	2.64%
Information and communication	55	34	15	54	158
	34.81%	21.52%	9.49%	34.18%	100.00%
	47.41%	13.60%	11.36%	20.69%	20.82%
Manufacturing	33	22	19	14	88
	37.50%	25.00%	21.59%	15.91%	100.00%
	28.45%	8.80%	14.39%	5.36%	11.59%
Other service activities	0	0	0	5	5
	0.00%	0.00%	0.00%	100.00%	100.00%
	0.00%	0.00%	0.00%	1.92%	0.66%
Professional, scientific and technical activities	2	31	18	64	115
	1.74%	26.96%	15.65%	55.65%	100.00%
	1.72%	12.40%	13.64%	24.52%	15.15%
Real estate activities	0	0	0	10	10
	0.00%	0.00%	0.00%	100.00%	100.00%
	0.00%	0.00%	0.00%	3.83%	1.32%
Transportation and storage	1	7	6	3	17
	5.88%	41.18%	35.29%	17.65%	100.00%

	0.86%	2.80%	4.55%	1.15%	2.24%
Water supply, sewerage, waste management and remediation activities	4	6	9	1	20
	20.00%	30.00%	45.00%	5.00%	100.00%
	3.45%	2.40%	6.82%	0.38%	2.64%
Wholesale and retail trade; repair of motor vehicles and motorcycles	10	0	0	29	39
	25.64%	0.00%	0.00%	74.36%	100.00%
	8.62%	0.00%	0.00%	11.11%	5.14%
Total	116	250	132	261	759
	15.28%	32.94%	17.39%	34.39%	100.00%
	100.00%	100.00%	100.00%	100.00%	100.00%

Each Table cell has three rows. The first reports *frequencies*; the second reports *row percentages* and the third reports *column percentages*

2. Balance Checks

Table A2: Balance Checks RCT1

Variable	Description	Treatment		Control		Difference	
		Mean	SD	Mean	SD	b	p
Currently Employed	Proportion of team members employed at the time of the training	0.68	0.39	0.72	0.42	-0.04	(0.566)
Currently Studying	Proportion of team members enrolled in an education program at the time of training	0.19	0.37	0.28	0.42	-0.09	(0.249)
Education Level	Highest educational level attained by team members (5=PhD, 4=MBA, 3=MSc, 2=BA, 1=high school, 0=otherwise; Team Average)	2.34	0.86	2.12	0.91	0.22	(0.191)
Experience: Entrepreneurial Founder or Employee	Number of years of experience working with companies other than focal as founder or employee (Team Average)	0.92	2.51	0.32	1.18	0.59	(0.110)
Experience: Entrepreneurial Instructor	Number of years of experience working with companies other than focal as Instructor or consultant (Team Average)	0.02	0.13	0.02	0.13	-0.00	(0.981)
Experience: Industry	Number of years of experience in industry (Team Average)	2.55	4.64	2.56	4.78	-0.01	(0.991)
Experience: Managerial	Number of years of managerial experience (Team Average)	2.03	3.34	1.22	3.32	0.81	(0.192)
Idea Stage	Dummy variable assuming value of 1 when the company has one business idea and 0 when the company has started working on the project but has not launched it on the market yet	0.63	0.49	0.65	0.48	-0.02	(0.807)
Lombardy	Dummy variable assuming value of 1 when the majority of team members comes from the Italian region of Lombardy and 0 otherwise	0.32	0.47	0.40	0.49	-0.08	(0.366)
Sector: Furniture	Dummy variable assuming value of 1 when the company operates in the furniture sector and 0 otherwise	0.25	0.44	0.25	0.43	0.01	(0.916)

Sector: Internet	Dummy variable assuming value of 1 when the company operates in the internet sector and 0 otherwise	0.44	0.50	0.51	0.50	-0.07	(0.467)
Sector: Retail	Dummy variable assuming value of 1 when the company operates in the retail sector and 0 otherwise	0.10	0.30	0.07	0.26	0.03	(0.549)
Team Size	Number of team members	2.85	1.36	2.72	1.31	0.13	(0.606)
Observations		59		57		116	

