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

Mu-Jeung Yang† Maclean Gaulin ‡ Nathan Seegert§

August, 2022
Preliminary and incomplete: Do not cite or circulate.

Abstract

Why do overconfident entrepreneurs fail to learn from frequent market feedback? Using two field experiments across almost 1,000 firms in Utah we follow for over a year, we explore the role of hindsight bias and causal misattribution. Both biases can potentially help sustain overconfidence psychologically, as hindsight bias creates a false memory of past mistakes while misattribution allows entrepreneurs to shift blame to external factors. We use two treatments to address biased memory and misattribution. First, under our "Error Reminder" treatment, entrepreneurs are shown past forecast errors to remove hindsight bias. Second, under our "Scientific Learning" treatment, we encourage entrepreneurs to develop a causal hypothesis about their firm and test this causal hypothesis empirically, to mitigate misattribution. We find that the Error Reminder treatment does not reduce overconfidence, because misattribution replaces hindsight bias to sustain overconfidence. In contrast, stronger engagement with hypothesis testing within scientific learning successfully reduces overestimation.

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

---

*We would like to thank Arnaldo Camuffo, Alfonso Gambardella, Josh Hausman, Holger Herz, Sara Holland, John Mondragon, Johannes Wieland, Matthew Wood, Todd Zenger and Florian Zimmermann and the seminar participants at the University of Oklahoma, 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. 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, mjyang@ou.edu, Department of Economics, University of Oklahoma, Cate Center One, 73072 Norman, Oklahoma

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

§nathan.seegert@eccles.utah.edu, David Eccles School of Business, University of Utah

1 Introduction

The existence of persistent entrepreneurial overconfidence is an enduring puzzle: Economists since Friedman (1953) and Becker (1962) as well as psychologists, such as Kahneman and Klein (2009) have argued that incentives and frequent feedback would induce entrepreneurs to learn and de-bias their beliefs. In contrast, recent work in behavioral economics on "motivated beliefs" argues that "wishful thinking" can impede learning, through biased memory or shifting blame for errors (see Benabou and Tirole (2016) for a survey of this literature).

We focus on two aspects of entrepreneurial overconfidence: overestimation of own sales growth and overprecision, defined as overconfidence about accuracy of own forecasts (Moore and Healy, 2008). To our knowledge, this study is the first mechanism field randomized control trial (RCT) (Ludwig et al., 2011; Congdon et al., 2017) providing direct evidence on the mechanisms psychologically sustaining entreprenuerial overconfidence.

Our randomized control trial is designed to (1) measure overconfidence and overprecision, (2) test the mechanisms that sustain them, (3) provide insights into the treatments that can overcome these mechanisms, and (4) develop a new methodology to calculate the welfare effects of biased entrepreneurial expectations, allowing for a motivating effect of overconfidence (Benabou and Tirole, 2002).

We collect unique and rich panel data from a set of approximately 1,000 entrepreneurs from Utah over the course of 13 months. These firms have a median workforce of 2 employees (excluding the founder), median age of 7 years, and 61% explicitly aspire to "profit maximization and growth." This contrasts with 24% of entrepreneurs in the Panel Study of Entrepreneurial Dynamics (PSED), who aspire to growth (Hurst and Pusley (2011)) and 12% of nascent entrepreneurs considering to start 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 overestimation and overprecision in our sample. Specifically, we find that entrepreneurs in the control group overestimate their next-month revenue growth by 5%, corresponding to a compounded annual revenue overestimation of 80%. Experience does not eliminate this bias: among entrepreneurs with firms that are at least 7 years old, the median monthly overestimation is still 4%. This persistent entrepreneurial overestimation is in contrast to no average overestimation among large firms in the Survey of Business Expectations (Barrero, 2022). 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. In addition, entrepreneurs are also overconfident about the accuracy of their estimates (overprecision). We asked entrepreneurs to report 80% confidence intervals for their revenue growth. These entrepreneurs reported 80% confidence intervals that are 21 percentage points narrower than statistical 80% confidence intervals, based on their realized revenue growth. This over precision by entrepreneurs is smaller and comparable to the 27.7 percentage point overprecision reported by Ben-David et al. (2012) for CFOs of major corporations. This overprecision also persists with experience as firms that are at least 7 years old exhibit overprecision by 23 percentage points.

Importantly, we show that entrepreneurs in the control group exhibit a high degree of hindsight bias, as entrepreneurs report that their past forecast errors were zero. We also show that biased memory and overestimation are systematically related. Specifically, individuals who recall making smaller forecast errors in the past also have higher degrees of overconfidence. These patterns are complementary to studies such as Zimmermann (2020) and Huffman, Raymond, and Shvets (2022) that have shown that biased memory sustains overplacement, defined as overconfidence about one's own rank relative to peers.

Our RCT design entailed randomizing firms into three groups; a control group, an error reminder treatment, and a scientific learning treatment. Firms remained in their group throughout the study and were unaware of the other groups. Our first treatment provides entrepreneurs with information about their past revenue forecast errors every month. This error reminder intervention targets biased memory, albeit in a light touch way.

Our second treatment prompts entrepreneurs to once a month develop and test hypotheses in a structured way to learn scientifically. This scientific learning intervention targets misattribution, which is an alternative mechanism that could sustain overconfidence. Misattribution is defined as the tendency to blame overestimation and the related underperformance on external factors rather than recognizing mistakes. Indeed, Hirshleifer (2001) argues that misattribution "causes individuals to learn to be overconfident rather than converging to an accurate self-assessment." Our treatment approach builds on the literature on structured management practices (Bloom and Van Reenen, 2007) and scientific learning in managerial (Yang et al., 2020) and entrepreneurial (Camuffo et al., 2019) decision-making.

Our RCT provides three results. First, our error reminder treatment is ineffective in reducing entrepreneurial overconfidence. This empirical result is consistent with the ineffectiveness of reminders of long sales histories by Bloom et al. (2022) on a large sample of internet entrepreneurs. We go further than Bloom et al. (2022) and provide evidence on why this is the case: entrepreneurs replace hindsight bias with misattribution to psychologically sustain overconfidence. Specifically, while misattribution is not significantly correlated with overestimation in the control group, it becomes highly significantly correlated with overestimation in the error reminder treatment group. This suggests that wishful thinking by entrepreneurs is indeed subject to limits as argued by models such as Caplin and Leahy (2019). At the same time, entrepreneurs seem willing to exert psychological effort to sustain overestimation via more use of misattribution in the face of objective information about past forecast errors.

Second, scientific learning can de-bias entrepreneurs if they engage. Our scientific learning treatment has two stages. 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 questions, seeAngrist et al. (1996), Angrist and Pischke (2009), Gerber and Green (2012). In the first hypothesis-development stage, entrepreneurs follow a structured script to explain the uniqueness of their business and to rigorously frame business problems. In the second testing stage, entrepreneurs are asked to test their hypotheses empirically. These stages target different biases. The first stage targets overprecision by encouraging entrepreneurs to consider potential outcomes. This first stage may also lead to more overconfidence because entrepreneurs may become more closely identified with their business (Belenzon et al., 2017) and put more weight on their contrarian perspective (Bernardo and Welch, 2004). The second stage targets overconfidence by encouraging entrepreneurs to consider the cause of different outcomes. It may also lead to more overprecision because entrepreneurs may feel more confident in their predictions. Consistent with these two stages, we find that entrepreneurs more strongly engaged with hypothesis development exhibit more overestimation and less overprecision. In addition, entrepreneurs more strongly engaged with hypothesis testing reduce their overestimation bias and increase overprecision. Overall, these results suggest that entrepreneurial overconfidence is not a fixed character trait and can be successfully influenced by structured practices.

Third, we document large profit gains from the scientific learning treatment for firms with "profit maximization and growth" as their main aspiration. Our estimates suggest that such opportunity-driven entrepreneurs boost their profits by an average of over \$100,000 per month, which is a very large effect, given that the average monthly profit for this group is roughly \$150,000. These profit results mirror large effects on revenue found by training programs based on scientific learning, such as Camuffo et al. (2019). At the same time, we find no significant effects for other entrepreneurs, including those with non-pecuniary main objectives, such as "personal or social goals" (Hurst and Pusley, 2011). These results are consistent with the zero or insignificant effects typically found 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 successful entrepreneurial training programs or subsides (Hurst and Pusley, 2011; Fairlie and Fossen, 2019) and reinforce the finding that interventions can be very effective in boosting high-growth entrepreneurship as shown by McKenzie (2017).

We use our experimental findings to develop a new methodology to assess the welfare effects of entrepreneurial overconfidence through hourly labor supply decisions. We begin with our measures of overestimation and realized profit. Then, we calculate entrepreneur-specific measures of the marginal benefit of entrepreneurship labor. We define this benefit as the present value of the expected rational marginal profit. This benefit measure not only corrects for biased expectations but also allow for the motivating effect of overconfidence, as in Benabou and Tirole (2002) and Compte and Postlewaite (2004).

Our welfare analysis suggests that most entrepreneurs are overworked. The median entrepreneur reports a marginal profit of \$3 per our. Our rational marginal profit measure implies a median loss for this entrepreneur of \$70 per hour. This is consistent with laboratory experimental evidence by Gish et al. (2019), who find that sleep deprivation causes inefficient decision-making by entrepreneurs. We show that in our context, entrepreneurial welfare could increase by roughly 20% of median monthly profits (\$1,000 per month) if we removed the median amount of overestimation. These welfare effects are a sizeable lower bound for the overall entrepreneurial welfare implications from overconfidence because labor supply is only one of many adversely affected margins (such as hiring and investment).

Our evidence on the welfare effects of entrepreneurial overconfidence stands in stark contrast to the theoretical predictions by Becker (1962), who stated that "irrational firms are limited by a budgetary constraint. (...) firms could not continually produce, could not "survive," outputs yielding negative profits, as eventually all the resources at their disposal would be used up." We find that instead of being competed out of the market, overconfident entrepreneurs tend to have excessively high labor supply, which increases accounting profits and therefore makes market survival of overconfident entrepreneurs *more* likely. However, in terms of welfare, each additional hour of work is very costly in terms of the opportunity cost of work, e.g., from direct disutility of work or the foregone utility of more leisure time. In other words, economic marginal profits from more work are negative, although accounting profits are positive. And because market survival depends on accounting profits, overconfident entrepreneurs are not forced to exit by market competition. Although this logic is consistent with the persistence in entrepreneurship despite low returns, as documented by Hamilton (2000), Moskowitz and Vissing-Jorgensen (2002) and Hall and Woodward (2010), ours is the first empirical paper to quantify the importance of overconfidence for 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 an empirical evaluation of the importance of overconfidence (Astebro et al., 2014).

Related Literature

Our study is related to at least three strands of literature in economics and entrepreneurship. The first literature is the theoretical and empirical work on understanding overconfidence and motivated beliefs, such as Benabou and Tirole (2002), Compte and Postlewaite (2004) Malmendier and Tate (2005), Ben-David et al. (2012), Malmendier and Tate (2015),Boutros et al. (2020), Benabou and Tirole (2016), Caplin and Leahy (2019), Zimmermann (2020), Huffman, Raymond, and Shvets (2022). The empirical studies most closely related to ours are Zimmermann (2020), which is a laboratory experiment and Huffman, Raymond, and Shvets (2022), which is an observational field study. Both of these studies focus on overplacement as opposed to overestimation and overprecision and neither of these studies conducts a field experiment or analyzes entrepreneurs.

Our work is also related to the literature on Management Practices (Bloom and Van Reenen, 2007), Strategy Practices (Yang et al., 2020) and Data-Driven Decision-Making (McElheran et al., 2022) and their role in expectation formation in firms, such as Altig et al. (2020), Coibion et al. (2020), Bloom et al. (2021), Barrero (2022) and Bloom et al. (2022). The study closest to ours is Bloom et al. (2022), which focuses on "noise", defined as squared forecast error, instead of overestimation and overprecision. Furthermore, the treatments in Bloom et al. (2022) are focused on providing incentives for accurate forecasts and providing training on forecast heuristics, rather than understanding the psychological mechanisms, that sustain overconfidence.

Our scientific learning treatment is close in spirit to recent work in entrepreneurship on scientific learning and experimentation, see Felin and Zenger (2009), Ries (2011), Kerr et al. (2014), Camuffo et al. (2019), Konings et al. (2022), Coali et al. (2022). The study closest to ours is the paper by Coali et al. (2022), who randomize a multiple week entrepreneurship training program on scientific learning for early stage startups. Beyond differences in the sample of firms (early startups vs relatively mature entrepreneurial firms) and treatment types (training sessions vs repeated structured nudges), they mainly provide indirect evidence on a de-biasing effect from scientific learning. Coali et al. (2022) identify a de-biasing effect by imposing a Heckman sample selection model and assume that any de-biasing effect is time-varying while 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.

2 Firm Setting and Recruiting

Our study was conducted from December 2020 to March 2022, with the core data collection and treatments being active from March 2021 to March 2022. Due to previous work we conducted for the Utah State governor and the Utah State Legislature, we had the cooperation of both government bodies as well as the State Chamber of Commerce. The cooperation and endorsement by these organizations was key to attract a deep and large pool of potential survey participants.

Our study context is attractive to study entrepreneurship for several reasons. Utah is among the most economically diverse states within the US, see Benway (2020). This enabled us to collect a sample of entrepreneurs from a variety of industries, instead of only sampling technology or e-commerce companies as in Bloom et al. (2022). Indeed the median firm in our sample reports no sales from e-commerce. Furthermore, the use of data on entrepreneurs from Utah is especially useful to study entrepreneurial overconfidence. Utah residents are consistently found to be very optimistic about the future in public polling[1]. If anything, a high level of optimism might be helpful for our study, because it suggests that any treatments that can successfully de-bias entrepreneurs in Utah might be even more effective in less optimistic states.

On the other hand, the time window of our experiment coincided with the ongoing COVID-19 pandemic, which was a time of elevated economic uncertainty, see Meyer et al. (2022). To some degree, this could be considered an attractive feature of our study, as entrepreneurship studies since Knight (1921) have argued that dealing with uncertainty and risk is a key function of entrepreneurship. At the same time, business uncertainty was on the decline since summer 2020 and stayed at a relatively stable level during our study, see Figure 1, which is from Meyer et al. (2022). Importantly, 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. Nevertheless, in our analysis, we take a cautious approach and include a full set of time fixed effects in any regression specification to control for the overall effects of changes in uncertainty due to the going COVID-19 pandemic.

2.1 Recruiting, Pilot Survey and Sample Characteristics

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.

