Why is Entrepreneurial Overconfidence (So) Persistent? Evidence from a Large-Scale Field Experiment*

Mu-Jeung Yang† Maclean Gaulin ‡ Nathan Seegert§

November, 2025

Abstract

Why do overconfident entrepreneurs fail to learn from frequent market feedback? We conducted two field experiments and followed nearly 1,000 firms for over a year to explore the role of hindsight bias and causal misattribution. Under our "error reminder" treatment, entrepreneurs are shown past forecast errors to remove hindsight bias. Under our "scientific learning" treatment, we encourage entrepreneurs to develop causal hypotheses and test these hypotheses empirically, to mitigate misattribution. We find that the error reminder treatment does not reduce overconfidence. In contrast, we find that stronger engagement with hypothesis testing within scientific learning successfully reduces overoptimism.

JEL: L26, D91, M21
Keywords: Entrepreneurial overconfidence, scientific learning, RCT

---

*We would like to thank Kevin Bryan, Arnaldo Camuffo, Alfonso Gambardella, Josh Hausman, Holger Herz, David Huffman, Sara Holland, Lubo Litov, John Mondragon, Collin Raymond, Jan Rivkin, Johannes Wieland, Matthew Wood, Todd Zenger and Florian Zimmermann and the seminar participants at the NBER Entrepreneurship Meeting 2022, the Empirical Management Conference 2022, the Academy of Management Annual Meetings 2023, the University of Oklahoma, the University of Utah, for helpful comments. The authors would like to acknowledge the support from the Utah Governor's Office of Management and Budget (GOMB), the Utah Governor's Office of Economic Development and the Utah Chamber of Commerce. Financial support from the Utah State Legislature is gratefully acknowledged. Additionally, Yang gratefully acknowledges summer support from the Dodge Family College of Arts and Sciences for Summer 2023 and financial support from the Leeds School of Business. None of the supporting entities reviewed the results of this study before circulation and the views expressed here are solely the authors' and do not reflect the views of the supporting institutions. The IRB number from the University of Utah is 00139342 and the AEA-RCT registry number is 0006860.

†Corresponding author, mj.yang@colorado.edu, Leeds School of Business, University of Colorado, Boulder; 995 Regent Drive, 419 UCB Boulder, Colorado 80309 USA

‡mac.gaulin@eccles.utah.edu, David Eccles School of Business, University of Utah

§n.seegert@northeastern.edu, D'Amore-McKim School of Business, Northeastern University

1 Introduction

Why do so many entrepreneurs remain overconfident despite strong feedback and incentives? Persistent entrepreneurial overconfidence is puzzling since standard learning models predict that such biases should diminish with experience (Friedman, 1953; Kahneman and Klein, 2009; Agrawal et al., 2024; Nanda, 2024). The recent literature on Bayesian entrepreneurship formalizes this prediction, assuming that entrepreneurs update beliefs rationally so that any initial overconfidence dissipates with experience (Agrawal et al., 2024; Nanda, 2024). Yet, we show that many entrepreneurs remain overly optimistic about their prospects even after years of operating in competitive markets. A growing behavioral literature argues that entrepreneurs may sustain biased beliefs through selective attention, motivated reasoning, or memory distortions (Chen et al., 2015; Schumacher et al., 2020; Benabou and Tirole, 2016). We build on this literature by testing two psychological mechanisms that may sustain overconfidence despite feedback and incentives: biased memory and misattribution. To do so, we run a 13-month randomized controlled trial with nearly 1,000 active entrepreneurs, which allows us to observe how these mechanisms operate in real firms.

We focus on two forms of entrepreneurial overconfidence: **overoptimism** about one's own sales growth, defined as consistently inflated predictions, and **overprecision**, defined as excessively narrow subjective confidence intervals, (Sharot, 2011; Moore and Healy, 2008; Landier and Thesmar, 2009). Our experimental design enables us to measure both biases at high frequency and identify interventions that can mitigate them (Ludwig et al., 2011; Congdon et al., 2017). We then develop a new methodology to quantify the welfare costs and potential motivational benefits of biased entrepreneurial expectations (Benabou and Tirole, 2002).

Our study follows approximately 1,000 entrepreneurs in Utah over a 13-month period, collecting monthly forecasts, realized outcomes, and survey measures of their beliefs. These are small, established firms. The median firm has two employees, excluding the founder, and a median age of seven years. A majority of founders (61%) report that they explicitly aspire to "profit maximization and growth." This is a notably higher share than in national benchmarks, such as the 40% in the US Census Bureau's 2016 Annual Survey of Entrepreneurs (ASE), 24% in the Panel Study of Entrepreneurial Dynamics (PSED) (Hurst and Pugsley, 2011), and 12% of nascent entrepreneurs considering starting a business because of a business opportunity as reported in the working paper version of Bennett and Chatterjee (2019).

We document high and persistent degrees of entrepreneurial overoptimism and overprecision in our sample. Entrepreneurs in the control group overestimate their next-month revenue growth by about 5%: an error that compounds to nearly 80% annually. This bias cannot be explained by Bayesian "apparent overconfidence" (Benoit and Dubra, 2011) or by persistent private information.[1] Experience does not eliminate this bias: among entrepreneurs with firms that are at least 7 years old, the median monthly overoptimism remains at 4.6%. This persistent overoptimism contrasts sharply with the behavior of large public firms, which show no such bias (Barrero, 2022, using the Survey of Business Expectations). Both sets of findings are consistent with Busenitz and Barney (1997), who show that entrepreneurs tend to be more overconfident than managers of large firms. Entrepreneurs are also systematically overprecise, that is, overconfident about the accuracy of their forecasts. We asked entrepreneurs to report 80% confidence intervals for their revenue growth. These entrepreneurs reported 80% confidence intervals that are 23.4 percentage points narrower than statistical 80% confidence intervals, based on their realized revenue growth. This overprecision by entrepreneurs is smaller but comparable to the 27.7 percentage point overprecision reported by Ben-David et al. (2012) for CFOs of major corporations. Like overoptimism, overprecision also persists with experience.

Turning to mechanisms that may sustain this persistent overconfidence, we document that entrepreneurs systematically misremember their past forecast errors. Entrepreneurs in the control group report a median past forecast error of zero, despite actual systematic bias in forecasts. We find a link between this biased memory and overoptimism. Those who

[1] See Appendix A.1 for our discussion of apparent overconfidence as in Benoit and Dubra (2011) and Appendix A.4 for our discussion of persistent private information.

"remember" fewer mistakes are also the ones who have the highest degrees of overoptimism. This pattern of selective memory is complementary to studies such as Zimmermann (2020) and Huffman, Raymond, and Shvets (2022) that show biased memory sustains overplacement, defined as overconfidence about one's own rank relative to peers.

To provide credible causal evidence on the mechanisms driving entrepreneurial overconfidence, we randomly assign entrepreneurs to one of three groups. Firms remain in their group throughout the study, are unaware of the other groups and get repeatedly treated every month. We have a control group, an *error-reminder* treatment, and a *scientific learning* treatment. The error-reminder treatment targets biased memory by presenting entrepreneurs with information about their past revenue forecasts. The scientific learning treatment targets misattribution by prompting entrepreneurs to develop and test hypotheses in a structured manner once a month, promoting scientific learning.

Misattribution may sustain overconfidence when entrepreneurs blame poor performance on bad luck rather than flawed judgment, allowing entrepreneurs to ignore negative feedback. Our second treatment encourages entrepreneurs to pay attention to negative feedback by inducing them to pre-specify and test hypotheses (Caplin and Leahy, 2019; Gagnon-Bartsch et al., 2021). For example, an entrepreneur might ignore lower-than-expected sales growth by hoping that "word of mouth" might eventually raise sales. Our scientific learning treatment encourages the entrepreneur to test the assumptions behind this hypothesis. In this case, the key assumption is that there exists a segment of customers who would purchase if they only knew about the entrepreneur's offerings. This assumption can then be tested via targeted advertising, and failure of such ads to increase demand focuses attention on the fact that the unobservable external factor of "word of mouth" is unlikely to raise sales. This mechanism directs attention toward negative feedback and facilitates Bayesian updating (Gagnon-Bartsch et al., 2021).

Our randomized control trial yields four main results. First, reminders of past errors fail to reduce overconfidence because entrepreneurs reinterpret evidence to preserve optimistic beliefs. Instead, entrepreneurs reinterpret the feedback. When faced with evidence of prior mistakes, they increasingly attribute shortfalls to external factors—a pattern consistent with "reality constraints" of wishful thinking (Caplin and Leahy, 2019). This empirical result is consistent with the ineffectiveness of reminders of long sales histories by Bloom et al. (2025) to reduce overoptimism in a large sample of internet entrepreneurs. We go further than Bloom et al. (2025) and provide evidence on why this is the case: entrepreneurs replace hindsight bias with misattribution to psychologically sustain overoptimism. Specifically, while misattribution is negatively correlated with overoptimism in the control group, it becomes significantly positively correlated with overoptimism in the error reminder treatment group. At the same time, entrepreneurs exert psychological effort to sustain overoptimism via more use of misattribution in the face of objective information about past forecast errors. This behavior is consistent with theories of motivated reasoning developed by Kunda (1990), Benabou and Tirole (2016), and Caplin and Leahy (2019).

Second, structured "scientific learning" practices can debias beliefs when entrepreneurs engage with them. The scientific learning treatment unfolds in two stages: a theory stage that asks founders to articulate testable hypotheses about their business, and a testing stage that guides them to confront those hypotheses with data. Entrepreneurs choose how much to engage with each stage.[2]

These two stages will affect overoptimism and overprecision differently. The theory stage is designed to emphasize the uncertainty of their theory and assumptions, through making implicit assumptions explicit. Our prior is that this will motivate entrepreneurs to pay attention to empirical tests of their hypotheses, consistent with Gagnon-Bartsch et al. (2021) and we expect it to reduce overprecision. At the same time, to help entrepreneurs differentiate competitively, our treatment emphasizes "contrarian views," which places more emphasis on potentially overconfident priors. This can increase overoptimism (Bernardo and Welch,

[2]To provide consistent causal estimates, we use our treatments as instruments for the endogenous variable of engagement, which is measured by the string length of free-form text responses to structured scientific learning questions, see Angrist et al. (1996), Angrist and Pischke (2009), Gerber and Green (2012).

2004). In contrast, the hypothesis testing stage directs attention to empirical implications of an entrepreneur's business model (Gagnon-Bartsch et al., 2021) and induces entrepreneurs to pay attention to both positive and negative feedback. Hypothesis testing, thereby, is likely to reduce overoptimism by directing attention equally to positive and negative feedback.

Consistent with these mechanisms, we find that entrepreneurs more strongly engaged with the theory stage exhibit more overoptimism and less overprecision. In addition, entrepreneurs who are more strongly engaged in hypothesis testing tend to reduce their overoptimism bias. Overall, these results suggest that entrepreneurial overconfidence is not a fixed trait but can be influenced and reduced by structured practices. We provide a battery of robustness checks for these main results. We confirm external validity of our causal estimates for the population of US entrepreneurs, using methods suggested by Andrews and Oster (2019). Our results are also robust with respect to incentive pay for accurate forecasts, sample attrition, different types of forecasts, hybrid entrepreneurship, or differential industry trends.

Third, these practices yield large profit gains for entrepreneurs whose primary goal is profit maximization. For these opportunity-driven entrepreneurs, monthly earnings rise from $300 monthly at the 15th percentile to over $45,000 monthly at the 85th percentile. These profit results mirror large effects on revenue found by training programs based on scientific learning, such as Camuffo et al. (2020). At the same time, we find no significant effects for entrepreneurs who do not pursue profit maximization and growth, including those with non-pecuniary main objectives, such as personal or social goals (Hurst and Pugsley, 2011). These results are consistent with the typically observed zero or insignificant effects in small business training programs (Lerner, 2009; Fairlie et al., 2015; McKenzie, 2021). Overall, our profit results suggest that identifying opportunity-driven entrepreneurs is key to the success of entrepreneurial training programs or subsidies (Hurst and Pugsley, 2011; Fairlie and Fossen, 2019), and reinforce the finding that interventions can be highly effective in boosting high-growth entrepreneurship, as shown by McKenzie (2017). This evidence suggests that debiasing efforts are most effective when incentives reinforce learning.

Fourth, overconfidence lowers welfare by inducing excessive labor supply, implying that entrepreneurs' apparent success can mask welfare losses. We use our experimental findings to develop a new methodology for assessing the welfare effects of entrepreneurial overconfidence. Our framework uses as its foundation the hourly labor supply decision. From this, we construct entrepreneur-specific measures of the marginal profit of entrepreneurial labor as the present value of the expected rational marginal benefit from working more hours, minus the opportunity cost of time. Using our data on forecast errors as well as actual profits, we correct for biased expectations and allow for a motivating effect of overconfidence, as in Benabou and Tirole (2002) and Compte and Postlewaite (2004).

Entrepreneurial welfare could increase by roughly 30% of median monthly profits, about $1,400 per month, if the median level of overoptimism were removed. This welfare gain arises because many entrepreneurs work too much, driven by overly optimistic expectations. The median entrepreneur perceives an expected marginal profit of about $3 per hour, whereas our bias-corrected rational benchmark implies a median loss of roughly $70 per hour. These findings align with evidence from Gish et al. (2019), who show that sleep deprivation impairs entrepreneurial decision-making. Our estimates represent a lower bound on the welfare costs of overconfidence because labor supply is only one of several margins, alongside hiring and investment, through which overconfidence distorts behavior.

This welfare analysis has broader implications for understanding market selection and the survival of entrepreneurial firms. Overconfident entrepreneurs are not competed out of the market; instead, their excessive labor supply inflates accounting profits and increases their likelihood of survival. Yet these additional hours yield little true economic value, as in Gish et al. (2019). Each extra hour of work generates positive accounting profits but negative marginal welfare once the opportunity cost of time is considered. As a result, overconfident entrepreneurs appear successful in the market even as they experience welfare losses. This mechanism helps explain the persistence of entrepreneurship despite low average returns documented by Hamilton (2000), Moskowitz and Vissing-Jorgensen (2002), and Hall and Woodward (2010). We add to this literature by quantifying how overconfidence affects entrepreneurial welfare. Most existing work either does not separate overconfidence from non-pecuniary rewards such as work flexibility or control (as in the studies by Hamilton, 2000; Moskowitz and Vissing-Jorgensen, 2002; Hall and Woodward, 2010) or does not offer a quantitative evaluation of the importance of overconfidence (Astebro et al., 2014). Our findings highlight a paradox of market selection: overconfidence helps entrepreneurship survive, even as it undermines entrepreneurs' welfare.

This paper makes three contributions. First, it provides direct experimental evidence on the psychological mechanisms that keep entrepreneurs overconfident, even when they face clear feedback and strong incentives. Second, we identify how structured scientific learning practices can causally reduce overconfidence, linking behavioral mechanisms to scalable managerial interventions. Third, we develop a new empirical framework to quantify the welfare implications of overconfidence through entrepreneurs' labor supply decisions, showing that excessive effort can increase survival while lowering welfare. Together, these results move beyond documenting behavioral biases to explaining their persistence, their responsiveness to targeted interventions, and their economic consequences.

2 Relation to Literature

Our study is related to at least three strands of the literature in entrepreneurship and economics. First, overconfidence and overoptimism have long been argued to be drivers of entrepreneurial entry decisions and explain persistence in entrepreneurship despite low returns and high idiosyncratic risk (for detailed surveys, see Astebro et al. (2014) and Zhang and Cueto (2017)). However, persistence of overoptimism has received much less attention than initial overoptimism, leading to excessive entry (Camerer and Lovallo, 1999). The economics literature has focused on persistence of "overplacement" (overconfidence of one's own rank relative to others) despite frequent feedback, both in the lab (Zimmermann (2020)) and for middle managers (Huffman, Raymond, and Shvets (2022)). Within empirical work on entrepreneurship, both Landier and Thesmar (2009) and Herbert (2023) have shown that a large sample of French entrepreneurs exhibits persistent overoptimism in a way similar to the sample of entrepreneurs we follow (see Appendix A.3). More broadly, our work is also related to recent work on treatments to address behavioral biases in entrepreneurship, either in the context of sales forecasting (Bloom et al., 2025) or technology adoption Gertler et al. (2022). However, none of these studies have analyzed practices that may reduce managerial or entrepreneurial overconfidence by addressing biased learning as a root cause. Our main contribution is therefore the conduct of a field experiment to address biased learning as the root cause of entrepreneurial overconfidence.

Second, our scientific learning treatment builds on the recent literature on structured practices, such as Bloom and Van Reenen (2007) and Bloom et al. (2021), and especially structured scientific approaches to managerial (Yang et al., 2025) and entrepreneurial decision-making, see Felin and Zenger (2009), Camuffo et al. (2020), Coali et al. (2022). The study closest to ours is the paper by Coali et al. (2022), which randomized a multiple-week entrepreneurship training program on scientific learning for early-stage startups. This focus on early-stage startups naturally prevents studies such as Camuffo et al. (2020) and Coali et al. (2022) from analyzing the persistence of entrepreneurial overconfidence. The sample in Camuffo et al. (2020) and Coali et al. (2022) entirely consists of early-stage startups, which prevents this work from analyzing the persistence of overconfidence in a similar fashion to this study. Beyond differences in the sample of firms (early startups vs relatively mature entrepreneurial firms) and treatment types (training sessions vs repeated structured nudges), Coali et al. (2022) mainly provide indirect evidence on a debiasing effect from scientific learning. They identify a debiasing effect by imposing a Heckman sample selection model and assume that any debiasing effect is time-varying, while the learning effects of scientific learning are constant. In contrast, we directly measure overestimation and overprecision as well as the potential sustaining mechanisms of hindsight bias and misattribution.

Third, related to the explosion of interest in entrepreneurial experimentation and learning (Ries (2011), Kerr and Nanda (2010), Kerr et al. (2014), Konings et al. (2022)), a recent literature on "Bayesian entrepreneurship" has emerged (Agrawal et al. (2024), Nanda (2024)), which mostly assumes that entrepreneurs learn in a costless and unbiased way, even if they may be initially overoptimistic (Agrawal et al. (2024)). Our work is more consistent with empirical evidence that entrepreneurs are systematically biased, even if they sometimes learn from their mistakes (Howell (2021)). At the same time, we show that it is not necessary for entrepreneurs to be completely unresponsive to feedback, but that persistent overoptimism can emerge if entrepreneurs learn less from negative than positive feedback, consistent with laboratory evidence by Amore et al. (2021). More broadly, theoretical work on Bayesian entrepreneurship typically assumes that Bayesian learning is not just normatively desirable, but positively accurate to describe actual entrepreneurs (Agrawal et al. (2024), Nanda (2024)). While we are sympathetic to the normative goal of unbiased Bayesian learning, empirical studies including evidence documented in this paper suggest that entrepreneurs are not naturally unbiased Bayesian learners. At the same time, our field experiment suggests that scientific learning practices can help to make entrepreneurs more Bayesian.

3 Firm Setting and Recruiting

We conducted the study between December 2020 and March 2022, with core data collection and treatments running from March 2021 to March 2022. The project was implemented in partnership with the Utah State Governor's Office, the Legislature, and the State Chamber of Commerce, which facilitated broad participation by local entrepreneurs.

Utah offers a diverse economic base across manufacturing, retail, health care, and technology sectors (Benway, 2020), enabling a broad sample of entrepreneurs.[3] Although the study coincided with the COVID-19 pandemic, aggregate business uncertainty had largely

[3] This contrasts with studies such as Bloom et al. (2025), which focus on technology or e-commerce companies, while in our study, the median firm reports no sales from e-commerce.

stabilized by early 2021 (Meyer et al., 2022).[4] All specifications include month fixed effects to absorb residual time variation, and robustness checks further include industry-by-month fixed effects.

3.1 Pilot Survey and Recruiting

Recruitment proceeded in two stages. In the first stage, we conducted a large pilot survey to all registered entrepreneurs in the State of Utah to collect baseline firm characteristics and identify those willing to participate in a year-long panel. Participating entrepreneurs were given an opportunity to win a prize including ten $1,000 gift cards, see Appendix A.2 for more details.[5] Our partnership with the state government, chamber of commerce, and incentives reduced sample selection and increased external validity, see section 7.1. Around 10,000 entrepreneurs completed our pilot survey and we verified business legitimacy through web or physical presence. About 4,000 firms agreed to be re-contacted, from which we drew a target sample of 1,000 entrepreneurs for the main experiment. To mitigate attrition, participants received small monetary incentives and forecast-accuracy bonuses during the study period.

In the second recruiting stage, we re-contacted interested entrepreneurs. We offered a $20 Amazon gift card for each completed survey, as well as an additional $50 for completing 6 surveys. Additionally, from October 2021 until March 2022, we offered a $5 bonus if participants forecasted their 4-week revenue growth within 5% of actual revenue growth during that time period, which effectively also reduced attrition.

[4] Meyer et al. (2022) report that 3 out of 5 measures of uncertainty had returned to their pre-pandemic levels by early 2021, which is when our study began. This is likely to be the consequence of the widespread availability of COVID-19 vaccines, which started to be rolled out in early 2021.

[5] Additionally, a local NBC affiliate ran a news segment on the evening of December 1, 2020, confirming that our survey was indeed legitimate, see Appendix A.2.

3.2 Sample characteristics

The final sample includes 1,067 firms spanning a wide range of industries as shown in Figure 1, including health care, retail, manufacturing, and information technology. The distribution of firm revenue, shown in the bottom panel of Figure 1, is roughly log normal with most firms being small to medium-sized. In Table 1 we show the median firm has monthly revenues of about \$15,000, while the average firm has much larger revenues at \$144,000. The median firm in our sample has 2 employees and is 7 years old, which confirms that most firms in our sample have already learned whether their business is viable, see Kerr and Nanda (2010), Haltiwanger et al. (2013).

Most firms in our sample have profit and growth as a goal. Rows 4-6 of Table 1 display long-term business goals of entrepreneurs to analyze how important non-pecuniary motives for running a business are in our sample (Hamilton, 2000; Hurst and Pugsley, 2011). Table 1 shows that only 12% of firms in our sample pursue non-pecuniary motives, such as "Personal or social goals other than profit or growth." In contrast, 61% elect "Profit maximization and growth" as their main goal.

4 Measurement and Documenting Biases

4.1 Measurement of main outcomes

Our main outcomes are forecast errors for monthly revenue growth. This requires us to measure growth forecasts and realized revenues each month.

4.1.1 Revenue Growth

We ask businesses to report their revenues over the last 4 weeks and use this data to construct realized monthly sales growth. This is necessary in our setup, as administrative data collected by the government is not accessible to us and would mostly not provide sales information on a monthly frequency. Anecdotal evidence from our sample firms suggests that the businesses used their own accounting books to provide us with these revenue numbers. For example, one entrepreneur wrote: "I set aside 30 minutes or so and pull out my financial data and start to work."[6,7]

Our use of entrepreneurial revenue growth also has the advantage of being robust to permanent misreporting at the individual level. For example, suppose for tax-purposes entrepreneurs under-report revenues $X_{i,t}$ by a constant fraction $u_i > 0$ as in Hurst et al. (2014), so reported revenues are $\tilde{X}_{i,t} = (1-u_i) \cdot X_{i,t}$. This under-reporting will be "differenced out" by considering revenue growth.[8]

4.1.2 Forecasts

Figure 2 displays the survey screen we use to elicit monthly growth forecasts. We ask respondents to forecast revenues over the next "four weeks" and to provide upper and lower confidence bounds for this forecast. Importantly, we verify that respondents' best forecast about revenues corresponds to their business's growth goals and ask firms to report business goals in case the two differ. We use business goals as our baseline a measure of growth forecasts, since businesses naturally have an incentive to generate accurate business growth goals.[9]

Appendix A.3 provides a more detailed validation analysis of the sales growth forecast data. There, we show that forecasts predict sales growth, as well as evidence consistent with the view that entrepreneurial expectations are close to rational, but for persistent overoptimism.

