

Information, Goal Setting and Performance: Experimental Evidence

Haoran He*, Marcus Roel**, Ronghai Su***, Qian Weng****

Abstract

We investigate how information about relative performance shapes goal setting and subsequent performance. Combining a field experiment in a university course with a complementary laboratory experiment, we exogenously vary whether individuals receive baseline relative-performance feedback before or after setting non-binding goals. We document that uninformed individuals anchor goals primarily on subjective, often inaccurate beliefs. Providing information systematically corrects these distortions: individuals who underestimate their relative ability revise their goals upward, whereas overestimators adjust their targets downward. However, the effect of information on performance is limited – it only improves performance for underestimators in the field experiment. To explore the underlying mechanism, we leverage laboratory evidence to examine voluntary goal resetting. We find that delayed feedback affects final goals only when individuals reset them.. Accordingly, treatment differences in final goals are substantially attenuated for underestimators but remain sizable for overestimators. Once this goal-adjustment process is accounted for, upward revisions among underestimators improve performance, whereas downward revisions among overestimators reduce challenge and impair ultimate performance.

* Business School, Beijing Normal University. No. 19 Xijiekouwai Street, Haidian District, 100875, Beijing, China. +86 10 5880 7847, haoran.he@bnu.edu.cn

** School of Economics, University of Nottingham Ningbo China & Centre for Decision Research and Experimental Economics China (CeDEx China), Ningbo, China. marcus.roel@nottingham.edu.cn

*** College of P.E. and Sports, Beijing Normal University. sronghai@163.com

**** Corresponding author: School of Labor and Human Resources, Renmin University of China. No. 59 Zhongguancun Street, Haidian District, Beijing 100872, Beijing, China. +86 10 82502205, qian.weng@ruc.edu.cn

Acknowledgments: The laboratory experiment described in this paper was registered on the American Economic Association randomized controlled trial (AEA RCT) registry website (AEARCTR-0015822) in April 2025 prior to data collection. The field experiment and the laboratory experiment are approved by the Institutional Review Board of Beijing Normal University Business School (BNU-BS-IRB 2022-018) and University of Nottingham Ningbo (FHSS-202425-052), respectively. We thank Sherry Xin Li, Shravan Luckraz, and conference and seminar participants at Beijing Normal University, Peking University, Renmin University of China, Shandong University, Shaanxi Normal University, Southeast University, Zhejiang Gongshang University, Zhejiang University of Finance and Economics, 2019 ESA Asia-Pacific Meeting, 2021 ESA World Meeting, and the Chinese Economists Society 2022 Annual Conference for helpful comments and discussions. We thank every member of our team of research assistants for their hard work and care, especially head research assistant Yu Cao, Shuai Chen and Lei Yue. Financial support from the Humanities and Social Science Fund of Ministry of Education of China (No. 21YJA790059), the National Natural Science Foundation of China (No. 72131003, 72273148, and 72473010), the National Program for Special Support of Top-notch Young Professionals, and the University of Nottingham Ningbo, China are gratefully acknowledged. All remaining errors are those of the authors. All authors made equal contributions. All authors share first authorship, but are listed alphabetically by convention.

Keywords: information; goal setting; belief accuracy; relative performance; field experiment; laboratory experiment

JEL classification: D90, C91, C93

Declarations of interest: none

1. Introduction

Goal setting is widely recognized as a potent commitment device to mitigate self-control problems by providing individuals with explicit internal performance benchmarks (Heath et al., 1999; Koch and Nafziger, 2011; Hsiaw, 2013). To effectively stimulate effort, these goals, regardless of academic grades, athletic records, or corporate revenues, must be specific, challenging but attainable; goals that are vague, overly easy or perceived as impossible typically fail to motivate (Locke, 1968; Locke and Latham, 1985). However, setting optimal goals is difficult in practice. When facing novel or complex tasks, individuals often suffer from imperfect information regarding the productivity of their effort and their true relative ability within a competitive environment, especially when goals are commonly set in advance of the actual tasks. Under such uncertainty, acquiring accurate information becomes indispensable for resolving information asymmetry and enabling individuals to evaluate their standing and recalibrate their expectations. This paper seeks to address: How does providing objective information about relative performance shape the choice of goals and subsequent performance for individuals with different prior belief accuracy?

We address this question by combining a field experiment with a complementary laboratory experiment. We first conduct a field experiment within a university physical education course, leveraging a relative grading scheme where students' grades are determined by their rank in a physical endurance task (holding a plank). We then implement a laboratory experiment that mirrors these institutional features using a cognitive memorization task. In both settings, subjects first perform a baseline test, after which they are asked to predict their rank within their respective peer group. Subjects are then asked to set a non-binding¹ and performance-based² goal for their final test rank. Our experimental manipulation is the exogenous timing of feedback: the baseline test performance information is randomly revealed either *before* or *after* the goal-setting stage. This design allows us to isolate how the resolution of uncertainty directly informs goal formation and subsequent performance across heterogeneous initial belief calibration. Furthermore, in the laboratory setting, we directly

¹ Goals can broadly be categorized into two often related dimensions – binding or non-binding, and self-set or exogenously given goals. Binding vs. non-binding refer to whether monetary rewards (punishment) related to the goal achievement (failure) are involved or not. Self-set vs. exogenously given refer to by whom the goals are set. Exogenously given binding goals are mostly used in the context of principal-agent problems where monetary rewards are used to elicit effort from workers (e.g., Goerg and Kube, 2012; Gosnell et al., 2020; Kuhn and Yu, 2021). Recent studies have also seen cases in which principals set non-binding goals for agents (e.g., Corgnet et al., 2015, 2018; Smithers, 2015; Fan and Gómez-Miñambres, 2020; Sauer et al., 2018). Self-set non-binding goals are often studied in situations involving self-control problems (e.g., van Lent and Souverijn, 2020, Clark et al., 2020, on course grade; Sackett et al., 2014, on running time; Uetake and Yang, 2018, on weight loss; Brookins et al., 2017, on principal-agent problem; Harding and Hsiaw, 2014, on energy conservation).

² In terms of the target, goals can also be broadly categorized into performance-based or task-based goals. Performance-based goals are set for particular performance outcomes, whereas task-based goals are set for the completion of particular tasks. We chose performance-based goals for the following reasons. Firstly, performance-based goals are more suitable to address our research question on how information impacts goal-setting level, and consequently are more directly linked with our outcome of interest. Secondly, they are also easier to be executed in our experimental setting (introduced below).

measure goal adjustment in the treatment where feedback is provided *after* goal setting by asking subjects, in a post-experimental survey, whether they had revised their initially chosen goal after receiving the delayed feedback and, if so, what the revised goal were. This elicitation allows us to observe voluntary goal resetting and to examine whether information affects final performance through a goal-recalibration channel.

Our conceptual framework illustrates how and when information about one's relative baseline performance shapes goal setting and, consequently, final performance. In a simple dual-self model, the student as goalsetter chooses a relative-performance goal to motivate a future doer, for whom goals act as an internal commitment device. The framework implies that performance is increasing in goals when goals are moderate, but declines when goals become excessively ambitious, so that optimal goals are challenging yet attainable. Because the optimal goal depends on the student's belief about her relative ability, inaccurate beliefs lead to miscalibrated goals: underestimators set goals that are too low, overestimators set goals that are too high, while students with accurate beliefs set appropriate targets. Providing information about actual relative performance therefore improves goal calibration by inducing underestimators to raise goals, overestimators to lower them, and leaving accurate predictors unaffected; in turn, this leads to higher subsequent performance for under- and overestimators.

Our empirical analysis yields three main sets of results. First, we find that while uninformed subjects anchor their goals primarily on subjective and often inaccurate predictions, providing relative performance information causes a significant reweighting toward factual performance. Feedback provision effectively realigns distorted goals with actual capabilities by correcting miscalibrated beliefs. Specifically, underestimators revise their pessimistic goals upward, while overestimators adjust their overly ambitious goals downward, a pattern that remains consistent across both field and laboratory settings.

Second, using the goal-revision measure from the laboratory experiment, we provide evidence on a goal-resetting mechanism that highlights the within-subject effect of information on goal setting. Among subjects who receive feedback only *after* setting their initial goal, over half report that they voluntarily revised that goal once the feedback was provided. Both the likelihood and the magnitude of this resetting are positively related to the size of the performance-prediction discrepancy, with underestimators revising goals upward and overestimators revising them downward. Once this adjustment is taken into account, treatment differences in final goals for underestimators are substantially attenuated but remain sizable for overestimators. This pattern suggests that delayed feedback can largely replicate the effect of immediate information for underestimators, whereas the timing of information continues to matter for overestimators.