Table A3: Balance Checks RCT2

Variable	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	(0.055)
Background: Other	Team members with no economics background (%)	0.22	0.36	0.20	0.33	0.02	(0.716)
Background: STEM	Team members with a STEM (Science Technology Engineering Mathematics) background (%)	0.38	0.4	0.49	0.41	-0.11	(0.029)
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	0.73	0.37	0.75	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)	85.8	14.38	85.67	16.16	-0.13	(0.947)
Idea Value: Mean	Estimated value of the project (mean, 0 to 100)	66.34	14.47	64.44	16.84	1.90	(0.341)
Idea Value: Min	Minimum estimated value of the project (0 to 100)	46.87	19.57	43.21	22.93	3.66	(0.175)
Idea Value: Range	Estimated value of the project (range, 0 to 100)	38.93	18.49	42.46	20.99	-3.54	(0.159)
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 start up 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	0.55	0.50	0.54	0.50	0.01	(0.899)
Months to Revenue	Number of months to revenue	11.49	5.78	11.33	5.83	0.16	(0.828)
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	19.13	20.82	22.83	20.39	-3.70	(0.158)
Team Size	Number of team members	2.25	1.46	2.27	1.36	-0.02	(0.893)
Observations		125		125		250	

Table A4: Balance Checks RCT3

Variable	Description	Treatment Mean	Treatment SD	Control Mean	Control SD	Difference b	Difference p
Age	Age (Team Average)	30.58	9.07	30.48	7.09	0.10	(0.942)
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 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” and ”Multiple elements are taken into account when making a decision for our company, pros and cons are carefully evaluated in every situation"	4.27	0.65	4.39	0.56	-0.13	(0.233)
Background: Economics	Team members with Economics background (%)	0.19	0.32	0.21	0.36	-0.02	(0.790)
Background: Other	Team members with no Economics/STEM background (%)	0.56	0.43	0.44	0.46	0.12	(0.130)
Background: STEM	Team members with a STEM (Science Technology Engineering Mathematics) background (%)	0.25	0.37	0.35	0.45	-0.10	(0.162)
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"	3.42	0.53	3.32	0.64	0.10	(0.336)
Currently Studying	Number of team members enrolled in an education program at the time of training	0.26	0.30	0.22	0.30	0.04	(0.429)
Education	Highest educational level attained by team members (5=PhD, 4=MBA, 3=MSc, 2=BA, 1=high school, 0=otherwise; Team Average)	1.86	0.88	2.07	1.08	-0.21	(0.222)
Experience: Business Plan	Dummy taking value of 1 if the team had years of experience in business plan design, 0 otherwise	0.27	0.37	0.35	0.42	-0.08	(0.283)
Experience: Entrepreneurial	Number of years of entrepreneurial experience (Team Average)	1.81	4.38	1.71	3.35	0.11	(0.873)
Experience: Industry	Number of years of experience in industry (Team Average)	2.88	5.65	2.99	5.01	-0.11	(0.910)
Experience: Managerial	Number of years of managerial experience (Team Average)	1.56	2.71	1.73	3.74	-0.18	(0.758)
Gender (Female)	Proportion of women in the team	0.68	0.39	0.75	0.35	-0.07	(0.309)
Hours: Total Weekly	Weekly hours dedicated to the company (Team Average)	11.33	9.95	11.62	12.32	-0.28	(0.885)
Idea Maturity	Maturity of the idea (in months)	9.95	9.54	12.16	11.63	-2.21	(0.236)
Idea Potential	Independent assessment of the value of the idea (two evaluators, average) based on five criteria: innovation, feasibility, sustainability, team competence, market size	48.85	12.05	49.17	12.77	-0.33	(0.881)
Idea Value: Mean	Estimated value of the project (mean)	66.24	18.89	63.54	16.06	2.69	(0.379)
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”	2.79	0.86	2.71	0.98	0.08	(0.604)
Later Stage	Dummy variable taking value of 1 if the firm is at a more advanced stage than others, 0 otherwise	0.14	0.35	0.10	0.31	0.03	(0.553)
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 person's 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 satisfying job" and "To get a really good job you have to have high-level acquaintances"	3.85	0.67	3.78	0.70	0.07	(0.556)