[1]See for example Gallup polling: https://news.gallup.com/poll/189140/utah-residents-positive-state-economy.aspx. It should also be noted that optimism is a related but distinct psychological concept, defined as the degree an individual thinks that "good things will happen".

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 2. 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".

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 2. 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.

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 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 a follow up survey. 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 3 shows the distribution of firms across 2-digit NAICS industries in our initial sample of 1,067 companies in March 2021. As expected, our sample includes firms from a wide variety of industries, including health care, retail and even manufacturing and information technology.[2]. Additionally, Figure 4 displays the firm size distribution of our sample in terms of revenue. Most of the firms in the sample are small to medium sizes, and are therefore well approximated by a log normal distribution. Table 1 displays key summary statistics for the initial March 2021 sample. The first row shows that the median firm has monthly revenues of about \$15,000, while the average firm has much larger revenues at \$144,000, which suggests the presence of a few very large firms in our sample. 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).

An important question raised by studies such as Hamilton (2000) and Hurst and Pusley (2011), is whether entrepreneurs have non-pecuniary motives for running a business, in which case, they might not be motivated to forecast their revenues well. Rows 4-6 of Table 1 display long-term business goals, as stated by the entrepreneurs. In response to the question "What are your businesses' long term goals?", we offered three possible responses: (1) "Profit maximization and growth", (2) "Enough profit to sustain livelihood, (..) but no growth plans" and (3) "Personal or social goals other than profit or growth". The results in Table 1 show that only 12% of firms in our sample have explicitly non-percuniary motives for running a business. In contrast, 61% elect "Profit maximization and growth" as their main goal.

[2]There exists a regional technology cluster in Utah, called "Silicone Slopes". Obviously, there is much better data to study industries, such as internet entrepreneurship, such as Bloom et al. (2022). On the other hand, the variety of industry backgrounds is a strength of our sample.

Therefore most of our sample can be characterized as "opportunity-driven entrepreneurs" (Kerr et al., 2018). We also note that although 27% of entrepreneurs describe their long-term goals as "Enough profit to sustain livelihood", these firms still have an incentive to make somewhat accurate sales forecasts, to ensure that the sustainability of their firm is not in danger.

2.2 Survey Incentives to minimize Sample Attrition

To reduce the impact of sample attrition and preserve the panel dimension of our data as much as possible, we provided respondents incentives for continued participation. Specifically, the beginning of their survey screen included the following text:

> **"What are the possible benefits from being in the project?**
> You will help Utah to recover from the pandemic and get Utahns get back to work. Additionally, as an expression of our gratitude for your time, you will receive a $20 gift card if you complete the survey. This study is also a year-long survey of Utah businesses. For every 6 surveys you participate in, you will receive an additional $50 gift card."

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 increased participation incentives.

3 Measurement and Documenting Biases

3.1 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.

3.1.1 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 the monthly frequency. Bloom et al. (2022) overcome this issue by teaming up with a online payment processing firm, but this restricts their sample to firms with significant e-commerce presence, while the median e-commerce share of sales for our sample is zero.

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.". 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). 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, we show that all of our core results are robust to normalizing sales growth rates to a 28 day period.

Our use of entrepreneurial revenue growth has also two additional advantages. On the one hand, using revenue growth makes measurement robust to permanent misreporting at the individual level. For example, suppose 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 automatically "differenced out" by considering revenue growth.[3] On the other hand, another potential issue might be that entrepreneurs over-report their revenue growth as a result of "social desirability bias" in our survey and the related wish to seem more successful than they are. However, such misreporting would distort reported revenue growth upwards and thereby bias results against us finding positive forecast errors

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

and overestimation.

3.1.2 Forecasts

Figure 5 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.[4]

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:

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

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 6 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

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

expectations of revenue growth. In other words, equation (1) holds for entrepreneurial expectations, which is in stark contrast to stock market expectations as shown by Shiller (1981). At the same time Figure 6 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:

$$g_{i,t+1} = b \cdot g^f_{i,t+1} + D_i + e_{i,t+1} \tag{2}$$

where $e_{i,t+1}$ is a mean zero, iid error term and D_i is a firm fixed effect. Regression (2) 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 $b < 1$.[5]

The first row in Table 2 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

[5] To see this, we can solve (2) for the forecast error $g_{i,t+1} - g^f_{i,t+1}$ and take expectations to obtain: $E\left[g_{i,t+1} - g^f_{i,t+1}\right] \propto \left(\frac{1}{b} - 1\right) \cdot E[g_{i,t+1}]$ which is positive if $E[g_{i,t+1}] > 0$.

that expectations are indeed rational. This result in Table 2 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 mean reversion effects in revenue growth, as lagged growth remains statistically significant.

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. (2022).

3.2 Documenting Biases

We follow Moore and Healy (2008) and Astebro et al. (2014) and distinguish between three types of overconfidence. Overestimation is overconfidence about the growth rate of the entrepreneur's own business. This is the main measure of overconfidence we use and we measure it using forecast errors. In contrast, overprecision refers to overconfidence about the accuracy of own forecasts, see also Moore et al. (2015). A third type of overconfidence is overplacement and it refers to overconfidence about the own rank relative to peers. We do not analyze this type of overconfidence, as it has received much attention in the current literature, see Camerer and Lovallo (1999), Zimmermann (2020), Huffman, Raymond, and Shvets (2022).

We begin our documentation of biases in Figure 7, which shows the distribution of forecast errors in solid blue. There are two benchmarks in this figure. The first benchmark is the vertical red dashed line for zero forecast error. Measured by this benchmark, entrepreneurs are systematically overconfident. The median forecast error for the entrepreneurs in the control group is 5% (before October 2021). This is a very large forecast error, which implies an annual overestimation of sales growth by almost 80%. Furthermore, this forecast error is persistent in the sense that more experienced entrepreneurs are not de-biased. Entrepreneurs with firms that are at least 7 years old still exhibit a median monthly forecast error of 4%. This is substantial, given that over 80% of newly founded firms fail within their first 7 years (Fairlie and Miranda, 2017) and these experienced firms are therefore among the top 20% of startups.

The second benchmark is the grey dashed line, which displays the distribution of forecast errors if entrepreneurs used simple adaptive expectations. Under adaptive expectations, entrepreneurs simply extrapolate their current revenue growth rate to the growth rate next month. As can be seen in Figure 7, the distribution of forecast errors under simple adaptive expectations is symmetric around zero. This is in contrast to the distribution of entrepreneurial forecast errors, which is skewed positively. In other words, overestimation is not just a simple mean shift, but the result of disproportionately many overconfident forecasts.

Moving from overestimation to overprecision, we begin by denoting by $P_{x,i}$ the percentile x of monthly growth across all months for entrepreneur i and by $P^f_{x,i,t}$ 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, $\sigma^f_{g,i,t} \approx \frac{P^f_{90,i,t}-P^f_{10,i,t}}{2.65}$. These approximations are not important for any of our results, but they facilitate the interpretation of results. The degree of overprecision (or precision error) can therefore be defined as

$$\begin{aligned} \omega_{i,t} &= \sigma_{g,i} - \sigma^f_{g,i,t} && (3) \\ &= \left(\frac{1}{2.65}\right) \cdot \left[P_{90,i} - P_{10,i} - \left(P^f_{90,i,t} - P^f_{10,i,t}\right)\right] && (4) \end{aligned}$$

The distribution of entrepreneurial overprecision is displayed in Figure 8, again focusing on the control group and the time period before the introduction of incentives for accurate forecasting. Figure 8 shows that the vast majority of entrepreneurs in our sample exhibits overprecision, that is, the stated confidence intervals of their monthly growth forecasts are much narrower than the dispersion of the growth outcomes over time. The extent of overconfidence in stated forecast accuracy is also very big, as the median precision error is 21 percentage points. This is a bit smaller than the 27.7 percentage point overprecision error reported by Ben-David et al. (2012) for CFOs of public corporations, but is still quite comparable. This also suggests that the entrepreneurs in our sample are not unusually overprecise. Our findings on precision error are also consistent with separate literatures in psychology and economics, that document the robustness of overprecision. Indeed, Moore et al. (2015) describe overprecision as the most robust form of overconfidence and quite distinct from overestimation and overplacement. 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)).

We are also interested, in whether entrepreneurs who exhibit especially large forecast errors also tend to be excessively certain about their forecasts. Such a phenomenon has been documented in psychology by Kruger and Dunning (1999). Panel A of Table 3 analyzes the correlation of overestimation and overprecision and confirms the presence of the Dunning-Kruger effect in our sample of entrepreneurs. The dependent variables are either forecast errors for overestimating or underestimating entrepreneurs, or noise, defined as the absolute value of forecast errors. We find, that while both, large positive and large negative forecast errors are correlated with overprecision, the correlation of overestimation and overprecision is stronger than the correlation of underestimation and overprecision. Thus, while entrepreneurs who provide the worst forecasts tend to be more certain about these forecasts, it is especially very overconfident entrepreneurs, who are very certain about their forecast ability.

3.3 Mechanisms potentially sustaining overconfidence

We focus on two mechanisms that could potentially sustain overconfidence, even in the presence of frequent market feedback. The first of these mechanism is biased memory, which has been theoretically related to overconfidence by Benabou and Tirole (2002). 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 overestimation or overprecision. The specific form of biased memory we have in mind is hintsight bias: the tendency of individuals to be excessively certain about their past ability to forecast. Hintsight bias is able to sustain overconfidence, since subjects can justify that they do not need from their past forecast mistakes, if they did not make any mistakes.

To document the presence of biased memory in the control group, we ask participants with provide us with an estimate of their forecast error for the past month. We then contrast this recalled forecast error with the actual forecast error for the control group, we previously showed in Figure 7. The results are shown in Figure 9, where we keep the color of the realized forecast error in blue and add the distribution of recalled forecast error in black. As can be seen from Figure 9, control entrepreneurs' recalled forecast errors are much more concentrated around zero. Indeed, the median recalled forecast error is zero. At the same time, the modal forecast error is slightly larger than zero, suggesting that many entrepreneurs recognize that they might be overconfident, but think that the degree of their overconfidence is very small.

A more formal approach to show the link between biased memory and overconfidence is provided in Panel B of Table 3. In this table, we measure biased memory as the absolute value of the recalled forecast error for the last month for the control group. Higher values of this absolute recalled error correspond to lower levels of hindsight bias. The main finding of Panel B in Table 3 is that lower absolute values of recalled forecast error are systematically correlated with more overestimation. In other words, more hindsight bias and more overestimation are linked at the individual level the same way that biased memory and overplacement are linked in Zimmermann (2020) and Huffman, Raymond, and Shvets (2022).

A second mechanism that might sustain overconfidence is causal misattribution (henceforth "misattribution"). For our purposes we define this mechanism as blaming external factors for own overconfidence or underperformance of forecasts. The basis for the measurement of misattribution is 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 overestimation). 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 their 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.

4 Experimental Design

Before discussing the details of our treatments, it is worth noting the general design idea. Our treatments are different than either training treatments, by studies such as Camuffo et al. (2019) or Bloom et al. (2022) or one-time nudges. Instead, the best way to describe our treatments is "light touch, but persistent." Our treatments are nudges and therefore light touch, as we cannot force experimental subjects to engage with the treatments we provide. At the same time, they are persistent, as we only randomize the treatment assignment once, in March 2021 and keep entrepreneurs in their respective treatment or control group. As a result, they are nudged repeatedly (once every month) for 13 months to engage with our treatments.

4.1 Error Reminder Treatment

To counter the effects of biased memory, our first treatment reminds subjects 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 btwo 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.

4.2 Scientific Learning Treatment

This treatment includes the simple Error Reminder treatment, but adds additional layers to address the potential misattribution of causality. Specifically, we develop a treatment that primes entrepreneurs to develop and test hypotheses, the same way an empirical economist would. This approach is related to recent work applying scientific learning to different contexts, such as CEO decision-making (Lafley et al. (2012),Felin and Zenger (2017), Yang et al. (2020)), teaching students to think scientifically (Ashraf et al. (2022)) and of course entrepreneurial experimentation (Felin and Zenger (2009),Ries (2011), Camuffo et al. (2019), 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)

2. Pre-postmortem

3. Hypothesis testing, based on theory.

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[6]. (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?

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

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 basic idea of the scientific learning treatment is to nudge entrepreneurs to explicitly state the assumptions behind their "theory of competitive advantage". We define "competitive advantage" as "a strength your company has, which distinguishes you from your competitors and which is hard to copy". Such a competitive advantage is a key element even for elevator pitches for startups seeking VC investment, see Lerner et al. (2012). Our scientific learning treatment starts with inspiring entrepreneurs to think about their competitive advantage by considering contrarian beliefs, as such beliefs are often helpful for firms seeking competitive advantage (Felin and Zenger, 2009; Lafley et al., 2012). In this context, it should be noted that a framing encouraging entrepreneurs to seek out contrarian beliefs is likely to be associated with more overconfidence, as in the theoretical model of Bernardo and Welch (2004).[7] However, the elements following the initial framing, such as the definition of problems, potential solution approachs, statement of implicit assumptions and proposed tests of assumptions potentially have an attenuating effect on overconfidence. Appendix 1 provides more details on the questions we ask as well as the specific sandwich shop example we use to illustrate possible answers. For each of these questions, we ask respondents to provide written responses and use the length of these written texts to measure engagement with the treatment.

[7] Indeed in their model, overconfidence is defined as a higher reliance of entrepreneurs on private signals as opposed to publicly observable actions of other entrepreneurs.

The second part of the scientific learning treatment is the practice of "Pre-Postmortem", advocated by Klein (2007) and others. This practice is used to inspire respondents to anticipate potential problems, which in turn is intended to reduce overconfidence. Specifically, we ask:

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

To mirror the questions we use to measure misattribution, we respondents can use two checkboxes with a text entry: (1) "Reasons internal to the company (please specify)" and (2) "Reasons external to the company (please specify)". When measuring engagement with this pre-postmortem, we focus on pre-definition of internal reasons for failure, since this most likely offsets the misattribution bias we discussed in section 3.3.

The third part of the scientific learning treatment happens in the next survey round, in which we follow up with the theory part that entrepreneurs filled out previously. Specifically, at the beginning of the survey, we ask enrepreneurs in the scientific learning treatment:

> 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?

With the possible responses "No" and "Yes (please specify)." As before, we use the responses on this textbox to measure engagement with hypothesis testing.

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 described business as "high-end furniture and the use of exotic woods". She described 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 recognized 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." So she developed three alternative approaches to address this problem:

1. "One plan is to use targeted advertising in order to reach a wider audience."

2. "Another plan could be to create a cheaper alternative to the fine woodworking products I offer."