Finally, we demonstrate that prior to accounting for goal resetting, the effect of information on performance is limited – it only improves performance for underestimators in the field experiment. After accounting for feedback-driven goal resetting in the laboratory experiment, we show that

underestimators experience performance gains as they recalibrate their goals upward. In contrast, overestimators suffer a performance decline as lowering goals in response to negative feedback may reduce motivation below the level required for performance improvement.

Our study is primarily related to three strands of literature. First, this paper relates to the growing body of research on the impact of relative performance feedback on effort exertion and academic performance. A number of field experiments demonstrate that individuals care directly about rank, so that disclosing it boosts performance across the distribution even in the absence of tangible rewards (e.g., Azmat and Iriberry, 2010; Tran and Zeckhauser, 2012; Dobrescu et al., 2021). However, the effects of rank-based information are notably mixed, because the behavioral response heavily depends on students' specific position within the performance distribution (e.g., Gill et al., 2019; Brade et al., 2022), and how the revelation of relative standing interacts with their prior beliefs about their ranks (e.g., Azmat et al., 2019; Brade et al., 2022).³

Second, our research connects to the literature evaluating the effectiveness of “soft” behavioral interventions, such as goal setting, commitments, coaching, and reminders, in higher education. The empirical evidence in this literature is highly mixed. While some low-touch coaching and online goal-setting exercises struggle to translate behavioral shifts into tangible academic benefits (e.g., Dobronyi et al., 2019; Oreopoulos et al., 2022; Oreopoulos and Petronijevic, 2023), other studies highlight specific conditions under which these nudges work (e.g., Clark et al., 2020; van Lent and Souverijn, 2020; Brade et al., 2024).⁴ Relatedly, Altan et al. (2026) examine how goal setting and feedback interact in a non-pecuniary environment, showing that feedback increases effort only when goals are set by the employer, but not when they are self-set. This finding highlights a critical gap: despite the ubiquity of relative grading schemes and the recognized potential of “soft” commitment devices, evidence remains scarce regarding how relative feedback interacts with an individual's internal process of goal formation, as a prominent example of soft intervention, and potential goal revision after the feedback.

Finally, this paper links to the literature examining the dynamic relationship between performance feedback, goal revision, and subsequent behavior. A strand of psychological research demonstrates that individuals adjust their goals in response to performance feedback: while positive feedback typically prompts upward goal revision, negative feedback leads individuals to lower or maintain their previous

³ Gill et al. (2019) demonstrate a U-shaped effort response to rank feedback that individuals work hardest after being placed first or last. Azmat et al. (2019) find that providing relative feedback reduces overall performance, driven primarily by students who initially underestimate their relative position in the absence of information and subsequently lower performance upon receiving “good news.” Brade et al. (2022) document that relative feedback can enhance performance only among students who are placed above average, as the feedback corrects their prior underestimation of relative abilities.

⁴ Clark et al. (2020) show that task-based goals are significantly more effective than performance-based goals in enhancing performance. van Lent and Souverijn (2020) find that while mentor-guided goal setting improves grades, pushing students to raise goals can actually backfire. Brade et al. (2024) demonstrate that well-designed soft commitments combined with regular reminders can yield substantial long-term improvements in student success in terms of increase in graduation rate and reduction in time to graduation, particularly for those prone to procrastination.

goals to protect self-efficacy (e.g., Williams et al., 2000; Ilies and Judge, 2005; Krenn et al., 2013; Theobald et al., 2025). Crucially, however, even when goals are downwardly revised, they typically do not reduce below their newly feedback-informed performance levels, thereby preserving a positive discrepancy to maintain motivation (Williams et al., 2000). In the economics literature, van Lent (2019) shows that for students offered the opportunity to revise goals, goal revision is unrelated to the initial goal-performance gap, but this opportunity helps those whose grades are very different from goals by preventing a subsequent drop in performance. More recently, Kaiser et al. (2024) demonstrate that non-binding goals can remain effective self-regulation devices on effort provision even when individuals are allowed to revise goals and they revise goals downwardly, and when revised goals are made salient through reminders. Compared with these studies, our focus is on how objective relative-performance feedback corrects miscalibrated beliefs, both between and within subjects, affects goal formation, and generates heterogeneous downstream performance effects depending on prior belief accuracy.

2. Experimental Design

2.1 Field experiment

2.1.1 Experimental setting

We conducted the field experiment in a physical education course (PEC)⁵ titled “strength and conditioning training course (SCTC)” during the 2018-2019 academic year at Beijing Normal University, a leading Chinese University with over 20,000 students. This optional course carried one academic credit and contributed to students’ GPA.⁶ Designed to improve general fitness, the SCTC spanned an entire semester of 16 weeks with one training session of 1.5 hours every week.

The course was structured as follows. The first two weeks introduced the course arrangements and grading criteria and provided general training. The subsequent 12 weeks were divided into three modules of equal length, each targeting a different physical capacity - aerobic endurance, speed, and power. The course concluded with two final sessions of general training. Grades for each module were determined independently right after the module. At the end of the course, one of the three modules was randomly selected (the same module for all students) to determine the overall course grade.

The same professional PE instructor offered three identical classes in the fall semester and four in the spring semester. Enrollment ranged from 25 to a maximum of 30 students per class,, yielding in total 196 students in the seven classes that participated in the original field experiment. The course content was well-established and remained stable for several years.

⁵ The university offered more than 20 PECs, including fitness courses (quality development/expansion, strength and conditioning training, bodybuilding), sports courses (volleyball, basketball, tennis, badminton, football, soccer, swimming, taekwondo), and leisure courses (basic ballet, ballroom dance, aerobics, yoga, aikido, ladies’ self-defense).

⁶ Students typically needed 145-165 credits (varying across majors) to complete their bachelor degree spanning four years, with PECs required to be taken in four out of eight semesters.

2.1.2 Treatments and test grading scheme

We conducted our experiment in module 1, which focused on enhancing core muscle endurance. Individual test sessions were arranged on a weekly basis to evaluate the physical capacity targeted, which was followed by a joint training session. The designated test task was the plank, as it is one of the common tasks used to measure endurance. After the initial test session (referred to as “baseline test” hereafter), we implemented a 1×2 between-subjects intervention, in which we varied whether students were informed of their *relative* performance of the test prior to goal setting.

Test scores were determined based on a *relative* grading scheme commonly used in similar courses in Chinese universities. Specifically, we ranked students’ plank time (measured in seconds, in descending order) for each gender in each class, and mapped each ranking into a decile interval (hereafter “rank”) by dividing the ranking by the gender-specific class size. Each rank was then assigned a corresponding score in accordance with the university course grading system (between 60 and 100) (see Question 1 in Appendix A.1).⁷

Specifically, the treatments are as follows:

- *NoInfo*: students were asked to set goals for their final test performance, the rank and the associated score, *before* being informed of their baseline test performance.
- *Info*: identical to *NoInfo*, except that students were asked to set their goals *after* being informed of their baseline test performance.

Within each class and gender, we employed a stratified random assignment of students to one of the two treatments.⁸ Since the experiment spanned eight weeks across two semesters, this stratified approach helped to ensure that any temporal factors across classes, such as fluctuations in temperature or air quality, were balanced across treatments. As students joined the course from 19 of the university’s 28 schools and were typically unfamiliar with each other, interactions with one another were limited, minimizing concerns about information spillover or treatment contamination.⁹

2.1.3 Experimental procedure and timeline

⁷ The structure of scores, which are not all equally spaced, follows common practice in university physical education courses in China.

⁸ The original field experiment also included a control treatment, *NoGoal*, in which students were not asked to set a goal. Because our focus is how the timing of information provision affects self-set goals, the main analysis in the paper centers on the theoretically aligned *NoInfo* vs. *Info* comparison. Results incorporating the *NoGoal* treatment are reported in Appendix F.

⁹ According to responses to the second survey (discussed below), students knew the goals of only about 2% of their classmates.