Months to Revenue	Number of months to revenue	12.42	11.2	14.63	10.51	-2.21	(0.244)
Piedmont	Dummy variable taking value of 1 when the majority of team members comes from the Italian region of Piedmont, and 0 otherwise	0.49	0.50	0.52	0.50	-0.03	(0.732)
Probability Pivot Idea	Probability of changing the business idea	30.69	22.96	32.42	26.56	-1.73	(0.691)
Probability Pivot Other	Probability of changing other component of the business model	51.6	22.46	52.58	26.13	-0.98	(0.817)
Probability Pivot Problem	Probability of changing the problem and customer segment	33.75	22.68	34.42	25.02	-0.66	(0.873)
Probability Termination	Probability of terminating the project	12.95	16.27	17.31	21.52	-4.36	(0.192)
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"	4.21	1. 00	3.95	1.04	0.26	(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 person's abilities and the position he/she holds” and ”I think I am often less cautious than other people"	4.04	1.10	3.98	0.90	0.06	(0.715)
Scientific Intensity: 1 Theory	Theory development score (0-5 scale)	2.87	1.34	3.02	1.21	-0.15	(0.514)
Scientific Intensity: 2 Hypothesis	Hypothesis development score (0-5 scale)	2.12	1.64	1.97	1.5	0.15	(0.587)
Scientific Intensity: 3 Test	Test score (0-5 scale)	1.33	1.72	1.29	1.68	0.04	(0.906)
Scientific Intensity: 4 Evaluation	Evaluation score (0-5 scale)	.85	1.5	.94	1.62	-0.09	(0.750)
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”	5.43	1.08	5.56	0.95	-0.13	(0.460)
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 can't 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"	4.97	0.83	5.23	0.86	-0.26	(0.074)
Startcup	Dummy variable taking value of 1 if the firm takes part to a local competition, 0 otherwise	0.11	0.31	0.18	0.39	-0.07	(0.246)
Team Size	Number of team members	2.48	1.62	2.19	1.38	0.28	(0.282)
Observations		65		67		132	

Table A5: Balance Checks RCT4

Variable	Description	Treatment		Control		Difference	
		Mean	SD	Mean	SD	b	p
Age	Age (Team Average)	35.76	8.43	36.37	9.20	-0.61	(0.579)
Background: Economics	Team members with Economics background (%)	0.14	0.29	0.15	0.29	-0.01	(0.847)
Background: Other	Team members with no economics background (%)	0.08	0.11	0.09	0.16	-0.01	(0.405)
Background: STEM	Team members with a STEM (Science Technology Engineering Mathematics) background (%)	0.29	0.39	0.36	0.43	-0.07	(0.191)
Business Age	Age of the business (years)	2.48	3.22	3.28	5.17	-0.80	(0.133)
Confidence	Agreement on a 1-5 scale with the following statements (Team Average): ”We are confident in our entrepreneurial skills”, ”We are sure we are deploying the best strategy for our business”, ”We are confident in our ability to manage our business”, ”We master the competences necessary for our venture”, ”We are sure there is no better business model for our idea"	3.41	0.69	3.34	0.76	0.07	(0.436)
Education	Highest educational level attained by team members (5=PhD, 4=MBA, 3=MSc, 2=BA, 1=high school, 0=otherwise; Team Average)	2.67	0.80	2.58	0.79	0.10	(0.333)
Experience: Entrepreneurial	Number of years of entrepreneurial experience (Team Average)	3.81	3.41	4.58	5.86	-0.78	(0.191)
Experience: Industry	Number of years of experience in industry (Team Average)	6.66	6.31	7.66	7.51	-1.00	(0.244)
Experience: Managerial	Number of years of managerial experience (Team Average)	5.88	5.12	6.15	6.02	-0.27	(0.694)
Experience: Work	Number of years of work experience (Team Average)	12.99	7.86	13.51	8.53	-0.52	(0.610)
Gender (Female)	Proportion of women in the team	0.58	0.42	0.50	0.44	0.08	(0.119)
Hours: % Innovation monthly	Working hours dedicated to the design of new products or services in the last month (January 2019, %)	39.24	33.8	36.84	34.59	2.41	(0.570)
Hours: % Innovation yearly	Working hours dedicated to the design of new products or services in the last year (2018, %)	45.92	32.98	40.02	32.68	5.90	(0.148)
Hours: Total Weekly	Weekly hours dedicated to the company (Team Average)	31.51	18.29	29.61	17.12	1.89	(0.389)
Idea Value: Mean	Estimated value of the project (mean, 0 to 100))	66.83	16.80	66.62	20.22	0.21	(0.929)
Idea Value: Range	Estimated value of the project (range, 0 to 100))	39.26	21.70	38. 00	21.94	1.26	(0.642)
Probability Expansion	Probability of expanding the business outside of the current industry or market	68.32	27.09	66.59	28.12	1.73	(0.613)
Probability Pivot Idea	Probability of making a radical change to the business	45.78	27.86	42.12	26.99	3.66	(0.283)
Probability Pivot Problem	Probability of changing the problem and customer segment	38.13	25.86	40.55	26.26	-2.43	(0.453)
Scientific Intensity	Scientific Intensity	2.56	1.23	2.35	1.29	0.20	(0.200)
Team Size	Number of team members	1.80	2.09	1.90	2.21	-0.09	(0.725)
Turnover Annual	Annual turnover (2018) £	57142.92	170000	83133.84	230000	-25990.92	(0.290)
Turnover Monthly	Monthly turnover (January 2019) £	5806.46	20261.11	7044.92	28286.50	-1238.46	(0.684)
Observations		133		128		261	