[6] This suggests that looking up data from accounting records was easier for most entrepreneurs than misreporting such data and that survey participants had less incentive to misreport earnings on our survey than e.g. on tax forms, as discussed by Hurst et al. (2014).

[7] Additionally, although respondents did sometimes not exactly respond in 4 week intervals to our invitations to fill out the survey, the median time between responses to two subsequent surveys is 31 days in our sample. In Appendix A.12, we show that all of our core results are robust to normalizing sales growth rates to a 28-day period.

[8] To see that, note that $g_{i,t+1} = \frac{\tilde{X}_{i,t+1} - \tilde{X}_{i,t}}{\tilde{X}_{i,t}} = \frac{(1-u_i) \cdot [X_{i,t+1} - X_{i,t}]}{(1-u_i) \cdot X_{i,t}} = \frac{[X_{i,t+1} - X_{i,t}]}{X_{i,t}}$

[9] Section 6 analyzes robustness of our main results to this choice.

4.2 Types of Overconfidence

We distinguish different types of overconfidence. Overoptimism refers to systematically inflated expectations about the growth of one's own business and is measured by positive average forecast errors (Sharot, 2011; Landier and Thesmar, 2009). Overprecision refers to excessively narrow subjective confidence intervals, see also Moore et al. (2015).

Healy and Moore (2007) discuss two additional and distinct types of overconfidence. Overestimation is overconfidence about "actual ability, performance, level of control, or chance of success,"[10] while overplacement refers to overconfidence about the own rank relative to peers. We do not analyze these types of overconfidence, since overestimation has been shown to be strongly context dependent (Healy and Moore, 2007) and overplacement is subject to difficulties separating true overconfidence from Bayesian updating as pointed out by Benoit and Dubra (2011).[11]

Entrepreneurs systematically overestimate their short-term revenue growth. Panel (A) of Figure 3 shows the distribution of forecast errors in solid blue. The median forecast error for the entrepreneurs in the control group is 5% (before October 2021). This represents a significant forecast error, indicating an annual overoptimism of sales growth by nearly 80%. Among firms at least seven years old, the median error remains +4.6 percent, despite these firms representing the most experienced 20 percent of startups in the U.S. (Fairlie and Miranda, 2017). Appendix A.4 confirms that these patterns are robust to allowing for persistent private information under Bayesian updating.

To measure overprecision, let $P_{x,i}$ denote the realized xth percentile of monthly growth for entrepreneur i, and $P^f_{x,i,t}$ the corresponding subjective percentile reported at month t. To facilitate interpretation, but not estimation, of our results, we can approximate the monthly

[10] Don Moore distinguishes overestimation as being about own ability, while "optimism is distinctively about a forecast for the future, and whether you think good things or bad things are going to happen to you," see jasoncollins.blog/posts/the-three-faces-of-overconfidence.

[11] We show in Appendix A.1, that this issue does not apply to measures of forecast errors, which we use to measure overoptimism. We also note that Benoit and Dubra (2011) explictly acknowledge that their analysis does not apply to overprecision, as they state that "Our analysis (...) is not directly applicable to overconfidence in the precision of estimates."

volatility growth rates $\sigma_{g,i} \approx \frac{P_{90,i}-P_{10,i}}{2.65}$ and $\sigma^f_{g,i,t} \approx \frac{P^f_{90,i,t}-P^f_{10,i,t}}{2.65}$, under the assumption of normal distribution of growth rates. The degree of overprecision (or precision error) can therefore be defined as

$$\begin{aligned} \omega_{i,t} &= \sigma_{g,i} - \sigma^f_{g,i,t} \\ &= \left(\frac{1}{2.65}\right) \cdot \left[P_{90,i} - P_{10,i} - \left(P^f_{90,i,t} - P^f_{10,i,t}\right)\right]. \end{aligned} \tag{1}$$

The vast majority of entrepreneurs in our sample exhibit overprecision. The distribution of entrepreneurial overprecision for the control group before the introduction of incentives for accurate forecasting is displayed in Panel (B) of Figure 3. This figure shows that the stated confidence intervals of entrepreneurs' monthly growth forecasts are much narrower than the dispersion of the growth outcomes. The median precision error is 23.4 percentage points and is comparable to the 27.7 percentage point overprecision error reported by Ben-David et al. (2012) for CFOs of public corporations.[12]

4.3 Mechanisms sustaining overconfidence: theory and measurement

To understand the mechanisms that might sustain overconfidence, we begin with the baseline of Bayesian updating. An entrepreneur i seeks to forecast the growth rate of her business $g_{i,t+1}$, which is driven by a venture-specific, permanent effect g_i and iid random noise $\nu_{i,t+1} \sim N(0, \sigma^2_\nu)$, so that

$$g_{i,t+1} = g_i + \nu_{i,t+1}. \tag{2}$$

To forecast growth rate $g_{i,t+1}$, the entrepreneur forms an expectation, denoted by $\mu_{i,t+1}$, based on a prior $\mu_{i,0} \sim N(\mu_0, \sigma^2_{\mu,0})$, current sales growth $g_{i,t}$ and for simplicity, we assume

[12] Our findings on precision error are also consistent with separate literatures in psychology and economics, which document the robustness of overprecision. Furthermore, various field studies in economics and finance document the presence of overprecision for large firms (Altig et al. (2020), Barrero (2022)) and CFOs of large public companies (Ben-David et al. (2012), Boutros et al. (2020)).

that the variances are known. The optimal Bayesian forecast for sales $\mu_{i,t+1}$, is given by

$$\mu_{i,t+1} = \alpha_t \cdot g_{i,t} + (1 - \alpha_t) \cdot \mu_{i,t}, \tag{3}$$

with $\alpha_t \in [0, 1]$[13] as the weight placed on feedback or the observed growth rate, rather than the prior. Under fully Bayesian learning, $\alpha_t \to 1$ for $t \to \infty$, so that no weight is placed on priors in the limit. To allow for learning that is distorted by behavioral biases, we first assume that the weight α_t does not converge to 1, setting $\alpha \in (0, 1)$ to a constant value. However, even with this assumption, the entrepreneur's expectations will be unbiased in steady state, i.e., $E[\mu_{i,t} - g_{i,t}] = 0$.[14]

4.3.1 Biased memory

The first mechanism sustaining overoptimism is biased memory, which has been theoretically related to overconfidence by Benabou and Tirole (2002).[15] We focus on hindsight bias, in which recalled forecast errors are biased toward zero. In the context of (3), this is a case where $\alpha = 0$, so that feedback is completely ignored by the entrepreneur. The entrepreneur keeps their initial forecast, which we assume is overoptimistic, so that $\mu_{i,0} > E[g_{i,t}]$. The resulting steady-state forecast error is

$$E[\mu_{i,t} - g_{i,t}] = (\mu_{i,0} - E[g_{i,t}]) \cdot 1_{\{\alpha=0\}}. \tag{4}$$

Hindsight bias sustains overconfidence by allowing individuals to disregard past forecast errors, since they recall little or no evidence of having been wrong.

[13] The exact expression for $\alpha_t = \frac{t/\sigma_\nu^2}{t/\sigma_\nu^2 + \sigma_{\mu,0}^2}$

[14] To see this point, subtract future growth $g_{i,t+1}$ from both sides of (3), substituting in for $g_{i,t+1}$ from (2) on the left hand side and rearranging gives the error correction form: $(\mu_{i,t+1} - g_{i,t+1}) - (\mu_{i,t} - g_{i,t}) = \alpha \cdot (g_{i,t} - \mu_{i,t}) - (\epsilon_{i,t+1} - \epsilon_{i,t})$. To derive the long-run forecast error, take expectations of the error correction form, impose the steady state condition $E[\mu_{i,t+1} - g_{i,t+1}] = E[\mu_{i,t} - g_{i,t}]$, and solve for the steady-state forecast error $E[\mu_{i,t} - g_{i,t}]$.

[15] Additionally, previous empirical work by Zimmermann (2020) and Huffman, Raymond, and Shvets (2022) has documented the connection of overconfidence and biased memory, albeit in the context of overplacement and not for overoptimism or overprecision.

To document the presence of biased memory in the control group, we ask participants to provide us with an estimate of their forecast error for the past month.[16] The distribution of recalled forecast errors is shown in Panel (C) of Figure 3. The median recalled forecast error is zero, consistent with hindsight bias. More evidence on the direct link between biased memory and overconfidence is provided in Table 2. There, we measure biased memory as the absolute value of the recalled forecast error for the last month for the control group. The main finding of Table 2 is that lower absolute values of recalled forecast error (more hindsight bias) are correlated with more overoptimism. In other words, more hindsight bias and more overoptimism are linked at the individual level in the same way that biased memory and overplacement are linked in Zimmermann (2020) and Huffman, Raymond, and Shvets (2022).

4.3.2 Misattribution

A second mechanism that might sustain overconfidence is causal misattribution (henceforth "misattribution"). For our purposes, we define this mechanism as ignoring negative feedback. Formally, define positive feedback as $(g_{i,t} - \mu_{i,t})^+ = \max\{0, g_{i,t} - \mu_{i,t}\}$ and negative feedback as $(g_{i,t} - \mu_{i,t})^- = \min\{0, g_{i,t} - \mu_{i,t}\}$. Under misattribution, the entrepreneur places more weight on positive as opposed to negative feedback, or $\alpha_P > \alpha_N$. The long-run forecast error is then given by

$$E\left[\mu_{i,t} - g_{i,t}\right] = \left(\frac{\alpha_P - \alpha_N}{\alpha_N}\right) \cdot E\left[(g_{i,t} - \mu_{i,t})^+\right] \cdot 1_{\{\alpha_P \neq \alpha_N, \alpha_N > 0\}} \tag{5}$$

with $(g_{i,t} - \mu_{i,t})^+ \geq 0$ by definition.[17] Equation (5) shows that under misattribution, entrepreneurs will exhibit overoptimistic forecasts in steady state, as they underweight neg-

[16] We also experimented with giving control group members their realized growth and asking them to recall their forecasted growth. This is a measurement approach that has been used in psychology to show that subjects are often unable to correctly recall ex ante expectations, once ex post results are known, see Kahneman (2011), Chapter 9.

[17] To derive the long-run forecast error, we first begin with the error correction form $E\left[(\mu_{i,t+1} - g_{i,t+1}) - (\mu_{i,t} - g_{i,t})\right] = \alpha_P \cdot E\left[g_{i,t} - \mu_{i,t})^+\right] + \alpha_N \cdot E\left[(g_{i,t} - \mu_{i,t})^-\right]$, then impose $E\left[\mu_{i,t+1} - g_{i,t+1}\right] = E\left[\mu_{i,t} - g_{i,t}\right]$ and rearrange and solve for $E\left[\mu_{i,t} - g_{i,t}\right]$.

ative feedback. Importantly, under misattribution, steady-state overoptimism can emerge even without initially overoptimistic priors, as argued by Gervais and Odean (2001).

To document the presence of misattribution, we ask a follow-up question to information about past forecast errors. In the control group, we ask entrepreneurs to recall their forecast errors in the past month. As we discuss in more detail below, for the treatment groups, we report the past forecast errors directly. For all participants, we ask respondents to provide a justification for these forecast errors. In particular, for the control group, the survey screen displays the following question:

> "You indicated that you missed your expected revenue growth during the past four weeks by "X" percent. What is the most likely reason for this miss?"

We provided two checkboxes with a text entry: (1) "Reasons internal to the company (please specify)" and (2) "Reasons external to the company (please specify)." Our measure of misattribution begins with focusing on firms that blame external factors for underperformance (or overoptimism). Since it is possible that indeed external factors led to a surprising underperformance, we then calculate the median forecast error in the same industry (2-digit NAICS) for the same time. If the median firm outperformed its forecast while the focal firm underperformed by blamed external forces, we classify this as misattribution. It should be noted that this is a very conservative measure of misattribution and we will only be able to capture some very extreme cases of misattribution. This is made clear by the fact that only about 3% of observations in the control group exhibit this misattribution. Nevertheless, it turns out that this measure is very informative about the psychological mechanism that help sustain overconfidence.

5 Experimental Design

The interventions are designed as light-touch but persistent nudges. Treatment assignment was randomized once in March 2021 and remained fixed for the duration of the study.

Entrepreneurs received monthly prompts over 13 consecutive months.

5.1 Error Reminder Treatment

Our first treatment targets biased memory by reminding entrepreneurs of their past forecast error. Specifically, instead of being asked to recall the forecast error of the last month, we display the following text:

> In the last survey, you predicted that your revenue growth would g_P% over the coming four weeks. Based on your reported revenue for these four weeks $\$X_1$ and the revenue you reported in last month's survey (which is $\$X_0$), your revenue growth for these four weeks was $g_A = \frac{\$X_1 - \$X_0}{\$X_0}$. This implies a forecast error of $g_P - g_A$%. What is the most likely reason for your deviation from your goal?

with two checkboxes with a text entry: (1) "Reasons internal to the company (please specify)" and (2) "Reasons external to the company (please specify)." This treatment has the goal to directly replace biased memory with correct information about the last forecast error. The section containing questions about the forecast for the upcoming month immediately follows the treatments, to ensure that subjects have the past forecast error in mind when they make their further predictions.

In terms of the model in Section 4.3.1, the treatment corresponds to the case of $\alpha > 0$, while the control group can be formalized as $\alpha = 0$, so that the theoretical prediction of the treatment effect for the error reminder treatment is

$$E\left[\mu_{i,t} - g_{i,t} \middle| \alpha > 0\right] - E\left[\mu_{i,t} - g_{i,t} \middle| \alpha = 0\right] = -\left(\mu_{i,0} - E[g_{i,t}]\right) < 0. \tag{6}$$

In other words, the error reminder treatment is predicted to reduce overoptimism.

5.2 Scientific Learning Treatment

The scientific learning treatment builds on the error-reminder intervention by adding structured steps that target misattribution and encourage attention to negative feedback. Molden and Higgins (2012) point out that motivated reasoning does "not so much lead people to ignore the sometimes disappointing reality they face as it inspires them to exploit the uncertainties that exist in this reality to their favor." Reducing ambiguity about how a business operates can encourage entrepreneurs to update beliefs symmetrically by learning from both negative and positive feedback (Gagnon-Bartsch et al., 2021). In the model of Section 4.3.2, this corresponds to moving from asymmetric updating in the control group ($\alpha_P > \alpha_N$) toward equal weighting in the treatment group ($\alpha_P = \alpha_N$), which implies

$$E\left[\mu_{i,t} - g_{i,t} \middle| \alpha_P = \alpha_N\right] - E\left[\mu_{i,t} - g_{i,t} \middle| \alpha_P > \alpha_N\right] = -\left(\frac{\alpha_P - \alpha_N}{\alpha_N}\right) \cdot E\left[(g_{i,t} - \mu_{i,t})^+\right] < 0. \tag{7}$$

Thus, by promoting balanced learning from all feedback, the treatment should reduce forecast errors. Our design builds on recent work applying structured scientific learning to managerial and entrepreneurial contexts, including CEO decision-making (Lafley et al., 2012; Felin and Zenger, 2017; Yang et al., 2025), teaching students to think scientifically (Ashraf et al., 2022), and entrepreneurial experimentation (Felin and Zenger, 2009; Ries, 2011; Camuffo et al., 2020; Felin et al., 2020; Konings et al., 2022). Further implementation details are provided in the Appendix.

The scientific-learning treatment consists of three structured components. First, entrepreneurs are prompted to engage in "accuracy-driven reasoning" by thinking about their unique advantage and contrarian beliefs that may differ from conventional wisdom (Kunda, 1990). These beliefs have been shown to be helpful for innovative startups (Felin and Zenger, 2009; Lerner et al., 2012; Lafley et al., 2012). They are then encouraged to specify the main obstacle to realizing that advantage. Because emphasizing contrarian reasoning can strengthen pre-existing priors, this step may temporarily reinforce overconfident beliefs

(Bernardo and Welch, 2004). Second, entrepreneurs are encouraged to anticipate and diagnose potential failures ex ante, what Klein (2007) and Kahneman and Klein (2009) call pre-postmortem analysis. This analysis targets and counteracts overconfidence. Third, entrepreneurs are asked to report whether and how they tested the hypotheses or solutions identified earlier, reinforcing iterative scientific learning. We measure engagement by the length of participants' written responses to each treatment component.

We illustrate engagement with the scientific-learning treatment with a concrete example. To preserve the privacy of the company, we refer to the firm as "Bennett Woodworks" and its founder as "Olivia." Olivia describes her business as "high-end furniture" manufacturing and the unique idea of her business as utilizing "exotic woods to create wood art. This is an area of woodworking that isn't done by many woodworkers." Olivia identifies "The biggest problem is that the majority of customers in this market generally don't spend a lot of money for collectible products so I limit myself in this regard." Her expected monthly sales growth is $\mu_{i,t} = 5\%$, but in the last months, sales growth has been $g_{i,t} = 0\%$. Therefore her forecast error is $\mu_{i,t} - g_{i,t} = 5\%$ or a negative feedback of $g_{i,t} - \mu_{i,t} = -5\%$. Olivia might ignore this negative feedback, because she believes that demand is low, not because there are not enough interested potential customers, but because these customers do not know about Olivia's offerings yet. Once more, potential customers learn about her product through "word of mouth," she may believe sales will pick up. Guided by our treatment, she develops a plan: "use targeted advertising in order to reach a wider audience." To test the hypothesis, Olivia decides to run targeted ads for her merchandise on Facebook. This test did not generate much demand, which is inconsistent with the existence of a large pool of potential customers for her current products. This realization makes it more likely for Olivia to learn from the underperformance of her monthly sales goal.

Firms are similar in age, employees, revenue growth, and forecast error across the treatments and control group.

To conclude this section on the treatment design, we show balance tests in Table 3.

We find firms are similar in age, employees, revenue growth, and forecast error across the treatments and control group.

6 Results

6.1 Error Reminder Treatment

The error reminder treatment has a limited effect on overoptimism and overprecision. In Table 4, we report the difference in forecast error, noise, and precision error for the error reminder treatment relative to the control group. All of the estimates are relatively small in magnitude and none are statistically significant at the 5% level, though precision error is statistically significant at the 10% level. This limited response contrasts with theoretical predictions linking biased memory to overconfidence (Benabou and Tirole, 2002) and with laboratory and field evidence documenting similar mechanisms for overplacement (Zimmermann, 2020; Huffman et al., 2022). The difference likely reflects that our setting targets overoptimism and overprecision rather than overplacement, the relative-rank beliefs.

We analyze the association between misattribution and overoptimism as a potential explanation for why the error reminder treatment is ineffective. In Table 5, we find that for the control group, where biased memory persists, misattribution is negatively correlated with overoptimism. In both treatment groups, once biased memory is removed, the correlation between misattribution and overoptimism is positive and significant. This pattern suggests that entrepreneurs substitute misattribution for hindsight bias to preserve overconfident beliefs. The relationship is not mechanical: the sign reversal across groups rules out a purely statistical explanation.

Furthermore, the results in Table 5 suggest the error reminder treatment had an effect on entrepreneurs, even if this treatment did not debias them. Removing hindsight bias forced entrepreneurs to find another way to rationalize the validity of overconfident forecasts in the face of evidence for underperformance. Such behavior is consistent with a model of "motivated reasoning," in which decision-makers exert mental effort to sustain overconfidence even in the face of reminders of past errors (Caplin and Leahy, 2019). Importantly, highly deliberate people might be better at such self-delusion instead of being less likely to be biased (Kahan, 2013). In contrast, the failure of the error reminder treatment to debias entrepreneurs is only surprising from a simplistic interpretation of "System 1 biases," which are related to biased intuition and heuristics, see Kahneman (2011), Benabou and Tirole (2016). Only in a very mechanistic view of such System 1 biases, would nudges to remove biased memory successfully lead to reducing overoptimism and overprecision biases.[18]

A remaining possibility is that entrepreneurial overoptimism and overprecision are stable personality traits rather than malleable cognitive patterns. The next section provides evidence inconsistent with this view.

6.2 Scientific Learning: Access

In contrast to the error reminder treatment, the scientific learning treatment requires much more attention and effort to be effective. Because treated entrepreneurs could choose not to engage with the material, the design allows for one-sided noncompliance (Angrist and Pischke, 2009; Gerber and Green, 2012).[19] We therefore begin with an Intent-to-Treat (ITT) analysis. The effects from the ITT can best be understood as reflecting the effect of access to (or the option to engage with) scientific learning. In the following section, we instrument for engagement intensity to recover causal effects of actual participation on overconfidence.

The ITT results show that access to scientific learning increased overconfidence, consistent with it placing more emphasis on "contrarian ideas" (Bernardo and Welch, 2004). Table 6 collects our baseline results of access to scientific learning on overconfidence. The first column shows that scientific learning increases overoptimism. This result is consistent

[18] This type of model is simplistic in that the deliberative System 2 is rational, as for example in rational inattention models such as Gabaix (2014). In contrast, Kahneman (2011) provides a more sophisticated system 2 model, which is also biased and may distort information processing in a self-serving way.

[19] Non-compliance is one-sided because entrepreneurs in the control group are unable to access the Scientific Learning treatment.

with a reinforcement of overconfident priors as a result of considering "contrarian ideas," as we discussed in section 5.2.

At the same time, access to scientific learning reduces overprecision, as seen in column 3 of Table 6. While the reduction in overprecision is far from completely debiasing entrepreneurs, it does reduce overprecision by about 15% on average ($0.1524 = 3.6/23.4$). This reduction follows naturally from the treatment's emphasis on clarifying the assumptions underlying their growth plans. Specifically, we ask entrepreneurs "What would have to be true for each of the two plans you listed in the last question, to achieve your growth goal for the next month?" By prompting entrepreneurs to ask what conditions must hold for their strategies to succeed, the exercise increases awareness of uncertainty and directs attention to potential unknowns.

Finally, column 4 of Table 6 shows no systematic effect of access to scientific learning on measured misattribution. The null result likely reflects that our misattribution measure is intentionally conservative, capturing only the most extreme instances of attributing underperformance to external factors.

6.3 Scientific Learning: Engagement and Impact of Different Practices

In this section, we document evidence on the causal effects of engagement with our scientific learning treatment. Engagement is measured as string length of the free form text responses we collected with each question of the scientific learning treatment. Engagement with scientific learning is endogenous because entrepreneurs who benefit more from engagement will also tend to engage more. To address this endogeneity, we follow common practice and use the randomly assigned scientific learning treatments as an IV. On the one hand, only entrepreneurs in the treatment group will have access to scientific learning, which directly makes the scientific learning treatment a relevant instrument to predict engagement. On the other hand, the scientific learning treatment is randomly assigned, which implies that it will impact overconfidence and other firm outcomes only through scientific learning engagement and not through other channels. In other words, the IV exclusion restriction will be met. For more details, see Angrist et al. (1996), Angrist and Pischke (2009), Gerber and Green (2012). All measures of engagement are normalized to have a standard deviation of one, for ease of exposition.