Before the start of our experiment in week 3,¹⁰ the instructor informed students of the course arrangements. In week 1, apart from introducing the basic course structure and grading criteria, students were informed that the course was part of an education reform project approved by the university's academic affairs office to improve teaching quality.¹¹ As a result, they would be required to complete a number of online surveys, administered through wx.cn (China's equivalent to Qualtrics.com) and distributed via individualized WeChat¹² messages by a project assistant unrelated to the course. Given the university's course selection rules, informing students in the first week gave them the opportunity to opt out from the course and, consequently, the surveys. Students were assured that survey responses would remain confidential from the instructor, the teaching assistants and other students. As education reform projects were frequently implemented at the university, students were unlikely to recognize being involved in a scientific study. This ensures that our experiment keeps the features of a natural field experiment (Harrison and List, 2004).

In week 2, the instructor introduced the objectives and arrangements of the first module. Specifically, students were informed that each week would consist of a test session and a training session. The test sessions were conducted individually at a separate time chosen by the student before the training sessions. During the tests, the instructor or teaching assistant ensured that each student maintained the required body posture of the plank and recorded the duration held. No performance-related feedback was provided to the student at the time of the test.¹³ The purpose of the training sessions was to improve core-muscle strength and endurance. The training phase included various training tasks, such as crunches, leg lifts, etc.

The instructor then announced and posted this module's grade composition (see Appendix A.1 question 2) and the *relative* grading scheme. Afterwards, the project assistant added each student on WeChat for future communication.

At the beginning of the first training session, the instructor and teaching assistants measured students' baseline physical fitness indicators, including height, weight, body fat rate, and metabolic rate.¹⁴ On the day after the first training session, students were sent the first survey via personalized links through which the intervention was implemented (see Appendix A.1 for the goal-setting related

¹⁰ There are two reasons to start the experiment in week 3: first, since students could freely withdraw from the course in the first two weeks, starting the intervention in the third week minimized potential attrition or self-selection problems; second, week 3 represented the earliest feasible time point given the course arrangements. Only one registered student dropped out after week 1.

¹¹ The SCTC's instructor had received a grant to conduct an education reform project aiming at improving course quality before we decided to embed the experiment in this course.

¹² WeChat is China's dominant instant-communication program, which supports text-messaging, photo and file sharing, digital payments, etc.

¹³ While students could probably estimate their absolute performance fairly accurately – or simply re-test themselves privately – they crucially did not know anything about other students' absolute performance and hence hardly knew their relative performance prior to receiving feedback.

¹⁴ Height and weight were measured by a height and weight meter; body fat rate and metabolic rate were measured by a body fat meter.

questions). In the survey, among other questions, students were first asked to predict their baseline test performance – score and rank. Prediction accuracy was incentivized with 1, 0.5, or 0 points depending on whether the prediction was exact, off by one, or off by more than one rank. Subsequent questions varied by treatment. In the *NoInfo* treatment, students were first asked to set goals for the final test performance, and then presented with a grade card displaying their actual baseline test performance.¹⁵ In the *Info* treatment, the sequence of the two questions was reversed: students received baseline test performance and then were asked to set goals for the final test performance.

On the day after the training sessions 2, 3 and 4, students were sent their grade cards with updated test scores and reminders of their goals via WeChat message (see Appendix A.2 for an example message). After the final training session, students were sent a second survey, collecting information about socioeconomic background (see Appendix A.3), among other questions. Figure 1 summarizes the timeline of the experiment.

< Figure 1 about here >

To mitigate concerns about experimenter demand effects, we incorporated some blindness into the experimental design. Both intervention and performance feedback were implemented via WeChat by an independent education reform project assistant. Neither the instructor nor the teaching assistants were informed of the treatment assignment or survey responses during the course. Moreover, to ensure rigorous experimental control, there were no direct interactions between the students and the experimenters; all communications were administered with carefully pre-scripted written messages.

2.2 (Follow-up) laboratory experiment

2.2.1 Why conducting a (follow-up) laboratory experiment?

While the field experiment provides high external validity and captures naturally occurring goal-setting behavior under real incentives, it also faces limitations inherent to the field setting. First, the production functions of physical and cognitive performance may differ, making it unclear whether findings from the physical task are informative about the effects of goal setting in cognitive tasks. Second, if students in the *NoInfo* treatment reset their goals after receiving feedback on their actual baseline test performance, the estimated treatment effects from the field experiment should be interpreted as lower-bound estimates of the true effect of information. Third, due to practical constraints on expanding the sample size, the statistical power of the field experiment is somewhat limited (see Appendix B for power analysis), rendering its results suggestive.

¹⁵ Each question was presented on a separate page to prevent students from returning to previous pages and modify responses of previous questions.

To complement the field evidence, we conducted a controlled laboratory experiment using a cognitive task, explicitly elicited whether subjects reset their goals, and employed a larger sample size. Importantly, the laboratory experiment closely preserved the key design features of the field experiment, including a real-effort task, *relative* grading scheme, incentive structures, and informational timing of goal setting.

2.2.2 Treatments

The treatments are identical to those in the field experiment. We varied whether a goal for the final test rank was set *before* (*NoInfo*) or *after* (*Info*) the information about baseline relative performance was provided. Treatments were randomly assigned to subjects within each session, stratified by baseline performance and by belief accuracy (underestimate, accurate, overestimate).

2.2.3 Experimental task and test reward scheme

To mirror the longitudinal structure of the field experiment, we implemented a repeated cognitive task across two rounds. Each round consisted of a memorization stage and a test stage.

In the memorization stage, subjects were presented with a map featuring a specific region of the African continent. It displayed the locations, names and capitals of 12 countries in round 1 and 36 different countries in round 2 (in English). Subjects were asked to memorize as many items as possible - without the aid of pen and paper or any other devices - in 8 and 24 minutes, respectively, in the two rounds. It required subjects to concentrate on both the names and locations on the map.

The test stage followed immediately. Subjects were provided with the same map, but names of countries and capitals were hidden and replaced with identification numbers.¹⁶ Subjects were asked to type the names of countries and capitals corresponding to the numbers into empty entry fields, one country (capital) per page, in ascending order. To prevent strategic editing, subjects were unable to return to previous entries once they had proceeded to the next entry. The time limits were 6 and 18 minutes, respectively, in the two rounds.¹⁷

We selected this task for the following reasons. First, given the volume of items to be memorized and the length of time commitment required, this task was designed to be cognitively demanding. For most subjects this task would therefore eventually become tiresome and boring. As such, it makes an excellent setting to test whether a well-set goal can mitigate the disutility of effort and sustain effort, and allow for learning and improvements over repeated trials. Second, subjects typically held inaccurate

¹⁶ Countries were numbered sequentially starting from the top-left position (designated as number 1). The numbering proceeded horizontally to the right end of the row. Upon reaching the right edge of the map, numbering continued from the leftmost position of the subsequent row.

¹⁷ Appendix Figure D.1 shows the distribution of number of pages subjects submitted, which suggests that the majority managed to submit all pages. In other words, the amount of time allocated to each round did not impose a ceiling effect on their performance. Instead, the number of correct answers reflects memorization capacity and motivation, similar to how the plank task reflects physical capacity and motivation.

beliefs about their relative performance in such tasks, making the provision of rank information highly relevant for setting motivating goals. Third, by shifting from a physical to a cognitive task while maintaining the underlying incentive structure, we can investigate whether the behavioral effects of relative performance information are robust across different task domains.¹⁸

Several practical design features were implemented to ensure internal validity. First, to minimize the impact of prior knowledge, we utilized the map of Africa because Chinese participants are relatively unfamiliar with African countries compared to European countries or US states. Second, to ensure orthographic precision and linguistic consistency, the task was conducted in English, because a single standardized name can be translated into Chinese in multiple character combinations, which might obscure the evaluation. In addition, English is the primary medium of instruction at the university where the experiment was conducted, ensuring all subjects possessed the necessary language proficiency to complete the experiment.

A *relative* reward scheme was designed to parallel the *relative* grading scheme used in the field experiment. Each correct answer - country or capital - was worth one point. For an answer to be correct, the spelling had to be 100% correct as shown in the map, though case insensitive. Any wrong or unfilled entry was scored 0 points. Ranks for each round were established by comparing a subject's total points against all other subjects in that session. Following the literature suggesting that gender differences in simple cognitive test performance are marginal compared to those in physical domains (e.g., Niederle and Vesterlund, 2007), we utilized gender-neutral rankings. No longer constrained by the university course grading system, we set a linear earning schedule across the percentile intervals (see Appendix A.4).¹⁹ To match the time allocated to each round, the earnings in round 2 for each corresponding rank were three times that of round 1. These weightings roughly mapped to the grade composition between the first three tests and the final test in the field experiment. Earnings were calculated in Experimental Currency Units (ECUs) with 1 ECU equal to 0.2 Chinese yuan (CNY).