3. "alternatively, still create high quality products but redesign them to be cheaper to manufacture and then offer them at a lower price point."

For each of these three plans, she specified first conditions under which these plans would work and then proposed ways to test them. The plans and associated outcomes are displayed in Figure 10. For example, the usage of targeted advertising would only help if there are enough customers who are willing to pay high prices for high quality collectibles. To test the hypothesis that there are enough local customers for her products, she decides to run ads for her merchandise on Facebook (called "targeted advertising" in Figure 10). As can be seen in Figure 10, this test did not generate much demand, which left Olivia doubtful about her hypothesis as well as about the reliability of the signal in her feedback (Zellweger and Zenger (2021)).

On the other hand, the success of new, cheaper products will crucially depend on the price elasticity of the demand curve she is facing. To test the demand elasticity she is facing, Olivia decides to begin by offering a discount of some of her products (called "cut costs and prices of existing products"). This led to a substantial increase in demand, as seen in the second branch in Figure 10. However, Olivia was still unsure whether the demand boost was based on the discount alone, or whether the initial pricing of her products was too high. In other words, Olivia now faces an important identification problem. As a result, she creates new, cheaper products to offer, which is met with consistently higher sales, as can be seen in the last branch of Figure 10. The new products provide direct and reliable feedback about her hypothesis and about the price elasticity of demand, which is why Olivia ends up adding the cheaper products to her permanent product offerings.

To conclude this section on the treatment design, we show balance tests in Table 4. It should be noted that March 2021 was the first month of treatments, but the fact that balance tests are still confirming insignificant differences across treatment and control groups validates our randomization.

5 Results

5.1 Error Reminder Treatment

The baseline results of the Error Reminder treatment are displayed in Table 5. There we document that the Error Reminder treatment is basically ineffective in addressing either overestimation or overprecision, although the latter results are significant at a 10% level. This ineffectiveness of the Error Reminder treatment is consistent with findings by Bloom et al. (2022), who provide incentives to review past sales to a sample of internet entrepreneurs and find that these treatments are ineffective in reducing sales forecast errors. These results also seems surprising, not only given theoretical attention on the link between biased memory and overconfidence e.g. in Benabou and Tirole (2002), but also empirical evidence from lab experiments as in Zimmermann (2020) and field settings such as Huffman, Raymond, and Shvets (2022). However, it is worth emphasizing that both Zimmermann (2020) and Huffman, Raymond, and Shvets (2022) mainly focus on overplacement, while we study overestimation and overprecision.

To further investigate why the Error Reminder treatment is ineffective, we analyze the relation between misattribution and overestimation in Table 6. As can be seen, misattribution is basically uncorrelated with overestimation in the control group, which suffers from bias memory. However, in both treatment groups in which we removed biased memory, misattribution is highly correlated with overestimation. This is consistent with the replacement of hintsight bias to sustain overconfidence with misattribution. Importantly the correlation of misattribution and overestimation is not mechanical, as the two variables are insignificantly correlated with the opposite sign in the control group. Furthermore, the results in Table 6 suggest the Error Reminder treatment has an effect on entrepreneurs, even if this treatment did not de-biased 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 beliefs," in which highly deliberate people might be better at self-delusion, as documented by Kahan (2013). In contrast, the failure of the Error Reminder treatment to de-bias entrepreneurs is only surprising from the perspective of "System 1 biases," which are the result of 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 overestimation and overprecision biases.

At the same time, another possible explanation of the failure of the Error Reminder treatment to reduce overestimation and overprecision, is that these are permanent character traits of entrepreneurs, which cannot at all be impacted by nudge treatments of the kind we use in this study. The next section will provide evidence that is inconsistent with this view.

5.2 Scientific Learning: Access

In contrast to the Error Reminder treatment, the Scientific Learning treatment requires attention and effort to be effective. In this context, participants in the Scientific Learning treatment group have the choice to not engage at all with the material, which in turn means that there is the possibility of "one sided non-compliance" (Angrist and Pischke, 2009; Gerber and Green, 2012)[8]. We therefore begin our investigation of Scientific Learning treatment effects with an Intent-to-Treat (ITT) analysis. The effects from the ITT can best

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

be understood as reflecting the effect of access to (or the option to engage with) Scientific Learning. The next section will use Instrumental Variables (IV) to analyze the causal impact of engagement with Scientific Learning on overconfidence.

Table 7 collects our baseline results of access to Scientific Learning on overconfidence. Surprisingly, access to Scientific Learning increases overestimation, as documented in the first column of Table 7. However, this result is only surprising when primarily equating Scientific Learning with hypothesis testing. Instead, a large part of our Scientific Learning treatment consists of structured hypothesis or theory development in the service of finding a potential source of unique value. Such theorizing is likely to increase overconfidence, due to at least two different channels. On the one hand, entrepreneurs are more likely to place more emphasis on their own private information as opposed to "conventional wisdom", as in the model of Bernardo and Welch (2004). In fact, the overall framing of the scientific learning treatment starts by asking entrepreneurs to think about "contrarian" ideas. On the other hand, the identification of an entrepreneur with the firm is likely to increase, similar to the effects documented in Belenzon et al. (2017). Higher identification with the firm in turn, increases the perceived cost of underperformance, which is consistent with "stakes dependent belief" effects discussed in Benabou and Tirole (2016).

At the same time, access to Scientific Learning reduces overprecision, as seen in column 3 of Table 7. While the reduction in overprecision is far from completely debiasing entrepreneurs, it does reduce overprecision by about 15% on average ($0.1524 = 3.4/22.3$). This effect can also be understood from the logic of our Scientific Learning treatment, which asks 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?" In other words, entrepreneurs are asked to consider all the key conditions that have to be met for their plans to support revenue growth targets, which in turn can make them aware of the tenuousness of many assumptions.

Although the Scientific Learning treatment was designed to reduce misattribution, column 4 of Table 7 shows that it was ineffective in achieving this goal. This might be related to the fact that our measure of misattribution is very conservative and therefore captures only a few very strongly misattributing entrepreneurs.

5.3 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. Since the degree of engagement is endogenous, we follow common practice and use random assignment to treatments as IV, since it will be systematically correlated with engagement but will otherwise be exogenous, 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 8 shows the causal impact of overall engagement with Scientific Learning on overconfidence. Overall engagement in turn is measured by the sum of string lengths with all three parts 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 overestimation by 1.3 percentage points per month. In other words, more overall engagement leads to higher levels of overestimation. At the same time, a one standard deviation higher overall engagement with Scientific Learning also reduces overprecision by 1.9 percentage points as seen in column 3 of Table 8. These effects are consistent with the ITT analysis of the last section. They also reflect the fact that in terms of text response, the hypothesis development section is much longer than either the pre-postmortem or the hypothesis development sections, so that the effects of hypothesis development we discussed in the last section will dominate these estimates.

To contrast the different parts of the scientific learning treatment, we construct relative engagement measures. For example, for hypothesis testing we take the string length of responses to the hypothesis testing question and subtract the string length of all responses to the hypothesis development questions. The resulting measure then tells us how much more entrepreneurs were engaged with hypothesis testing relative to hypothesis development. Similarly, we compute the string length of responses to the internal factors[9] cited for the pre-postmortem and subtract the string length of the hypothesis development section. All relative engagement measures are normalized to have a unit standard deviation.

The first two columns of Panel B in Table 8 show that relatively more engagement with hypothesis testing significantly reduces overestimation. A one standard deviation higher relative engagement with hypothesis development implies a reduction in overestimation of 2.25 percentage points per month. Over the entire sample, the control group has a median forecast error of 3.8% per month, so the a one standard deviation increase in relative engagement with hypothesis testing reduces this bias by almost 60% (0.59 = 2.25/3.8). However, this de-biasing of overestimation goes hand in hand with an increase in overprecision. As the second column of of Panel B in Table 8 documents, precision error increases by 3.34 percentage points for every standard deviation increase in relative engagement with hypothesis testing compared to theory. This result is consistent with the view that the conduct of empirical tests suggests to respondents that they understand the sources of uncertainty in their business very well. Although this might be true for the specific part of their business for which they conducted a hypothesis test, this better understanding of risk is unlikely to be true for all possible sources of risk in their business, so that in the end, they become more overly confident in their own forecasts.

The last two columns of Panel B in Table 8 show that the IV effects for relative hypothesis testing are not mechanical. There, we estimate IV effects of relative engagement with the pre-postmortem and do not find significant effects. This analysis suggests that the use of pre-postmortem in our Scientific Learning nudges does not differentially impact overconfidence,

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

over and above the effects of hypothesis development.

5.4 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 1, one potential concern with such an analysis is that the dynamics of the COVID-19 pandemic might impact our estimates. As in the test of the analysis, we therefore include 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 11 highlights our main result from this analysis. The figure shows the evolution of forecast errors over time for the Scientific Learning treatment. Exposure to Scientific Learning initially strongly increases forecast error and therefore overestimation, but this effect slowly fades over time. In other words, although our Scientific Learning treatment increases overestimation over the entire sample, entrepreneurs eventually learn to adjust their forecasts and learn that they have been overconfident.

Table 9 presents the formal regressions results underlying Figure 11 as well as additional results for other outcomes. Column 3 of Panel A in Table 9 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 overestimation. While Scientific Learning increases the overestimation bias, and this bias slowly fades over time, the same treatment reduces overprecision and this effects is persistent over time.

The learning effects associated with Scientific Learning also contrast with the dynamic effects of the Error Reminder treatment, which are reported in Panel B of Table 9. Consistent with our estimates in Table 5, the Error Reminder treatment has no effect on overestimation. 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 9 also suggests that the Error Reminder treatment reduces overprecision, but not as strongly as the Scientific learning treatment. 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.

5.5 Profit Effects of Treatments

With this section we begin to investigate the broader welfare consequences of our treatments. Specifically, for 6 of the 13 months we collected data on monthly total operating costs and variable operating costs. We defined total operating costs as "expenses for the day-to-day running of your business, like rent or material costs." We also asked for variable operating costs using the question

> Breaking down your total operating costs from above into 'fixed' and 'variable' costs, how much are variable costs, which vary with sales volume? (Variable costs would include hourly wages, direct material costs, etc.)

Subtracting operating costs from revenues allows us to calculate monthly profits for all firms. We will estimate the effects of treatments on monthly profit levels, as we want to allow for the possibility that some of the firms in our sample make losses during the study time window.

Following Syverson (2011) we are especially interested in within-industry differences in firm performance, so we include industry fixed effects in the profit regressions. Furthermore, it is unlikely that our treatments will improve performance for firms that do not have profit maximization as main aspiration, as they might not seize opportunities to grow the business, see Hurst and Pusley (2011), Fairlie and Fossen (2019). We therefore use the data on the main goal of the business, which we collected during the December 2020 pilot survey as variable to be interacted with our treatments.

Table 10 reports the results from our profit regressions. It suggests that the Scientific Learning treatment systematically increased profits, especially at firms with profit maximization and growth as their main goal. The value of the estimated profit effects is large. Compared to the control group, Scientific Learning treatment group firms with profit maximization goals see their monthly operating profits increase by an average of $\$116,000 (116.66 = 151.07 - 34.41)$. This is a large effect, compared to the average monthly profit of $\$148,948$ for the profit-maximizing entrepreneurs in our sample. The effect is large, even compared to the average monthly revenue of $\$202,115$ for these firms. The effects on variable operating profit are even larger, as document in column 2 of Table 10. Additionally, the last two columns of Table 10 show that these results are neither driven by firm age or firm size, measured as revenue in the initial survey in March 2021. However, we also note that Table 10 also 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).

6 Robustness

6.1 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 in terms of purchasing too many or to few materials or hiring too many or to few employees.

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 5, 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 11 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 11, 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 12. As before, we measure engagement with the string length of free form text responses and instrument engagement with the random scientific learning treatment. As before, 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 12, our main results on from Table 8 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 8.

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. (2022), which varied amounts of up to $400 to reward entrepreneurs for forecasts within 10% of their actual revenue growth. Bloom et al. (2022) 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. As mentioned, systematic errors in business targets will lead to follow-on errors in materials purchases and hiring which will be very costly for entrepreneurs. There is also a variety of evidence, which suggests that the impact of incentives on behavioral biases is 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)).

6.2 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 official business targets to motivate employees and might therefore tend to me more optimistic than their best guess of revenue growth. On the other hand, excessively optimistic business targets 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 5. 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 13 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.

6.3 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 14 and show that if anything stronger results. This suggests that sample attrition is likely to bias results against us.

6.4 Part-time 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 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 15 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.

7 Extension: Welfare Analysis

In this section we develop a methodology to evaluate some of the welfare consequences from overconfident entrepreneurs. For this purpose, we focus on the intensive margin of labor supply from entrepreneurs, since this margin has been a key theoretical mechanism of how overconfidence impacts welfare since Benabou and Tirole (2002). The question we seek to answer is whether de-biasing an entrepreneur, for example using a scientific testing, would increase the welfare of that entrepreneur. The answer to this question is theoretically ambiguous, due to the confluence of two opposing forces. On the one hand, an overconfident entrepreneur might work more hours compared to a rational entrepreneur, despite a negative marginal profit. In this case, de-biasing the entrepreneur and reducing her work hours would increase her welfare. On the other hand, theoretical work since Benabou and Tirole (2002) and Compte and Postlewaite (2004) has shown that overconfidence can have positive welfare effects. For example, Benabou and Tirole (2002) have argued that the motivating effects of overconfidence might offset other behavioral biases, such as hyperbolic discounting. In this case, hyperbolic discounting leads to procrastination and low work hours despite positive marginal profits. A motivating effect of overconfidence can compensate the tendency to procrastinate and lead to more work and higher profit. In this context, de-biasing an entrepreneur would actually harm her welfare, as she returns to procrastinate work after reducing overconfidence.

The key empirical challenge in this context is to correct subjective estimates of marginal profit for the presence of overconfidence. Providing a comprehensive and assumption-free approach to achieve this is beyond the scope of this paper. Instead we develop a workable approach that can be used by researchers running their own field experiments using a few key additional survey questions, in combination with a small number of very strong assumptions.