Panel A of Table 7 shows the causal impact of overall engagement with scientific learning on overconfidence. We measure overall engagement as the total length of all three modules of the Scientific Learning treatment: hypothesis development, pre-postmortem and hypothesis testing. Column 1 shows that a one standard deviation increase of overall engagement with Scientific Learning increases overoptimism by 1.54 percentage points per month. At the same time, a one standard deviation higher overall engagement with Scientific Learning also reduces overprecision by 2.05 percentage points as seen in column 2 of Table 7. These patterns align closely with the ITT estimates discussed in the previous section. Next, we consider the effect of engagement with the theory module, which includes five of the seven parts of the scientific learning treatment.[20] The effects of theory engagement on forecast error and precision error are very similar to overall engagement, as can be seen in columns 3 and 4 of panel A in Table 7.

We construct relative engagement measures to contrast the distinct roles of each component. For example, to capture engagement with hypothesis testing, we take the string length of responses to the hypothesis testing question and subtract the string length of theory development module. The resulting measure then tells us how much more entrepreneurs were engaged with hypothesis testing relative to theory. Similarly, we compute pre-postmortem relative to theory as the string length of responses to the internal factors cited for the pre-postmortem and subtract the string length of the hypothesis development section.[21] All relative engagement measures are normalized to have a unit standard deviation.

[20]For more details on the modules see 5.2 and Appendix A.5.

[21]As previously mentioned, we focus on internal factors in the pre-postmortem, since they should counter the misattribution bias of blaming external forces for overoptimism or underperformance.

Overoptimism declines significantly with greater engagement with hypothesis testing relative to theory, as shown in columns 1 and 2 of Panel B in Table 7. A one standard deviation higher relative engagement with hypothesis development implies a reduction in overoptimism of 2.61 percentage points per month. Over the entire sample, the control group has a median forecast error of 4.39% per month, so the a one standard deviation increase in relative engagement with hypothesis testing reduces this bias by almost 60% (0.59 = 2.61/4.39). However, this debiasing of overoptimism goes hand in hand with an increase in overprecision. We find that precision error increases by 3.53 percentage points for every standard deviation increase in relative engagement with hypothesis testing compared to theory, as shown in column 2 of Panel B in Table 7. This result is consistent with the view that conducting empirical tests gives entrepreneurs a sense that they understand their business risks well. That perception may be accurate for the specific question they tested, but it likely does not extend to other sources of risk. As a result, they become overly confident in their own forecasts.

We fail to find significant effects on forecast error and precision error when we use the relative engagement with pre-postmoterem engagement, as shown in the last two columns of Panel B in Table 7. This result indicates that the pre-postmortem components of Scientific Learning do not independently affect overconfidence beyond the influence of theory development. These results also imply that the IV effects for relative hypothesis testing are not mechanical.

6.4 Profit Effects of Treatments

In this section, we begin to investigate the broader welfare consequences of our treatments by looking at their effect on profits. To do this, we first calculate profits using data on monthly revenues minus total operating costs.[22] Following Syverson (2011) we are especially interested in within-industry differences in firm performance, so we include industry fixed effects in the profit regressions. We also expect treatment effects to depend on firms' underlying objectives.

[22]We defined total operating costs as "expenses for the day-to-day running of your business, like rent or material costs."

In particular, firms that do not prioritize profit maximization may be less likely to act on growth opportunities (see Hurst and Pugsley (2011); Fairlie and Fossen (2019)). To capture this heterogeneity, we interact treatment indicators with a variable identifying firms whose main goal is profit maximization or growth, as reported in the December 2020 pilot survey.

The Scientific Learning treatment substantially increased profits among firms that report profit maximization as their primary goal, as shown in Table 8. The value of the estimated profit effects is large and heterogeneous. Compared to the control group, scientific learning treatment group firms with profit maximization goals see their monthly operating profits increase by an average of $\$101,000 (101.32 = 132.00 - 30.86)$. This is a large effect, compared to the average monthly profit of $\$130,000$ for the profit-maximizing entrepreneurs in our sample. Quantile regressions shown in columns 2-4 show that these gains are concentrated at the top of the profit distribution. At the 15^{th} percentile profit-maximizing firms gain around \$380 (0.38 = 0.51-0.13) per month, which is economically significant, even if it is not statistically significant. At the 85^{th} percentile profit-maximizing firms gain around $\$46,000$ (46.29 = 53.44- 7.14) per month. Effects become much larger and statistically significant at the 90^{th} percentile. In other words, scientific learning potentially has a huge impact on profits for already very successful entrepreneurs. Decomposing this effect further, the last two columns of Table 8 highlight that this effect is driven by increased revenue and not by cost savings.

We caution that these profits effects are not representative for the average firm in our sample but instead suggest large gains for already profitable opportunity-driven entrepreneurs. Note that Table 8 fails to find evidence for profit effects of our treatments on entrepreneurial profits on average, which is consistent with small or statistically insignificant effects of small business training programs found in the literature, see Lerner (2009), Fairlie and Fossen (2019), McKenzie (2021).

7 Robustness

7.1 External Validity

A natural question for any randomized control trial is whether its results generalize beyond the study sample. RCTs always rely on voluntary participation in the experiment, which may result in sample selection bias in the trial sample (the sum of treatment and control groups) relative to the underlying population. In our context, one concern could be that entrepreneurs who expected to benefit more from participation actively selected into the sample. However, it is worth noting that trial participants were part of a larger survey on business sentiment, which makes the RCT less salient. This makes such sample selection unlikely.

While sample selection is unlikely, we quantify it following Andrews and Oster (2019). Specifically, we find that across several characteristics, the entrepreneurs in our sample differ from those in other surveys. Panel A of Table 9 compares eight founder and firm characteristics from our data to benchmarks from the Annual Survey of Entrepreneurs (Azoulay et al., 2020) and the Kauffman Firm Survey. We use these characteristics to quantify the bias. Let TE_i denote the treatment effect for entrepreneur i, P_S the trial population, and P the population of US entrepreneurs. Under the assumption of "small selection bias," Andrews and Oster (2019) show that the RCT participation bias can be written as

$$E_{P_S}\left[TE_i\right] - E_P\left[TE_i\right] \approx \Psi\left(E_{P_S}\left[C_i\right] - E_P\left[C_i\right]\right)' \gamma \tag{8}$$

where C_i is a vector of observable characteristics,γ is a parameter vector to be estimated, and Ψ scales the role of unobservables. When $\Psi = 1$, selection arises only through observables; when $\Psi = 2$, selection on unobservables is of similar sign and magnitude.

The parameter vector γ is obtained by estimating

$$Y_i = (1 - T_i) \cdot \alpha_0 + (1 - T_i) \cdot C_i' \gamma_0 + T_i \cdot \alpha_1 + T_i \cdot C_i' \gamma_1 \tag{9}$$

where Y_i is the dependent variable of interest (such as forecast error), T_i is an indicator for being in the treatment group, and C_i denotes observable characteristics. Panel B of Table 9 reports estimates of γ_0 and γ_1, along with implied bias terms from (8) when $\Psi = 1$. Summing the bias terms across characteristics yields the total selection-on-observables effect.

Accounting for observable selection slightly increases the estimated treatment effect—from 2.70 (Table 6) to 3.06. Thus, selection on observables works against observing treatment effects. Assuming equal selection on observables and unobservables ($\Psi = 2$) raises the selection-corrected effect to 3.40. To overturn the baseline result, selection on unobservables would need to be opposite in sign and more than seven times larger in magnitude than selection on observables ($\Psi(0) = -7.73$), an implausibly strong condition.

Panel C of Table 9 repeats the RCT participation bias quantification using hypothesis testing engagement as the outcome. This variable is of interest, as it forms the basis of our discussion on the impact of engagement with scientific learning on debiasing entrepreneurial overconfidence in Section 6. Furthermore, it is well-known that IV estimators can be understood as the ratio of the direct impact of the treatment on the outcome on the one hand and the first stage estimate capturing the impact of the treatment on the endogenous variable (here: engagement) on the other hand, see (Angrist and Pischke, 2009; Gerber and Green, 2012). Panel B of Table 9 already highlights the external validity of our ITT estimate, which corresponds to the reduced form of an IV regression of testing engagement on forecast error, using the scientific learning treatment as an instrument. Panel C shows the first-stage relationship is externally valid as well, reinforcing the external validity of both the reduced form and first-stage components of the IV in Table 7.[23]

7.2 Additional Robustness

We conduct a series of robustness checks to verify that our main findings are not driven by specific design features or sample composition.

[23]It should be noted that for testing engagement, it is true that $\gamma_0 = 0$ for all values, because entrepreneurs in the control group cannot engage with scientific testing.

First, we use variation in the incentives to report accurate sales forecasts and determine whether this may bias or add noise to our estimates. When we restrict the sample to those without incentivized forecasts, we find our results do not change, as presented in Appendix A.7. Second, we consider a different measure of sales forecast based on the entrepreneur's best guess to ensure the business target was not inflated for other reasons, such as motivating employees. When we use this alternative measure, we find our main results remain unchanged, as reported in Appendix A.8. Third, as in most field experiments with repeated surveys, attrition could bias estimates if dropouts differ systematically from those who remain. Re-estimating our models using data from the first half of the experiment yields results nearly identical to those reported in the baseline as presented in Appendix A.9. Fourth, we investigate whether our findings are influenced by the presence of hybrid entrepreneurs—individuals who hold outside employment and devote a limited amount of time to their ventures. Restricting the sample to entrepreneurs working at least 35 hours per week produces similar results, as presented in Appendix A.10.

Fifth, we test for differential industry-specific trends by including a set of industry-by-time fixed effects. These fixed effects control for differences that may arise, for example, from the recovery of the COVID-19 pandemic. We find the estimates are similar when these fixed effects are included, as presented in Appendix A.11. Sixth, our estimates may contain measurement error due to the timing of respondent reports, as not all respondents l report 28 days after their last report. To correct for this, we normalize the responses using the actual time windows. We find the results are unaffected by this change, as presented in Appendix A.12.

8 Extension: Welfare Analysis

In this section, we develop a methodology to evaluate the welfare consequences of entrepreneurial overconfidence. We focus on the intensive margin of labor supply, the work hours entrepreneurs choose. This margin has been a key theoretical mechanism through which overconfidence impacts welfare, as noted by Benabou and Tirole (2002). Our question is straightforward: Does debaising an entrepreneur make them better off?

The answer is theoretically ambiguous. Overconfident entrepreneurs may work longer hours even when the marginal return to labor is negative. For these individuals, correcting beliefs and reducing work effort raise welfare. However, classic models show the reverse can also occur. Overconfidence can serve as a commitment device that prompts individuals to exert effort they would otherwise postpone due to present bias or other forms of hyperbolic discounting (Benabou and Tirole, 2002; Compte and Postlewaite, 2004). In those cases, removing overconfidence may reduce labor supply and lower welfare. In short, overconfidence can either impose costly overwork or provide motivation to induce effort to overcome other biases.

The empirical challenge is to recover an entrepreneur's true marginal product of labor from self-reported beliefs that may be distorted. We develop a practical approach that researchers can implement in field settings using a small set of targeted survey questions and a few strong but transparent assumptions. We leave for future research approaches that rely on weaker assumptions, as this is outside the scope of our work.

8.1 Theory

To fix ideas, let $\pi^e_{S,i}(h_i)$ denote the expected present value of future profits for an entrepreneur i who works h_i hours per week. The opportunity cost of these hours is given by $w_{O,i} \cdot h_i$, where $w_{O,i}$ represents the entrepreneur's hourly outside option.

The net expected profit from labor supply h_i can then be written as

$$\begin{aligned}\Pi^e_{S,i}(h_i) &= \pi^e_{S,i}(h_i) - w_{O,i} \cdot h_i \\ &= \left[\pi^e_{R,i}(h_i) - w_{O,i} \cdot h_i\right] + \epsilon_i \cdot \pi^e_{R,i},\end{aligned} \tag{10}$$

where the last line uses the notation of $\pi^e_{R,i}(h_i)$ for the rational expected present value of future profits from entrepreneurial work and $\epsilon_i = \frac{\pi^e_{S,i} - \pi^e_{R,i}}{\pi^e_{R,i}}$ denotes the profit forecast error. We use $\Pi^e_{S,i}$ to denoted the expected subjective (biased) profit net of opportunity costs of time and $\Pi^e_{R,i}(h_i) = \pi^e_{R,i}(h_i) - w_{O,i} \cdot h_i$ as the expected rational (unbiased) profit net of opportunity costs. The change in expected profit from an incremental increase in labor supply can then be approximated

$$\Pi^e_{R,i}(h_{1,i}) - \Pi^e_{R,i}(h_{0,i}) \approx \left[\frac{\partial \Pi^e_{R,i}(h_{0,i})}{\partial h_i}\right] \cdot \left(\frac{dh_i}{d\epsilon_i}\right) \cdot (\epsilon_{1,i} - \epsilon_{0,i}) \tag{11}$$

where, $\frac{dh_i}{d\epsilon_i}$ is the labor supply response to increased profit expectation errors and $(\epsilon_{1,i} - \epsilon_{0,i})$ is a change in this forecast error.

Equation (11) the resulting change in entrepreneurial welfare, defined as expected profit net of the opportunity cost of time. The critical term in (11) is the rational expected marginal profit, $\left[\frac{\partial \Pi^e_{R,i}(h_{0,i})}{\partial h_i}\right]$. If this term is positive, then increased labor supply induced by overconfidence will increase welfare, as would be the case in the theoretical models of Benabou and Tirole (2002) and Compte and Postlewaite (2004). If this term is negative, then overconfidence drives entrepreneurs to overwork, reducing welfare.

This rational marginal profit term $\left[\frac{\partial \Pi^e_{R,i}(h_{0,i})}{\partial h_i}\right]$ can be calculated as

$$\frac{\partial \Pi^e_{R,i}(h_{0,i})}{\partial h_i} = \underbrace{\left[\frac{\partial \pi^e_{S,i}(h_{0,i})}{\partial h_i} - w_{O,i}\right]}_{\substack{\text{(i) Subjective} \\ \text{Marginal Profit}}} - \left\{ \underbrace{\frac{\pi^e_{R,i}(h_{0,i})}{\partial h_i / \partial \epsilon_i}}_{\substack{\text{(ii) Motivational} \\ \text{Effect}}} + \underbrace{\frac{\partial \pi^e_{R,i}(h_{0,i})}{\partial h_i} \cdot \epsilon_i}_{\substack{\text{(iii) Biased} \\ \text{Expectations}}} \right\}. \tag{12}$$

Equation (12) provides the core measurement framework for our welfare calculations. Before turning to the identifying assumptions, it is useful to unpack the economics behind each of the three terms that make up the rational marginal profit from more entrepreneurial work.

The first term on the right-hand side of (12) is the subjective marginal profit of work. For a profit-maxiziming rational entrepreneur, this term should be zero. However, this term could be non-zero, reflecting potentially behavioral frictions such as present-bias (Benabou and Tirole, 2002) or market frictions, such as credit or liquidity constraints that prevent optimal effort adjustment.

The second and third terms, in curly brackets, are the wedge between the rational and subjective marginal profit.

The second component in (12), labeled the motivational effect, captures how miscalibrated beliefs alter the entrepreneur's labor supply decision. When entrepreneurs overestimate the returns to effort, they may work more hours than they otherwise would. overconfidence can raise welfare by counteracting underinvestment in effort due to self-control problems or short-termism. In this sense, the motivational effect formalizes the intuition from Benabou and Tirole (2002) and Compte and Postlewaite (2004): optimistic beliefs can serve as an endogenous commitment device that sustains productive effort.

The third component in (12), labeled the biased expectations term, isolates the direct welfare loss from distorted beliefs. Even when effort choices are unchanged, overconfidence inflates the perceived marginal benefit of work and thus biases welfare evaluations upward. This term quantifies this discrepancy. This term is positive and increases with more overconfident entrepreneurs, when $\epsilon_i > 0$ is larger

Together, these components clarify why overconfidence can have ambiguous welfare effects. The motivational channel operates through behavior, potentially improving outcomes by increasing effort in otherwise slack regions of the choice space. The biased expectations channel operates through perception, overstating the true welfare value of those same choices. Consider two cases. When entrepreneurs are able to set their subjective marginal profits to zero (in the absence of frictions), the wedge is likely positive for overconfident entrepreneurs. As a result, rational marginal profits $\frac{\partial \Pi^e_{R,i}(h_{0,i})}{\partial h_i}$ will be negative, implying welfare losses from overconfidence. When subjective marginal profits are sufficiently positive, rational marginal profits will be positive as well, thereby implying welfare increases from more hours worked. This exercise highlights that whether debiasing improves or worsens welfare depends on which of these forces dominates in equilibrium.

8.2 Measurement

We begin by measuring the first term in (12), the subjective marginal profit of additional hours. To elicit this term, we adapt the approach of Altig et al. (2020) to the context of expected profits from additional labor supply. Following Barrero (2022), we ask entrepreneurs to report a subjective distribution of future profits under alternative work-hour scenarios. To account for risk preferences, we subsequently elicit certainty equivalents. Entrepreneurs are presented with the following choice:

> Consider a choice between working for 10 hours that would result in the uncertain profits you reported above and being an offered a contract for a fixed profit that would require 10 hours of your labor. What is the smallest amount of fixed profits in the contract that would encourage you to accept the fixed profit option over the uncertain profit option. (Note: We are trying to understand the cost of uncertainty, please do not consider the fact that you may not be able/willing to work an additional 10 hours).

The resulting certainty-equivalent measure converts the subjective expected profit distribution into risk-adjusted expected returns to additional work.

To measure the opportunity cost of time $w_{O,i}$, we directly ask respondents to value the forgone leisure associated with 10 additional work hours:

> Suppose you need to spend 10 more hours at work this week and have to forgo this time you would otherwise spend on a non-work activity you enjoy the most. This would be spending time with your family, relaxing, gardening etc. How much would you be willing to pay to avoid working these 10 hours?

This valuation provides a personalized measure of the implicit shadow wage of leisure, allowing us to compute each entrepreneur's subjective marginal profit net of opportunity costs.

For the remaining components of (12), we need to make a number of strong assumptions.

Assumption 1. *The labor supply response* $\frac{\partial h_i}{\partial \epsilon_i}$ *can be measured as the effect of higher growth targets on hours worked, using a direct survey question.*

Assumption 1 allows us to measure $\frac{\partial h_i}{\partial \epsilon_i}$ using the following survey question:

> Suppose, one month you decide to increase your revenue growth goal, just to motivate yourself and for no other reason. You increase your revenue goal for your business over the next four weeks by an additional 5%. How many additional hours do you think you would end up working per week to meet this new goal?

Although this survey question is less ideal than estimating labor supply elasticities with respect to overestimation, it has two advantages. On the one hand, the responses are entrepreneur-specific, thereby making pooling of data across entrepreneurs unnecessary. On the other hand, the question focuses on increased revenue growth goals, irrespective of potential demand shocks or other business opportunities.[24]

The next assumption allows us to use the estimated forecast errors ξ_i from our experiment to proxy for the profit forecast error ϵ_i.

Assumption 2. *The forecast error in expected marginal profits* ϵ_i *can be measured by the forecast error in revenue growth* ξ_i.

This assumption would for example be valid in a model of monopolistic competition with a constant returns to scale production technology as in Dixit and Stiglitz (1977), in which profits are proportional to revenues. Since almost all of our entrepreneurs are small to medium sized businesses, strategic interactions among oligopolistic firms are unlikely to be relevant, which makes a monopolistic competition assumption more attractive.

[24] An alternative approach that we pursued is to use labor supply estimates from the literature. Results are available upon request

Assumption 3. *The rational flow profit term* $\pi^e_{R,i}(h_{0,i})$ *can be approximated by average daily profits and the marginal rational profit term* $\frac{\partial \pi^e_{R,i}(h_{0,i})}{\partial h_i}$ *can be approximated by hourly profits.*

This last assumption will be valid, for example under rational expectations and a constant returns to scale production technology, which are very strong assumptions but which allow us to go back and forth between average and marginal changes. See Appendix A.13 for more details.

8.3 Welfare Results

We find substantial differences between rational and subjective marginal profits. In Figure 4, we present a kernel density estimate of the subjective marginal profit term as a grey solid line and the rational marginal profit in blue dashed line, based on equation (12).

Most entrepreneurs believe they are optimizing. The subjective marginal profit distribution has a lot of mass around zero. In addition, the median entrepreneur has a subjective marginal profit of \$2.90 per hour, close to zero and denoted by the vertical red dashed line.

Most entrepreneurs work too many hours and have a negative rational marginal profit. The rational marginal profit distribution has less mass concentrated around zero marginal profits and more mass in the left tail of the distribution. In fact, the median entrepreneur has a rational marginal profit of negative \$70 per hour, denoted by the vertical black dashed line. To put this number in context, it is similar to the \$50 median opportunity cost of an hour of additional work. Furthermore, the negative median marginal profit of hours worked we find, is consistent with laboratory evidence by Gish et al. (2019), who show that sleep deprivation can cause inefficient entrepreneurial decision making, such as the pursuit of worse business opportunities.

Entrepreneurs that have positive rational marginal profit measures exhibit limited overoptimism. Specifically, the wedge between the rational and subjective marginal profits to the right of zero is close to zero as the distributions are very similar.

Table 10 shows welfare results from debiasing entrepreneurs in our sample. The median entrepreneur exhibits a forecast error of roughly five percentage points. Eliminating this bias reduces excessive labor supply and raises welfare by \$1,417 per month. To put this in context, the increase in welfare is 30% of the median monthly accounting profit of $\$4,500$ in our sample. These average gains, however, mask considerable heterogeneity. As the table illustrates, welfare effects from debiasing vary sharply across entrepreneurs, ranging from large gains among those who substantially overwork to negligible or even negative effects among those whose optimism serves a motivational role.

9 Conclusion

This study provides the first mechanism field experiment to identify how entrepreneurial overconfidence is psychologically sustained. Our results extend the behavioral economics literature on motivated beliefs and wishful thinking to the high-stakes context of entrepreneurial sales forecasting (Benabou and Tirole, 2016). We find that sustained engagement with structured management practices, specifically scientific learning and hypothesis testing, can meaningfully reduce overconfidence (Bloom and Van Reenen, 2007; Camuffo et al., 2020; Yang et al., 2025). This evidence suggests that entrepreneurial overconfidence is not a fixed character trait, but instead a result of limited adoption of structured practices.

Our findings open up several possibilities for future research. One promising direction is to study whether scientific learning mitigates other behavioral biases commonly attributed to entrepreneurs, such as loss aversion (Kahneman and Tversky, 1979), the planning fallacy (Buehler et al., 1994), or sunk-cost sensitivity. Because scientific learning explicitly confronts uncertainty, it may also provide a natural mechanism for managing ambiguity (Knight, 1921) and cognitive complexity in entrepreneurial environments. Exploring these dimensions would deepen our understanding of both entrepreneurial decision-making and the broader welfare effects of structured learning.

A second direction is the exploration of the effects of scientific learning on entrepreneurial financing. Overconfidence may influence not only labor supply and effort but also the way entrepreneurs present their ventures to investors and how they fund these ventures. Theoretical arguments by Malmendier and Tate (2005, 2015) suggest that rational investors will disproportionately increase the cost of capital in response to managerial overconfidence, thereby reducing investment in profitable opportunities. A key question, therefore, is whether scientific learning can reduce overconfidence while simultaneously improving the credibility and persuasiveness of entrepreneurs in capital markets.

Finally, although there is a broad consensus that experimentation is crucial, especially for opportunity-driven entrepreneurship (Kerr and Nanda, 2010), there are several distinct approaches to such experimentation. Our intervention builds on scientific learning approaches applied in relatively mature firms (Lafley et al., 2012; Camuffo et al., 2020; Yang et al., 2025). In contrast, early-stage entrepreneurs often adopt the Lean Startup methodology (Ries, 2011), which emphasizes rapid customer validation through "minimum viable products," rather than formal hypothesis testing (Felin et al., 2019). Like Felin et al. (2019) and Cao et al. (2020), we recognize both the promise and the pitfalls of this approach. Determining which structured practices best foster disciplined experimentation in young firms remains an open empirical question.