2.2.4 Experimental procedure

At the beginning of each session, subjects were provided with comprehensive instructions introducing the structure of the experiment, the content of the cognitive task, and the nature of the *relative* reward scheme, all of which were ensured to be common knowledge. To assist comprehension of the interface, the task was illustrated using a part of the European map as an example. Then, subjects completed round 1 following the instructions, which served as the measure of their baseline performance.

¹⁸ It also bridges our findings to the literature of academic performance in higher education.

¹⁹ In the case of a tie in total points, all involved subjects were assigned the most favorable rank. For example, in a session of ten subjects, if two achieved identical top points, both would be ranked in the top 10% rank. The subject with the next highest points would then be assigned to their natural position, i.e., the 21%–30% rank.

After the test in round 1, we elicited subjects' predictions about their performance ranks in the previous test. Prediction accuracy was incentivized with rewards of 40, 20, or 0 ECUs depending on whether the predicted rank was exact, off by one, or off by more than one rank, respectively. These incentives were scaled to broadly match the point weightings used in the field experiment.

Following the prediction, subjects were asked to set goals for their performance rank in round 2 test. The timing of the information provision of their baseline test performance varied by treatment: In the *NoInfo* treatment, students received their actual round 1 test rank (together with the discrepancy from their prediction) *after* completing the goal-setting question, whereas in the *Info* treatment, students received their actual rank *before* the goal-setting question (See Appendix A.5 for the instructions). To ensure goal salience, we also displayed the goal prominently at the top center of the screen throughout the memorization and test stages of round 2.

After completing round 2, we administered a post-experimental survey to collect data on subjects' socioeconomic background such as gender, academic major, year of study, disposable income, GPA in last semester (in 0-100), number of siblings and birth order, prior participation in laboratory experiments, as well as experiment-specific feedback. Crucially, we asked subjects in the *NoInfo* treatment whether feedback about round 1 test had prompted a revision of their initial goals for round 2 test, and if so, what the updated goal would be.

The experiment was programmed and conducted using z-Tree (Fischbacher, 2007) at the University of Nottingham Ningbo China between April and September 2025. We recruited a total of 299 students via ORSEE over 15 sessions, and all recruited subjects completed the experiment. Subjects were paid via WeChat digital cash after the experiment, with average earnings of 98.5 CNY including a 36 CNY show-up fee.²⁰ The experiment on average lasted about 85 minutes.

3. Conceptual framework and predictions

We present a simple model to illustrate how information about relative baseline test performance shapes goal setting and final test performance for individuals with different prior belief accuracy. To build intuition, we focus on the main mechanisms; technical details, proofs, further discussion of modeling choices, and an equilibrium extension are deferred to Appendix C.

We model the student's decision-making process using a two-stage, dual-self framework. In stage 1, the student acts as a goalsetter, choosing a performance target g for her relative performance y in the final test, based on her expected relative ability. In stage 2, the student acts as a doer. Taking the previously set goal as given, she chooses her relative performance y , which yields a material benefit of y while incurring a cost $c(y)$, which is increasing and strictly convex in y . Relative performance can

²⁰ The average hourly payoff is 69.5 CNY, which corresponds to the hourly wage rate of college students in Ningbo China. At an exchange rate of 0.14 (September, 2025), this translates to 9.73 USD. This is sufficiently high compared to the hourly wage of 40 CNY for student research assistants in this university.

be thought of as the difference between the decision maker's absolute performance and the average (absolute) performance of other students in her comparison group. If her relative performance falls short of her goal, she additionally incurs a psychological cost $k(g - y)$, which is increasing in $g - y$ yet concave; there is no psychological benefit if the goal is met or exceeded, i.e., $k(g - y) = 0$ for $y \geq g$.²¹ We focus on losses only, both for simplicity and because allowing gains would generate implausible incentives to set very low goals.

The two selves differ in how they weigh material cost against material benefit. The divergence, or the “wedge”, arises from present-biased preferences in the spirit of Strotz (1955), Laibson (1997), and O'Donoghue and Rabin (1999): relative to the goalsetter, the doer undervalues future benefits compared to immediate effort costs of performance. To overcome this self-control problem, goals serve as an internal commitment device. Following the standard approach in the goal-setting literature, we model goals as reference points for evaluating future performance (Heath et al., 1999; Wu et al., 2008; Suvorov and van de Ven, 2008; Koch and Nafziger, 2011; Hsiaw, 2013; Clark et al., 2020). By incurring a psychological cost when performance falls short of the reference point, goals partially close the “wedge” and motivate the doer to reach higher performance than she otherwise would. To significantly simplify our framework, we adopt a reduced-form representation that captures this dynamic without carrying explicit discount factors or timing of payoffs: the goalsetter chooses g to maximize $y^*(g)$ while the doer chooses y to maximize $y - k(g - y) - c(y)$.

3.1 The doer's problem

Solving by backward induction, the doer in stage 2 takes g as given and solves²²

$$\max_y y - k(g - y) - c(y)$$

If we disregarded the psychological cost term, the problem would be standard, with an optimal solution characterized by $c'(y_l) = 1$, where y_l represents the “low” performance level the doer implements in the absence of any goals. From the goalsetter's perspective, y_l is too low: she would prefer the doer to perform at a higher level, and could use a goal to achieve this. A goal $g > y_l$ introduces an additional incentive to raise performance whenever $y < g$.

Figure 2 illustrates the doer's problem for a representative case with three different levels of goals ($g_1 < \hat{g} < g_3$), which are all above y_l . To graphically represent the doer's decision, we plot the two opposing forces she faces. On one side is the material tradeoff of performance, captured by combining the marginal cost and the marginal material benefit into a single net marginal cost curve, $(c(y) - y)'$,

²¹ Formally, $k(z)$ is continuous with $k(z) = 0$ for $z \leq 0$, $k(z) > 0$, $k'(z) > 0$, and $k''(z) < 0$ for $z > 0$; $k'(0^+)$ is bounded. Formally, $c(y)$ is derived from a cost function over absolute performance that satisfies standard regularity conditions; see Appendix C.2 for details.

²² Where the doer's problem admits multiple optima, we assume she selects the highest performance level.

which equals 0 at y_l . On the other side is the additional incentive to raise performance stemming from the goal, depicted by the marginal psychological benefit curve, $k'(g - y)$, which is positive whenever performance falls short of the goal ($y < g$). Unlike a standard marginal benefit, it is *increasing* in y , because $-k(g - y)$ is convex in y . In other words, the psychological incentive grows stronger as performance approaches the goal from below. Once performance exceeds the goal, $k'(g - y)$ is zero.

< Figure 2 about here >

For the lowest goal g_1 , $k'(g_1 - y)$ exceeds the net marginal cost for all y up to g_1 , so the doer's optimal performance exactly meets the goal, $y^*(g_1) = g_1$. Raising the goal to $\hat{g}$ has two effects. First, the marginal psychological benefit curve shifts rightward, so its nonzero region now extends up to the higher goal. Second, at any given performance level, the curve now lies lower than before, since $-k(g - y)$ is convex in y and the shortfall $\hat{g} - y$ is larger. Despite this downward shift, $k'(\hat{g} - y)$ still exceeds the net marginal cost for all y up to $\hat{g}$, so the doer again just meets her goal, now at the higher performance $y^*(\hat{g}) = \hat{g}$. For the even more ambitious goal g_3 , the second effect dominates: the downward shift is large enough that $k'(g_3 - y)$ now crosses the net marginal cost curve from above at $y^*(g_3) < g_3$. The goal is no longer met, and performance also falls below that under $\hat{g}$, the most ambitious goal the doer will still meet. In short, an overly ambitious goal backfires: the shortfall becomes so large that the psychological incentive weakens and consequently performance falls.