7.1 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. These benefits of entrepreneurial labor supply compared to opportunity costs of $w_{O,i} \cdot h_i$, where $w_{O,i}$ is an hourly opportunity cost of work for i. The net expected profit from labor supply h_i can therefore be written as

$$\begin{aligned}\Pi^e_{S,i}(h_i) &= \pi^e_{S,i}(h_i) - w_{O,i} \cdot h_i \\ &= \left[\pi^e_{R,i}(h_i) - w_{O,i} \cdot h_i\right] + \epsilon_i \cdot \pi^e_{R,i}\end{aligned} \tag{5}$$

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 $\Pi^e_{R,i}(h_i) = \pi^e_{R,i}(h_i) - w_{O,i} \cdot h_i$ the expected rational (unbiased) profit net of opportunity costs. We show in the appendix, that expected profit changes from more labor supply can be approximated by

$$\Pi^e_{R,i}(h_{1,i}) - \Pi^e_{R,i}(h_{0,i}) \approx \left[\frac{\partial \Pi^e_{R,i}(h_{0,i})}{\partial h_i}\right] \cdot \left(\frac{dh_i}{d\epsilon_i}\right) \cdot (\epsilon_{1,i} - \epsilon_{0,i}) \tag{6}$$

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 (6) summarizes changes in net entrepreneurial welfare, defined as expected profit net of opportunity costs of time, as a result from changes in forecast errors, such as debiasing through intensive use of scientific hypothesis testing.

The key term in (6) is the rational expected marginal profit $\left[\frac{\partial \Pi^e_{R,i}(h_{0,i})}{\partial h_i}\right]$: if it 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). On the other hand, if this marginal profit is negative, then any additional work due to overconfidence will reduce welfare.

In the appendix, we show that this rational marginal profit term $\left[\frac{\partial \Pi^e_{R,i}(h_{0,i})}{\partial h_i}\right]$ can be calculated as

$$\frac{\partial \Pi^e_{R,i}(h_{0,i})}{\partial h_i} = \underbrace{\left[\frac{\partial \pi^e_{S,i}(h_{0,i})}{\partial h_i} - w_{O,i}\right]}_{\substack{\text{(i) Subjective} \\ \text{Marginal Profit}}} - \left\{ \underbrace{\frac{\pi^e_{R,i}(h_{0,i})}{\partial h_i / \partial \epsilon_i}}_{\substack{\text{(ii) Motivational} \\ \text{Effect}}} + \underbrace{\frac{\partial \pi^e_{R,i}(h_{0,i})}{\partial h_i} \cdot \epsilon_i}_{\substack{\text{(iii) Biased} \\ \text{Expectations}}} \right\} \qquad (7)$$

Equation (7) is our main measurement tool driving our welfare calculations. Before relating it to the needed measurement assumptions, it is worth discussing the economic intuition for (7). Rational marginal profit from more entrepreneurial work consists of three components. The first term on the right-hand side of (7) is the subjective marginal profit of work $\left[\frac{\partial \pi^e_{S,i}(h_{0,i})}{\partial h_i} - w_{O,i}\right]$. For a profit-maxiziming rational entrepreneur, this term should be zero. However, this term could be non-zero, reflecting potentially behavioral frictions such as weak willpower (Benabou and Tirole (2002)) or market frictions, such as credit constraints.

Our main concern is that overconfident entrepreneurs might perceive themselves as profit maximizing with a subjective marginal profit of zero, but in reality their expectations might be biased by overconfidence. Therefore the "wedge" term on the curly brackets in (7) corrects the subjective marginal profit for two effects. On the one hand, the term $\frac{\pi^e_{R,i}(h_{0,i})}{\partial h_i / \partial \epsilon_i}$ corrects for the motivational effect of overconfidence. If entrepreneurs are very responsive to overconfidence ($\partial h_i / \partial \epsilon_i$ is large), then this term will be smaller, as any perceived positive marginal profit will lead to a large increase in labor supply which will thereby reduce marginal profits under diminishing returns. Therefore, under very elastic labor supply, subjective marginal profit measures do not need to be corrected much. On the other hand, the term $\frac{\partial \pi^e_{R,i}(h_{0,i})}{\partial h_i} \cdot \epsilon_i$

corrects for biased expectations of the marginal benefits from work, using information on the forecast error ϵ_i. The more overconfident entrepreneurs are ($\epsilon_i > 0$ is larger) the more subjective marginal profits need to be corrected for this overconfidence.

To summarize the welfare effects in (7), note that even if subjective marginal profits are zero, the term in the curly brackets in (7) 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. At the same time, if subjective marginal profits in (7) are sufficiently positive, rational marginal profits will be positive as well, thereby implying welfare increases from more hours worked.

7.2 Measurement

We begin by measuring the first term in (7), the subjective marginal profit of more hours, $\left[\frac{\partial \pi^e_{S,i}(h_{0,i})}{\partial h_i} - w_{O,i}\right]$. We use the methodology of Altig et al. (2020) applied to expected profits from more work. Figure 12 shows the survey screen that was shown to participants to measure the present value of the benefits of additional work. After the questions in Figure 12, we asked entrepreneurs to convert these expected values into certainty-equivalent units to remove the influence of risk aversion by asking the question

> 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).

To measure the opportunity cost of time $w_{O,i}$, we ask respondents:

> 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?

For the remaining components of (7), 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.

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.

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.

Figure 13 shows the key data components entering in our calculation of the wedge term in equation (7). Panel (A) of Figure 13 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 (7). 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.

7.3 Welfare Results

The distributions of our measures of expected marginal profit are displayed in Figure 14. The distribution in grey is a kernel density estimate of the subjective marginal profit term $\frac{\partial \pi^e_{S,i}(h_{0,i})}{\partial h_i} - w_{O,i}$. It has a median value of \$2.90 per hour which is close to the red dashed zero line that is added as a reference point in Figure 14. This suggests that the median entrepreneur in our sample believes herself to be optimizing.

The blue distribution in Figure 14 reports the rational marginal profit, based on equation (7). It differs from the subjective distribution in that it has less mass concentrated around zero marginal profits and more mass in the left tail of the distribution - where entrepreneurs exhibit marginal losses from more work. Indeed, the median rational profit for our sample of entrepreneurs is \$70 per hour, which is sizable, but not unrealistically so. To put this number into perspective, the median opportunity cost of an hour of additional work is \$50 in our sample and therefore quite comparable in magnitude. 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. Figure 14 highlights that our rational marginal profit measures most correct the estimates of entrepreneurs who believe themselves to work an optimal amount. In other words, our correction does not reduce the number of entrepreneurs to believe themselves to have very high marginal profits, because these entrepreneurs do not exhibit much overestimation in our data.

Figure 15 illustrates the heterogeneity of the welfare effects from equation (7). The x-axis displays values of rational marginal profit $\left[\frac{\partial \Pi^e_{R,i}(h_{0,i})}{\partial h_i}\right]$ from the 25^{th} to the 75^{th} percentiles of the values in the data. The y-axis displays values for the labor response per percentage point monthly growth goal $\left(\frac{dh_i}{d\epsilon_i}\right)$, ranging from the 25^{th} to the 75^{th} percentiles for this variable. The combined plot gives values for the term $\left[\frac{\partial \Pi^e_{R,i}(h_{0,i})}{\partial h_i}\right] \cdot \left(\frac{dh_i}{d\epsilon_i}\right)$ to convey the heterogeneity of the implied welfare loss per week as result of different combinations of entrepreneurial labor supply responses and rational marginal losses. Welfare losses can range from almost zero to almost \$1,200 per week.

8 Conclusion

This study provides the first mechanism field RCT investigating the channels through which entrepreneurial overconfidence is psychologically sustained. Our findings are broadly consistent with the recent Behavioral economics literature on motivated beliefs and wishful thinking (Benabou and Tirole, 2016), applied to the important field setting of entrepreneurial sales forecasts. We find that relatively intensive engagement with structured practices (Bloom and Van Reenen, 2007; Camuffo et al., 2019; Yang et al., 2020), for scientific testing can successfully de-bias entrepreneurs. This 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. For example, how does scientific learning impact other potential behavioral biases of entrepreneurs, such as loss aversion (Kahneman and Tversky, 1979), the planning fallacy (Buehler et al., 1994), the sunk cost fallacy, and others. Furthermore, scientific learning is by its very nature a natural approach to deal with ambiguity (Knight, 1921) and complexity. The analysis of these additional dimensions will not only offer a better understanding of the entrepreneurial decision-making but also a broader appreciation of the effects of scientific learning.

Another avenue for future research is the exploration of the effects of scientific learning on entrepreneurial fundraising. Indeed, entrepreneurial overconfidence might not just have a motivating effect on effort and labor supply, but also the ability of entrepreneurs to persuade investors to fund them. A key question is therefore whether scientific learning can reduce entrepreneurial overconfidence, while at the same time providing entrepreneurs with the tools to better convince investors of the future potential of their startup.

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. In this study, we followed previous work by Lafley et al. (2012), Camuffo et al. (2019) and Yang et al. (2020) and used scientific learning, in the context of relatively mature firms. However, the most popular practitioner approach for early stage entrepreneurship is the "Lean Startup" methodology, see Ries (2011). This methodology is an alternative set of structured practices, which emphasize early customer validation of product ideas through "minimum viable products", without the emphasis on stating and testing assumptions as in the scientific learning approach (Felin et al., 2019). Like Felin et al. (2019) and Cao et al. (2020), we are cognizant of potential pitfalls of the lean startup approach to early stage entrepreneurship. However, we also believe that the effectiveness of different structured practices for early stage startups is ultimately an empirical question, which is left for future research.

References

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

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.

Barrero, J. 2022. The micro and macro of managerial beliefs. *Journal of Financial Economics* .

Becker, G. 1962. Irrational behavior and economic theory. *Journal of Political Economy* **70(1)** 1–13.

Belenzon, S., A. Chatterji, B. Daley. 2017. Eponymous entrepreneurs. *American Economic Review* **107(6)** 1638–1655.

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.

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. 2022. Rationalizing entrepreneurs' forecasts. *Working Paper, Stanford University* .

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. 2019. 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* .

Coali, A., A. Gambardella, E. Novelli. 2022. Understanding probabilistic reasoning in innovation. *Bocconi University Working Paper* .

Coibion, O., Y. Gorodnichenko, T. Ropele. 2020. Inflation expectations and firm decisions: New causal evidence. *Quarterly Journal of Economics* **135** 165–219.

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.

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.

Friedman, M. 1953. Essays in positive economics. *University of Chicago Press* .

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 Publishers* .

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.

Hirshleifer, D. 2001. Investor psychology and asset pricing. *Journal of Finance* **56(4)** 1533–1597.

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. Pusley. 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, S., W. Kerr, T. Xu. 2018. Personality traits of entrepreneurs: A review of recent literature. *Foundations and Trends in Entrepreneurship* **14(3)** 279–356.

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* .

Kruger, J., D. Dunning. 1999. Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology* **77(6)** 1121–1134.

Lafley, A., R. Martin, J. Rivkin, N. Siggelkow. 2012. Bringing science to the art of strategy. *Harvard Business Review* .

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.

McElheran, K., E. Brynjofsson, M. Yang. 2022. Data in action: Data-driven decision-making in US manufacturing. *University of Toronto Working Paper* .

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.

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.

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

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.

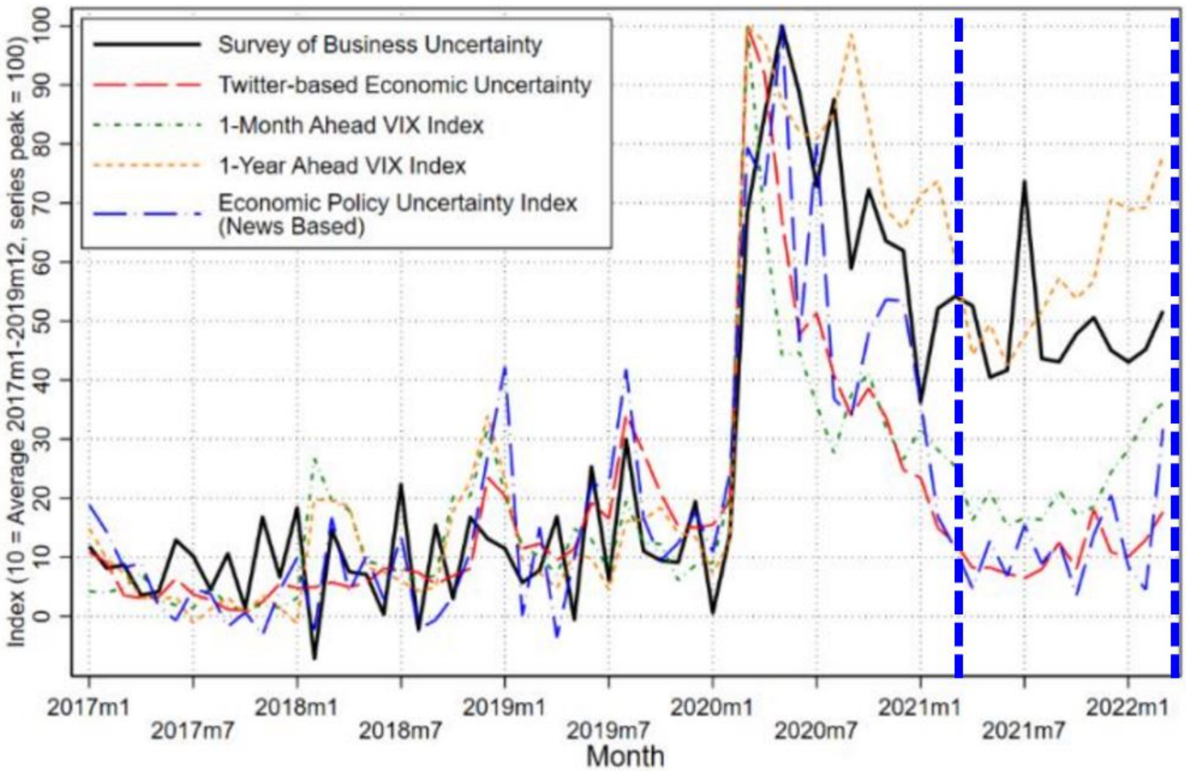
Yang, M., M. Christensen, N. Bloom, R. Sadun, J. Rivkin. 2020. How do CEOs make strategy? *NBER Working Paper 27952* .

Zellweger, T., T. Zenger. 2021. Entrepreneurs as scientists: A pragmatist approach to producing value out of uncertainty. *Academy of Management Review* .

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

Figures and Tables

Figure 1: Growth and business uncertainty during survey period



Note: Figure 3A from Meyer et al. (2022) with overlaid vertical dashed blue lines to indicate study time window.

Figure 2: 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.