References

Agrawal, A., A. Camuffo, A. Gambardella, J. S. Gans, E. Scott, S. Stern. 2024. Bayesian Entrepreneurship. *MIT Working Paper* .

Altig, D., J. Barrero, N. Bloom, S. Davis, B. Meyer, N. Parker. 2020. Surveying Business Uncertainty. *Journal of Econometrics* .

Amore, D., O. Garofalo, V. Martin-Sanchez. 2021. Failing to Learn from Failure: How Optimism Impedes Entrepreneurial Innovation. *Organization Science* .

Andrews, I., E. Oster. 2019. A Simple Approximation for Evaluating External Validity Bias. *Economics Letters* **178** 58–62.

Angrist, J., G. Imbens, D. Rubin. 1996. Identification of Causal Effects Using Instrumental Variables. *Journal of the American Statistical Association* **91(434)** 444–455.

Angrist, J., J. Pischke. 2009. Mostly Harmless Econometrics. *Princeton University Press* .

Ashraf, N., A. Banerjee, V. Nourani. 2022. Learning to Teach by Learning to Learn. *MIT Working Paper* .

Astebro, T., H. Herz, R. Nanda, R. Weber. 2014. Seeking the Roots of Entrepreneurship: Insights from Behavioral Economics. *Journal of Economic Perspectives* **28(3)** 49–70.

Azoulay, P., D. Jones, D. Kim, J. Miranda. 2020. Age and High-Growth Entrepreneurship. *American Economic Review: Insights* **2(1)** 65–82.

Barrero, J. 2022. The Micro and Macro of Managerial Beliefs. *Journal of Financial Economics* .

Ben-David, I., J. Graham, C. Harvey. 2012. Managerial Miscalibration. *Quarterly Journal of Economics* **128(4)** 1547–1584.

Benabou, R., J. Tirole. 2002. Self-Confidence and Personal Motivation. *Quarterly Journal of Economics* **117(3)** 871–915.

Benabou, R., J. Tirole. 2016. Mindful Economics: The Production, Consumption, and Value of Beliefs. *Journal of Economic Perspectives* **30(3)** 141–164.

Bennett, V., R. Chatterjee. 2019. The entrepreneurial process: Evidence from a nationally representative survey. *Strategic Management Journal* 1–31.

Benoit, J., J. Dubra. 2011. Apparent Overconfidence. *Econometrica* **79(5)** 1591–1625.

Benway, D. 2020. Measuring Economic Diversity: The Hachman Index, 2018. *https://gardner.utah.edu/wp-content/uploads/Hachman-Brief-Apr2020.pdf* .

Bernardo, A., I. Welch. 2004. On the Evolution of Overconfidence and Entrepreneurs. *Journal of Economics and Management Strategy* **10(3)** 301–330.

Bloom, N., M. Cordreanu, R. Flechter. 2025. Rationalizing Entrepreneurs' Forecasts. *NBER Working Paper 33384* .

Bloom, N., T. Kawakubo, C. Meng, P. Mizen, R. Riley, T. Senga, J. Van Reenen. 2021. Do Well Managed Firms Make Better Forecasts? *NBER Working Paper 29591* .

Bloom, N., J. Van Reenen. 2007. Measuring and Explaining Management Practices Across Firms and Countries. *Quarterly Journal of Economics* **122(4)** 1351–1408.

Boutros, M., I. Ben-David, J. Graham, C. Harvey, J. Payne. 2020. The Persistence of Miscalibration. *NBER Working Paper 28010* .

Buehler, R., D. Griffin, M. Ross. 1994. Exploring the "planning fallacy": Why people underestimate their task completion times. *Journal of Personality and Social Psychology* **67(3)** 366–381.

Busenitz, L., J. Barney. 1997. Differences between entrepreneurs and managers in large organizations: Biases and Heuristics in Strategic Decision-Making. *Journal of Business Venturing* **12** 9–30.

Camerer, C., R. Hogarth. 1999. The Effects of Financial Incentives in Experiments: A Review and Capital-Labor-Production Framework. *Journal of Risk and Uncertainty* **19(1)** 7–42.

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

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

Cao, R., R. Koning, R. Nanda. 2020. Biased Sampling of Early Users and the Direction of Startup Innovation. *Working Paper, Harvard Business School* .

Caplin, A., J. Leahy. 2019. Wishful Thinking. *NBER Working Paper 25707* .

Chen, G., C. Crossland, S. Luo. 2015. Making the same mistake all over again: CEO overconfidence and corporate resistance to corrective feedback. *Strategic Management Journal* .

Coali, A., A. Gambardella, E. Novelli. 2022. Understanding Probabilistic Reasoning in Innovation. *Bocconi University Working Paper* .

Compte, O., A. Postlewaite. 2004. Confidence-Enhanced Performance. *American Economic Review* **94(5)** 1536–1557.

Congdon, W., J. Kling, J. Ludwig, S. Mullainathan. 2017. Social Policy: Mechanism Experiments and Policy Evaluations. *in: Banerjee, A. and Duflo, E. (eds) Handbook of Economic Field Experiments* **2** 389–426.

Daniel, K., D. Hirshleifer. 2015. Overconfident Investors, Predictable Returns, and Excessive Trading. *Journal of Economic Perspectives* **29(4)** 61–88.

Dixit, A., J. Stiglitz. 1977. Monopolistic Competition and Optimum Product Diversity. *American Economic Review* **67(3)** 297–308.

Enke, B., U. Gneezy, B. Hall, D. Martin, V. Nelidov, T. Offerman, J. Van de Ven. 2022. Cognitive Biases: Mistakes or Missing Stakes? *Review of Economics and Statistics* .

Fairlie, R., F. Fossen. 2019. Defining Opportunity versus Necessity Entrepreneurship: Two Components of Business Creation. *NBER Working Paper 26377* .

Fairlie, R., D. Karlan, J. Zinman. 2015. Behind the GATE Experiment: Evidence on Effects of and Rationales for Subsidized Entrepreneurship Training. *American Economic Journal: Economic Policy* **7(2)** 125–161.

Fairlie, R, J. Miranda. 2017. Taking the Leap: The Determinants of Entrepreneurs Hiring their First Employee. *Journal of Economics and Management Strategy* **26(1)** 3–34.

Feldman, M. 1987. Bayesian learning and convergence to rational expectations. *Journal of Mathematical Economics* **16(3)** 297–313.

Felin, F., A. Gambardella, S. Stern, T. Zenger. 2019. Lean Startup and the Business Model: Experimentation Revisited. *Long Range Planning* .

Felin, T., A. Gambardella, T. Zenger. 2020. Value Lab: A Tool for Entrepreneurial Strategy. *Management and Business Review* .

Felin, T., T. Zenger. 2009. Entrepreneurs as theorists: on the origins of collective beliefs and novel strategies. *Strategic Entrepreneurship Journal* **3(2)** 127–146.

Felin, T., T. Zenger. 2017. The Theory-Based View: Economic Actors as Theorists. *Strategy Science* **2(4)** 211–287.

Folta, T., F. Delmar, K. Wennberg. 2010. Hybrid Entrepreneurship. *Management Science* **56(2)** 253–269.

Friedman, M. 1953. Essays in Positive Economics. *University of Chicago Press* .

Gabaix, X. 2014. A Sparsity-Based Model of Bounded Rationality. *Quarterly Journal of Economics* **129(4)** 1661–1710.

Gagnon-Bartsch, T., M. Rabin, J. Schwartzstein. 2021. Channeled Attention and Stable Errors. *HBS Working Paper* .

Gaulin, M., N. Seegert, M. Yang, S. Alder, A. Looney. 2021. Doing Good rather than Doing Well: What Stimulates Personal Data Sharing and Why? *University of Utah Working Paper* .

Gerber, A., D. Green. 2012. Field Experiments: Design, Analysis and Interpretation. *W.W. Norton & Company* .

Gertler, P., S. Higgins, U. Malmendier, W. Ojeda. 2022. Why small firms fail to adopt profitable opportunities. Tech. rep.

Gervais, S., T. Odean. 2001. Learning to Be Overconfident. *Review of Financial Studies* **14(1)** 1–27.

Gish, J., D. Wagner, D. Gregoire, C. Barnes. 2019. Sleep and entrepreneurs' abilities to imagine and form initial beliefs about new venture ideas. *Journal of Business Ventures* **34(6)** 105943.

Hall, R., S. Woodward. 2010. The Burden of the Nondiversifiable Risk of Entrepreneurship. *American Economics Review* **100(3)** 1163–1194.

Haltiwanger, J., R. Jarmin, J. Miranda. 2013. Who Creates Jobs? Small vs. Large vs. Young. *Review of Economics and Statistics* **95(2)** 347–361.

Hamilton, B. 2000. Does entrepreneurship pay? An empirical analysis of the returns to self-employment. *Journal of Political Economy* **108(3)** 604–631.

Healy, P., D. Moore. 2007. Bayesian Overconfidence. *Working Paper, UC Berkeley* .

Herbert, C. 2023. Learning from Errors in Entrepreneurship. *Working Paper, University of Toronto* .

Hestermann, N., Y. Yaouanq. 2021. Experimentation with Self-Serving Attribution Biases. *American Economic Journal: Microeconomics* **13(3)** 198–237.

Howell, S. 2021. Learning from feedback: Evidence from new ventures. *Review of Finance* .

Huffman, D., C. Raymond, J. Shvets. 2022. Persistent Overconfidence and Biased Memory: Evidence from Managers. *American Economic Review* .

Hurst, E., G. Li, B. Pugsley. 2014. Are Household surveys like tax forms? Evidence from income underreporting of the self-employed. *Review of Economics and Statistics* **96(1)** 19–33.

Hurst, E., B. Pugsley. 2011. What Do Small Businesses Do? *Brookings Papers on Economic Activity* **43(2)** 73–142.

Kahan, D. 2013. Ideology, Motivated Reasoning, and Cognitive Reflection. *Judgment and Decision Making* **8(4)** 407–424.

Kahneman, D. 2011. Thinking, Fast and Slow. *Farrar, Straus and Giroux Publishing* .

Kahneman, D., G. Klein. 2009. Condtions for Intuitive Expertise: A Failure to Disagree. *American Psychologist* **64(6)** 515–526.

Kahneman, D., A. Tversky. 1979. Prospect Theory: An Analysis of Decision under Risk. *Econometrica* **47(2)** 263–292.

Kerr, W., . Nanda, M. Rhodes-Kropf. 2014. Entrepreneurship as Experimentation. *Journal of Economic Perspectives* **28(3)** 25–48.

Kerr, W., R. Nanda. 2010. Banking Deregulations, Financing Constraints and Firm Entry Size. *Journal of the European Economic Association* **8(2-3)** 582–593.

Klein, G. 2007. Performing a project premortem. *Harvard Business Review* .

Knight, F. 1921. Risk, Uncertainty and Profit. *Houghton Mifflin, Boston.* .

Konings, R., S. Hasan, A. Chatterjee. 2022. Experimentation and Start-up Performance: Evidence from A/B Testing. *Management Science* .

Kunda, Z. 1990. The Case for Motivated Reasoning. *Psychological Bulletin* **108(3)** 480–498.

Lafley, A., R. Martin, J. Rivkin, N. Siggelkow. 2012. Bringing Science to the Art of Strategy. *Harvard Business Review* .

Landier, A., D. Thesmar. 2009. Financial Contracting with Optimistic Entrepreneurs. *Review of Financial Studies* **22(1)** 117–150.

Lerner, J. 2009. Boulevard of Broken Dreams: Why Public Efforts to Boost Entrepreneurship and Venture Capital Have Failed–and What to Do about It . *Princeton University Press* .

Lerner, J., A. Leamon, F. Hardymon. 2012. Venture Capital, Private Equity, and the Financing of Entrepreneurship. *Wiley Publishing Company* .

Ludwig, J., J. Kling, S. Mullainathan. 2011. Mechanism Experiments and Policy Evaluations. *Journal of Economic Perspectives* **25(3)** 17–38.

Malmendier, U., G. Tate. 2005. CEO Overconfidence and Corporate Investment. *Journal of Finance* **60(6)** 2661–2700.

Malmendier, U., G. Tate. 2015. Behavioral CEOs: The Role of Managerial Overconfidence. *Journal of Economic Perspectives* **29(4)** 37–60.

Manso, G. 2016. Experimentation and the Returns to Entrepreneurship. *Review of Financial Studies* **29(9)** 2319–2340.

McKenzie, D. 2017. Identifying and Spurring High-Growth Entrepreneurship: Experimental Evidence from a Business Plan Competition. *American Economic Review* **107(8)** 2278–2307.

McKenzie, D. 2021. Small business training to improve management practices in developing countries: re-assessing the evidence for 'training doesn't work' . *Oxford Review of Economic Policy* **37(2)** 276–301.

Meyer, B., E. Milhaylov, J. Barrero, S. Davis, D. Altig, N. Bloom. 2022. Pandemic-Era Uncertainty. *Journal of Risk and Financial Management* **15(8)** 338–352.

Molden, D., E. Higgins. 2012. Motivated Thinking. *The Oxford handbook of thinking and reasoning* .

Moore, D., P. Healy. 2008. The Trouble with Overconfidence. *Psychological Review* **115(2)** 502–517.

Moore, D., E. Tenney, U. Haran. 2015. Overprecision in Judgement. *Keren, G. and Wu, G. (eds). Wiley Blackwell Handbook of Judgment and Decision Making* .

Moskowitz, T., A. Vissing-Jorgensen. 2002. The Returns to Entrepreneurial Investment: A Private Equity Premium Puzzle? *American Economic Review* **92(4)** 745–778.

Nanda, R. 2024. Priors, Experiments, Learning and Persuasion in (Bayesian) Entrepreneurial Finance. *Working Paper, Harvard Business School* .

Ries, E. 2011. The Lean Startup: How today's entrepreneurs use continuous innovation to create radically successful businesses. *New York, NY: Crown Publishing Group* .

Schumacher, C., S. Keck, W. Tang. 2020. Biased interpretation of performance feedback: The role of CEO overconfidence. *Strategic Management Journal* .

Sharot, T. 2011. The optimism bias. *Current biology* **21(23)** R941–R945.

Shiller, R. 1981. Do Stock Prices Move Too Much to be Justified by Subsequent Changes in Dividends? *American Economic Review* **71(3)** 421–436.

Sutton, J. 1997. Gibrat's Legacy. *Journal of Economic Literature* **35(1)** 40–59.

Syverson, C. 2011. What determined productivity? *Journal of Economic Literature* **49(2)** 326–65.

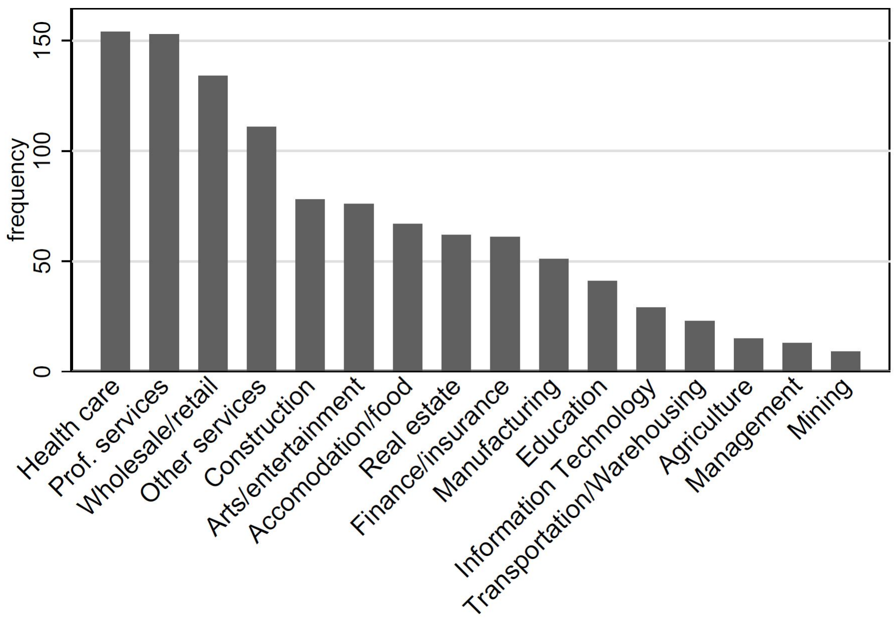
Yang, M., M. Christensen, N. Bloom, R. Sadun, J. Rivkin. 2025. How Do CEOs Make Strategy? *Management Science* .

Zhang, S.X., J. Cueto. 2017. The study of bias in entrepreneurship. *Entrepreneurship theory and Practice* .

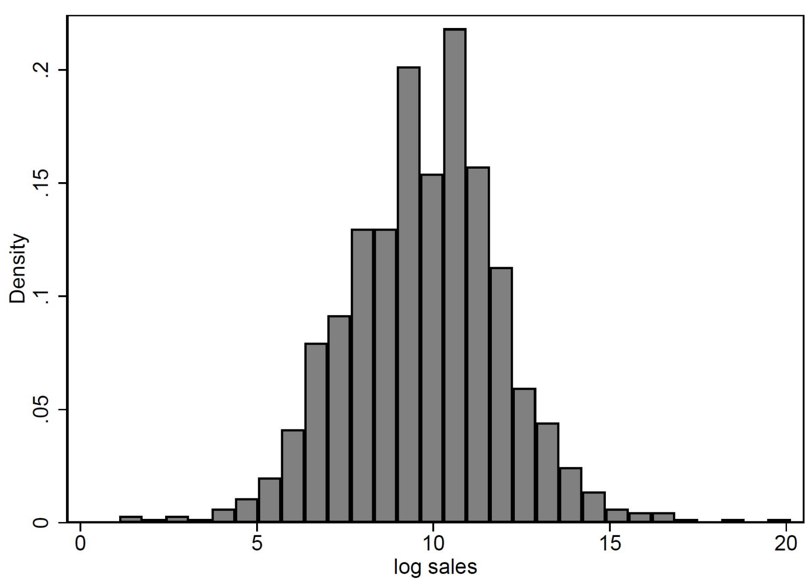
Zimmermann, F. 2020. The dynamics of motivated beliefs. *American Economic Review* **110(2)** 337–361.

Figures and Tables

Figure 1: Distribution of firms



(A) Distribution of firms across industries



(B) Firm size distribution

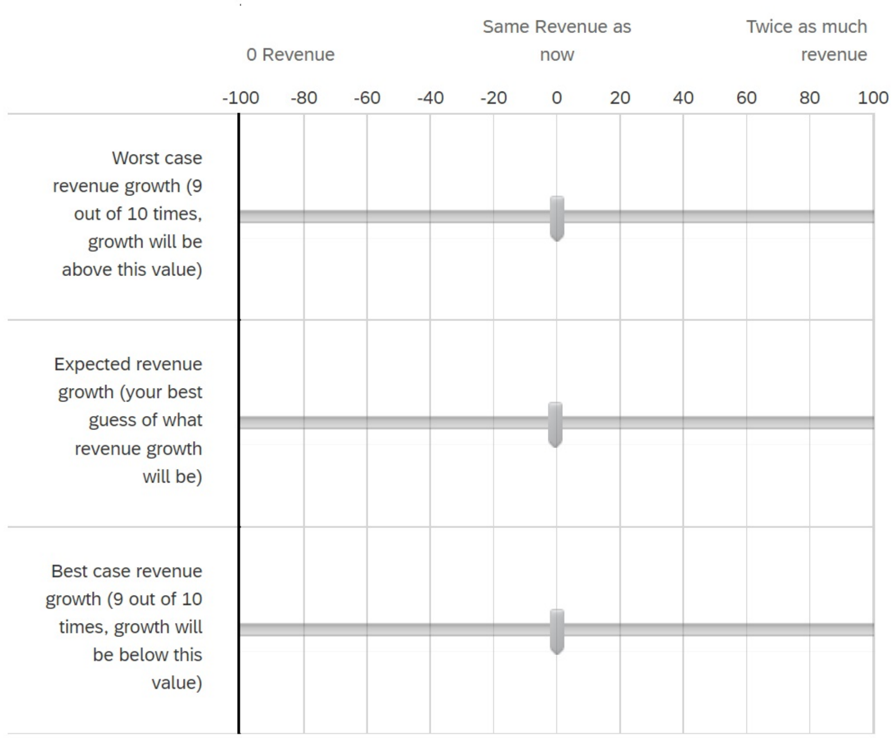
Note: Initial sample of 1027 firms in Utah in March 2021. Firm size is measured by the log of revenue in March 2021.

Figure 2: Measurement of forecasts

Please enter below the worst case revenue growth you worry about (bottom of the range), the growth you *actually* expect, and the best case revenue growth you hope for (top of the range).

We want to know the range of revenue growth you reasonably expect next four weeks (in percent, compared to this month), such that 9 times out of 10 you are certain that revenue growth over the next four weeks would be between this worst case and best case.

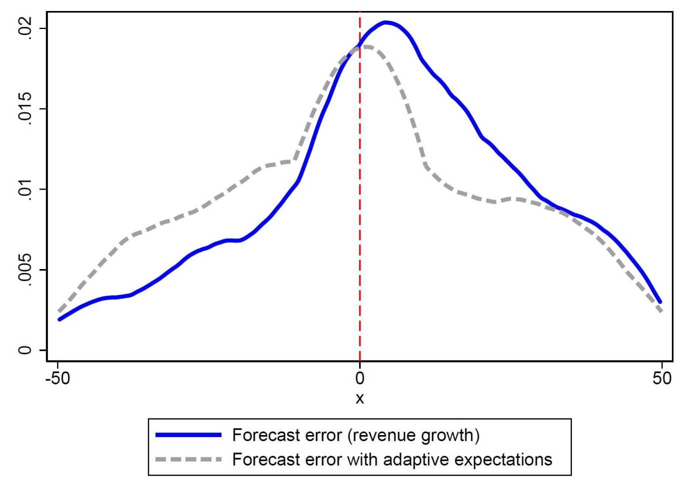
Attention: If your best guess (what you enter under "*Expected revenue growth*" below) is **within 5% of your actual revenue growth over the next 4 weeks**, we will **add an additional $5,-** to your Amazon giftcard you will receive for filling out this survey.



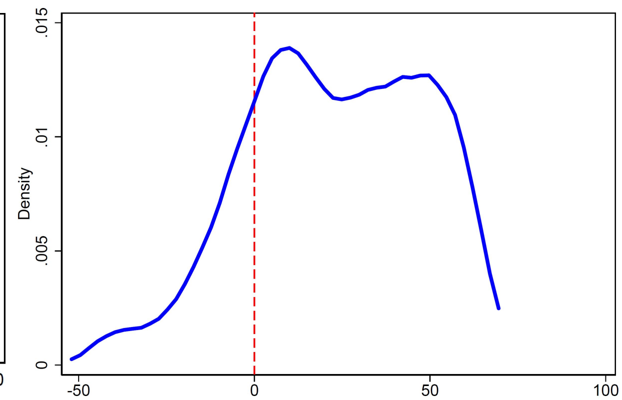
Are your goals for next four weeks' revenue growth different from the expected revenue from the previous question? (In other words, are your goals higher or lower than your expectations?)