The three goals in Figure 2 trace out the general pattern. Since the doer strictly prefers to reach y_l on material grounds alone, any goal below y_l leaves performance unchanged, so performance equals y_l for goals up to y_l . Above y_l , performance increases one-for-one with the goal up to the threshold goal $\hat{g}$ that maximizes performance, and decreases for $g > \hat{g}$. Formally:

Proposition 1: There exists a performance-maximizing goal $\hat{g} > y_l$. For $g \leq y_l$, $y^*(g) = y_l$; for $g \in (y_l, \hat{g}]$, $y^*(g) = g$; for $g > \hat{g}$, $y^*(g) \in (y_l, \hat{g})$ and is strictly decreasing in g .

The proposition mirrors the established result in the literature (Wu et al., 2008): performance is increasing in goals for low-to-moderate targets but ultimately decreases for goals that are excessively high.

3.2 The goalsetter's problem and the role of information

Given $y^*(g)$, the goalsetter in stage 1 solves $\max_g y^*(g)$. By proposition 1, the optimal goal is $g^* = \hat{g}$: the most ambitious goal that the doer will still meet. It captures the notion of a “challenging but

attainable” target: any lower goal under-motivates, while any higher goal triggers a drop in performance.

To set the optimal goal g^* , the goalsetter must correctly perceive the net marginal cost curve. Her relative ability determines this curve: higher ability relative to others shifts it rightward because reaching any given performance level y is less costly, while lower ability shifts it leftward. Miscalibrated beliefs about relative ability lead the goalsetter to acting on the wrong cost curve, yielding a goal that is too low when ability is underestimated and too high when it is overestimated.

A critical feature of our design is that all students learn their true baseline test rank before the final test, so the doer always chooses performance on the true net marginal cost curve. What varies across treatments is only whether the goalsetter has access to this information. In the *Info* treatment, she sets her goal knowing her true relative ability; in the *NoInfo* treatment, she sets her goal under her initial (potentially incorrect) belief. Figure 3 illustrates how information corrects these misperceptions, shaping goal setting and subsequent performance for students with different prior beliefs.

Underestimators (Figure 3a). A student who underestimates her relative ability perceives the net marginal cost curve as $(c_0(y) - y)'$, lying to the left of the true curve $(c_1(y) - y)'$. Setting goals under this pessimistic belief in the *NoInfo* treatment, she chooses g_0 , the performance-maximizing goal given the perceived cost curve. Information shifts the perceived curve rightward to its true location, raising the performance-maximizing goal to g_1 , which is subsequently achieved by the doer in the *Info* treatment ($y^*(g_1) = g_1$). However, given a goal of g_0 , the doer in the *NoInfo* treatment finds it optimal to just meet her goal ($y^*(g_0) = g_0$), performing below the *Info* treatment doer even though both act on the same true cost curve.

Overestimators (Figure 3b). A student who overestimates her relative ability perceives the net marginal cost curve as $(c_0(y) - y)'$, lying to the right of the true curve $(c_1(y) - y)'$. Setting goals under this overly optimistic belief, she chooses g_0 . Information shifts the perceived curve leftward to its true location, lowering the performance-maximizing goal to g_1 , which is subsequently achieved by the doer in the *Info* treatment ($y^*(g_1) = g_1$). However, given a goal of g_0 , the doer in the *NoInfo* treatment – who must be acting on the true cost curve – finds it too costly to fully meet her goal. Her final performance, $y^*(g_0)$, hence, lies below both her unrealistic goal g_0 and the optimal goal g_1 achieved in the *Info* treatment.

The final case – accurate beliefs – requires little explanation. If initial beliefs are correct, information is redundant.

< Figure 3 about here >

The three predictions below summarize the above discussion. We state predictions for both relative and absolute performance: since *Info* and *NoInfo* students share the same sessions, others' performance is identical across treatments, so any treatment difference in relative performance has the same sign as the corresponding difference in absolute performance.

Prediction 1. *Students who underestimate their relative baseline performance (a) set higher goals and (b) have higher relative and absolute performance in the Info than in the NoInfo treatment.*

Prediction 2. *Students who hold accurate beliefs about their relative baseline performance (a) set the same goals and (b) have the same relative and absolute performance in the Info as in the NoInfo treatment.*

Prediction 3. *Students who overestimate their relative baseline performance (a) set lower goals and (b) have higher relative and absolute performance in the Info than in the NoInfo treatment.*

4. Results

The key variables collected from the experiment include the predicted baseline test rank, actual baseline and final test ranks and absolute performance measures – the plank duration in the field experiment and the number of correctly identified items in the laboratory experiment, as well as rank-based goals set by the subjects. Definition and summary statistics for these and other control variables are reported in Appendix Tables D.1 and D.2, respectively. Balance checks, reported in Table D.3, confirm that treatment random assignment was effectively successful.

Our empirical analysis proceeds in three steps. We first document how information provision of the baseline test performance influences goal-setting behavior in both the field and laboratory experiments. Next, we leverage laboratory evidence to study goal resetting as a behavioral mechanism through which students respond to information. Finally, we examine whether information-induced adjustments in goals translate into changes in performance.²³

4.1 Information and Goal Setting

As our analysis predominantly focuses on the effect of information on goal-setting behavior across different belief-accuracy types, we first document the distribution of their misperceptions. In the field experiment, 35.9% of the subjects underestimate their baseline performance, 14.5% make accurate predictions, and 49.6% overestimate. In the laboratory experiment, underprediction is more prevalent:

²³ Since goal setting serves as the intermediate mechanism in our conceptual framework, analyses regarding goal setting and resetting are considered secondary.

54.4% of subjects underestimate, 18.5% hold accurate beliefs, and 27.2% overestimate.²⁴ Figure 4 provides a more disaggregated depiction of belief (in)accuracy by plotting the distribution of the difference between actual and predicted test ranks, with ranks ranging from 1 to 10 and higher rank representing better performance. Overall, the distribution is bell-shaped: accurate beliefs account for the single largest group, while larger deviations from the truth appear less frequently. Importantly, the figure also indicates substantial scope for an information intervention to align beliefs with actual performance.

< Figure 4 is about here >

In Appendix Figure D.2, we plot the distribution of goals by treatment for both field and laboratory experiments. We find that while goals in the field are systematically over-optimistic and exhibit significant treatment differences (Fisher's exact test, p -value = 0.016), goals in the laboratory are notably less top-heavy and remain statistically indistinguishable across treatments (Fisher's exact test, p -value = 0.314).

Table 1 reports the analysis of how the effect of information about baseline test performance on goal setting varies by belief accuracy. The dependent variable is the goal rank, treated as a continuous variable. The key independent variables include a treatment dummy (*Info*), indicators for belief-accuracy types with the accurate type in the *NoInfo* treatment as the reference category, and interactions of the treatment dummy and the type indicators. The top panel presents ordinary least squares (OLS) regression results, and the bottom panel presents the implied differential effects of information provision for each of the three accuracy types.

Columns (1) and (2) report results from the field experiment, and columns (3) and (4) report results from the laboratory experiment. As goals are set in terms of rank in both settings, we also pool the data to obtain more precise estimates (columns 5 and 6).²⁵ Columns (7) and (8) report results from considering goal resetting in the laboratory experiment. All specifications control for baseline test ranks to capture potential inherent individual ability heterogeneity. We also control for group fixed effects, defined at the semester – class – gender level for the field experiment and session level for the laboratory experiment, respectively, to account for group unobserved heterogeneity affecting goals. Even-numbered specifications further add individual characteristics, and the overlapping variables for the

²⁴ Because the subgroup with accurate beliefs is small and theoretically should experience no effect of information on behavior, analysis regarding this group is considered exploratory.

²⁵ We pool the data only for harmonized rank-based outcomes, for which the field and laboratory experiments implement identical treatments and are designed to identify identical underlying mechanism. Although the two experiments differ in year, cohort, task environment, and implementation details, pooling improves statistical power and precision. We interpret the pooled estimates as complementary aggregate evidence, i.e., weighted-average treatment effects across two internally randomized study settings, while continuing to present the field and laboratory results separately.

pooled data (see Table 1 notes for the list of variables). Standard errors are clustered at the group level to allow for within-group correlation of unobservables.

We start with discussing the results in the bottom panel of Table 1, which provides the direct test of the theoretical predictions. In the field experiment, providing information increases goals among underestimators by approximately 1.4 to 1.6 ranks relative to the underestimators in the *NoInfo* treatment. Conversely, information reduces goals among overestimators by a similar magnitude. The laboratory experiment yields qualitatively and quantitatively similar effects. For subjects holding accurate beliefs, the effect of information is not statistically significant in all settings. The analysis of the pooled data from the field and the laboratory confirms the robustness of these findings. Taken together, these results lend support to Predictions 1(a), 2(a), and 3(a).