Notes. Some values of some of the variables were missing (about 2% of the data). We filled in these missing values with the median value of the variable.

Table A6: Balance Checks Full Sample

Variable	Description	Treatment		Control		Difference	
		Mean	SD	Mean	SD	b	p
Business Age	Age of the business (years)	0.86	2.23	1.11	3.39	-0.25	(0.229)
Education	Highest educational level attained by team members (5=PhD, 4=MBA, 3=MSc, 2=BA, 1=high school, 0=otherwise; Team Average)	2.24	0.88	2.21	0.91	0.04	(0.594)
Experience: Managerial	Number of years of managerial experience (Team Average)	2.13	3.37	2.22	4.18	-0.08	(0.766)
Experience: Entrepreneurial	Number of years of entrepreneurial experience (Team Average)	4.13	5.53	4.29	6.09	-0.16	(0.706)
Experience: Industry	Number of years of experience in industry (Team Average)	3.37	4.46	3.33	5.13	0.04	(0.908)
Team Size	Number of team members	2.23	1.75	2.20	1.71	0.03	(0.835)
Turnover: Monthly	Monthly turnover EUR	2025.81	12242.8	2391.91	16775.41	-366.10	(0.731)
Observations		382		377		759	

3. Additional analysis

Table A7a: Seemingly unrelated regression

VARIABLES	(1) Termination Cross-Section Full Sample	(2) Pivoting once Cross-Section Full Sample	(3) Revenue Cross-Section Full Sample
Intervention	0.057	0.059	2377.59
	(0.000)	(0.000)	(0.001)
Constant	0.076	-0.019	-1126.29
	(0.053)	(0.648)	(0.768)
Observations	759	759	759
Dummies for instructors and RCTs	Yes	Yes	Yes
Clustered Errors	Intervention Instructor RCT	Intervention Instructor RCT	Intervention Instructor RCT

Robust pval in parentheses. Our balance checks have shown that three variables in two RCTs were unbalanced between the treatment and control group despite randomization. To address this issue, we conservatively included these variables as controls in all specifications. Specifically, Model (3) controls for “Background: Economics” and ”Background: STEM”, Model (4) controls for ”Self-regulation”, Model(1) controls for the interaction between each RCT dummies and variable that was un-balanced in that specific RCT. However, results are similar when not controlling for these variables.