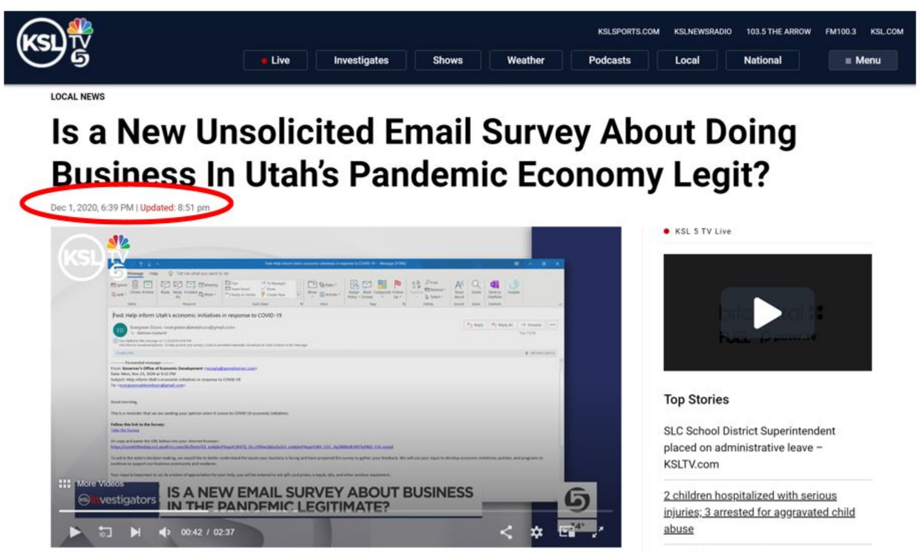
Your input is important to us! As a token of appreciation for your help, you will be entered to win one of ten $1,000 gift cards, a kayak, skis, and other outdoor equipment.

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

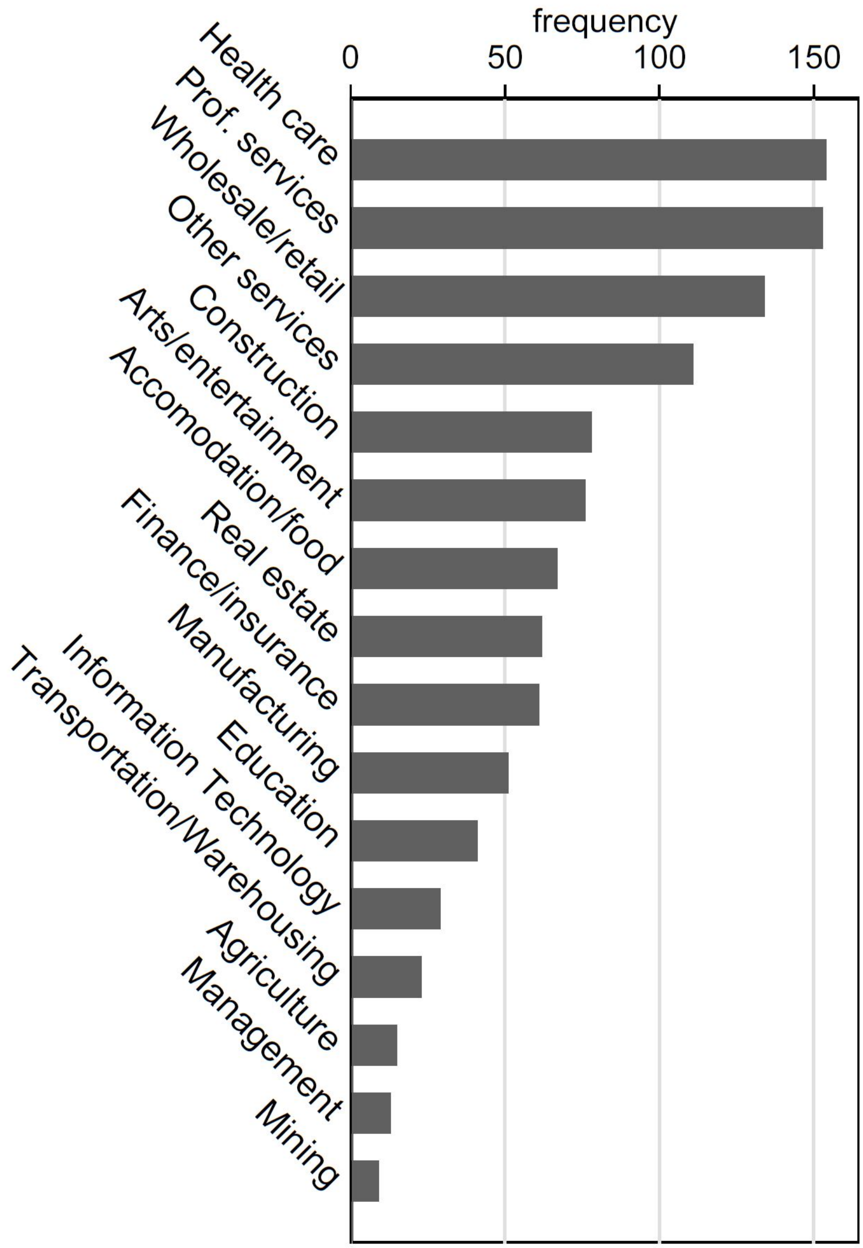
(A) Pilot survey contact email



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

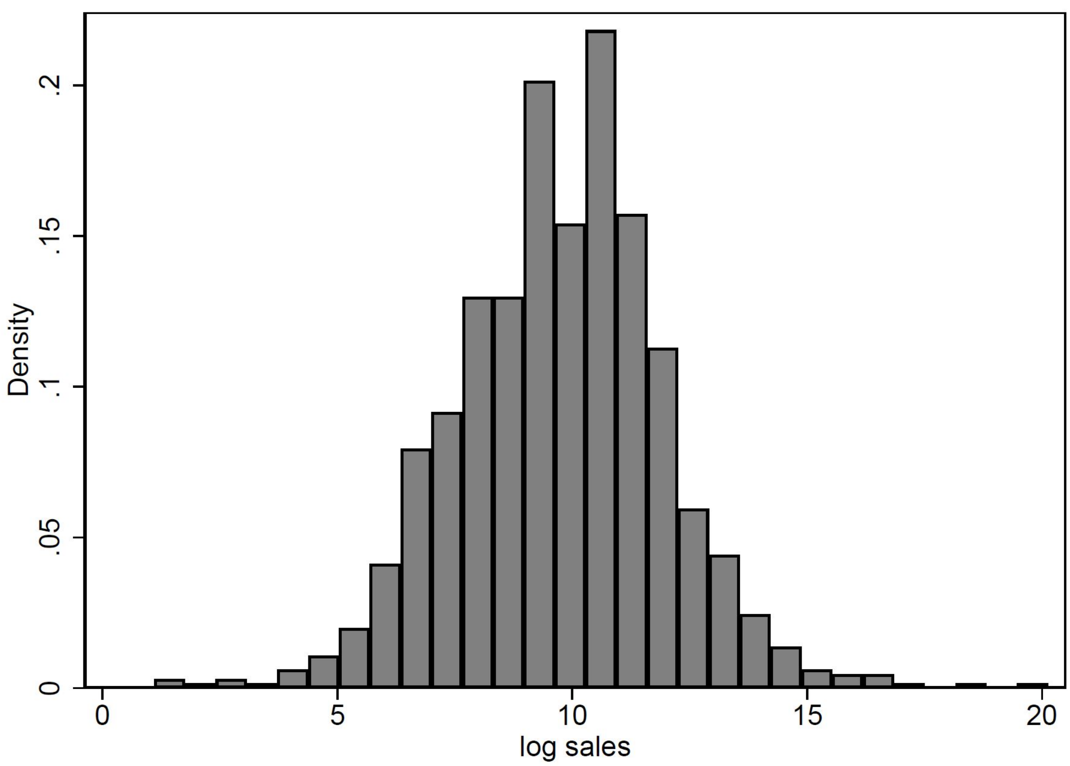
Note: Figures show elements of initial recruiting of participants in December 2020.Full video of evening news coverage of the pilot survey available at: https://ksltv.com/450121/is-a-new-unsolicited-email-survey-about-doing-business-in-utahs-pandemic-economy-legit/

Figure 3: Distribution of firms across industries



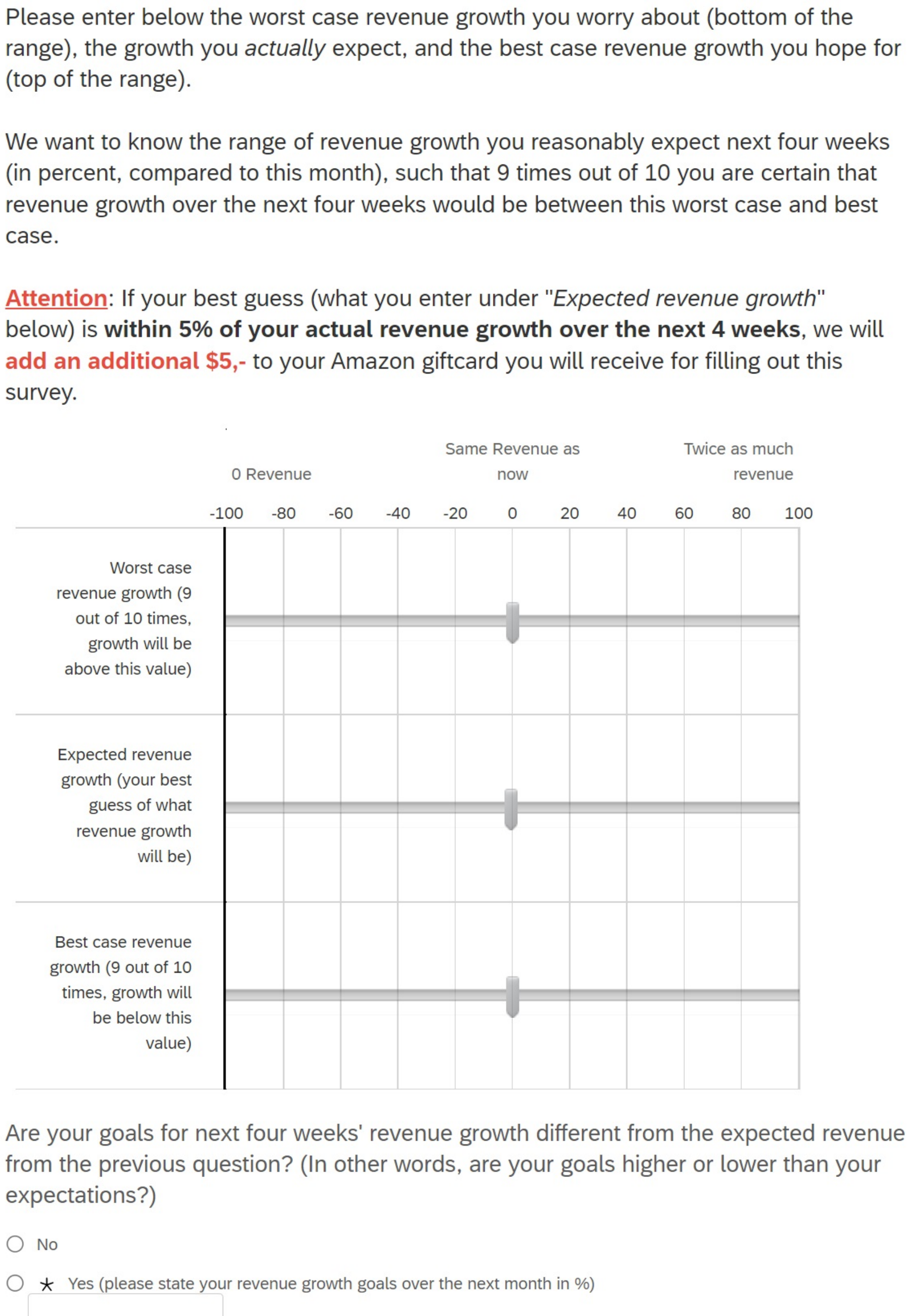
Note: Initial sample of 1027 firms in Utah in March 2021.

Figure 4: Firm size distribution in initial month (March 2021)



Note: Firm size is measured by the log of revenue in March 2021.

Figure 5: 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.

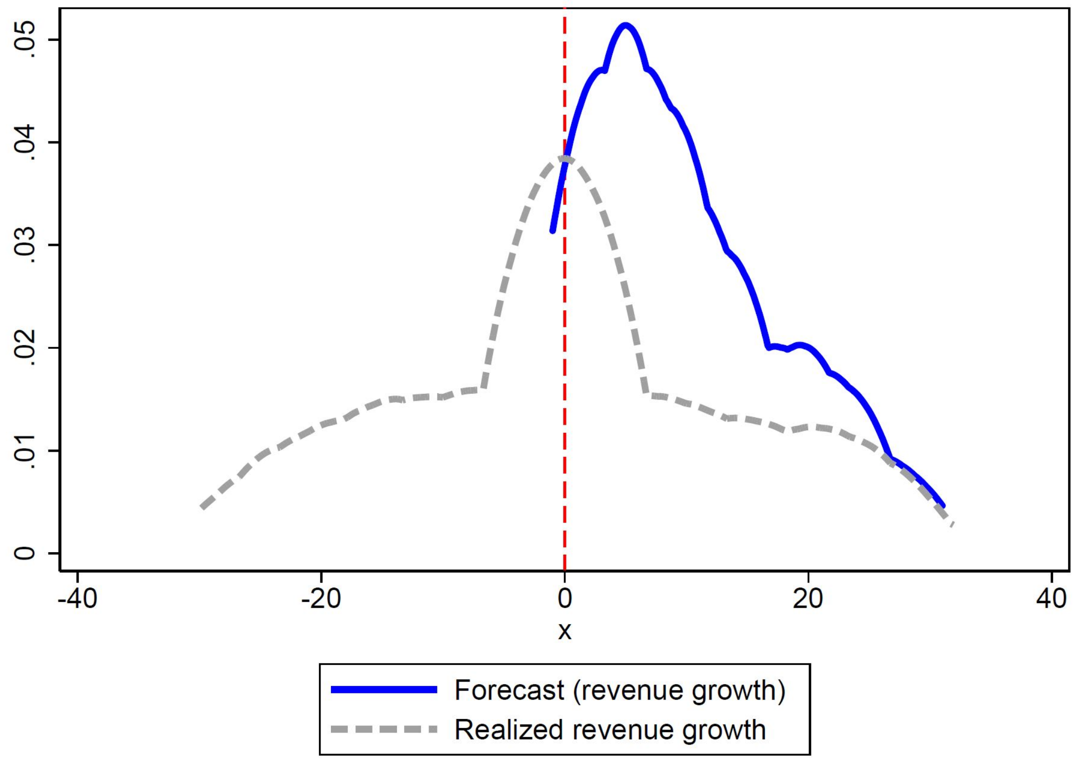
Attention: If your best guess (what you enter under "*Expected revenue growth*" below) is **within 5% of your actual revenue growth over the next 4 weeks**, we will **add an additional $5,-** to your Amazon giftcard you will receive for filling out this survey.