- ○ No
- ○ * Yes (please state your revenue growth goals over the next month in %)

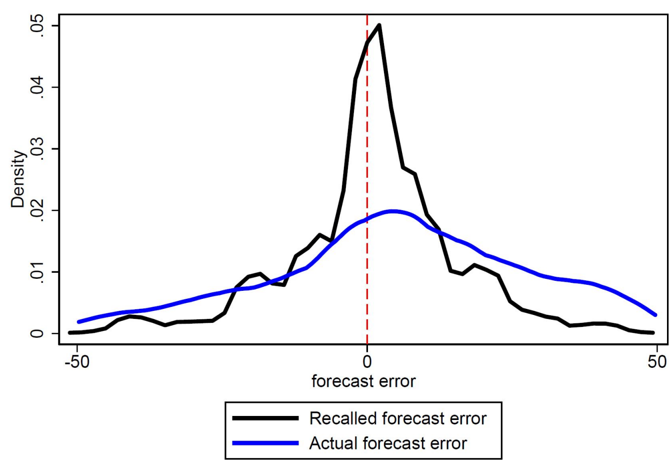
Note: Survey screen to elicit monthly revenue growth forecasts and uncertainty about forecasts. Incentive payments were introduced in October 2021 (7 months into the study and 6 months before the end of the study).



(a) Forecast error



(b) Precision error



(c) Recall error

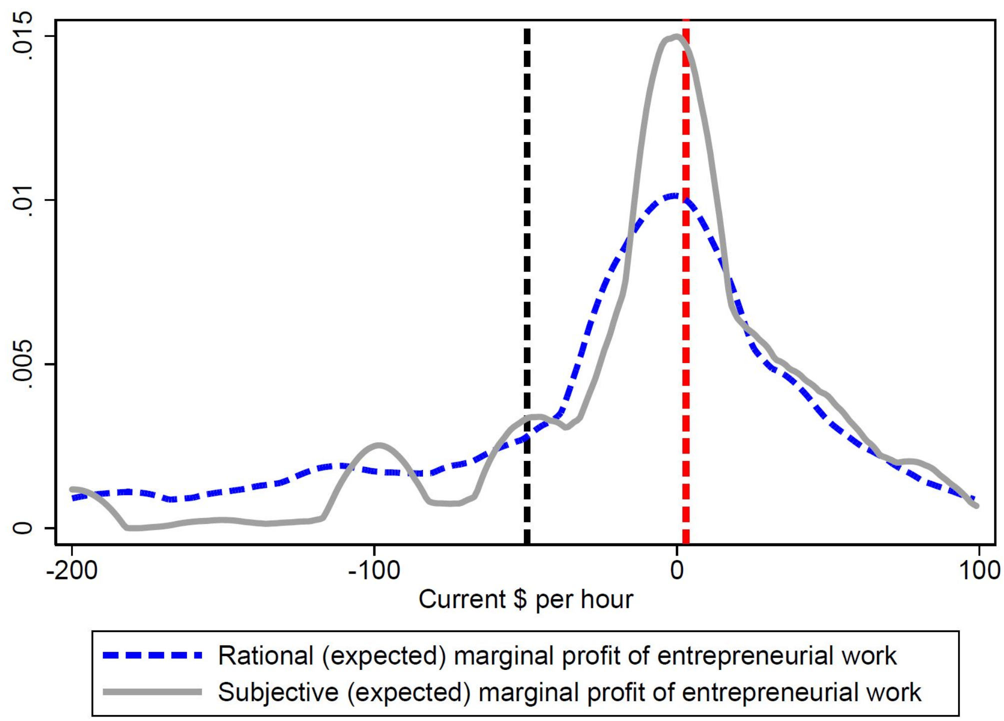
Figure 3: Distributions of errors in the control group

Panel (A): Forecast error $\xi_{i,t+1}$ is measured as difference between monthly revenue growth forecast $g^f_{i,t+1}$ and actual revenue growth $g_{i,t+1}$: $\xi_{t+1} = g^f_{i,t+1} - g_{i,t+1}$. Adaptive expectations uses lagged actual sales growth as forecast for the next month $g^{ada}_{i,t+1} = g_{i,t}$. Adaptive expectations forecast error is therefore calculated as adaptive expectation forecast minus actual monthly revenue growth $\xi^{ada}_{i,t+1} = g_{i,t} - g_{i,t+1}$.

Panel (B): Let $P_{x,i}$ denote the percentile x of monthly growth, and $P^f_{x,i}$ the subjective percentile x of monthly growth at month t for firm i. Under normal distribution of growth rates, the following approximation holds: $\sigma_{g,i} \approx \frac{P_{90,i}-P_{10,i}}{2.65}$, where $\sigma_{g,i}$ is the monthly volatility of growth rates. Similarly, $\sigma^f_{g,i,t} \approx \frac{P^f_{90,i}-P^f_{10,i}}{2.65}$ The precision error is then defined as $\omega_{i,t} = \sigma_{g,i} - \sigma^f_{g,i,t}$.

Panel (C): The solid black line displays recalled forecast error $\xi^{rec}_{i,t+1}$. The blue line shows forecast error from Panel (A). Data for all figures uses only periods before introduction of incentive payments for prediction accuracy.

Figure 4: Distribution of present value of expected marginal profits across firms



Note: Subjective expected marginal profit for individual entrepreneur i is defined as difference between the certainty-equivalent present value of expected profit increases due to 1 hour more work per week $\frac{\partial}{\partial h}\pi^e_{S,i}(h_{0,i})$, minus the opportunity costs of that 1 hour work increase $w_{0,i}$: $\frac{\partial}{\partial h}\Pi^e_{S,i}(h_{0,i}) = \frac{\partial}{\partial h}\pi^e_{S,i}(h_{0,i}) - w_{0,i}$, with $h_{0,i}$ denoting current hours worked per week. The rational marginal profit then corrects the subjective marginal profit for motivating effects of overconfidence (or demotivating effects of underconfidence): $\frac{\partial}{\partial h}\Pi^e_{R,i}(h_0) = \frac{\partial}{\partial h}\Pi^e_{S,i}(h_{0,i}) - \frac{\partial \pi_\epsilon}{\partial h}$. Rational marginal losses are bounded below using the opportunity cost of time. For more details, see text.

Table 1: Summary statistics, March 2021 (1,077 responses)

	Mean	Std	25^{th} Perc	Median	75^{th} Perc
Revenue ($)	144,919.6	578,587	2,800	15,000	60,000
Employees	10.09	26.6	0	2	8
Firm age (years)	12.77	13.72	4	7	17.5
Profit max & Growth?[1]	.61	.49	0	1	1
Livelihood?[2]	.27	.45	0	0	1
Non-pecuniary?[3]	.12	.33	0	0	0
Revenue growth (%)	14.78	52.83	-20	0	42.86
Forecast error[4] (%)	-2.08	49.03	-36	3	38.33

[1] Indicator for stated objective "Profit maximization and Growth."
[2] Indicator for stated objective "Enough profit to sustain livelihood, but no growth plans."
[3] Indicator for stated objective "Personal or social goals other than profit and growth."
[4] Forecast error $\xi_{i,t+1}$ is measured as difference between monthly revenue growth forecast $g^f_{i,t+1}$ and actual revenue growth $g_{i,t+1}$: $\xi_{t+1} = g^f_{i,t+1} - g_{i,t+1}$.

Table 2: Overconfidence and Biased Memory

	Forecast Error	Noise	Overprecision
Abs. value of recalled Forecast Error	-0.2467***	0.3749***	0.0479
	(0.0823)	(0.0541)	(0.0578)
Constant	5.2972***	30.2750***	23.5945***
	(1.2872)	(1.2269)	(1.3153)
Time FE	YES	YES	YES
R-squared	0.0094	0.0390	0.0036
Number of firms	429	429	465
Number of observations	1,519	1,519	1,763

Notes: The precision error is then defined as $\omega_{i,t} = \left(\frac{P_{90,i} - P_{10,i}}{2.65}\right) - \left(\frac{P^f_{90,i} - P^f_{10,i}}{2.65}\right)$, where $P_{x,i}$ denotes the percentile x of monthly growth, and $P^f_{x,i}$ the subjective percentile x of monthly growth at month t for firm i. Forecast error $\xi_{i,t+1}$ is measured as difference between monthly revenue growth forecast $g^f_{i,t+1}$ and actual revenue growth $g_{i,t+1}$: $\xi_{t+1} = g^f_{i,t+1} - g_{i,t+1}$. Noise is the absolute value of forecast errors. Overestimation error are all values for which $\xi_{i,t+1} > 0$, while underestimation error is the absolute value of $\xi_{i,t+1}$ conditional on $\xi_{i,t+1} < 0$. Absolute value of recalled forecast error is measured using reported forecast error for current month from memory. Sample only considers periods before the introduction of incentives. Standard Errors are clustered at the firm level.

Table 3: Balance Tests of Randomization

	A: **Error Reminder**			B: **Scientific Learning**		
	Treatment (ERT)	Control (CON)	Difference (CON−ERT)	Treatment (SLT)	Control (CON)	Difference (CON−SLT)
Firm age	12.93	12.59	-0.340	12.85	12.59	-0.263
			(.7318)			(.7966)
Employees	10.01	9.236	-0.770	11.37	9.236	-2.136
			(.6643)			(.3055)
Revenue ($, 000)	106	126	19	217	126	-90*
			(.5371)			(.0698)
Revenue growth (%)	11.91	17.96	6.055	13.77	17.96	4.195
			(.1559)			(.3645)
Forecast Error (%)	-1.524	-5.127	-3.603	1.578	-5.127	-6.705
			(.4440)			(.1755)

Notes: Firm age is measured as reported years since founding. Revenue measures monthly revenue for the month of March 2021. Revenue growth is measured between April and March 2021. Forecast error $\xi_{i,t+1}$ is measured as difference between revenue growth forecast from March 2021 to April 2021 $g^f_{i,t+1}$ and actual revenue growth $g_{i,t+1}$: $\xi_{t+1} = g^f_{i,t+1} - g_{i,t+1}$. P-values reported in parentheses.

Table 4: (No) Impact of Error Reminder Treatment

	Forecast Error	Noise	Precision Error
Error Reminder Treatment	0.6497	-0.8622	-2.7852*
	(1.0268)	(1.1915)	(1.4708)
Constant	0.2202	34.7826***	24.8517***
	(0.6847)	(0.7948)	(1.0310)
Time FE	YES	YES	YES
R-squared	0.0033	0.0065	0.0119
Number of firms	926	926	951
Number of observations	6,222	6,222	7,905

Notes: Forecast error $\xi_{i,t+1}$ is measured as difference between monthly revenue growth forecast $g^f_{i,t+1}$ and actual revenue growth $g_{i,t+1}$: $\xi_{t+1} = g^f_{i,t+1} - g_{i,t+1}$. Noise is defined as the absolute value of the forecast error $|\xi_{t+1}|$ The precision error is defined as $\omega_{i,t} = \left(\frac{P_{90,i}-P_{10,i}}{2.65}\right) - \left(\frac{P^f_{90,i}-P^f_{10,i}}{2.65}\right)$, where $P_{x,i}$ denotes the percentile x of monthly growth, and $P^f_{x,i}$ the subjective percentile x of monthly growth at month t for firm i. Standard Errors are clustered at the firm level.

Table 5: Correlation of Misattribution and Overoptimism

	Forecast Error		
	Control Group	Error Reminder Treatment Group	Scientific Learning Treatment Group
Misattribution (negative)	-10.5095**	43.3881***	42.1397***
	(5.1439)	(2.2550)	(2.5979)
Constant	0.5041	-0.5155	1.4757*
	(0.6945)	(0.7959)	(0.7973)
Time FE	YES	YES	YES
R-squared	0.0053	0.0458	0.0475
Number of firms	480	446	322
Number of observations	3,255	2,967	1,988

Notes: Negative misattribution is measured as firms that state they underperformed their forecasted revenue growth due to reasons external to the firm, while being in industries in which the median firm outperformed their forecasted revenue growth. Forecast error $\xi_{i,t+1}$ is measured as difference between monthly revenue growth forecast $g^f_{i,t+1}$ and actual revenue growth $g_{i,t+1}$: $\xi_{t+1} = g^f_{i,t+1} - g_{i,t+1}$.

Table 6: Causal impact of Access to Scientific Learning Treatment (Intend-to-Treat/ITT effect)

	Forecast Error	Noise	Precision Error	Misattribution
Scientific Learning Treatment	2.7014***	-0.2703	-3.6175**	0.0077
	(1.0321)	(1.3184)	(1.6756)	(0.0054)
Constant	0.2164	34.8045***	24.8132***	0.0270***
	(0.6857)	(0.7946)	(1.0306)	(0.0032)
Time FE	YES	YES	YES	YES
R-squared	0.0062	0.0057	0.0153	0.0266
Number of firms	802	802	827	802
Number of observations	5,243	5,243	6,647	5,243

Notes: Forecast error $\xi_{i,t+1}$ is measured as difference between monthly revenue growth forecast $g^f_{i,t+1}$ and actual revenue growth $g_{i,t+1}$: $\xi_{t+1} = g^f_{i,t+1} - g_{i,t+1}$. Noise is defined as the absolute value of the forecast error $|\xi_{t+1}|$ The precision error is defined as $\omega_{i,t} = \left(\frac{P_{90,i} - P_{10,i}}{2.65}\right) - \left(\frac{P^f_{90,i} - P^f_{10,i}}{2.65}\right)$, where $P_{x,i}$ denotes the percentile x of monthly growth, and $P^f_{x,i}$ the subjective percentile x of monthly growth at month t for firm i. Negative misattribution is measured as firms that state they underperformed their forecasted revenue growth due to reasons external to the firm, while being in industries in which the median firm outperformed their forecasted revenue growth. Standard Errors are clustered at the firm level.

Table 7: Causal impact of Engagement with Scientific Learning

	A: **Overall and Theory Engagement**			
	Forecast Error	Precision Error	Forecast Error	Precision Error
Overall Engagement with Scientific Learning	1.5402** (0.5990)	-2.0576** (0.9536)		
Theory Engagement			1.5737** (0.6119)	-2.0872** (0.9673)
Time FE	YES	YES	YES	YES
Constant?	YES	YES	YES	YES
Stock-Yogo Weak Identification Test	242.11	272.60	262.61	293.48
Kleibergen-Paap Underidentification Test	155.27	174.74	164.09	183.93
Number of firms	802	827	802	827
Number of observations	5,243	6,647	5,243	6,647

	B: **Testing and Pre-Postmortem Engagement**			
	Forecast Error	Precision Error	Forecast Error	Precision Error
Testing relative to Theory	-2.6176** (1.0334)	3.5301** (1.6314)		
Pre-Postmortem relative to Theory			-17.4865 (12.2391)	21.4321 (15.7083)
Time FE	YES	YES	YES	YES
Constant?	YES	YES	YES	YES
Stock-Yogo Weak Identification Test	140.90	140.17	2.99	3.63
Kleibergen-Paap Underidentification Test	109.03	107.16	2.98	3.61
Number of firms	802	791	802	791
Number of observations	5,243	5,012	5,243	5,012

Notes: Forecast error $\xi_{i,t+1}$ is measured as difference between monthly revenue growth forecast $g^f_{i,t+1}$ and actual revenue growth $g_{i,t+1}$: $\xi_{t+1} = g^f_{i,t+1} - g_{i,t+1}$. The precision error is defined as $\omega_{i,t} = \left(\frac{P_{90,i}-P_{10,i}}{2.65}\right) - \left(\frac{P^f_{90,i}-P^f_{10,i}}{2.65}\right)$, where $P_{x,i}$ denotes the percentile x of monthly growth, and $P^f_{x,i}$ the subjective percentile x of monthly growth at month t for firm i. Engagement is measured by length of response (string length) to free-form textboxes, in which we ask about the reasoning behind responses to scientific learning questions. Overall scientific learning engagement consists of the normalized (zero mean, unit variance) engagement score in theory, pre-postmortem and testing. Theory consists of basic idea of the business, definition of problems preventing the idea from being more successful, solution approaches, definition of conditions under which the core idea of the business will be more successful, and definition of tests to validate or falsify conditions for success. Pre-postmortem consists of internal firm conditions that might imply underperformance next month. Testing captures description of empirical tests that firm conducted to test conditions for success of the core business idea. All measures are normalized (zero mean, unit variance). Samples exclude firms with Error Reminder Treatment. Standard Errors are clustered at the firm level.

Table 8: Intend-to-Treat Profit Effects

	Profit	Profit	Profit	Profit	Revenue	Cost
			(in \$1,000 per month)			
	Average	15^{th} Perc.	85^{th} Perc.	90^{th} Perc.	90^{th} Perc.	90^{th} Perc.
Error Reminder Treatment	-25.2571	-0.0083	-6.3400	-12.3300	-29.0000	-5.9000
	(17.1421)	(0.2431)	(4.0484)	(7.6810)	(21.5677)	(5.3549)
Profit/Growth Max	56.3772*	-0.1175	29.6990**	47.6700	85.5000**	50.7000***
	(30.7368)	(0.3453)	(12.3565)	(35.7309)	(41.5708)	(17.9003)
Error Reminder Treatment	7.2831	0.3675	5.3350	13.1180	22.2840	3.8000
× Profit/Growth Max	(44.5475)	(0.4658)	(16.7580)	(48.6157)	(56.4405)	(30.7168)
Scientific Learning Treatment	-30.8628	-0.1317	-7.1450*	-11.0420	-24.0000	-4.0000
	(19.5803)	(0.2568)	(4.0724)	(7.7881)	(21.8361)	(10.7765)
Scientific Learning Treatment	132.0058**	0.5192	53.4450	148.1030**	188.5000**	80.6140
× Profit/Growth Max	(66.9236)	(0.5366)	(47.8103)	(66.8933)	(74.4116)	(70.8327)
	45.0087***	-0.2417	11.3400**	21.7880**	60.0000*	32.2000*
	(13.8271)	(0.2609)	(4.5927)	(9.8744)	(33.4032)	(16.5686)
Time FE	YES	YES	YES	YES	YES	YES
Industry FE	YES	YES	YES	YES	YES	YES
R-squared	0.07	0.00	0.02	0.05	0.06	0.08
Number of firms	1067	1067	1067	1067	1067	1067
Number of observations	4,223	4,223	4,223	4,223	4,223	4,223

Notes: All numbers as in 1,000 \$ per month. Profit is measured as the difference between operating revenues and operating costs. The variable "Profit/Growth Max" is an indicator that is one if the firm stated that its main objectives are profit maximization and growth in the pilot survey in December 2020. Column 1 is an OLS regression, while columns 2-6 are quantile regressions with the quantile defined in the column header. Full set of time fixed effects are included to control for changes due to the COVID-19 pandemic. Industry fixed effects are at the 2-digit NAICS level. Standard errors are clustered at the firm level.

Table 9: External Validity: Correcting for RCT Participation Bias

		Female Founder	Founder Age	Founder Married	Founder Hours	Goal: Growth	Firm Age	Firm Employment	Firm Revenue
		A: Founder and Firm Characteristics X_i							
RCT Sample		0.33	46.8	0.87	31.76	0.6	12.37	8	140.68
All US entrepreneurs		0.25	41.80	0.79	41.81	0.38	18.50	21.62	511.72
		B: External Validity of Forecast Error ITT							
	γ_1	-0.9126	0.0754	5.1242	0.0121	2.3500	0.0422	-0.1833	0.0064
	γ_0	-0.3659	0.1493	2.0714	-0.0717	1.3441	-0.1943	-0.1912	0.0118
Outcome y_i:	Bias term(s)	-0.0461	-0.3696	0.2442	-0.8413	0.2222	-1.4495	-0.1075	1.9981
Forecast Error	Baseline ITT			**2.70**					
	Bias-corrected ITT (select. on observ. only)			**3.05**					
	Bias-corrected ITT (select. on observ. & unobserv.)			**3.40**					
	$\Psi(0)$ (for $ITT = 0$)			**$-$7.73**					
		C: External Validity of Testing Engagement ITT							
	γ_1	-0.2836	0.0116	-0.1131	-0.0011	0.1921	-0.0155	0.0005	0.0003
	γ_0				0				
Outcome y_i:	Bias term(s)	-0.0239	0.0580	-0.0090	0.0109	0.0424	0.0950	-0.0063	-0.0969
Testing Engagement	Baseline ITT			**$-$0.99**					
	Bias-corrected ITT (select. on observ. only)			**$-$1.07**					
	Bias-corrected ITT (select. on observ. & unobserv.)			**$-$1.14**					
	$\Psi(0)$ (for $ITT = 0$)			**$-$14.21**					

Notes: Panel A displays average firm characteristics in our RCT sample and the same characteristic among all US entrepreneurs. Representative sample data on martial status of entrepreneurs is from the Kauffman Firm Survey, 2004 and data on founder age is from Azoulay et al. (2020). All other variables are from the the Annual Survey of Entrepreneurs, 2016 by the US Census Bureau. For panels B and C, estimates γ, are obtained from a regression of outcome y_i on the characteristic C_i in the column headers. γ_1 is for the scientific learning treatment group and γ_0 for the control group respectively. Following Andrews and Oster (2019), the bias term corresponding to each variable is calculated as $(\gamma_1 - \gamma_0) \cdot (\bar{C}_{P_S} - \bar{C}_P)$, where $\bar{X}_{P_S}$ is the average in the RCT (or "trail") sample and $\bar{X}_P$ is the average in the population of all US entrepreneurs. In panel C, all values for $\gamma_0 = 0$, since firms in the control group cannot engage with scientific learning by definition of being in the control group. The overall bias correction is the sum of all individual bias-correction terms. The first bias-corrected ITT term assumes that there is only selection on observables. The second bias-corrected ITT term assumes that bias on unobservables is in the same direction and of the same magnitude as selection on observables. $\Psi(0)$ quantifies the direction and magnitude that needs to be assumed for selection on unobservables to overturn our estimated treatment effects. For example: $\Psi(0) = -2$ means that selection of unobservables needs to move in the opposite direction of the selection on observables and has to be double the magnitude to imply a zero treatment effect.

Table 10: Welfare effects of Debiasing and Scientific Learning Treatment

	A: Debiasing		B: Scientific Learning Treatment	
	$ per month	% of median monthly profit	$ per month	% of median monthly profit
40^{th} Percentile	$4,786.56	106.42%	−$2155.03	−47.91%
Median	**$1417.02**	**31.50%**	**−$637.98**	**−14.18%**
60^{th} Percentile	$232.38	5.16%	−$104.62	−2.32%
75^{th} Percentile	−$378.00	−8.40%	$170.18	3.78%
85^{th} Percentile	−$2313.00	−51.42%	$1041.37	23.15%

Notes: Debiasing is defined as removing the median of the average monthly overoptimism error of 5% per month, in the control group, before the introduction of forecasting incentives in October 2021. Scientific Learning Treatment counterfactual is adding an average monthly forecast error of 2.7% (from estimates in Table 6). All welfare calculations are on a monthly basis. Median monthly entrepreneurial profit in the sample is roughly $4,500.

A Appendix

A.1 Apparent Overplacement and Forecast Errors

Benoit and Dubra (2011) offer a critique of empirical work on overplacement by arguing that this work measures "apparent overconfidence": despite the population updating correctly using Bayes' Law and being bias-free, survey measures of self-ranking may indicate overplacement spuriously. In this section, we show that their critique requires a discrete type space and does not apply to overestimation, as measured in this paper. To make this point, we provide a counterexample to the conjecture that their Bayesian model can explain overestimation. This counterexample is building on an example of apparent overplacement, provided by Benoit and Dubra (2011) but translated to our context. Specifically, consider a population of entrepreneurs, which can be of three types: low-growth, medium-growth and high-growth, denoted τ_L, τ_M, τ_H. Each of the three types is equally likely, implying priors for the types of $p_{0,\tau} = 1/3$. Firm growth for entrepreneurs is a function of their type and can either be high or low, with low growth firms not growing ($g_L = 0$) and high growth growing by 10% ($g_H = 0.1$). The corresponding random variable of firm growth is denoted G. Each of the three types differ in their probabilities of not growing, with $P(g_L|\tau_L) = 0.5875$, $P(g_L|\tau_M) = 0.5625$, $P(g_L|\tau_M) = 0.05$. As a result, the probability of no growth across the population of entrepreneurs is 40% $\left(\frac{1}{3} \cdot 0.5875 + \frac{1}{3} \cdot 0.5625 + \frac{1}{3} \cdot 0.05 = 0.4\right)$.