Correlation matrix of residuals:

	Termination	Pivoting once	Revenue
Termination	1.0000		
Pivoting once	-0.0469	1.0000	
Revenue	0.0153	0.0095	1.0000

Table A7b: Copula regression

VARIABLES	(1) Termination Full Sample	(2) Pivoting only once Full Sample
Pivoting only once	-1.073	
	(0.000)	
Intervention	0.363	0.290
	(0.000)	(0.001)
Constant	-0.260	-1.050
	(0.034)	(0.000)
Observations	759	759
Clustered Errors	Intervention Instructor RCT	Intervention Instructor RCT

Wald test of theta=0: chi2(1) = 3.71248
Prob > chi2 = 0.0540

Robust pval in parentheses. Our balance checks have shown that three variables in two RCTs were unbalanced between the treatment and control group despite randomization. To address this issue, we conservatively included these variables as controls in all specifications. Specifically, Model (3) controls for "Background: Economics" and "Background: STEM", Model (4) controls for "Self-regulation", Model(1) controls for the interaction between each RCT dummies and variable that was un-balanced in that specific RCT. However, results are similar when not controlling for these variables.

We use Frank copula to estimate the joint distribution of our binary variables termination and pivoting once, for two main reasons. First, Frank copula is symmetric and, while being similar to the Gaussian copula, it is well suited to deal with the dependence between binary responses (Winkelmann, 2012; Radice et al., 2016; Lin and Chaganty, 2021). Second, the STATA package for Frank copula allows to compute the total average marginal effects of Intervention on the joint probability, which is useful to address the role of Intervention.

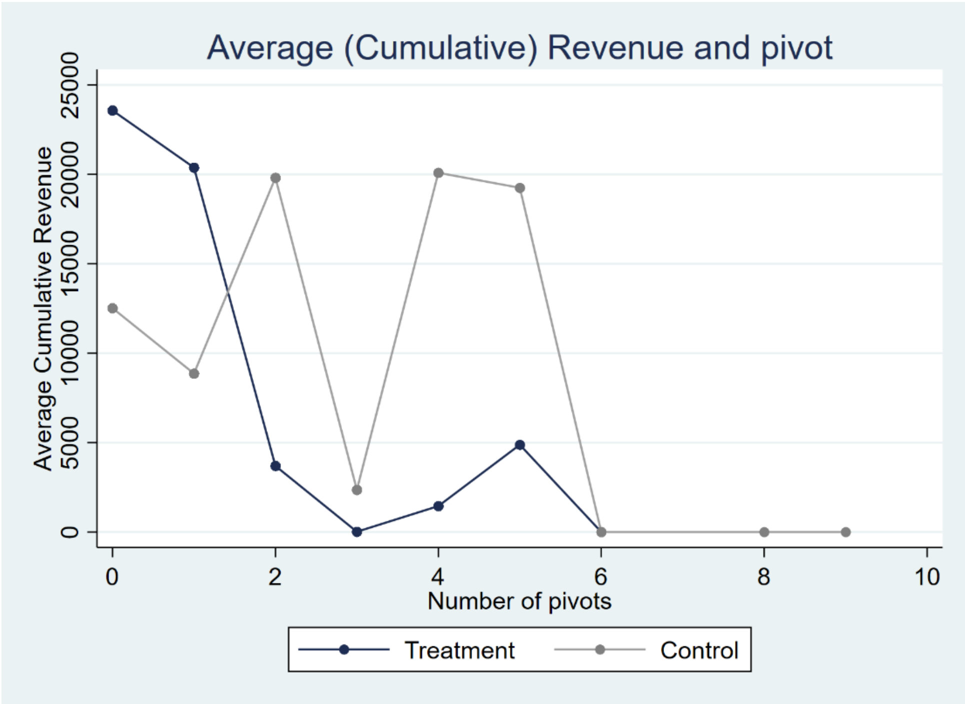
Table A8: Instrumenting Pivoting once

VARIABLES	(1) Termination Cross-Section Full Sample	(1) Revenue Cross-Section Full Sample
Pivoting only once	1.173	83,489.59
	(0.019)	(0.090)
Constant	0.333	500.00
	(0.157)	(0.157)

Observations	759	759
Dummies for instructors and RCTs	Yes	Yes
Clustered Errors	Intervention Instructor RCT	Intervention Instructor RCT

Robust pval in parentheses. Our balance checks have shown that three variables in two RCTs were unbalanced between the treatment and control group despite randomization. To address this issue, we conservatively included these variables as controls in all specifications. Specifically, Model (3) controls for “Background: Economics” and ”Background: STEM”, Model (4) controls for ”Self-regulation”, Model(1) controls for the interaction between each RCT dummies and variable that was un-balanced in that specific RCT. However, results are similar when not controlling for these variables.