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 %)

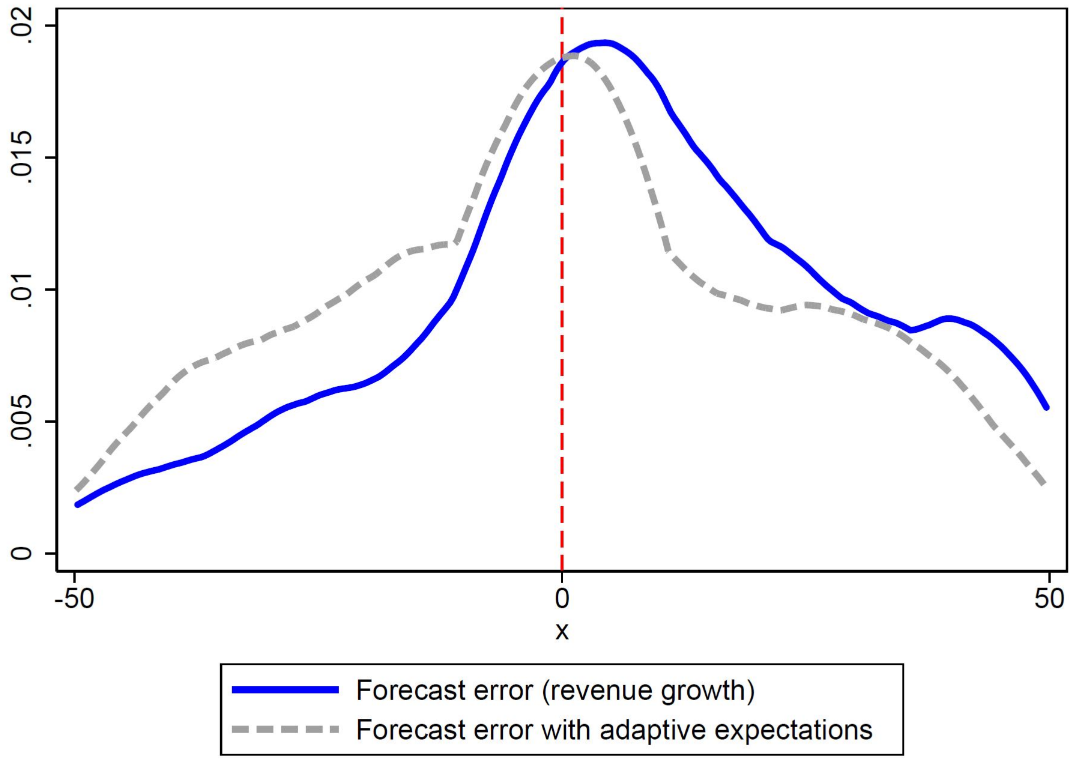
Note: Survey screen to elicit monthly revenue growth forecasts and uncertainty about forecasts. Incentive payments were introduced in October 2021 (7 months into the study and 6 months before the end of the study).

Figure 6: Distribution of growth rates and forecasts in control group



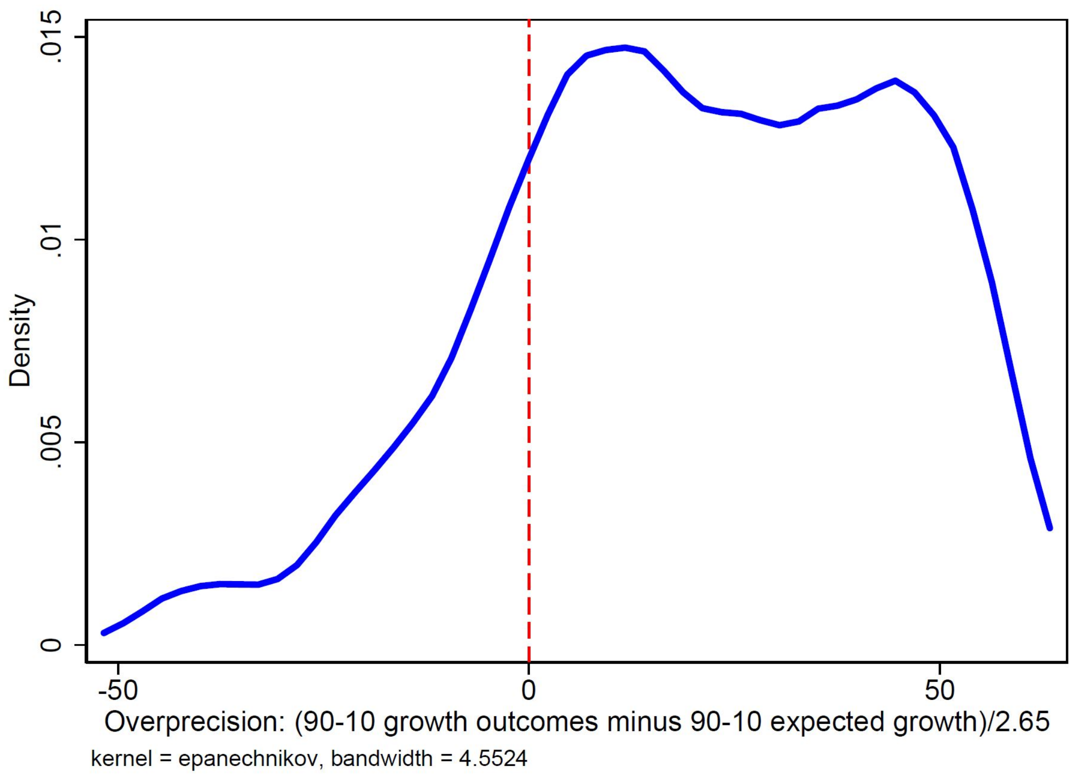
Note: Forecasts in blue are $g^f_{i,t+1}$, while the grey dashed line shows actual revenue growth $g_{i,t+1}$. Data for the figure uses only periods before introduction of incentive payments for prediction accuracy.

Figure 7: Overestimation in the control group



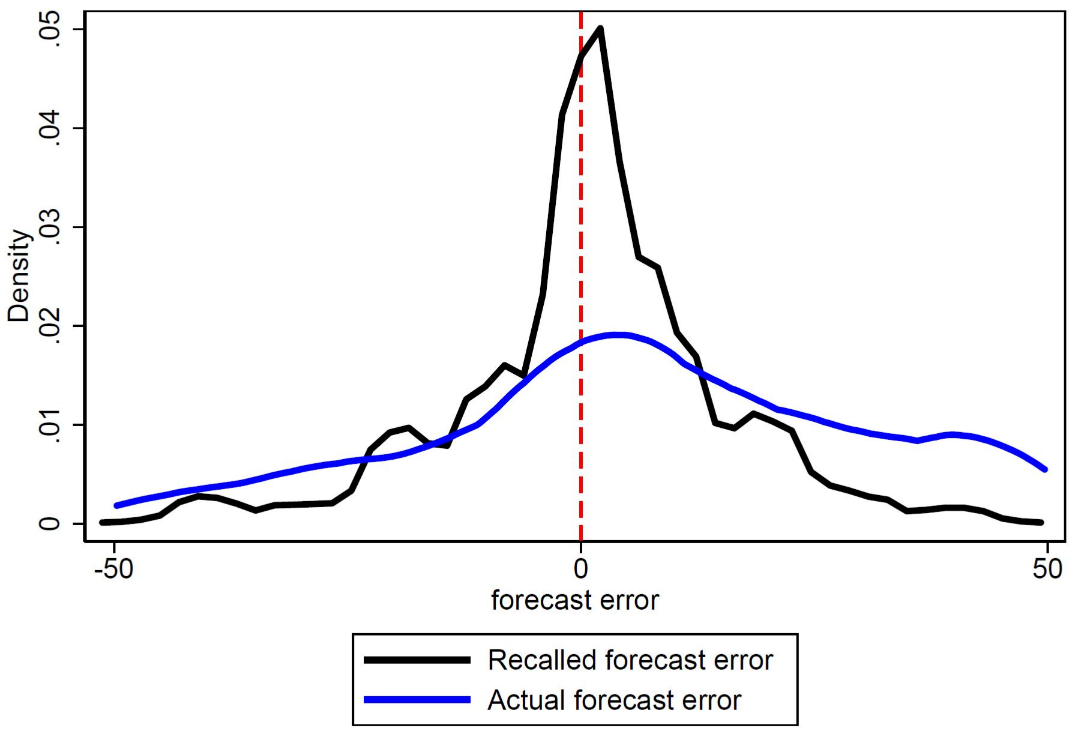
Note: Forecast error $\xi_{i,t+1}$ is measured as difference between monthly revenue growth forecast $g^f_{i,t+1}$ and actual revenue growth $g_{i,t+1}$: $\xi_{t+1} = g^f_{i,t+1} - g_{i,t+1}$. Adaptive expectations uses lagged actual sales growth as forecast for the next month $g^{ada}_{i,t+1} = g_{i,t}$. Adaptive expectations forecast error is therefore calculated as adaptive expectation forecast minus actual monthly revenue growth $\xi^{ada}_{i,t+1} = g_{i,t} - g_{i,t+1}$. Data for the figure uses only periods before introduction of incentive payments for prediction accuracy.

Figure 8: Precision error in the control group



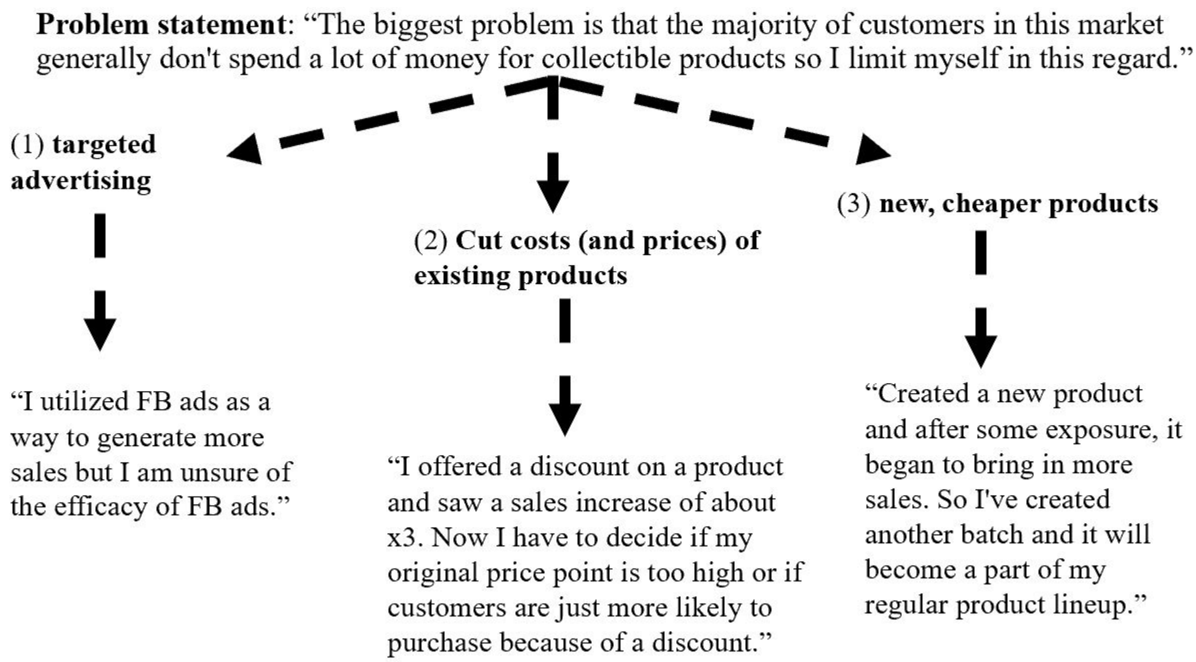
Note: Let $P_{x,i}$ denote the percentile x of monthly growth, and $P^f_{x,i}$ the subjective percentile x of monthly growth at month t for firm i. Under normal distribution of growth rates, the following approximation holds: $\sigma_{g,i} \approx \frac{P_{90,i}-P_{10,i}}{2.65}$, where $\sigma_{g,i}$ 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 $\omega_{i,t} = \sigma_{g,i} - \sigma^f_{g,i,t}$. Data for the figure uses only periods before introduction of incentive payments for prediction accuracy.