Benoit and Dubra (2011) show that in this example the majority of entrepreneurs might consider themselves to be above average, and therefore exhibit overplacement in the sense of Moore and Healy (2008), despite the fact that all entrepreneurs correctly use Bayes' Law to update their beliefs and are therefore bias-free. For this purpose, consider the subgroup of entrepreneurs who did not have zero growth, but instead experienced a 10% growth rate. According to Bayes' Law, their posteriors for being of the different types are

$$\begin{aligned} P(\tau_L|g_H) &= \frac{\frac{1}{3} \cdot 0.5875}{\frac{1}{3} \cdot 0.5625 + \frac{1}{3} \cdot 0.5875 + \frac{1}{3} \cdot 0.05} = 0.2291 \\ P(\tau_M|g_H) &= \frac{\frac{1}{3} \cdot 0.5625}{\frac{1}{3} \cdot 0.5625 + \frac{1}{3} \cdot 0.5875 + \frac{1}{3} \cdot 0.05} = 0.243 \\ P(\tau_H|g_H) &= \frac{\frac{1}{3} \cdot 0.05}{\frac{1}{3} \cdot 0.5625 + \frac{1}{3} \cdot 0.5875 + \frac{1}{3} \cdot 0.05} = 0.527 \end{aligned} \tag{A.1}$$

In other words, 60% of the entrepreneurs (who had a growth rate of 10%) think that they are more likely than not to be in the top third of the population. Specifically, they think that the probability of being in the top third is 0.527 and therefore higher than 1/2. A researcher conducting a representative survey in this population of entrepreneurs would therefore find widespread overplacement, despite everybody forming their beliefs rationally. A crucial issue is that survey respondents are asked to select one type among a discrete number of types and therefore select the type that is most likely. In the terminology of Benoit and Dubra (2011), beliefs are "median-rationalizable."

However, within this same example, the use of growth forecast errors shows that this population of entrepreneurs is bias-free, as one would expect under Bayesian updating. To see this, we first calculate the expected growth rates for the three types, which are given by

$$
\begin{aligned}
E[G|\tau_L] &= (1-0.5875)\cdot 0.1 = 0.04125 \\
E[G|\tau_M] &= (1-0.5675)\cdot 0.1 = 0.04375 \\
E[G|\tau_H] &= (1-0.05)\cdot 0.1 = 0.095
\end{aligned} \tag{A.2}
$$

These expected growth rates in (A.2), conditional on type can now be combined with the posterior probabilities for type in (A.1) to calculate growth forecast, given that the last growth rate was g_H: $g^f_H = E[G|g_H]$

$$
\begin{aligned}
g^f_H &= \sum_{\tau\in\{\tau_L,\tau_M,\tau_H\}} P(\tau|g_i = 0.1)\cdot E[g_i|\tau] \\
&= 0.07
\end{aligned} \tag{A.3}
$$

And similarly for the firms that experienced no growth $g^f_L = E[G|g_L]$

$$
\begin{aligned}
g^f_L &= \sum_{\tau\in\{\tau_L,\tau_M,\tau_H\}} P(\tau|g_i = 0)\cdot E[g_i|\tau] \\
&= 0.0446
\end{aligned} \tag{A.4}
$$

Ex post, there will be four possible forecast errors, $\xi_{k,l} = g^f_k - g_l$ with $k,l \in L,H$ namely

$$
\begin{aligned}
\xi_{H,H} &= 0.07 - 0.1 = -0.03 \\
\xi_{H,L} &= 0.07 - 0 = 0.07 \\
\xi_{L,H} &= 0.0446 - 0.1 = -0.0553 \\
\xi_{L,L} &= 0.0446 - 0 = 0.0446
\end{aligned} \tag{A.5}
$$

Averaging over the group with high-growth as previous realized outcome, we get[25]

$$
\begin{aligned}
\bar{\xi}_H &= \sum_{l,\tau}[\xi_{H,l}\cdot P(g_l|\tau)]\cdot P(\tau|g_H) \\
&= [0.07\cdot 0.5975 - 0.03\cdot 0.4125]\cdot 0.2291 \\
&+ [0.07\cdot 0.5625 - 0.03\cdot 0.4375]\cdot 0.243 \\
&+ [0.07\cdot 0.05 - 0.03\cdot 0.95]\cdot 0.5277 \\
&\approx 0
\end{aligned} \tag{A.6}
$$

In other words, the same 60% of entrepreneurs that exhibited apparent overplacement, do not exhibit overestimation as measured by average forecast error. In this group, firms with positive forecast error cancel out firms with negative forecast error. This suggests that the critique of Benoit and Dubra (2011) does not apply generally to average forecast error as measure of overconfidence.

[25] Similarly, for firms with low-growth as previous realized outcome the average forecast error could be computed as $\bar{\xi}_L = \sum_{l,\tau}[\xi_{L,l}\cdot P(g_l|\tau)]\cdot P(\tau|g_H) \approx 0$

A.2 Recruiting, Pilot Survey and Sample Characteristics

As discussed in the main text, recruitment to participate in the survey proceeded in two steps. In the first step, we ran a large pilot survey during which we collected information on business characteristics and asked whether entrepreneurs would be interested in participating in a long-run study. In the second step, we re-contacted interested entrepreneurs for the actual study and provided incentives to reduce sample attrition over time.

The pilot study was conducted in December 2020 in cooperation with the Utah State Chamber of Commerce, which provided us with access to their internal email list of businesses in the state. Our recruiting email was sent to businesses on behalf of the Governor's Office of Economic Development as well as the Utah State Chamber of Commerce and the University of Utah, see Panel A of Figure A.3. Importantly, our recruitment strategy was based on our field experiment in Gaulin et al. (2021). In this study we showed that moral engagement through recruitment letter framing can significantly boost participation in COVID-19 testing and is complementary with monetary incentives. Consequently, we urged entrepreneurs to participate, to help the state recover from the "ongoing health and economic crisis" and promised that "We will use your input to develop economic initiatives, policies and programs to support our business community and residents." Only after this moral engagement framing did we offer randomized prizes, such as ten $1,000 gift cards and non-pecuniary rewards as a "token of our appreciation for your help." Importantly, this recruitment strategy did not mention anything about the RCT we planned, making selection into the RCT based on perceived benefits impossible.

Additionally, due to the unsolicited nature of our recruitment email, some of the potential respondents contacted a local NBC affiliate, which ran an news segment on the evening of December 1, 2020, confirming that our survey was indeed legitimate, see panel B of Figure A.3. The combination of our moral engagement-based recruitment strategy and the evening newscast build a lot of credibility for our data collection, which we believe reduced sample selection bias, since only few entrepreneurs selected into the study based solely on monetary incentives. This is in contrast to studies using convenience samples, such as Amazon Mechanical Turk workers. All these factors make it likely that our RCT results will generalize beyond the specific RCT sample we collected. We confirm the external validity of our treatment effects formally in section 7.1

Among the key variables we obtained in the pilot survey were questions about entrepreneurs' business goals and whether respondents are interested in participating in follow-up research. Since the initial email list of the Utah Chamber of Commerce includes only business owners, we directed our survey towards entrepreneurs. Around 10,000 entrepreneurs completed our pilot survey, and we used a research assistant to ensure that almost all of these are verified businesses with a website or a physical address. After the pilot survey, about 4,000 entrepreneurs agreed to be re-contacted for follow up "year-long" surveys. In March 2021, we started recontacting 3,000 businesses, to target a final sample of about 1,000 entrepreneurs. During the study we also offered the remaining 1,000 businesses a chance to participate, to replenish our sample and offset the effects of sample attrition.

Figure A.1: Key elements of recruiting

Dear Utah Business,

COVID-19 is an ongoing health and economic crisis in Utah, and a strong recovery is only possible with your help.

To aid in the state's decision-making, we would like to better understand the issues your business is facing and have prepared this survey (link below) to gather your feedback. We will use your input to develop economic initiatives, policies, and programs to continue to support our business community and residents.

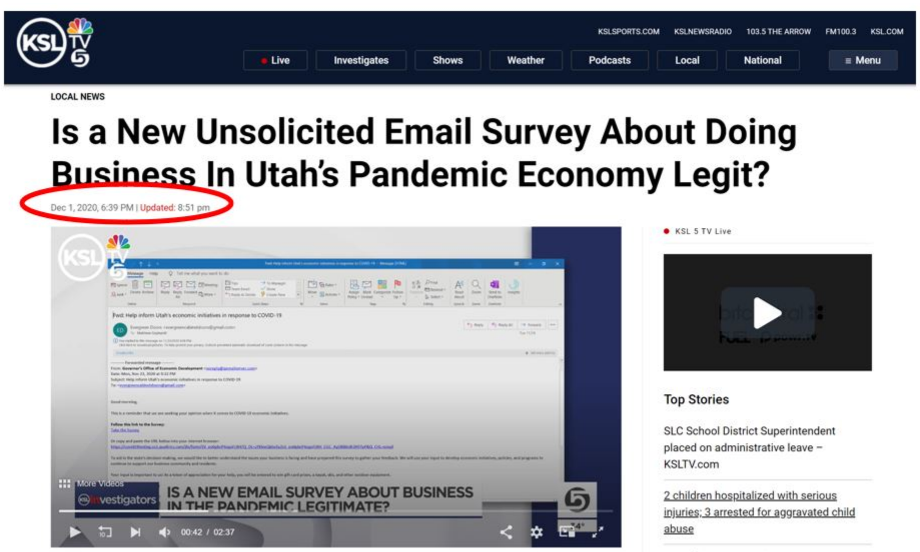
Your input is important to us! As a token of appreciation for your help, you will be entered to win one of ten $1,000 gift cards, a kayak, skis, and other outdoor equipment.

Many thanks,

Q. Val Hale
Executive director of the Governor's Office of Economic Development

Taylor Randall
Dean, David Eccles School of Business, University of Utah

(A) Pilot survey contact email



(B) Evening news coverage by local NBC affiliate (Dec 1, 2020)

Note: Figures show elements of initial recruiting of participants in December 2020.Full video of evening news coverage of the pilot survey available at: https://ksltv.com/450121/is-a-new-unsolicited-email-survey-about-doing-business-in-utahs-pandemic-economy-legit/

A.3 Validating the forecast data

Figure 2 displays the survey screen we use to elicit monthly growth forecasts. We ask respondents to forecast revenues over the next "four weeks" and to provide upper and lower confidence bounds for this forecast. Importantly, we verify that respondents' best forecast about revenues correspond to their business' growth goals and ask firms to report business goals in case the two differ. Our baseline analysis will use business goals as measure of growth forecasts, since businesses naturally have an incentive to generate accurate business growth goals.[26]

A natural question our data collection raises is whether entrepreneurial forecasts are mostly noise or whether they reflect meaningful effort to forecast future growth. The main challenge in addressing this question is that forecasted variable (revenue growth) is well known to be very noisy itself (Sutton (1997)). One influential approach to evaluate the noisiness of forecasts follows Shiller (1981) and compares the total variation in the forecasted variable and the forecasts. To fix ideas, let $g_{i,t+1}$ denote the monthly growth rate from t to $t+1$ for entrepreneur i and $g^f_{i,t+1} = E_{i,t}[g_{i,t+1}]$ the forecasted growth rate at time t. Since $g^f_{i,t+1}$ are (subjective) conditional expectations, they should be smoother than the variable they are forecasting, or:

$$Var[g_{i,t+1}] > Var[g^f_{i,t+1}] \tag{A.7}$$

To evaluate this inequality in our data, we focus on the control group before October 2021, i.e. before the introduction of incentives for accurate forecasts. We do this to make sure that entrepreneurial expectations are unaffected by any of our interventions and provide an undiluted picture on the validity of entrepreneurial expectations. Figure A.2 displays the distribution of revenue growth and forecasted revenue growth over the same time horizon. It highlights that actual revenue growth tends to be much more dispersed than entrepreneurial expectations of revenue growth. In other words, equation (A.7) holds for entrepreneurial expectations, which is in stark contrast to stock market expectations as shown by Shiller (1981). At the same time Figure A.2 already foreshadows the importance of overconfidence in our sample, as only very few growth forecasts are negative, while many growth outcomes are.

An alternative and more formalized way to evaluate the validity of entrepreneurial expectations is to use the following OLS regression:

$$g_{i,t+1} = a + b \cdot g^f_{i,t+1} + D_i + D_t + e_{i,t+1} \tag{A.8}$$

where $e_{i,t+1}$ is a mean zero, iid error term, a is a constant, D_t are time fixed effects and D_i is a firm fixed effect. Regression (A.8) nests at least three relevant benchmarks for expectations formation. First, under $b = 0$ growth forecasts $g^f_{i,t+1}$ could be complete noise or suffer from large amounts of classical measurement error. Alternatively, revenue growth could more generally be unforecastable, i.e. a random walk - possibly with a firm-specific drift D_i. Second, on the other extreme, entrepreneurial expectations could be completely rational and unbiased with $b = 1$. In this case, entrepreneurs would make no systematic forecasting mistakes, even if their forecasts might be very noisy. Third, somewhat between rational expectations and useless forecasts are adaptive expectations, as proposed for example by Muth (1960). In the simplest case of adaptive expectations, $b = 1$ and $g^f_{i,t+1} = g_{i,t}$, i.e. entrepreneurial forecasts do not include more information than is included in past sales growth. In contrast to these three benchmarks, overconfident entrepreneurial expectations are implied if $b < 1$.[27]

[26] Section 6 analyzes robustness of our main results to this choice.

[27] To see this, we can solve (A.8) for the forecast error $g_{i,t+1} - g^f_{i,t+1}$ and take expectations to obtain:

The first row in Table A.1 shows that entrepreneurial forecasts are systematically correlated with actual revenue growth. This suggests that revenue growth is no random walk and that entrepreneurial growth forecasts are not on average arbitrary guesses. Furthermore, when we include a full set of firm fixed effects, the coefficient estimate for b rises substantially towards the rational expectations benchmark of $b = 1$ and one cannot reject the hypothesis that expectations are indeed rational. This result in Table A.1 is consistent with the view that overconfidence is very persistent and that the use of firm fixed effects removed such persistent overconfidence. Put differently, entrepreneurial expectations are close to rational, but-for persistent overconfidence.

The last column adds lagged revenue growth as predictor alongside entrepreneurial expectations. This shows that entrepreneurial expectations contain information that goes beyond what is contained in data on lagged sales growth. On the flip side, this column also shows that entrepreneurial expectations do not fully incorporate lagged growth, as this variable remains statistically significant if it is included alongside expectations. This could be consistent with entrepreneurs failing to take account of mean reversion, but other explanations are possible, too.

These results motivate our focus on understanding biases in forecasts instead of the variance of forecasts, as measured by "noise," which is defined as the absolute value of forecast errors. As we conduct our analysis, we will report results on the impact of our treatments on noise, but leave a detailed analysis of this aspect for other research, including Bloom et al. (2025).

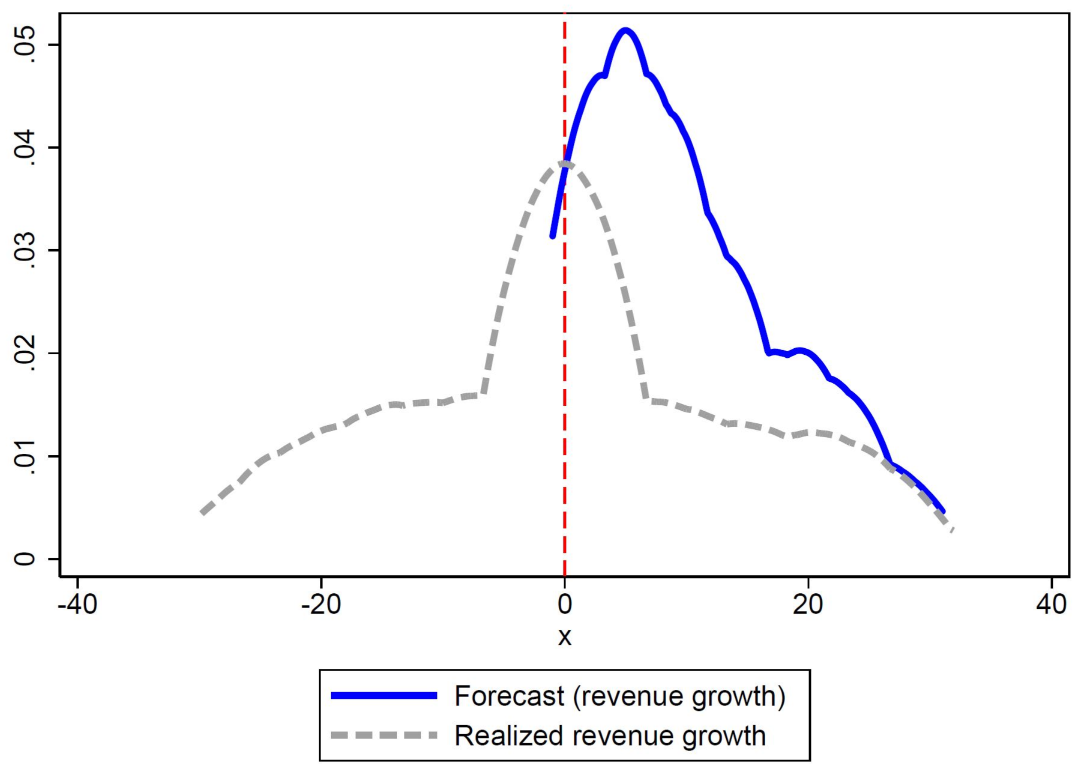
Table A.1: Benchmarking Entrepreneurial Expectations

	Revenue growth $g_{i,t+1}$			
	(OLS)	(OLS)	(AB)	(AB)
Forecast $g^f_{i,t+1}$	0.5978***	0.8455***		0.8074***
	(0.0860)	(0.1401)		(0.1931)
Lagged growth			-0.1893***	-0.1757***
			(0.0387)	(0.0420)
Constant	3.4754***	1.2076	9.7964***	2.8328
	(1.0483)	(1.3198)	(1.1009)	(2.1351)
Time FE?	YES	YES	YES	YES
Firm FE?	NO	YES	YES	YES
Number of firms	461	389	328	305
Number of observations	1,952	1,880	1,145	998

Notes: Dependent variable $g_{i,t+1}$ is revenue growth. Forecast is forecasted revenue growth $g^f_{i,t+1}$. Sample of observations before the introduction of the forecast accuracy incentive. Columns (3) and (4) use Arellano-Bond dynamic panel estimation. Standard Errors are clustered at the firm level.

$E\left[g_{i,t+1} - g^f_{i,t+1}\right] \propto \left(\frac{1}{b} - 1\right) \cdot E[g_{i,t+1}]$ which is positive if $E[g_{i,t+1}] > 0$.

Figure A.2: Distribution of growth rates and forecasts in control group



Note: Forecasts in blue are $g^f_{i,t+1}$, while the grey dashed line shows actual revenue growth $g_{i,t+1}$. Data for the figure uses only periods before introduction of incentive payments for prediction accuracy.

A.4 Persistent Private Information

In the main text we use simple forecast error, calculated as the difference between sales growth forecast and realized sales growth. However, this approach can be problematic, if there exists persistent private information. For example, an entrepreneur might consistently make very high growth forecast, since he knows about a business opportunity that might realize over the next several months, but every month it does not realize, his growth forecast spuriously looks like overoptimism. In this section we build on a model of Bayesian learning with persistent private information in Healy and Moore (2007) to correct for this issue. As we show below, correcting for persistent private information leads to even higher measured entrepreneurial overoptimism in the control group. At the same time, the theory developed in this section confirms that persistent overoptimism cannot be generated by Bayesian updating, as shown by example in Appendix A.1.

The firm growth rate $g_{i,t}$ for entrepreneur i in time period t can be modeled as

$$g_{i,t} = \theta_{i,t} + u_{i,t} \tag{A.9}$$

where $u_{i,t}$ is an iid error term with $u_{i,t} \sim N(0, \sigma_u^2)$, so that $\mu = E[g_{i,t}]$. In this context, $\theta_{i,t}$ is the forecastable part of firm growth with $\theta_{i,t} \sim N(\mu, \sigma_\theta^2)$. For simplicity, we assume that variances $\sigma_u^2, \sigma_\theta^2$ are known.

Entrepreneurs observe a noisy private signal $s_{i,t-1}$, which is unobserved by the econometrician and is given by

$$s_{i,t-1} = g_{i,t} + e_{i,t-1} \tag{A.10}$$

where $e_{i,t-1}$ is an iid error term with $e_{i,t-1} \sim N(0, \sigma_e^2)$. Under Bayesian updating growth forecasts will be

$$E[g_{i,t}|s_{i,t}] = \alpha \cdot \mu + (1-\alpha) \cdot s_{i,t-1} \tag{A.11}$$

with $\alpha = \frac{\sigma_e^2}{\sigma_e^2+\sigma_u^2+\sigma_\theta^2} \in (0,1)$. An econometrician analyzing the entrepreneur's forecasts does not observe the private information $s_{i,t-1}$. However, the econometrician will observe the growth outcome $g_{i,t}$, which is correlated with the private signal $s_{i,t-1}$.[28] This insight is key to address the presence of private information. Specifically, conditional on observing $g_{i,t}$, one can integrate out the private signal:

$$\begin{aligned} E_s\left[E[g_{i,t}|s_{i,t-1}]|g_{i,t}\right] &= E_s\left[\alpha \cdot \mu + (1-\alpha) \cdot s_{i,t-1}|g_{i,t}\right] \\ &= \alpha \cdot \mu + (1-\alpha) \cdot E_s\left[s_{i,t}|g_{i,t}\right] \\ &= \alpha \cdot \mu + (1-\alpha) \cdot E_s\left[(g_{i,t}+e_{i,t-1})|g_{i,t}\right] \\ &= \alpha \cdot \mu + (1-\alpha) \cdot g_{i,t} \end{aligned} \tag{A.12}$$

The last line in (A.12) shows that one can use the realized growth rate to condition on private information. Intuitively, conditioning on realized growth rates allows the econometrician to control for all possible private signals that are correlated with this growth. The simplest approach to do this is regress forecasts on contemporaneous growth rates and use the fitted forecast values as forecast measure when calculating forecast errors. This is the approach we pursue below.[29]

[28] If the private signal $s_{i,t-1}$ would be uncorrelated with the growth outcome $g_{i,t}$, a rational entrepreneur should not put any weight on it, i.e. $\alpha = 1$ in (A.11).

[29] One could in principle also use Machine Learning methods to construct this expectations. However, although these methods would have lower out-of-sample forecast errors, they would produce biased forecasts,

Before the empirical application, it is worthwhile showing in the context of our discussion of Benoit and Dubra (2011) in Appendix A.1, that persistent overoptimism will be an outcome if entrepreneurs deviate from Bayesian updating. To see this, we subtract the growth rate $g_{i,t}$ from both sides of (A.12) and take expectations:

$$E\left[E_s\left[E[g_{i,t}|s_{i,t}]|g_{i,t}\right] - g_{i,t}\right] = \alpha \cdot (\mu - E[g_{i,t}]) \tag{A.13}$$

Note that by definition, $\mu = E[g_{i,t}]$, so that equation (A.13) implies that $E\left[E_s\left[E[g_{i,t}|s_{i,t}]|g_{i,t}\right] - g_{i,t}\right] = 0$ or that should not be any persistent overoptimism once we average across entrepreneurs and time periods. This is consistent with the conventional wisdom that Bayesian learning converges to unbiased (or "rational") expectations, see Feldman (1987).