Figure A1: Revenue and Pivot



4. Attrition treatment

Unfortunately, not all firms continued to answer our interviews until the end of the program. Notoriously, attrition is common in field experiments and there is no established best approach to deal with it (Gerber & Green, 2012; Ghanem, Hirshleifer & Ortiz-Becerra, 2022). To contrast this tendency, we designed the program so that the core training was followed by a series of monthly events focused on relevant themes of interest to the study participants. The events included no treatment and were delivered in separate days but in the same way for treated and control firms. However, participation in these events was allowed only to firms that showed their continued engagement with the data collection. Nevertheless, some of these firms did not reach the last interview round. The motivation provided by entrepreneurs in expressing their unavailability to be interviewed was that, since the main training was over, their incentive to answer the interviews was lower. Overall, 22% of firms in our sample withdrew at different points of the program.

To verify that attrition did not affect our results, we check if there was not any significant difference between treated and control in their early withdrawal from the program. In Table A6, we estimate early withdrawal from the program as a function of the intervention, which we show has no significant impact. In our main analyses, we addressed attrition by inputting the missing values of those who left the study, making the conservative assumption that the performance of firms that left the program did not change after they left the program.

Table A9: Early Withdrawal from Program

	Early Withdrawn
	OLS Cross-section
VARIABLES	Full Sample
Intervention	-0.019
	(0.419)
Constant	0.343
	(0.174)
Observations	759
R-squared	0.179
Dummies for Instructors and RCTs	Yes
Clustered Errors	Intervention Instructor RCT

Robust pval in parentheses. All specifications control for the variables that were unbalanced between the treatment and control group despite randomization.

5. Scientific Intensity

In this section, we provide additional details about the coding scheme used to evaluate the adoption of the scientific approach by study participants. Following Camuffo et al. (2020), a team of research assistants conducted regular phone calls with the entrepreneurs participating to each of the RCTs. Calls followed a detailed protocol with a script including several open-ended questions which were used to measure the level of adoption of the scientific approach. In using open-ended questions, we follow an approach similar to the one described by Bloom and Van Reenen (2010) for their World Management Survey. Research assistants, during the phone interviews, asked open-ended questions whose content was then coded to understand to what extent participants formed, tested, and updated their belief according to the scientific approach in their entrepreneurial activities. The fact that participants did not know the coding scheme helped ensure the collection of unbiased information. Adoption of the scientific approach was assessed through behavioral observation scales -from 0 (lowest score) to 5 (highest score), across the four key components of the scientific approach described in the theory section of the paper: 1) Theory, 2) Hypotheses (for belief formation), 3) Tests (for belief testing), and 4) Evaluation (for belief updating).

In line with key literature in this area, we consider each component of the scientific approach as a multidimensional construct. For instance, the articulation of a theory rests on a variety of aspects, such as its clarity (use of logic), level of detail (conceptual sharpness, parsimony), the extent to which it is grounded on reality (novelty) and the extent to which it considers alternative explanations (generalizability). To adequately capture the multiple dimensions of each component, we identified some sub-components that measure the key aspects that define theory, hypotheses, tests, and evaluation. Noteworthy, each of these elements and sub-elements can greatly vary across entrepreneurs. One entrepreneur might have an extremely clear theory related to how his/her firm generates value for customers, while another might have a very murky explanation for his/her value creation process. All research assistants received extensive training prior to performing phone calls and conducting interviews. The research team designed and provided multiple training and practice sessions to the research assistants to clarify how to code the interviews and score them. These sessions also coding and scoring examples of mock interviews to create a template and a standard that research assistants could refer to. We provide an overview of the sub-components of the scientific approach and their related scores in Table A7 below. More details about research assistants and the coding procedure are available upon request. The measure of scientific intensity we used in our TSLS cross-sectional regressions is the average of the within-participant scores of scientific intensity over the entire observation period.