< Table 1 is about here >

To understand why performance information shifts goal-setting behavior, we examine the determinants of goals. That is, how subjects weigh subjective beliefs (predicted baseline test rank) vs. objective performance (actual baseline test rank) when setting goals. Table 2 reports the regression results of goal ranks on predicted baseline test rank, actual baseline test rank, information treatment dummy, and the interactions between ranks and information treatment dummy.

< Table 2 is about here >

Across columns (1)-(6), the top panel shows that subjects in the *NoInfo* treatment rely primarily on their predicted baseline ranks when setting goals. A one-rank higher prediction is associated with a 0.5-0.7 rank higher goals. In contrast, the coefficient on the actual baseline test rank is small, below 0.1 ranks, and statistically insignificant. This indicates that, without feedback, subjects anchor their goals mainly to their subjective beliefs rather than their true underlying ability.

Baseline performance information substantially shifts this weighting pattern. As shown by the interaction terms, once information is provided, actual baseline rank becomes a significantly stronger predictor of goal rank, whereas predicted baseline rank becomes a significantly weaker predictor. The linear combinations of the estimates reported in the bottom panel confirm this shift: in the *Info* treatment, both predicted and actual baseline ranks positively correlate with goals, but the association between actual baseline rank and goal rank is two to three times larger than the association between predicted baseline rank and goal rank (Wald tests, p -value < 0.01). Overall, Table 2 shows that subjects set goals based on the information they possess. The consistency between the field and laboratory results suggests that the underlying mechanism – the reweighting in goal determination of predicted

and actual baseline ranks in response to objective feedback – is robust across different effort domains and incentive stakes.

Results in Table 2 provide a clear rationale for the patterns observed in Table 1 top panel. Underestimators in the absence of information set significantly lower goals than accurate predictors (coefficient on *Underestimate*), and they upwardly adjust their goals in the presence of information because the information treatment forces them to replace their pessimistic subjective beliefs with superior objective facts (coefficient on *Info* + *Info* x *Underestimate*). Information causes goals of underestimators to converge toward goals of accurate predictors in the *Info* treatment, effectively correcting the initial underestimation-belief-driven distortions (Wald test of *Underestimate* + *Info* x *Underestimate* = 0, p -value > 0.10). The opposite holds for overestimators (Wald test, p -value > 0.10).²⁶ Our findings are in line with previous findings that individuals will lower or maintain their goal after negative feedback, and raise their goal after positive feedback (e.g., Ilies and Judge, 2005; Krenn et al., 2013).

Result 1: *Providing baseline rank information increases goals among underestimators, decreases goals among overestimators, while leaving the goals of accurate predictors largely unchanged. This pattern indicates that information improves the calibration of self-set goals by shifting them away from biased beliefs and toward objective performance. These results support Predictions 1(a), 2(a), and 3(a).*

4.2 Goal Resetting as a Response to Information

In contrast to exogenously imposed goals, which are typically sustained by an extrinsic incentive scheme, self-set goals motivate individuals simply by their intrinsic desires to fulfill self-committed targets. A potential concern in this context is the *stickiness* of such goals, that individuals may adjust their internal targets as new information becomes available without any external constraints. While our experimental design attempted to highlight goal salience by prominently showing goals on the grade cards after each test in the field and displaying them throughout the memorization and test stages of round 2 in the lab, a subject's true internal goal may remain slippery and unobservable.

Leveraging the post-experimental survey of our laboratory experiment, we investigate whether subjects in the *NoInfo* treatment reset goals, by asking whether they revised their goals upon receiving baseline performance feedback, and if so, what their new goals were. Because goal resetting is elicited

²⁶ Appendix Figure D.3 plots the distribution of goals against predicted and actual baseline test ranks by treatment and belief accuracy. It provides visual evidence supporting the findings that subjects generally anchor on the best available information to set goals and goals are set above their (perceived) baselines to motivate higher achievement – a tendency that is stronger in the field than in the laboratory, and that relative performance feedback corrects prior belief distortions, prompting subjects – especially those whose initial goals deviated significantly from actual performance – to revise and set more realistic goals.

ex-post, we consider this analysis as providing suggestive, within-subject evidence on the mechanism of goal formation, which complements the between-subject findings in Section 4.1. Overall, 66.7% of the 150 subjects in the *NoInfo* treatment stated that they revised their goals, but only 58.7% of all subjects set a goal different from their initial ones. The latter constitutes our definition of goal resetting in the subsequent analysis.

Table 3 presents the regression results for the determinants of whether subjects in the *NoInfo* treatment reset their goals. It shows in all specifications that the absolute gap between actual and predicted baseline test ranks (i.e., the prediction error) is a strong predictor of goal resetting. Subjects whose beliefs deviate more from their true performance are significantly more likely to revise their goals once performance information becomes available. This relationship holds symmetrically for underestimators and overestimators. In columns (3) and (4) where final test rank and its squared term are added, we find that neither significantly predicts the likelihood of goal resetting. This suggests that the reported revisions are consistent with a genuine response to baseline feedback rather than an *ex-post* justification for their final test outcomes. This is further confirmed by the joint insignificance of final test rank indicators in a regression with nonlinear performance measure and all control variables (F test, p -value = 0.210).

< Table 3 is about here >

Table 4 investigates the determinants of goal updating within the *NoInfo* treatment. The dependent variable is the rank difference between the revised goal and the original goal. For subjects who reported no revision, the rank difference is coded as zero. We find in columns (1) and (2) that the magnitude of goal updating is strongly and positively correlated with the difference between actual and predicted baseline test ranks. Specifically, each one-rank gap between realized and predicted baseline rank translated into approximately half a rank goal updating toward the actual rank. In columns (3) and (4), the correlation is allowed to differ between underestimators and overestimators. The results indicate that the magnitude of the adjustment is insignificantly different across types. As the sign of the prediction error is positive for underestimators and negative for overestimators, the estimates suggest that underestimators revise their goals upward, whereas overestimators revise their goals downward. When the sample is restricted to subjects who reset their goals in columns (5)-(6), the estimated effects naturally increase in magnitude, reflecting a more pronounced behavioral response.

< Table 4 is about here >

Together, these within-subjects results indicate goal adjustment as a behavioral response to information. Rather than passively adhering to initial goals, individuals in the *NoInfo* treatment also

revise them in systematic ways that reflect the correction of prior beliefs. These results are consistent with the findings from the psychology literature that positive feedback prompts upward goal revision, negative feedback leads individuals to lower or maintain their previous targets to protect self-efficacy (e.g., Williams et al., 2000; Ilies and Judge, 2005; Krenn et al., 2013; Theobald et al., 2025).

Building on this suggestive evidence, we next examine whether accounting for reported goal revision changes the interpretation of the main goal-setting results in columns (7) and (8) of Tables 1 and 2. These specifications recognize that subjects in the *NoInfo* treatment who reset their goals effectively “treated” themselves with information; we therefore reclassify them into the *Info* treatment and use their revised goals as the dependent variable.²⁷ While this reclassification introduces endogeneity, it nevertheless provides a useful test of whether delayed feedback can generate effects on final goal setting comparable to those of providing performance information before initial goal setting.

The results reveal clear heterogeneity across belief-accuracy types. For underestimators, the positive treatment differences in the bottom panel of Table 1 are substantially attenuated in columns (7) and (8). This attenuation is driven by those who initially set relatively low goals and revise their goals upward from a mean of 5.930 to 7.351, a level close to that in the *Info* treatment (7.287), whereas those who do not reset had already set relatively high initial goals before receiving feedback (mean 6.480). For overestimators, by contrast, the negative treatment differences remain sizable in columns (7) and (8). Here, the adjustment is concentrated among those who initially set less inflated goals and revise their goals sharply downward from a mean of 6.652 to 4.435, even below the mean goal in the *Info* treatment (5.244), whereas those who do not reset retain more overoptimistic initial goals (mean 6.941). Consistent with this pattern, the top panel of Table 2 shows that subjects who reset their goals after receiving delayed feedback place greater weight on actual baseline rank and less weight on predicted rank, while non-resetters remain more anchored to predicted rank. Taken together, these patterns suggest that the timing of information appears less consequential for underestimators, but remains more relevant for overestimators.