Figure 9: Biased memory (hindsight bias) in the control group



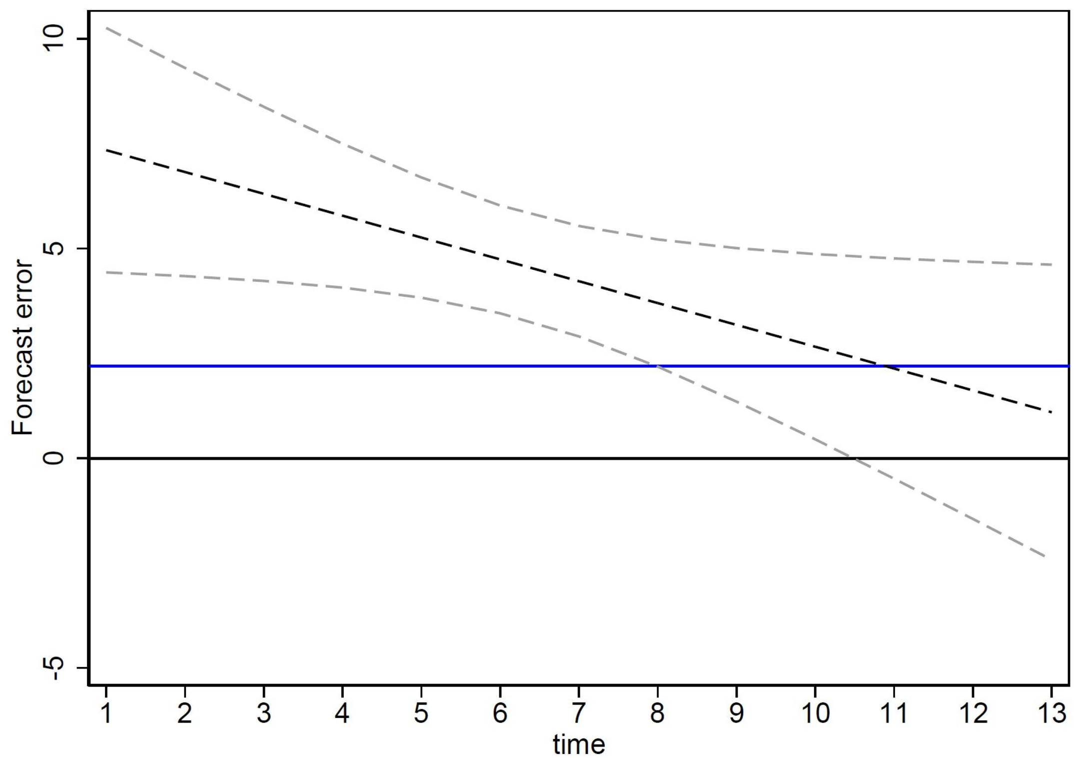
Note: Forecast error $\xi_{i,t+1}$ is measured as difference between monthly revenue growth forecast $g^f_{i,t+1}$ and actual revenue growth $g_{i,t+1}$: $\xi_{t+1} = g^f_{i,t+1} - g_{i,t+1}$. In contrast, the solid black line displays recalled forecast error $\xi^{rec}_{i,t+1}$. Data for the figure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

Figure 10: Scientific Learning Treatment Case for "Bennett Woodworks"



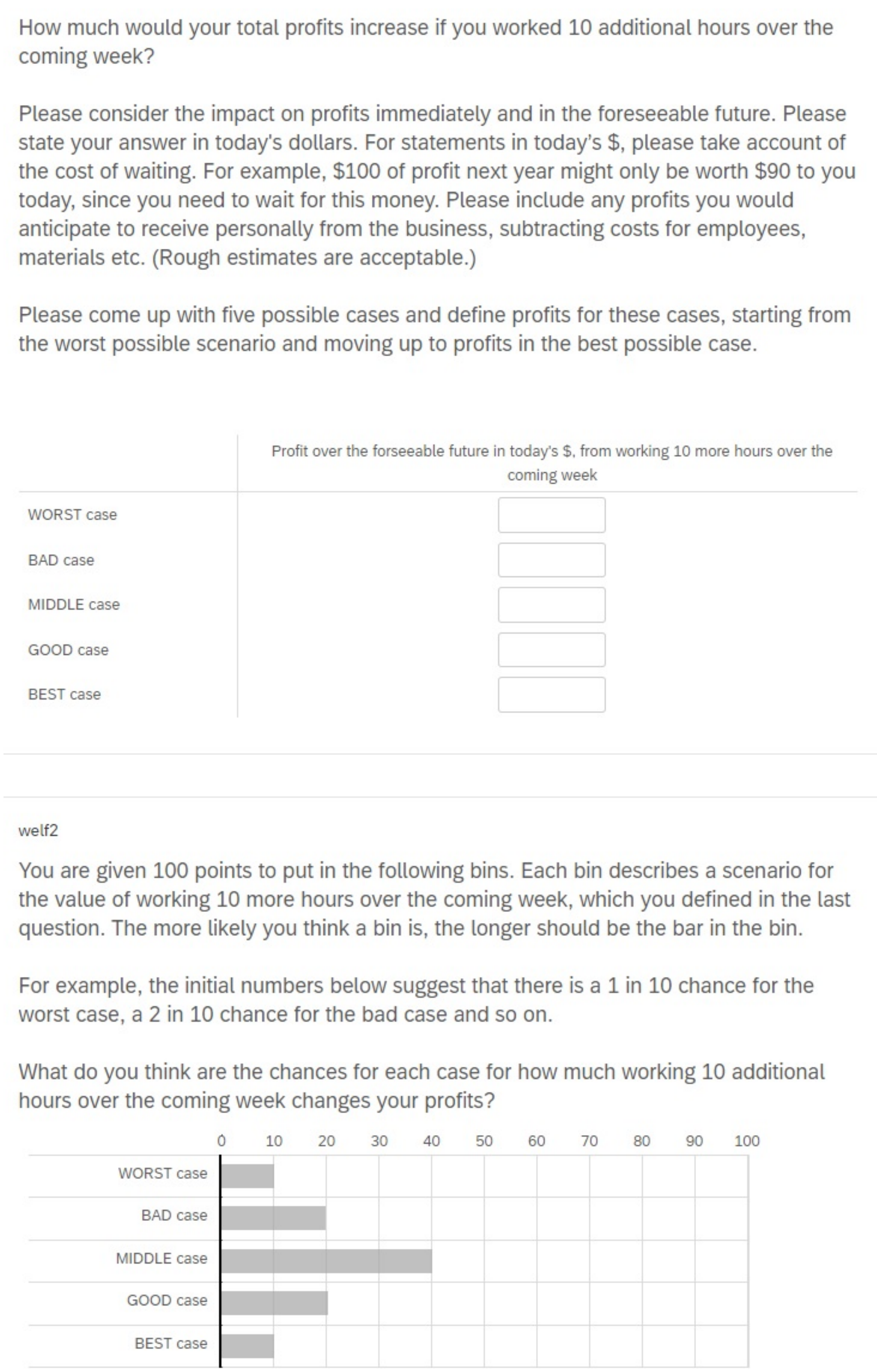
Note: Case example from the data for anonymized participant "Bennett Woodworks".

Figure 11: 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.

Figure 12: Measurement of the marginal expected benefit from more work hours


How much would your total profits increase if you worked 10 additional hours over the coming week?

Please consider the impact on profits immediately and in the foreseeable future. Please state your answer in today's dollars. For statements in today's $, please take account of the cost of waiting. For example, $100 of profit next year might only be worth $90 to you today, since you need to wait for this money. Please include any profits you would anticipate to receive personally from the business, subtracting costs for employees, materials etc. (Rough estimates are acceptable.)

Please come up with five possible cases and define profits for these cases, starting from the worst possible scenario and moving up to profits in the best possible case.

	Profit over the forseeable future in today's $, from working 10 more hours over the coming week
WORST case	
BAD case	
MIDDLE case	
GOOD case	
BEST case	

welf2

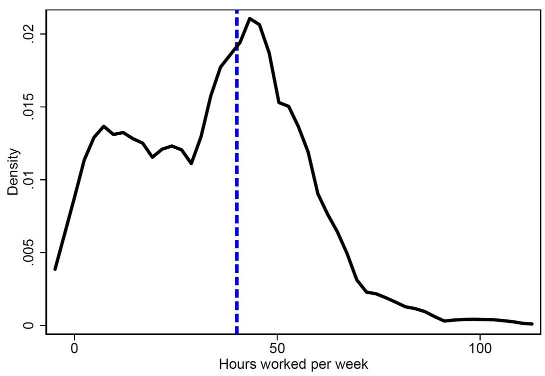
You are given 100 points to put in the following bins. Each bin describes a scenario for the value of working 10 more hours over the coming week, which you defined in the last question. The more likely you think a bin is, the longer should be the bar in the bin.

For example, the initial numbers below suggest that there is a 1 in 10 chance for the worst case, a 2 in 10 chance for the bad case and so on.

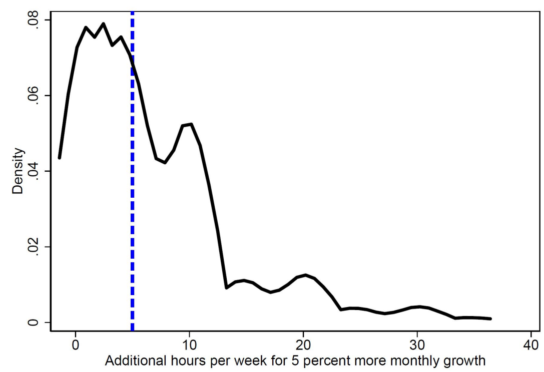
What do you think are the chances for each case for how much working 10 additional hours over the coming week changes your profits?

Note: Question measures the expected benefit of 10 more hours of work in terms of present value. After this question follows the measurement of the certainty equivalent value of the benefits.

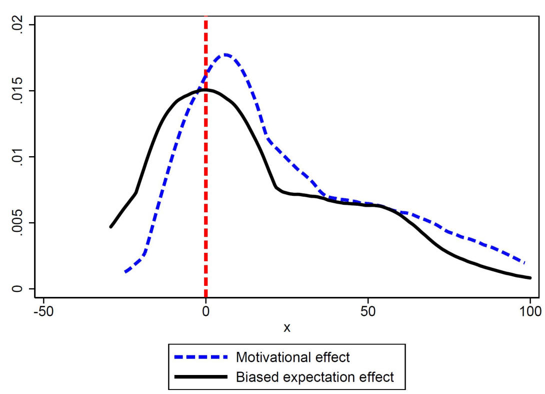
Figure 13: Data underlying correction of subjective marginal profits



(A) Hours worked per week



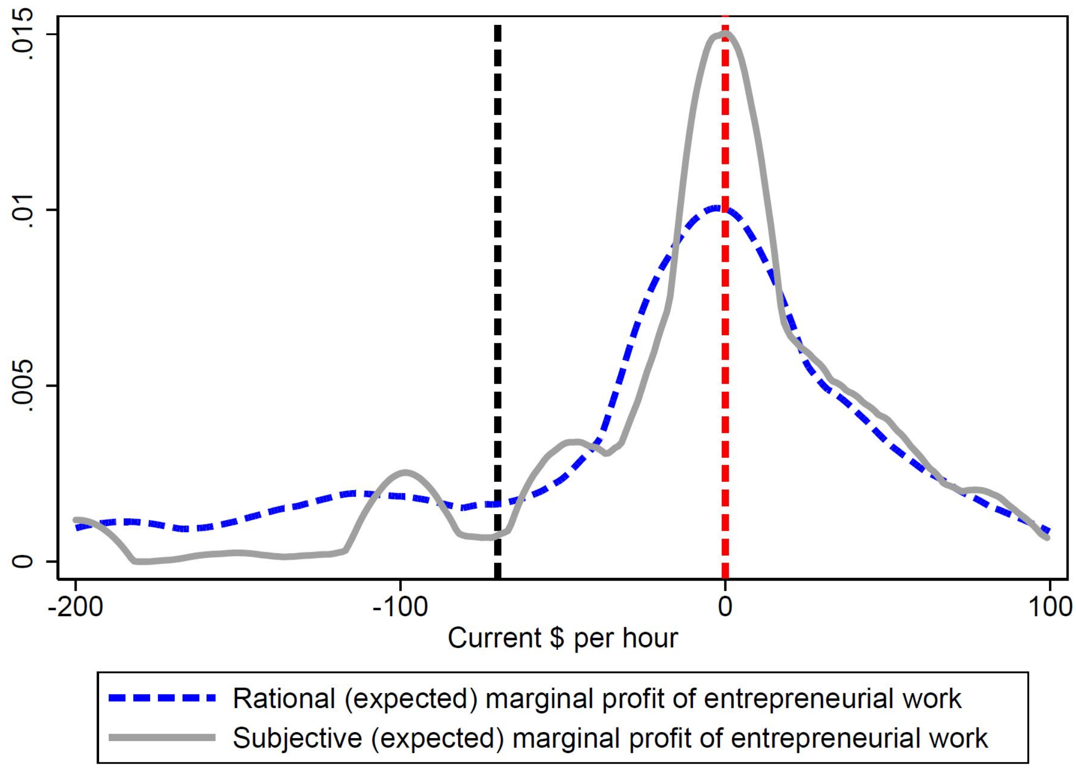
(B) Impact of growth targets on work hours



(C) Components of wedge

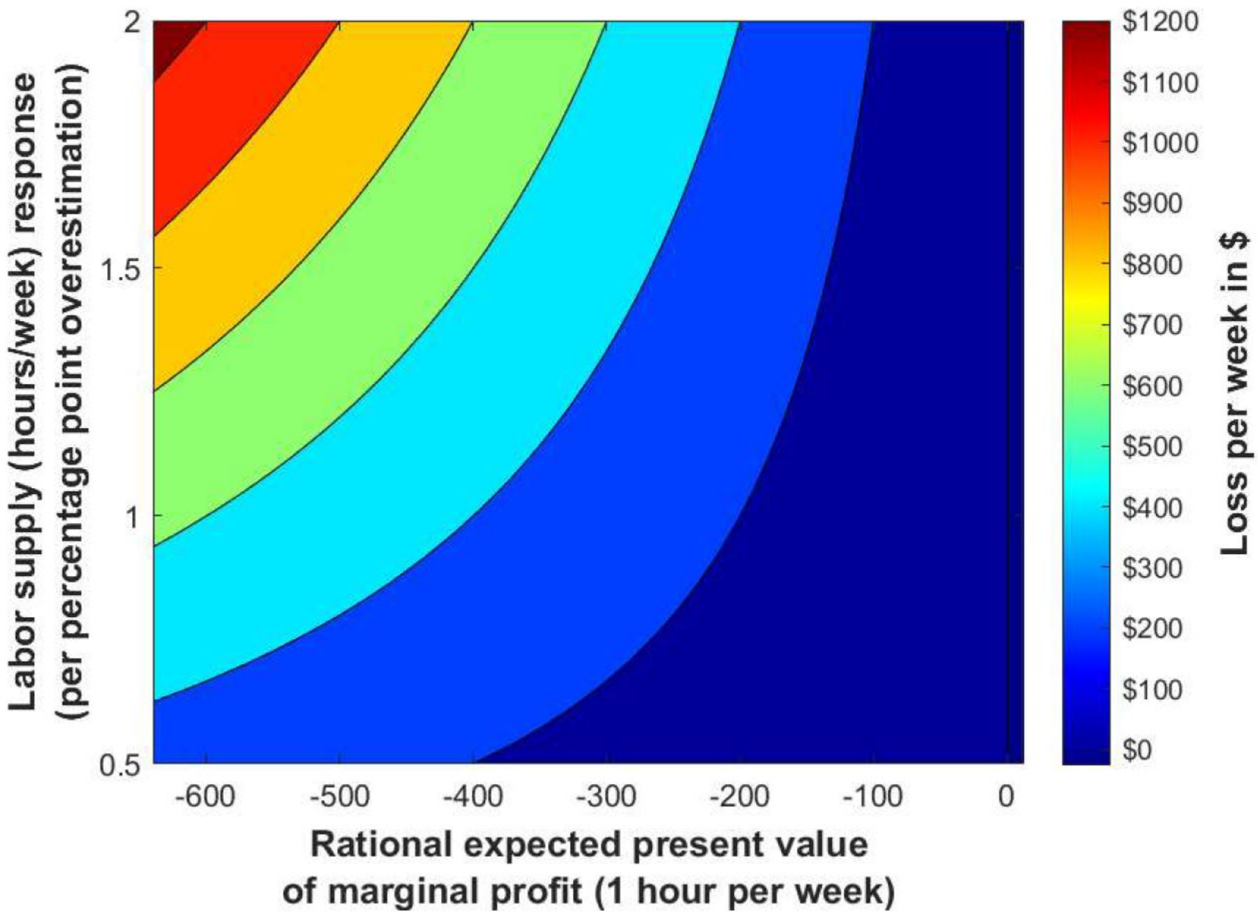
Note: Panel (A) shows reported work hours per week. Panel (B) shows individual estimate of additional work hours per week, required for 5 percentage point higher monthly growth. Panel (C) exhibits two terms of wedge between subjective and rational expected marginal profit of entrepreneurial work. In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

Figure 14: 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.