Table A10: Scientific Intensity components

Component	Sub-component	Definition	Score
Theory	Clarity of theory	The extent to which the theory is understandable	0 (no theory) or from 1 (not clear) to 5 (extremely clear)
Theory	Articulation of theory	The extent to which the theory is detailed	0 (no theory) or from (not detailed) to 5 (extremely detailed)
Theory	Consideration of alternatives	The extent to which the theory includes alternative possible options	0 (no theory) or from 1 (no consideration of alternatives) to 5 (careful consideration of many alternatives)
Theory	Theory based on evidence	The extent to which the theory is based on objective evidence	0 (no theory) or from 1 (theory not based on objective evidence) to 5 (extremely based on objective evidence)
Hypotheses	Explicitness of hypotheses	The extent to which the respondent can articulate the fundamental assumptions that make his/her business viable	0 (no hypotheses) or from 1 (not explicit hypotheses) to 5 (extremely explicit)
Hypotheses	Coherence of hypotheses	The extent to which hypotheses are coherent with the theory	0 (no hypotheses) or from 1 (not coherent) to 5 (extremely coherent)

Hypotheses	Level of details of hypotheses	The extent to which hypotheses clearly indicate the details of what the entrepreneur wishes to learn and how to measure it	0 (no hypotheses) of from 1 (not detailed) to 5 (extremely detailed)
Hypotheses	Falsifiability of hypotheses	The extent to which it is possible to clearly determine (after tests) whether the hypotheses are supported or not	0 (no hypotheses) or from 1 (not falsifiable) to 5 (extremely falsifiable)
Tests	Coherence of tests	The extent to which the test is coherent with the hypotheses	0 (no tests) or from 1 (not coherent) to 5 (extremely coherent)
Tests	Validity of tests	The extent to which the test has been conducted in a context similar to which the business operates	0 (no hypotheses) or from 1 (not valid) to 5 (extremely valid)
Tests	Representativeness of tests	The extent to which the test has been conducted with a sample that is representative of the broad group the firm targets	0 (no hypotheses) or from 1 (not representative) to 5 (extremely representative)
Tests	Rigorousness of tests	The extent to which the appropriate test and procedure for that type of test have been chosen for hypotheses-testing	0 (no hypotheses) or from 1 (not rigorous) to 5 (extremely rigorous)
Evaluation	Data-based assessment	The extent to which the evaluation is based on data	0 (no hypotheses) or from 1 (not based on data) to 5 (extremely based on data)
Evaluation	Coherence of measures	The extent to which the measure used are consistent with the learning objective the entrepreneur has in mind	0 (no hypotheses) or from 1 (not coherent) to 5 (extremely coherent)
Evaluation	Systematic evaluation	The extent to which the evaluation is based on systematically collected and analyzed data	0 (no hypotheses) or from 1 (not systematic) to 5 (extremely systematic)
Evaluation	Explanatory power of evaluation	The extent to which the evaluation results in clarity on the main findings from the test and their implications for the business	0 (no hypotheses) or from 1 (not explanatory) to 5 (extremely explanatory)

REFERENCES

Gerber, A. S., & Green, D. P. (2012). Field experiments: Design, analysis, and interpretation. *WW Norton.*

Ghanem, D., Hirshleifer, S. and Ortiz-Becerra, K., 2022. *Testing Attrition Bias in Field Experiments*, University of California Riverside, Department of Economics, (No. 202218).

Lin, H. & Chaganty, N.R. (2021). Multivariate distributions of correlated binary variables generated by pair-copulas. *Journal of Statistical Distributions and Applications*, 8(4), 1-14.

Winkelmann, R. (2012). Copula bivariate probit models: with an application to medical expenditures. *Health Economics*, 21(12), 1444–1455.

Radice, R., Marra, G. & Wojtyś, M. (2016). Copula regression spline models for binary outcomes. *Statistics and Computing*, 26(5), 981–995.