Result 2: *Self-set goals are endogenous and adjust systematically to newly revealed information. In the laboratory experiment, both the likelihood and the magnitude of goal resetting increase with the gap between predicted and actual baseline rank, with underestimators revising goals upward and overestimators revising goals downward. Once this adjustment is taken into account, treatment differences in final goals for underestimators are substantially attenuated, suggesting that delayed feedback can largely replicate the effect of immediate information on goals set for this group. For overestimators, however, treatment differences remain sizable, indicating that the timing of information continues to matter for this group.*

²⁷ Overall, 88 subjects were reassigned into the *Info* treatment – 57 underestimate, 8 accurate, and 23 overestimate, while 25, 20, and 17 of the respective types remain in the *NoInfo* treatment.

4.3 From Goals to Performance

Having established that information systematically shifts self-set goals, we next examine whether these adjustments in goals translate into changes in performance. We evaluate both relative performance in terms of final test rank in Table 5, and absolute performance in terms of plank duration in the field experiment or correct items in the laboratory experiment in Table 6. Since the absolute performance measures differ across the field and laboratory experiments, we only conduct pooled-data analysis for relative performance.²⁸ Our discussion predominantly focuses on the implied differential effects of information provision for each of the three accuracy types reported in the bottom panels, especially for subjects who under- or overestimate their baseline performance.

< Tables 5 and 6 are about here >

The results from a single experiment (columns (1)-(4) in Tables 5 and 6) shows that the impact of information on performance is limited, despite its strong influence on goals. In the field experiment, information provision significantly improves both the relative and absolute performance of underestimators. However, these effects are absent for overestimators and for all accuracy types in the laboratory sample. When pooling the datasets (columns (5)-(6) in Table 5), the estimated treatment effects on relative performance are close to zero and statistically insignificant, probably due to the larger size of the laboratory sample. These results in general provide evidence in support of Prediction 2(b) but against Predictions 1(b) and 3(b).

How can these diverging effects of feedback on goals and performance be reconciled? A potential explanation may lie in the behavioral mechanism identified in the previous section – goal resetting. If goal setting under more accurate information was to improve performance, and a significant proportion of subjects in the *NoInfo* treatment updated their goals upon receiving baseline performance feedback before performing the final test, the observed treatment differences would be attenuated toward zero. As documented in Section 4.2, more than half of the laboratory subjects in the *NoInfo* treatment indeed revised their goals as if they had been effectively “treated”, thereby lending support to this conjecture.

Consistent with this interpretation, when goal resetters are reclassified into the *Info* treatment in columns (7) and (8) of Table 5 and columns (5) and (6) of Table 6, the treatment effects for underestimators increase from those in columns (3) and (4) in both tables, and even become marginally significant in column (5) of Table 6. In contrast – inconsistent with this interpretation – performance in the final test markedly decreases for overestimators, and is statistically significant in three out of four estimates. It is noteworthy that these changes in performance align qualitatively with the direction of

²⁸ Analyses regarding absolute performance are hence relegated to be secondary.

goal adjustments – performance tends to increase for those who underestimate their baseline performance and upwardly revise their goals in response to information, and decrease for those who overestimate baseline performance and downwardly revise their goals. Performance does not change for subjects holding accurate beliefs.

While information clearly improves calibration of self-set goals, this need not imply that the resulting goals are performance-maximizing. In particular, the deterioration in performance among overestimators is consistent with the possibility that negative feedback induces an over-correction in goal choice: once overly optimistic beliefs are corrected, some students may lower their goals beyond the “challenging-but-attainable” region that yields maximal motivation. In other words, information narrows the gap between goals and baseline performance and may induce overestimators to set safer goals that lack sufficient challenge. This could weaken the motivational role of self-commitment and result in a worse final performance. This interpretation is of course suggestive, as we are unable to observe each individual’s performance-maximizing goal. Nevertheless, it highlights an important distinction between “better informed” goals in the sense of being aligned with objective ability and “better” goals in the sense of sustaining maximal motivation.

Our findings offer a potential explanation of the heterogeneous results found in the recent literature regarding the effect of relative performance feedback on educational performance. For example, while Azmat et al. (2019) find that underestimating students perform worse, Brade et al. (2022) find that underestimators perform better, and Dobrescu et al. (2021) show performance gains in general. Our data highlights the possibility that such accurate-type-based heterogeneity may arise from goal (re-)adjustment typically unobserved by researchers.

While we can only directly observe endogenous goal resetting in the laboratory survey, it is highly likely that a similar internal recalibration process also occurred among subjects in the field experiment. If that were the case, then the observed performance differences between treatments would have been the effectively attenuated estimates. Put differently, the true effect of setting an informed goal on physical performance is likely even larger. This aligns with the intuition that explicit goal-setting interventions may yield particularly pronounced effects in physical endurance tasks, where the marginal cost of effort is steeply increasing and external anchors are crucial for overcoming physiological limits, compared to cognitive tasks where internal expectations might be more fluid and easily reset.

Since our hypotheses hinge on the categorization of belief accuracy, we also verify whether our results hold when using a broader definition of this variable.²⁹ In this alternative definition, students are classified as accurate both when they correctly predicted their rank and when they mispredicted their rank by one rank (from above or below). As Tables D.4 and D.5 in Appendix D report, the results are generally robust. It is noteworthy that broadening the accurate category reduces the sample sizes of

²⁹ Since this analysis does not test model predictions, it is considered exploratory.

the underestimator and overestimator groups. Consequently, the previously significant treatment effect for underestimators in the field experiment (columns (1) and (2)) loses statistical significance in three out of four estimates, although the direction of the effect remains unchanged. Conversely, when goal reseters are reclassified into the *Info* treatment in columns (7) and (8) of Appendix Table D.4 and columns (5) and (6) of Appendix Table D.5, the negative treatment effects for overestimators become substantially larger in magnitude and statistical significance compared to the estimates in respective columns of Tables 5 and 6. This amplification suggests that students whose initial prediction deviated further from their actual baseline performance made larger downward revisions to their goals (as shown in Table 4), which ultimately translates into a more pronounced reduction in final test performance. To sum up, this analysis not only confirms the general robustness of our results with the “exact” accurate definition, but also helps explain the underlying behavioral mechanism through goal resetting.

Result 3: *Prior to accounting for goal resetting, the effect of information on performance is limited – it only improves performance for underestimators in the field experiment. After accounting for goal resetting, the performance effects are consistent with the direction of goal adjustment: underestimators improve their final performance as they revise their goals upward, whereas overestimators perform worse as they revise their goals downward, perhaps to a too low level that is better calibrated to baseline performance but insufficiently challenging to sustain motivation. These results support Prediction 2(b) but contradict Predictions 1(b) and 3(b).*

To further validate the behavioral assumptions underlying our conceptual framework, Appendix E explores whether self-set goals indeed serve as reference points in the evaluation of outcomes. Using self-reported satisfaction with final test performance as an empirical proxy for utility, we find that satisfaction is non-linearly related to the performance-goal discrepancy. Specifically, consistent with the prospect theory value function (Kahneman and Tversky, 1979), satisfaction increases at an increasing rate when performance approaches the goal from below (the loss domain), and increases at a decreasing rate when it exceeds the goal (the gain domain). Furthermore, the marginal psychological effect is significantly steeper in the loss domain than in the gain domain. This supplementary exploratory analysis provides direct empirical support for the premise that falling short of a goal imposes convex psychological costs, which is the core mechanism driving the motivational power of goal setting in our model.

5. Conclusion

This paper studies how information about relative performance shapes self-set goals and subsequent performance for individuals with different prior belief accuracy. Across both a field and a

complementary laboratory experiment, we exogenously vary whether individuals receive baseline relative performance feedback before or after setting non-binding goals.

We document that individuals who are not informed about their performance before goal setting anchor goals primarily on subjective, often inaccurate beliefs. Providing information systematically corrects these distortions: individuals who underestimate their relative ability revise their goals upward, while overestimators adjust their targets downward. However, the effect of information on performance is limited – it only improves performance for underestimators in the field experiment. To explore the underlying mechanism, we leverage laboratory evidence to examine voluntary goal resetting. We find that delayed feedback affects final goals only when individuals reset them. Accordingly, treatment differences in final goals are substantially attenuated for underestimators but remain sizable for overestimators. Once this goal adjustment process is taken into account, upward revisions among underestimators improve performance, whereas downward revisions among overestimators reduce challenge and impair ultimate performance. This distinction highlights a critical trade-off that better-informed goals are not necessarily performance-maximizing goals.