To investigate whether private information might drive forecast errors, we need to estimate forecasts, conditional on realized growth. The approach implied by theory is given by equation (A.12):

$$E_s\left[E[g_{i,t}|s_{i,t-1}]|g_{i,t}\right] = \alpha \cdot \mu + (1-\alpha) \cdot g_{i,t} \tag{A.14}$$

This can be estimated by regressing forecasts on a constant and realized growth rates $g_{i,t}$. A slightly more general version allows for firm fixed effects:

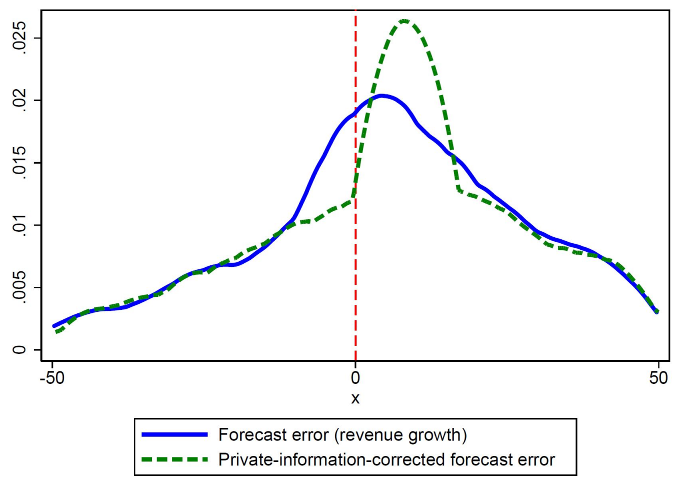
$$E_s\left[E[g_{i,t}|s_{i,t-1}]|g_{i,t}\right] = \alpha \cdot \mu_i + (1-\alpha) \cdot g_{i,t} \tag{A.15}$$

It is useful to recall that the Bayesian updating parameter is common across all entrepreneurs, if σ_e^2, σ_u^2 are common, since $\alpha = \frac{\sigma_e^2}{\sigma_e^2 + \sigma_u^2 + \sigma_\theta^2}$. To calculate private-information corrected forecasts we estimate (A.15) with firm fixed effects and then use (A.15) to solve for μ_i and then average across time periods to estimate it.

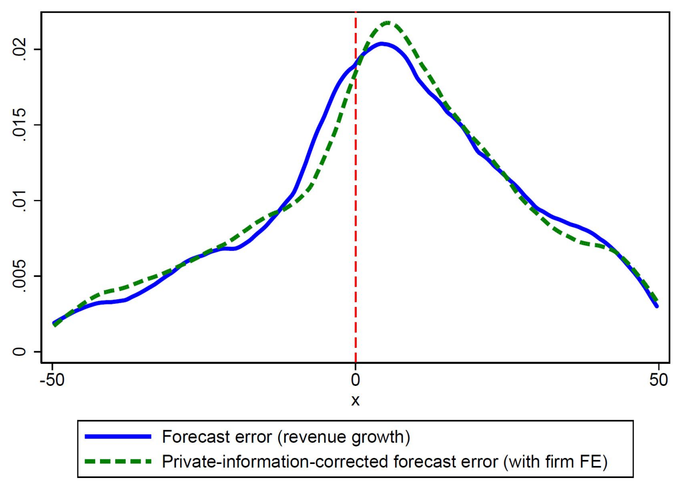
Figure A.3 contrasts our baseline measures of forecast error used in the main text in solid blue, with private-information-corrected forecast error in green dashed lines. Median overoptimism becomes stronger when accounting for private information: while the median overoptimism is 5% per month (in the control group before the introduce forecasting incentives in October 2021) for the raw forecast error measure we use in the main text. Using forecast errors with expectations that are corrected for private information results in median overoptimism of 6% per month. Private information about growth opportunities cannot explain the degree of entrepreneurial overconfidence we observe in the data and if anything makes overoptimism worse. Importantly the distribution of private-information-corrected forecast errors is very similar to the uncorrected forecast error distribution, especially if we account for firm fixed effects as in the bottom panel of Figure A.3.

for a rational Bayesian benchmark as modeled here.

Figure A.3: Foreacast Error and Private information



(A) Private-information-corrected forecast error



(B) Private-information-corrected forecast error with firm FE

Note: Forecast error $\xi_{i,t+1}$ is measured as difference between monthly revenue growth forecast $g^f_{i,t+1}$ and actual revenue growth $g_{i,t+1}$: $\xi_{t+1} = g^f_{i,t+1} - g_{i,t+1}$. Data for the figure uses only periods before introduction of incentive payments for prediction accuracy.

A.5 Details on Scientific Learning Treatment

Our treatment builds on recent work applying scientific learning to different contexts, such as CEO decision-making (Lafley et al. (2012), Felin and Zenger (2017), ?), teaching students to think scientifically (Ashraf et al. (2022)) and entrepreneurial experimentation (Felin and Zenger (2009), Ries (2011), Camuffo et al. (2020), Felin et al. (2020) and Konings et al. (2022)). On a high level, this treatment consists of three parts:

1. Structured problem-framing and hypothesis development ("theory" for short), to address motivation for reasoning and model misspecification.
2. Pre-postmortem
3. Hypothesis testing, based on theory, to address model misspecification and identification.

We detail each of of these three main parts in the following. Starting with hypothesis development (or theory), we follow Felin et al. (2020) and provide the following questions, which guide entrepreneurs along a structured script to formulate the theory of their firm[30]. (The bold headers are not displayed for survey respondents, but serve as guideposts for readers only.)

1. **Differentiation**: Do you have a unique idea or belief that differs from "conventional wisdom" in your industry? If you hold such a contrarian belief, what is it and how could it help with your growth goal?
2. **Problem-definition**: What are the most important problems that prevent your unique idea from being realized? Put differently, what are the reasons your belief is contrarian instead of being widely accepted in your industry?
3. **Problem-solving**: Please list two possible plans that might solve the problems that prevent your unique idea from being realized and which can help with your growth goals.
4. **Key conditions**: What would have to be true for each of the two plans you listed in the last question, to achieve your growth goal for the next month?
5. **Pre-definition of tests**: For each of the conditions you specified in the previous question, how would you test whether this condition is true?

   The detailed survey screens we displayed to the scientific learning treatment group are shown in Figure A.4 to A.8.

A specific case example is helpful in illustrating potential engagement with the scientific learning treatment. To preserve the privacy of the company, we will call it "Bennett Woodworks" with the fictitious founder name "Olivia." Olivia describes her business as "high-end furniture" manufacturing and the unique idea of her business as utilizing "exotic woods to create wood art. This is an area of woodworking that isn't done by many woodworkers." But she also recognizes that "The biggest problem is that the majority of customers in this market generally don't spend a lot of money for collectible products so I limit myself in this regard." Olivia then develops multiple hypotheses for the cause of this insufficient demand:

- H1: Demand might be too low, not because there are not enough potential customers with a sufficiently high willingness to pay, but because these customers do not know about Olivia's offerings.

[30]We would like to thank Todd Zenger, who gave us very useful feedback on this script.

- H2: Demand might be too low because prices on her existing products are too high.
- H3: Demand might be too low because Olivia's existing products target an unprofitable market segment.

Olivia develops three alternative approaches to test these different hypotheses:

- S1: "One plan is to use targeted advertising in order to reach a wider audience."
- S2: "Another plan could be to create a cheaper alternative to the fine woodworking products I offer."
- S3: "alternatively, still create high quality products but redesign them to be cheaper to manufacture and then offer them at a lower price point."

Each of these three plans addresses a different problem and corresponds to a different cause for why demand is too low. For example, the use of targeted advertising mostly helps if H1 is the main problem and not excessively high prices as in H2. On the other hand, the success of cutting costs and prices of existing products depends on the price elasticity of the demand curve for Olivia's existing products. Similarly, the success of new products to address H3, depends on demand for smaller but still high-quality furniture but tells her less about the demand for her existing products.

To test the hypothesis H1, Olivia decides to run ads for her merchandise on Facebook (called "targeted advertising" in Figure **??**). This test did not generate much demand, thereby suggesting that there might not be much latent demand for her products among her target customers. This reinforced the importance of the negative feedback she sees in her monthly sales.

To test hypothesis H2, Olivia decides to randomly cut prices on a few of her existing products via discounts. As a result, Olivia reports a 3-fold increased revenue for these products. However, as her quote in Figure **??** makes clear, she recognizes a potentially important identification problem: Olivia does not know whether her initial prices were too high or whether customers just responded to the availability of temporary discounts. However, this consideration makes clear that Olivia is focusing attention on what the testing tells her about her underlying theory, as advocated by Gagnon-Bartsch et al. (2021). Furthermore, this example illustrates that Olivia does not simply mindlessly repeat successful actions, but that our treatment is successful in inspiring her to understand the causal mechanisms at play. This addresses a key friction emphasized by Gagnon-Bartsch et al. (2021): "A person fails to discover a costly mistake only when he wrongly deems valuable data entirely useless and ignores it."

To test H3, Olivia creates new, cheaper products to offer to target the market segment of smaller-but-high-quality furniture. Her new products are met with consistently high sales, as can be seen in the last branch of Figure **??**. This test confirms Olivia's hypothesis that demand for smaller-but-high-quality furniture pieces is high, which is why Olivia ends up adding the cheaper products to her product offerings permanently.

To summarize, scientific learning helps address potential causal misattribution for Olivia in two ways. On the one hand, the theory development directs Olivia's attention towards different solutions for different causes of low demand, thereby encouraging her to think about various internal and external variables, which she might have otherwise ignored (Gagnon-Bartsch et al., 2021). The solutions in turn are based on Olivia's own cognitive problem-solving capabilities, which makes it harder to blame external factors if these solutions fail. On the other hand, testing Olivia's solution ideas provides evidence on internal and external causes of sales performance (Hestermann and Yaouanq, 2021). Olivia's tests fail to find evidence for the hypothesis that insufficient customer exposure is to blame for low demand, while they confirm her conjecture that prices might potentially be too high and that the wrong market segment was targeted. The latter two factors represent internal causes for low sales growth as opposed to external reasons.

Figure A.4: Scientific Learning Treatment Nudges

Part 1 (Hypothesis Development): (1) Differentiation

A "competitive advantage" is a strength your company has, which distinguishes you from your competitors and which is hard to copy.

Often such “competitive advantage” results from exploring previously untested ideas. **Do you have a unique idea or belief that differs from “conventional wisdom” in your industry? If you hold such a contrarian belief, what is it and how could it help with your growth goal?**

For example, you might own a sandwich shop and no other sandwich shop in your neighborhood might offer breakfast, because “conventional wisdom” is that there is not enough foot-traffic in the morning. A contrarian belief might be that many office workers are open to purchasing breakfast, but do not currently do so, because they want to avoid fatigue after eating a heavy and unhealthy breakfast sandwich.

Part 1 (Hypothesis Development): (2) Problem Framing

What are the most important problems that prevent your unique idea from being realized? Put differently, what are the reasons your belief is contrarian instead of being widely accepted in your industry?

In the sandwich shop example, among the problems preventing you from offering breakfast could be that you do not know demand by office workers for healthy breakfast options. Another problem might be that office workers do not know that healthy breakfast options are available for purchase.

Figure A.5: Scientific Learning Treatment Nudges

Part 1 (Hypothesis Development): (3) Hypothesis Generation

Please list two possible plans that might solve the problems that prevent your unique idea from being realized and which can help with your growth goals. Ideally, these two plans would be two different ways that help you solve a problem that other competitors in your market are not solving.

We recommend that these two plans include
(1) What advantage you intend to use or create to achieve your goal,
(2) What customer or market segment you will target
(3) A list the activities that you will use to deliver the intended results

Two questions other business executives have found helpful to come up with these two possible plans are the following:
(A) What does this company do especially well? How could that strength help to increase value for new potential customers or reduce costs to you?
(B) What are the underserved needs or needs that customers find hard to express, and what gaps have competitors left?

In the sandwich shop example, one plan might be to offer healthy breakfast smoothies with caffeine, which prevent customers from being tired after breakfast. The targeted customer segment are nearby office workers, which are more likely to be repeat customers. To deliver such smoothies you would need equipment and freshly purchased ingredients. One potential advantage might be your knowledge of tasty smoothie recipes.

Figure A.6: Scientific Learning Treatment Nudges

Part 1 (Hypothesis Development): (4) Key Assumptions

What would have to be true for each of the two plans you listed in the last question, to achieve your growth goal for the next month?

For each of the two plans, please make up a list of conditions, which you could potentially observe, and that can either assure you that your plan worked or make you confident that the plan did not work. Such a list of conditions can enable you to pay attention to the relevant business information in a targeted way and more accurately learn from your experiences.

One way to express this is an IF-THEN statement: IF your conditions are met, THEN your profit increases because of the problem the plan solves.

In the sandwich shop example, your condition might be "IF I can at least attract 45 office workers at $5 per breakfast smoothie every weekday, THEN offering breakfast is profitable". One way this condition might fail is that there are not enough office workers interested to purchase breakfast smoothies every weekday.

Figure A.7: Scientific Learning Treatment Nudges

Part 1 (Hypothesis Development): (5) Pre-Definition of Tests

For each of the conditions you specified in the previous question, how would you test whether this condition is true?

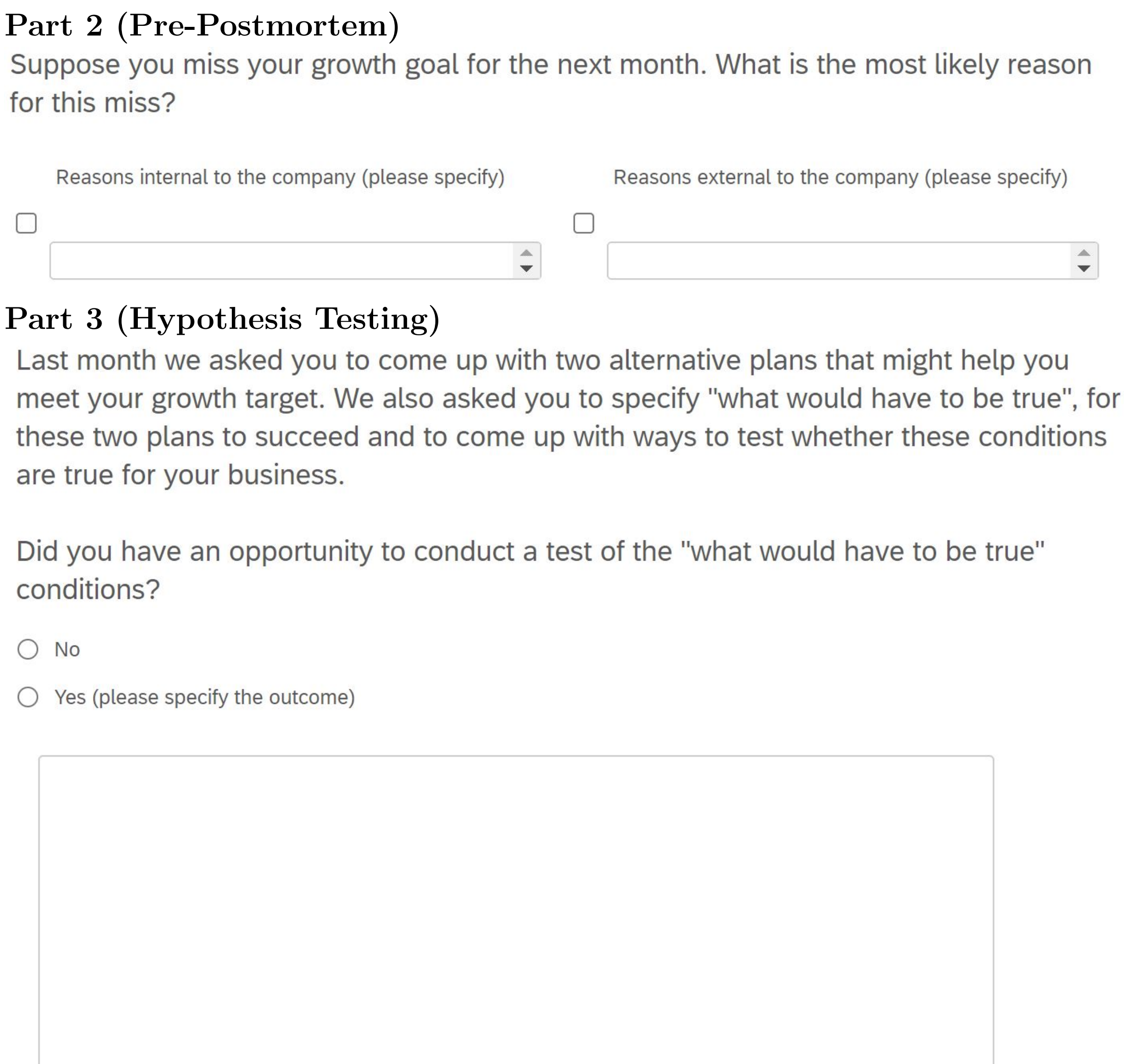
A "test" involves figuring out if the underlying REASON your plan works is correct or incorrect, just like a “business scientist” would. Understanding the reason your plan works can be important to ensure that you can repeat your success and do not rely on “luck”. It will also ensure you that you solved the problem that prevents other firms from doing the same.

Let’s return to the sandwich shop example with the condition "IF I can at least attract 45 office workers at $5 per breakfast smoothie every weekday, THEN offering breakfast is profitable". Your test might involve offering healthy breakfast smoothies with caffeine and advertise these healthy options in neighboring office buildings. Keeping track of how many of your breakfast smoothie customers are office workers and how many of your office workers are repeat customers can then tell you if you can repeat your success.

For more detail, see this article (which will open in a new tab and not interrupt your survey responses on this tab).

Note: The link on this page leads to an online version of Lafley et al. (2012), which is a general audience introduction to Scientific Learning for managers.

Figure A.8: Scientific Learning Treatment Nudges



Part 2 (Pre-Postmortem)

Suppose you miss your growth goal for the next month. What is the most likely reason for this miss?

Reasons internal to the company (please specify)

Reasons external to the company (please specify)

Part 3 (Hypothesis Testing)

Last month we asked you to come up with two alternative plans that might help you meet your growth target. We also asked you to specify "what would have to be true", for these two plans to succeed and to come up with ways to test whether these conditions are true for your business.

Did you have an opportunity to conduct a test of the "what would have to be true" conditions?

No

Yes (please specify the outcome)

A.6 Learning Dynamics

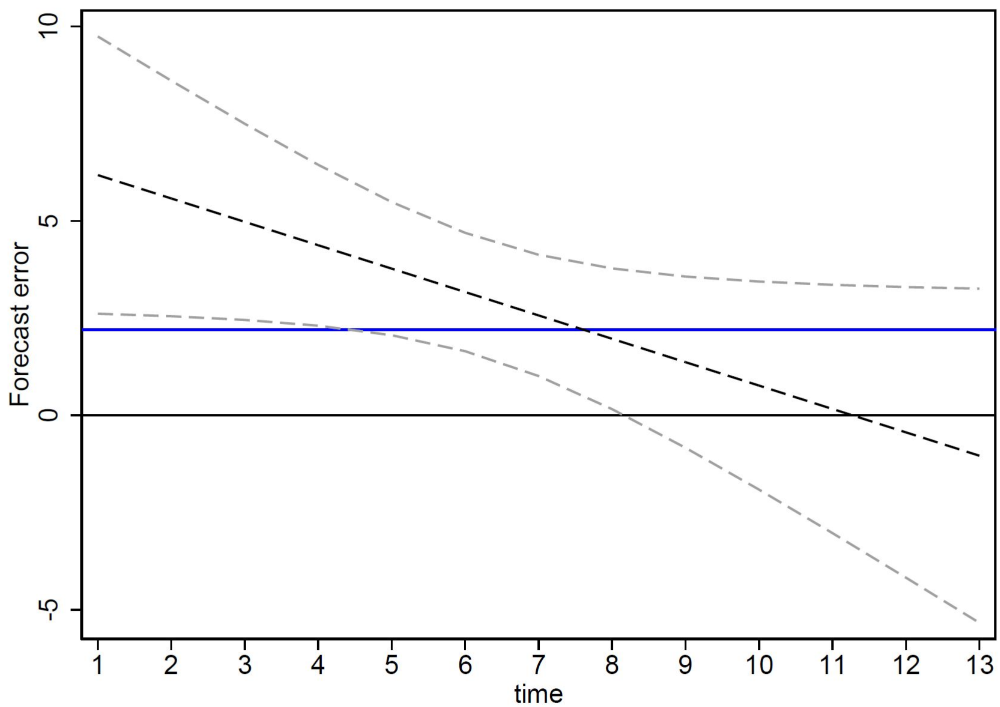
One of the strengths of our field experiment is the collection of relatively long panel data for the 13 months of our study. This allows us to go beyond average treatment effects to document how forecast and precision errors change over time. As we discussed in the context of Figure **??**, one potential concern with such an analysis is that the dynamics of the COVID-19 pandemic might impact our estimates. To address this concern, we include a full set of time fixed effects. However, to still estimate effects of how treatments impacted changes in forecast errors, we estimate interactions of treatment indicators with linear time trends.

Figure A.9 highlights our main result from this analysis. The figure shows the evolution of forecast errors over time for the scientific learning treatment. Access to scientific learning initially strongly increases forecast error and therefore overoptimism, but this effect slowly fades over time. In other words, although our scientific learning treatment increases overoptimism over the entire sample, entrepreneurs eventually learn to adjust their forecasts and learn that they have been overconfident.

Table A.2 presents the formal regressions results underlying Figure A.9 as well as additional results for other outcomes. Column 3 of Panel A in Table A.2 is of special interest, as it shows the impact of the scientific learning treatment on overprecision. These results stand in contrast to the dynamics just discussed for overoptimism. While Scientific Learning increases the overoptimism bias, and this bias slowly fades over time, the same treatment reduces overprecision and this effect is persistent over time.

The dynamic effects associated with scientific learning also contrast with the effects of the error reminder treatment, which are reported in Panel B of Table A.2. Consistent with our estimates in Table 4, the error reminder treatment has no effect on overoptimism. However, there is some evidence that entrepreneurs exposed to the error reminder treatment are systematically reducing noise - or the size of their forecast errors. Column 3 of Panel B in Table A.2 also suggests that the error reminder treatment reduces overprecision. This suggests that part of the reduction in overprecision in the scientific learning treatment is due to the fact that entrepreneurs in this treatment are shown their past forecast error as well and that they knowledge of these forecast errors makes them reduce the confidence in their own forecasting ability.

Figure A.9: Treatment effect of Scientific Learning on forecast error over time



Note: Dependent variable on the y-axis is forecast error $\xi_{i,t+1}$ and is measured as difference between monthly revenue growth forecast $g^f_{i,t+1}$ and actual revenue growth $g_{i,t+1}$: $\xi_{t+1} = g^f_{i,t+1} - g_{i,t+1}$. Figure plots the sum of the average treatment effect and the interaction effect of treatment and a linear time trend, controlling for a full set of time fixed effects to control for the impact of changes in uncertainty due to the COVID-19 pandemic. Time horizon is one year between March 2021 and March 2022.