Figure 15: Marginal entrepreneurial welfare as function of labor supply and rational marginal profit



Note: Labor supply is measured in hours per week in response to a one percentage point increase in revenue growth goals. Rational expected present value of marginal profit is measured per hour. Isoprofit levels show loss in $ per month. Extreme values of each axis are roughly the 25^{th} and 75^{th} percentile values in the data.

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 (%)	16.57	50.41	-20	0	42.86
Forecast error[4] (%)	1.18	38.71	-36	2.93	35.71

[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: Benchmarking Entrepreneurial Expectations

	Revenue growth $g_{i,t+1}$			
	(OLS)	(OLS)	(AB)	(AB)
Forecast $g^f_{i,t+1}$	0.6525***	0.9056***		0.8807***
	(0.0920)	(0.1523)		(0.2128)
Lagged growth			-0.1682***	-0.1578***
			(0.0380)	(0.0414)
Constant	4.0779***	1.4278	11.5546***	2.8846
	(1.1179)	(1.6100)	(1.1121)	(2.5113)
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.

Table 3: Relation of Biases in the Control Group

	A: **Overprecision and Size of Forecast Errors**		
	Noise	Underestimation Error	Overestimation Error
Overprecision	0.2223***	0.1759***	0.2457***
	(0.0260)	(0.0302)	(0.0368)
Constant	24.8544***	25.4345***	26.3589***
	(0.9065)	(1.0216)	(1.2682)
Time FE?	YES	YES	YES
R-squared	0.0568	0.0385	0.0662
Number of firms	456	409	368
Number of observations	1,871	1,066	762
	B: **Overconfidence and Biased Memory**		
	Forecast Error	Noise	Overprecision
Abs. value of recalled Forecast Error	-0.1810***	0.2787***	0.0135
	(0.0676)	(0.0393)	(0.0544)
Constant	6.1254***	26.0734***	21.6098***
	(1.0610)	(0.9233)	(1.2176)
Time FE?	YES	YES	YES
R-squared	0.0077	0.0381	0.0026
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 4: Balance Tests of Randomization

	A: **Error Reminder**		
	Treatment (ERT)	Control (CON)	Difference (CON−ERT)
Firm age (years)	12.93	12.59	-0.340
			(.7318646)
Employees	10.01	9.236	-0.770
			(.6643342)
Revenue ($)	106729.2	126303.9	19574.6
			(.537192)
Revenue growth (%)	14.07	19.41	5.343
			(.1898272)
Forecast Error	1.664	-1.082	-2.746
			(.4567501)
	B: **Scientific Learning**		
	Treatment (SLT)	Control (CON)	Difference (CON−SLT)
Firm age (years)	12.85	12.59	-0.263
			(.7966671)
Employees	11.37	9.236	-2.136
			(.3055529)
Revenue ($)	217034.7	126303.9	-90730.8*
			(.0698629)
Revenue growth (%)	15.55	19.41	3.868
			(.383219)
Forecast Error (%)	3.805	-1.082	-4.886
			(.2131215)

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 5: (No) Impact of Error Reminder Treatment

	Forecast Error	Noise	Precision Error
Error Reminder Treatment	0.3913	-0.4564	-2.4601*
	(0.8373)	(0.8835)	(1.3596)
Constant	2.2009***	29.3348***	22.3465***
	(0.5736)	(0.5827)	(0.9502)
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 6: Correlation of Misattribution and Overconfidence

	Forecast Error		
	Control Group	Error Reminder Treatment Group	Scientific Learning Treatment Group
Misattribution (negative)	-4.8136	37.6604***	35.2447***
	(3.7477)	(1.6371)	(1.8385)
Constant	2.3568***	1.1715*	3.0842***
	(0.5870)	(0.6277)	(0.6778)
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 7: Causal impact of Access to Scientific Learning Treatment

	Forecast Error	Noise	Precision Error	Misattribution
Scientific Learning Treatment	2.3250***	-0.0633	-3.4080**	0.0093
	(0.8688)	(0.9769)	(1.5463)	(0.0062)
Constant	2.1982***	29.3526***	22.3113***	0.0322***
	(0.5744)	(0.5825)	(0.9498)	(0.0037)
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 8: Causal impact of Engagement with Scientific Learning

	A: **Overall Scientific Learning Engagement**			
	Forecast Error	Noise	Precision Error	Misattribution
Overall Engagement with Scientific Learning	1.3256***	-0.0361	-1.9384**	0.0053
	(0.5067)	(0.5559)	(0.8802)	(0.0035)
Constant	-1.9267	29.6116***	23.8479***	0.0701***
	(1.8637)	(1.0530)	(1.0724)	(0.0133)
Time FE?	YES	YES	YES	YES
Stock-Yogo Weak Identification Test	242.11	242.11	272.60	242.11
Kleibergen-Paap Underidentification Test	155.27	155.27	174.74	155.27
Number of firms	802	802	827	802
Number of observations	5,243	5,243	6,647	5,243

	B: **Relative Scientific Learning Engagement**			
	Forecast Error	Precision Error	Forecast Error	Precision Error
Testing relative to Theory	-2.2529***	3.3406**		
	(0.8737)	(1.5003)		
Pre-Postmortem relative to Theory			-15.0501	20.2815
			(10.4173)	(14.6563)
Constant	-1.4778	23.6396***	-4.3572	27.7019***
	(1.8407)	(1.1907)	(2.7941)	(3.2127)
Time FE?	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 $g_{i,t+1}$: $\xi_{t+1} = g^f_{i,t+1} - g_{i,t+1}$. 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. 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.

Table 9: Learning Dynamics

	A: **Scientific Learning**		
	Forecast Error	Noise	Precision Error
Scientific Learning Treatment	5.6711***	1.8493	-5.2319***
	(1.7997)	(1.4556)	(1.7799)
Scientific Learning Treatment $\times$ linear time trend	-0.5207**	-0.2976	0.2620
	(0.2509)	(0.1860)	(0.1594)
Constant	2.1958***	29.3512***	22.3147***
	(0.5744)	(0.5824)	(0.9503)
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	1.8023	1.7702	-3.1740**
	(1.7590)	(1.2858)	(1.5910)
Error Reminder Treatment $\times$ linear time trend	-0.2185	-0.3447**	0.1015
	(0.2346)	(0.1588)	(0.1389)
Constant	2.2021***	29.3368***	22.3459***
	(0.5736)	(0.5826)	(0.9505)
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.

Table 10: Learning Treatments and Profit

	Total Profit	Variable Profit	Variable Profit	Variable Profit
	(in $1,000 per month)			
Error Reminder Treatment	-31.8411	-42.6833*	-68.5640***	-62.6487***
	(19.9159)	(23.1049)	(26.0712)	(22.8363)
Profit/Growth Max	59.1784*	63.9049	61.0493*	64.3056*
	(35.8780)	(38.8361)	(36.6914)	(35.1780)
Error Reminder Treatment	8.8032	24.5502	61.7233	54.6969
× Profit/Growth Max	(51.7722)	(59.2500)	(56.8854)	(54.8137)
Scientific Learning Treatment	-34.4184	-43.8539*	-45.3188	-42.7961
	(21.9638)	(26.3186)	(28.7500)	(26.2936)
Scientific Learning Treatment	151.0791**	182.7351**	183.2371**	162.2805**
× Profit/Growth Max	(76.5658)	(86.0943)	(83.7946)	(81.2556)
Firm Age			8.2793***	7.2015***
			(2.0498)	(1.9343)
Initial Revenues (in $1,000)				0.0091***
				(0.0033)
Constant	48.5510***	65.0694***	-35.4313	-26.2859
	(16.1563)	(18.8877)	(30.0503)	(28.4085)
Time FE?	YES	YES	YES	YES
Industry FE?	YES	YES	YES	YES
R-squared	0.06	0.07	0.12	0.17
Number of firms	1067	1063	1062	1062
Number of observations	4,223	4,137	4,134	4,134

Notes: Profit numbers as in 1,000 $ per month. Total profit is measured as the difference between operating revenues and operating costs. Variable profit is measured as the difference between operating revenues and variable 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. Firm age is number of years since founding. Initial revenues are revenues at initial survey in March 2022. 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 11: 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	-0.4555	-2.0143*	0.4331
	(1.5847)	(1.0567)	(0.9080)
Scientific Learning Treatment	3.2133***	0.4320	-4.0786***
	(1.0980)	(1.0637)	(1.5616)
Scientific Learning Treatment × Incentive Treatment	-2.2300	-1.2645	1.5124
	(1.7498)	(1.2582)	(1.0546)
Constant	2.2006***	29.3460***	22.3268***
	(0.5738)	(0.5825)	(0.9500)
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.

Table 12: Robustness: Incentive Treatments

	Forecast Error	Precision Error	Forecast Error	Precision Error
Scientific Learning Engagement	2.0502***	-2.5992***		
	(0.7087)	(0.9904)		
Scientific Learning Engagement × Incentive Treatment	-1.5746	1.3047**		
	(0.9709)	(0.6017)		
Testing relative to Theory			-3.5126***	4.2481***
			(1.2498)	(1.6253)
Testing relative to Theory × Incentive Treatment			2.7117	-2.1013**
			(1.6691)	(1.0161)
Time FE?	YES	YES	YES	YES
Stock-Yogo Weak Identification Test	297.63	405.94	83.05	113.02
Kleibergen-Paap Underidentification Test	440.33	600.25	156.20	209.80
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 $g_{i,t+1}$: $\xi_{t+1} = g^f_{i,t+1} - g_{i,t+1}$. 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. 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. 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.

Table 13: 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
Scientific Learning Engagement	1.4921***	-1.8794**		
	(0.5691)	(0.9053)		
Testing relative to Theory			-2.4378***	3.0207**
			(0.9386)	(1.4510)
Time FE?	YES	YES	YES	YES
Stock-Yogo Weak Identification Test	192.46	208.28	132.25	161.15
Kleibergen-Paap Underidentification Test	131.58	142.18	103.96	123.05
R-squared	-0.00	0.00	-0.01	-0.02
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 $g_{i,t+1}$: $\xi_{t+1} = g^f_{i,t+1} - g_{i,t+1}$. 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. 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. 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.

Table 14: Robustness: Sample of observations in the first half of the study

	Forecast Error	Precision Error	Forecast Error	Precision Error
Scientific Learning Engagement	2.0502***	-2.5992***		
	(0.7087)	(0.9904)		
Testing relative to Theory			-3.5126***	4.2481***
			(1.2498)	(1.6253)
Time FE?	YES	YES	YES	YES
Stock-Yogo Weak Identification Test	323.54	392.47	102.75	130.47
Kleibergen-Paap Underidentification Test	186.98	221.22	85.45	106.14
R-squared	0.00	0.01	-0.01	-0.04
Number of firms	770	803	770	803
Number of observations	3,143	3,651	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. 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. 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.

Table 15: Robustness: Sample of entrepreneurs working at least 35 hour per week in focal business

	Forecast Error	Precision Error	Forecast Error	Precision Error
Scientific Learning Engagement	1.7105***	-2.8702***		
	(0.6393)	(1.0337)		
Testing relative to Theory			-3.5300**	5.4909***
			(1.3847)	(1.9447)
Time FE?	YES	YES	YES	YES
Stock-Yogo Weak Identification Test	99.09	112.64	46.13	70.95
Kleibergen-Paap Underidentification Test	68.26	77.27	39.67	58.44
R-squared	-0.00	-0.01	-0.02	-0.11
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 $g_{i,t+1}$: $\xi_{t+1} = g^f_{i,t+1} - g_{i,t+1}$. 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. 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. 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.

Table 16: 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	\$2737.88	60.87%	-\$2263.58	-50.32%
Median	**\$990.23**	**22.01%**	**-\$818.69**	**−18.20%**
60^{th} Percentile	\$160.29	3.56%	−\$132.52	−2.94%
75^{th} Percentile	−\$67.49	−1.50%	\$55.79	1.24%
85^{th} Percentile	−\$787.39	−\$17.50	\$650.99	14.47%

Notes: Debiasing is defined as removing the median of the average monthly overestimation error of 2.81% per month, in the control group. Scientific Learning Treatment counterfactual is adding an average monthly forecast error of 2.3% (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 Figures and Tables

A.1 Normalizing revenue growth outcomes

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.1: Correlation of Misattribution and Overconfidence

	Forecast Error		
	Control Group	Error Reminder Treatment Group	Scientific Learning Treatment Group
Misattribution (negative)	-3.1621	33.3731***	30.8073***
	(3.3055)	(1.5597)	(1.6708)
Constant	3.2062***	2.0486***	3.9441***
	(0.5559)	(0.5865)	(0.6386)
R-squared	0.0042	0.0470	0.0480
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 A.2: Treatment Effect of Scientific Learning

	Forecast Error	Forecast Error
Scientific Learning Treatment	2.0889**	5.6573***
	(0.8238)	(1.6300)
Scientific Learning Treatment × linear time trend		-0.5553**
		(0.2244)
Constant	3.1060***	3.1034***
	(0.5430)	(0.5431)
Time FE?	YES	YES
R-squared	0.0058	0.0066
Number of firms	802	802
Number of observations	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 $g_{i,t+1}$: $\xi_{t+1} = g^f_{i,t+1} - g_{i,t+1}$. 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.2 Details on Scientific Learning Treatment

Table A.3: Causal impact of Engagement with Scientific Learning

	Forecast Error	Forecast Error	Forecast Error
Overall Engagement with Scientific Learning	1.1910**		
	(0.4798)		
Testing relative to Theory		-2.0241**	
		(0.8241)	
Pre-Postmortem relative to Theory			-13.5219
			(9.5036)
Constant		0.4053	-2.1818
		(1.5581)	(2.4673)
Time FE?	YES	YES	YES
Stock-Yogo Weak Identification Test	242.11	140.90	2.99
Kleibergen-Paap Underidentification Test	155.27	109.03	2.98
R-squared	-0.00	-0.00	-0.43
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.

Figure A.1: 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.2: 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.3: 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.4: 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.5: 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)