Table A.2: Learning Dynamics of ITT Effects

	A: **Scientific Learning**		
	Forecast Error	Noise	Precision Error
Scientific Learning Treatment	6.5663***	1.2875	-5.4189***
	(2.2014)	(1.9743)	(1.9133)
Scientific Learning Treatment	-0.6015*	-0.2424	0.2587
$\times$ linear time trend	(0.3084)	(0.2519)	(0.1673)
Time FE?	YES	YES	YES
R-squared	0.01	0.01	0.02
Number of firms	802	802	827
Number of observations	5,243	5,243	6,647
	B: **Error Reminder**		
	Forecast Error	Noise	Precision Error
Error Reminder Treatment	2.4386	1.7209	-3.4182**
	(2.1476)	(1.7574)	(1.7080)
Error Reminder Treatment	-0.2770	-0.3999*	0.0900
$\times$ linear time trend	(0.2889)	(0.2176)	(0.1456)
Time FE?	YES	YES	YES
R-squared	0.00	0.01	0.01
Number of firms	926	926	951
Number of observations	6,222	6,222	7,905

Notes: Forecast error $\xi_{i,t+1}$ is measured as difference between monthly revenue growth forecast $g^f_{i,t+1}$ and actual revenue growth $g_{i,t+1}$: $\xi_{t+1} = g^f_{i,t+1} - g_{i,t+1}$. Noise is defined as the absolute value of the forecast error $|\xi_{t+1}|$ The precision error is defined as $\omega_{i,t} = \left(\frac{P_{90,i}-P_{10,i}}{2.65}\right) - \left(\frac{P^f_{90,i}-P^f_{10,i}}{2.65}\right)$, where $P_{x,i}$ denotes the percentile x of monthly growth, and $P^f_{x,i}$ the subjective percentile x of monthly growth at month t for firm i. All specifications include a full set of time fixed effects to control for changes due to the COVID-19 pandemic. Standard errors are clustered at the firm level.

A.7 Incentives for Accurate Forecasts

One potential issue with our analysis is that entrepreneurs might have insufficient incentives to report accurate forecasts. We believe that this is unlikely for several reasons. On the one hand, we showed in section 3.1 that forecasts are systematically correlated with growth outcomes, which they should not be if they are just noise. On the other hand, our main analysis focuses on growth forecasts from explicit business targets. Any inaccurate business targets would result in misallocation of resources, for example by hiring too many or to few employees and purchasing too many or to few materials etc.

However, instead of just relying on the plausibility of these conceptual points, we incorporated explicit performance pay for accurate forecasts into our analysis. Specifically, from October 2021 until March 2022 we provided a bonus of an additional \$5 if revenue growth forecasts were within 5% of reported revenue growth over the next 4 weeks. We chose 5% since this was the median overestimation of the control group in the first few months of the study. This bonus payment was both, salient and credible. As shown in Figure 2, we use bright red color on the survey screen to highlight the bonus payment. Additionally, survey respondents had been part of the survey for 6 months at this point and knew that we would follow through with any promised payments. The incentive payment for accurate forecasts applied to all firms in our sample, because rather than being interested in the impact of incentives for accurate forecasts per se, we are interested whether higher incentives for forecast accuracy differentially affect treatment as opposed to control firms. If there is no interaction effect between the additional forecast accuracy incentives and our treatments, then our estimated treatment effects are by definition similar, with or without incentives for forecast accuracy. In contrast, if there are significant interaction effects, then treatment effects systematically differ if firms have more incentives for accurate forecasts, which would imply that our main analysis might not generalize to firms with more incentives to forecast more accurately.

Table A.3 reports our findings from the introduction of incentive pay for accurate forecasts. The variable "Incentive Treatment" is a dummy that is one after the introduction of our bonus payment for accurate forecasts. This allows us to estimate the effect of interest similar to a difference-in-difference specification. As can be seen in Table A.3, none of the interaction effects are significant at conventional levels. We therefore fail to find evidence that our results might not be valid for samples of firms with larger incentives for forecast accuracy.

Additionally, we re-estimate our key findings regarding engagement with scientific learning in Table A.4. As before, we measure engagement with the string length of free form text responses and instrument engagement with the random scientific learning treatment. We evaluate the importance of incentives by interacting scientific learning engagement with the incentive treatment variable. The corresponding instrument for this interaction variable is the interaction of the incentive treatment and the scientific learning treatment. As Table A.4, our main results on from Table 7 continue to hold. Importantly, none of the interaction effects of the incentive treatment and engagement with scientific learning or testing are significant for overestimation. There is some evidence that the incentive treatments attenuated the effect of scientific learning engagement on precision error, but the overall results are very similar to Table 7.

A possible objection to this conclusion might be that our incentives were not high stakes enough to matter. This point is reinforced by the incentive treatment used Bloom et al. (2025), which varied amounts of up to \$400 to reward entrepreneurs for forecasts within 10% of their actual revenue growth. Bloom et al. (2025) find that higher incentives induce entrepreneurs to reduce their biases. However, we believe that the use of business targets mitigates this issue, because entrepreneurs already have a strong incentive to avoid systematic forecast errors in business targets. There is also a variety of evidence, which suggests that the impact of incentives on behavioral biases is

Table A.3: Interaction of Learning Treatments and Incentive Pay

	Forecast Error	Noise	Precision Error
Error Reminder Treatment	0.5764	0.3551	-2.6568*
	(1.0628)	(0.9480)	(1.3811)
Error Reminder Treatment × Incentive Treatment	-0.5852	-2.3423	0.3687
	(1.9445)	(1.4377)	(0.9481)
Scientific Learning Treatment	3.6546***	0.0481	-4.2918**
	(1.3196)	(1.4378)	(1.6908)
Scientific Learning Treatment × Incentive Treatment	-2.3814	-0.8401	1.5206
	(2.1399)	(1.6936)	(1.1074)
Time FE?	YES	YES	YES
R-squared	0.00	0.01	0.01
Number of firms	1248	1248	1282
Number of observations	8,210	8,210	10,371

Notes: Forecast error $\xi_{i,t+1}$ is measured as difference between monthly revenue growth forecast $g^f_{i,t+1}$ and actual revenue growth $g_{i,t+1}$: $\xi_{t+1} = g^f_{i,t+1} - g_{i,t+1}$. Noise is defined as the absolute value of the forecast error $|\xi_{t+1}|$ The precision error is defined as $\omega_{i,t} = \left(\frac{P_{90,i}-P_{10,i}}{2.65}\right) - \left(\frac{P^f_{90,i}-P^f_{10,i}}{2.65}\right)$, where $P_{x,i}$ denotes the percentile x of monthly growth, and $P^f_{x,i}$ the subjective percentile x of monthly growth at month t for firm i. Incentive treatment is a dummy that is one from October 2021 onwards. All specifications include a full set of time fixed effects to control for changes due to the COVID-19 pandemic. Standard errors are clustered at the firm level.

limited. Camerer and Hogarth (1999) provided an early survey for laboratory experiments and recent work by Enke et al. (2022) has shown that even incentives that correspond to a month's pay are mostly unsuccessful in de-biasing participants in lab experiments. Furthermore, there are many empirical studies of high-stakes field settings that consistently document biases at highly educated and trained subjects, such as stock traders (Daniel and Hirshleifer (2015)), CEOs of major corporations (Malmendier and Tate (2005), Malmendier and Tate (2015)) and CFOs of public firms (Ben-David et al. (2012), Boutros et al. (2020)).

Table A.4: Robustness: Incentive Treatments

	Forecast Error	Precision Error	Forecast Error	Precision Error
Theory Engagement	2.4213***	-2.8172**		
	(0.8819)	(1.1051)		
Theory Engagement × Incentive Treatment	-1.8077	1.4202**		
	(1.2132)	(0.6510)		
Testing relative to Theory			-3.9947***	4.4696**
			(1.4952)	(1.7604)
Testing relative to Theory × Incentive Treatment			2.9645	-2.1517**
			(2.0391)	(1.0686)
Time FE?	YES	YES	YES	YES
Stock-Yogo Weak Identification Test	158.43	193.84	51.36	65.22
Kleibergen-Paap Underidentification Test	186.19	221.66	85.45	106.14
Number of firms	802	827	802	827
Number of observations	5,243	6,647	5,243	6,647

Notes: Forecast error $\xi_{i,t+1}$ is measured as difference between monthly revenue growth forecast $g^f_{i,t+1}$ and actual revenue growth $g_{i,t+1}$: $\xi_{t+1} = g^f_{i,t+1} - g_{i,t+1}$. The precision error is defined as $\omega_{i,t} = \left(\frac{P_{90,i}-P_{10,i}}{2.65}\right) - \left(\frac{P^f_{90,i}-P^f_{10,i}}{2.65}\right)$, where $P_{x,i}$ denotes the percentile x of monthly growth, and $P^f_{x,i}$ the subjective percentile x of monthly growth at month t for firm i. Theory consists of basic idea of the business, definition of problems preventing the idea from being more successful, solution approaches, definition of conditions under which the core idea of the business will be more successful, and definition of tests to validate or falsify conditions for success. Testing captures description of empirical tests that firm conducted to test conditions for success of the core business idea. All measures are normalized (zero mean, unit variance). Sample excludes firms with Error Reminder Treatment. Standard Errors are clustered at the firm level.

A.8 Use of Business Targets as Forecasts

Another potential issue with our analysis is the use of business targets as main proxy for forecasts. Entrepreneurs might use formal business targets to motivate employees and might therefore tend to be more optimistic than their best guess of revenue growth. On the other hand, business targets that are unrealistically high might induce employees to exert less effort rather than more.

To address potential issues with the use of business targets, we explicitly asked respondents to differentiate between their best forecast for revenue growth and business targets as we highlighted in the discussion of Figure 2. To test the robustness of our main analysis to the use of business targets, we focus on the sample of firms for which business targets and the best forecast are the same.

The results in Table A.5 show that our main results about the effect of engagement with scientific learning and testing relative to theory are robust to the use of business targets.

Table A.5: Robustness: Sample of entrepreneurs for which business goal and best guess for revenue growth is the same

	Forecast Error	Precision Error	Forecast Error	Precision Error
Theory Engagement	1.6805**	-2.0147**		
	(0.6824)	(0.9871)		
Testing relative to Theory			-2.7061**	3.2117**
			(1.1120)	(1.5722)
Time FE?	YES	YES	YES	YES
Stock-Yogo Weak Identification Test	221.54	235.55	132.25	161.15
Kleibergen-Paap Underidentification Test	145.59	155.39	103.96	123.05
Number of firms	742	786	742	786
Number of observations	4,316	5,078	4,316	5,078

Notes: Forecast error $\xi_{i,t+1}$ is measured as difference between monthly revenue growth forecast $g^f_{i,t+1}$ and actual revenue growth $g_{i,t+1}$: $\xi_{t+1} = g^f_{i,t+1} - g_{i,t+1}$. The precision error is defined as $\omega_{i,t} = \left(\frac{P_{90,i} - P_{10,i}}{2.65}\right) - \left(\frac{P^f_{90,i} - P^f_{10,i}}{2.65}\right)$, where $P_{x,i}$ denotes the percentile x of monthly growth, and $P^f_{x,i}$ the subjective percentile x of monthly growth at month t for firm i. Theory consists of basic idea of the business, definition of problems preventing the idea from being more successful, solution approaches, definition of conditions under which the core idea of the business will be more successful, and definition of tests to validate or falsify conditions for success. Testing captures description of empirical tests that firm conducted to test conditions for success of the core business idea. All measures are normalized (zero mean, unit variance). Sample excludes firms with Error Reminder Treatment. Standard Errors are clustered at the firm level.

A.9 Sample Attrition

Most of the incentives in our study were provided to reduce sample attrition. Nevertheless, sample attrition cannot be avoided. From April 2021 to August 2021, we averaged 920 responses per month, which fell to 850 responses from September 2021 to March 2022. In other words, the degree of sample attrition was quite moderate.

To evaluate to what degree sample attrition might drive our results, we focus on the sample time frame from March 2021 to August 2021 and re-estimate our main results of the impact of Scientific Learning on overconfidence. The results are presented in Table **??** and show that stronger results. This suggests that sample attrition is likely to bias results against us finding any effect.

Table A.6: Scientific Learning impact for subsample of first half of sample time (March 2021 to August 2021)

	Forecast Error	Noise	Precision Error
Scientific Learning Treatment	3.2381***	0.4392	-4.0866***
	(1.0989)	(1.0646)	(1.5622)
Constant	3.2332***	30.0402***	20.9498***
	(0.7073)	(0.6468)	(0.9653)
Time FE?	YES	YES	YES
R-squared	0.0051	0.0034	0.0135
Number of firms	770	770	803
Number of observations	3,143	3,143	3,651

Notes: Forecast error $\xi_{i,t+1}$ is measured as difference between monthly revenue growth forecast $g^f_{i,t+1}$ and actual revenue growth $g_{i,t+1}$: $\xi_{t+1} = g^f_{i,t+1} - g_{i,t+1}$. Noise is defined as the absolute value of the forecast error $|\xi_{t+1}|$ The precision error is defined as $\omega_{i,t} = \left(\frac{P_{90,i}-P_{10,i}}{2.65}\right) - \left(\frac{P^f_{90,i}-P^f_{10,i}}{2.65}\right)$, where $P_{x,i}$ denotes the percentile x of monthly growth, and $P^f_{x,i}$ the subjective percentile x of monthly growth at month t for firm i. All specifications include a full set of time fixed effects to control for changes due to the COVID-19 pandemic. Standard errors are clustered at the firm level.

A.10 Hybrid Entrepreneurship

Another potential concern might be that many of the entrepreneurs in our sample are only devoting limited attention to the business we are surveying. This could be the case, if they pursue their business primarily to supplement their income through flexible "gig work" or "hybrid entrepreneurship" Folta et al. (2010) or for the option value of the business (Manso, 2016). A related possibility would be that the entrepreneurs have several businesses and only devote limited time to every single one of them. To address this concern, we collected data on how many hours per week the entrepreneurs devote to the business we survey. About 70% state that they devote more than 35 hours per week to the surveyed business. We therefore re-run our main results on the sample of entrepreneurs devoting at least 35 hours per week to the surveyed business.

Table A.7 shows that our main results about how engagement with overall scientific learning or testing relative to theory, remains robust in the full-time work sample.

Table A.7: Robustness: Sample of entrepreneurs working at least 35 hour per week in focal business

	Forecast Error	Precision Error	Forecast Error	Precision Error
Theory Engagement	2.1508***	-3.1594***		
	(0.7717)	(1.1347)		
Testing relative to Theory			-4.3124***	5.9451***
			(1.6395)	(2.1153)
Time FE?	YES	YES	YES	YES
Stock-Yogo Weak Identification Test	116.49	132.12	46.13	70.95
Kleibergen-Paap Underidentification Test	76.36	86.51	39.67	58.44
Number of firms	518	540	518	540
Number of observations	2,981	3,750	2,981	3,750

Notes: Forecast error $\xi_{i,t+1}$ is measured as difference between monthly revenue growth forecast $g^f_{i,t+1}$ and actual revenue growth $g_{i,t+1}$: $\xi_{t+1} = g^f_{i,t+1} - g_{i,t+1}$. The precision error is defined as $\omega_{i,t} = \left(\frac{P_{90,i}-P_{10,i}}{2.65}\right) - \left(\frac{P^f_{90,i}-P^f_{10,i}}{2.65}\right)$, where $P_{x,i}$ denotes the percentile x of monthly growth, and $P^f_{x,i}$ the subjective percentile x of monthly growth at month t for firm i. Theory consists of basic idea of the business, definition of problems preventing the idea from being more successful, solution approaches, definition of conditions under which the core idea of the business will be more successful, and definition of tests to validate or falsify conditions for success. Testing captures description of empirical tests that firm conducted to test conditions for success of the core business idea. All measures are normalized (zero mean, unit variance). Sample excludes firms with Error Reminder Treatment. Standard Errors are clustered at the firm level.

A.11 Differential Industry Trends

Our main study data was collected from March 2021 to March 2022, shortly after vaccines for COVID-19 became widely available in all 50 US states. However it is well-known that the COVID-19 pandemic and associated (voluntary) social distancing affected industries differently. For example, in-person services and restaurants were negatively affected by the pandemic, while technology and e-commerce was positively affected. It might therefore be plausible that recovery dynamics after the pandemic differ across industries as well. To control for potential differential industry trends, we add a full set of industry-by-time fixed effects for our main IV analysis in Table A.8. The results in Table A.8 that our main results are unchanged even if we control for differential industry time trends.

Table A.8: Robustness: Differential industry trends

	Forecast Error	Precision Error	Forecast Error	Precision Error
Theory Engagement	1.4703**	-2.1114**		
	(0.6140)	(0.9636)		
Testing relative to Theory			-2.4343**	3.4316**
			(1.0248)	(1.5613)
Industry-by-Time FE?	YES	YES	YES	YES
Stock-Yogo Weak Identification Test	258.55	289.64	139.81	176.22
Kleibergen-Paap Underidentification Test	164.68	183.98	109.84	134.64
Number of firms	802	827	802	827
Number of observations	5,243	6,647	5,243	6,647

Notes: Forecast error $\xi_{i,t+1}$ is measured as difference between monthly revenue growth forecast $g^f_{i,t+1}$ and actual revenue growth $g_{i,t+1}$: $\xi_{t+1} = g^f_{i,t+1} - g_{i,t+1}$. The precision error is defined as $\omega_{i,t} = \left(\frac{P_{90,i}-P_{10,i}}{2.65}\right) - \left(\frac{P^f_{90,i}-P^f_{10,i}}{2.65}\right)$, where $P_{x,i}$ denotes the percentile x of monthly growth, and $P^f_{x,i}$ the subjective percentile x of monthly growth at month t for firm i. Theory consists of basic idea of the business, definition of problems preventing the idea from being more successful, solution approaches, definition of conditions under which the core idea of the business will be more successful, and definition of tests to validate or falsify conditions for success. Testing captures description of empirical tests that firm conducted to test conditions for success of the core business idea. All measures are normalized (zero mean, unit variance). Sample excludes firms with Error Reminder Treatment. Standard Errors are clustered at the firm level.

A.12 Normalizing revenue time window

As discussed in section 3, we asked participants to make revenue forecasts for the next 4 weeks, or roughly 28 days. To calculate realized revenue growth, we used reported monthly revenues in the main text. However, the median time between subsequent survey responses was about 31 days instead of 28 days.

A simple way to address this issue is to normalize the revenue growth outcomes to a 28 day time window. For this purpose, we use data on the reported revenues in combination with data on the number of days between responses to calculate the implied average daily revenue growth at the business. With these average daily revenue growth, we can then recalculate revenue growth to a 28-day horizon and then re-calculate forecast error.

The following tables show that all of our main results are robust to this rec-calculation of forecast errors.

Table A.9: Causal impact of Engagement with Scientific Learning

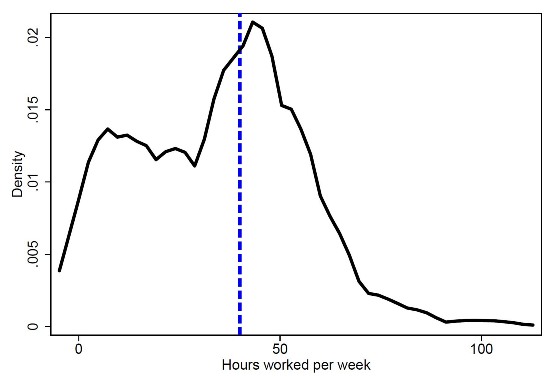
	Forecast Error	Forecast Error	Forecast Error
Overall Engagement with Scientific Learning	1.2369**		
	(0.5616)		
Testing relative to Theory		-2.0910**	
		(0.9521)	
Pre-Postmortem relative to Theory			-13.5516
			(9.7729)
Time FE?	YES	YES	YES
Stock-Yogo Weak Identification Test	238.43	134.77	3.05
Kleibergen-Paap Underidentification Test	154.81	106.25	3.10
Number of firms	802	802	802
Number of observations	5,243	5,243	5,243

Notes: Forecast error $\xi_{i,t+1}$ is measured as difference between monthly revenue growth forecast $g^f_{i,t+1}$ and actual revenue growth $g_{i,t+1}$: $\xi_{t+1} = g^f_{i,t+1} - g_{i,t+1}$. Engagement is measured by length of response (string length) to free-form textboxes, in which we ask about the reasoning behind responses to scientific learning questions. Scientific learning engagement consists of the normalized (zero mean, unit variance) engagement score in story, pre-postmortem and testing. Theory consists of basic idea of the business, definition of problems preventing the idea from being more successful, solution approaches, definition of conditions under which the core idea of the business will be more successful, and definition of tests to validate or falsify conditions for success. Pre-postmortem consists of internal firm conditions that might imply underperformance next month. Testing captures description of empirical tests that firm conducted to test conditions for success of the core business idea. All measures are normalized (zero mean, unit variance). Sample excludes firms with Error Reminder Treatment. Standard Errors are clustered at the firm level.

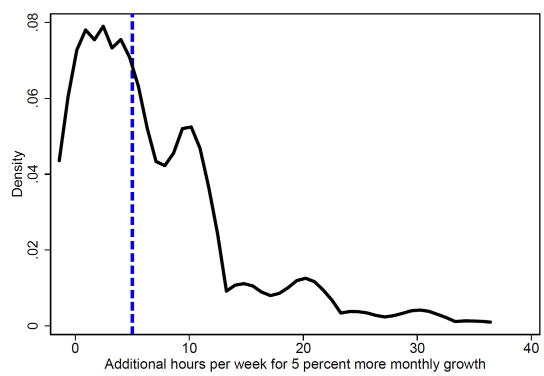
A.13 More details on data underlying the welfare analysis

Figure A.10 shows the key data components entering in our calculation of the wedge term in equation (12). Panel (A) of Figure A.10 displays the distribution of weekly work hours, which has a median of 40. Panel (B) reports the distribution of weekly hours responses to meet an additional 5% revenue growth goal, with a median of 5 hours per week or an additional hour per week for each percentage point higher sales growth per month. Panel (C) then shows the results of calculating the two components of the wedge term in (12). Overall, both terms are of similar importance and both terms exhibit a fat tail of values that are positive, suggesting large effects of calculating the wedge. To be conservative, we apply the correction implied by the wedge term only to entrepreneurs, which exhibit overestimation on average during the 13 months of our experiment.

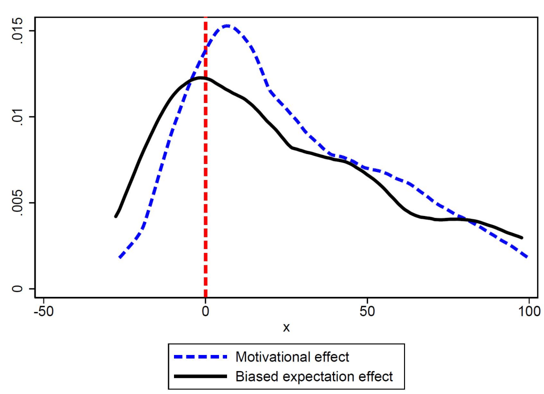
Figure A.10: Data underlying correction of subjective marginal profits



(A) Hours worked per week



(B) Impact of growth targets on work hours



(C) Components of wedge

Note: Panel (A) shows reported work hours per week. Panel (B) shows individual estimate of additional work hours per week, required for 5 percentage point higher monthly growth. Panel (C) exhibits two terms of wedge between subjective and rational expected marginal profit of entrepreneurial work. In panels (A)-(B) the median value is displayed with the vertical dashed lines.