Our findings highlight that the behavioral impact of relative performance feedback depends crucially on individuals' underlying belief accuracy. Consequently, evaluations of goal-setting interventions that rely solely on average performance effects risk masking these divergent responses. Enhancing the efficacy of such behavioral nudges requires making a critical distinction between promoting “realistic” goals and fostering “motivating” goals. Future research could estimate individual effort-provision curves by mapping various self-selected goals to performance, and test whether and how correcting miscalibrated beliefs can be optimized to drive individuals to set goals within their optimal motivating range.

References

- Altan, B., Guney, B., and Uler, N. (2026). Goal Setting, Goal-Setter Identity, and Employer Feedback in Non-Pecuniary Environments. Working paper.
- Azmat, G., Bagues, M., Cabrales, A., and Iriberry, N. (2019). What you don't know can't hurt you? A natural field experiment on relative performance feedback in higher education. *Management Science*, 65(8), 3714-3736.
- Azmat, G. and Iriberry, N. (2010). The importance of relative performance feedback information: Evidence from a natural experiment using high school students. *Journal of public economics*, 94(7-8), 435-452.
- Brade, R., Himmler, O., and Jäckle, R. (2022). Relative performance feedback and the effects of being above average—field experiment and replication. *Economics of Education Review*, 89, 102268.
- Brade, R., Himmler, O., Jäckle, R., and Weinschenk, P. (2024). Helping Students to Succeed—The Long-Term Effects of Soft Commitments and Reminders. mimeo
- Brookins, P., S. J. Goerg, and Kube, S. (2017). Self-chosen goals, incentives, and effort. Unpublished manuscript.
- Clark, D., Gill, D., Prowse, V. and Rush, M. (2020). Using goals to motivate college students: Theory and evidence from field experiments. *Review of Economics and Statistics*, 102(4), 648-663.
- Corgnet, B., Gómez-Miñambres, J. and Hernán-Gonzalez, R. (2015). Goal setting and monetary incentives: When large stakes are not enough. *Management Science*, 61(12), 2926-2944.
- Corgnet, B., Gómez-Miñambres, J. and R. Hernán-González. (2018). Goal setting in the principal–agent model: weak incentives for strong performance. *Games and Economic Behavior*, 109, 311–326.
- Dobrescu, L. I., Faravelli, M., Megalokonomou, R., and Motta, A. (2021). Relative performance feedback in education: Evidence from a randomised controlled trial. *The Economic Journal*, 131(640), 3145-3181.
- Dobronyi, C. R., Oreopoulos, P. and Petronijevic, U. (2019). Goal setting, academic reminders, and college success: a large-scale field experiment. *Journal of Research on Educational Effectiveness*, 12(1), 38-66.
- Fan, J. and Gómez-Miñambres, J. (2020), Nonbinding goals in teams: A real effort coordination experiment. *Manufacturing & Service Operations Management* 22(5), 1026-1044.
- Fischbacher, U. (2007). z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental economics*, 10(2), 171-178.
- Gill, D., Kissová, Z., Lee, J. and Prowse, V. (2019). First-place loving and last-place loathing: how rank in the distribution of performance affects effort provision. *Management Science* 65, 494–507.
- Goerg, S. J. and Kube, S. (2012). Goals (th)at work, goals, monetary incentives, and workers' performance. Discussion Paper Series of the Max Planck Institute for Research on Collective Goods (2012_19).

- Gosnell, G. K., List, J. A. and Metcalfe, R. D. (2020). The impact of management practices on employee productivity: A field experiment with airline captains. *Journal of Political Economy*, 128(4), 1195-1233.
- Harding, M. and Hsiaw, A. (2014). Goal setting and energy conservation. *Journal of Economic Behavior & Organization*, 107, 209–227.
- Harrison, G.W., and List, J.A. (2004). Field experiments. *Journal of Economic Literature*, 42(4), 1009-55.
- Heath, C., Larrick, R. P. and Wu, G. (1999). Goals as reference points. *Cognitive Psychology*, 38(1), 79–109.
- Hsiaw, A. (2013). Goal-setting and self-control. *Journal of Economic Theory*, 148(2), 601-626.
- Ilies, R., and Judge, T. A. (2005). Goal regulation across time: the effects of feedback and affect. *The Journal of applied psychology*, 90(3), 453–467.
- Kahneman, D. and Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2), 263-292.
- Kaiser, J. P., Koch, A. K., and Nafziger, J. (2024). Does goal revision undermine self-regulation through goals? An experiment. *Experimental Economics*, 27, 604–636.
- Koch, A.K. and Nafziger, J. (2011). Self-regulation through goal setting. *Scandinavian Journal of Economics*, 113(1), 212-227.
- Krenn, B., Würth, S., and Hergovich, A. (2013). The impact of feedback on goal setting and task performance. *Swiss Journal of Psychology*.
- Kuhn, P. J. and Yu, L. (2021). Kinks as goals: Accelerating commissions and the performance of sales teams (No. w28487). National Bureau of Economic Research.
- Laibson, D. (1997). Golden eggs and hyperbolic discounting. *The Quarterly Journal of Economics*, 112(2), 443-478.
- Locke, E. A. (1968). Toward a theory of task motivation incentives. *Organizational Behavior and Human Performance*, 3, 157–189.
- Locke, E. A. and Latham, G. P. (1985). The application of goal setting to sports. *Journal of Sport Psychology*, 7, 205–222.
- Niederle, M. and Vesterlund, L. (2007). Do women shy away from competition? Do men compete too much? *The quarterly journal of economics*, 122(3), 1067-1101.
- O'Donoghue, T. and Rabin, M. (1999). Doing it now or later. *American economic review*, 89(1), 103-124.
- Oreopoulos, P., Patterson, E.W., Petronijevic, U. and Pope, N.G. (2022). Low-touch attempts to improve time management among traditional and online college students, *Journal of Human Resources*, 57(1), pp. 1–43.

- Oreopoulos, P., and Petronijevic, U. (2023). The promises and pitfalls of using (mostly) low-touch coaching interventions to improve college student outcomes. *The Economic Journal*, 133(656), 3034-3070.
- Sackett, A. M., Wu, G., White, R. J. and Markle, A. (2014). Harnessing optimism: How eliciting goals improves performance. Working Paper. Available at SSRN: <https://ssrn.com/abstract=2544020> or <http://dx.doi.org/10.2139/ssrn.2544020>
- Sauer, S. J., Rodgers, M. and Becker, W. J. (2018). The effects of goals and pay structure on managerial reporting dishonesty. *Journal of Accounting, Ethics and Public Policy*, 19(3).
- Smithers, S. (2015). Goals, motivation and gender. *Economics Letters*, 131, 75–77.
- Suvorov, A. and Van de Ven, J. (2008). Goal setting as a self-regulation mechanism. Working Paper. Available at SSRN 1286029.
- Strotz, R. H. (1955). Myopia and Inconsistency in Dynamic Utility Maximization, *Review of Economic Studies* 23, 165–180.
- Theobald, M., Lin, W. M., Sakaki, M., Murayama, K., and Brod, G. (2025). How (much) do people revise their goals after success and failure? The Triple-A model of goal revision. *Educational Psychologist*, 1–21.
- Tran, A. and Zeckhauser, R. (2012). Rank as an inherent incentive: Evidence from a field experiment. *Journal of Public Economics*, 96(9-10), 645-650.
- Uetake, K. and Yang, N. (2018). Success breeds success: weight loss dynamics in the presence of short-term and long-term goals (No. 170002).
- van Lent, M. and Souverijn, M. (2020). Goal setting and raising the bar: A field experiment. *Journal of Behavioral and Experimental Economics*, 87, 1-16.
- van Lent, M. (2019). Goal setting, information, and goal revision: A field experiment. *German Economic Review*, 20(4), e949-e972.
- Williams, K. J., Donovan, J. J., and Dodge, T. L. (2000). Self-regulation of performance: Goal establishment and goal revision processes in athletes. *Human Performance*, 13(2), 159-180.
- Wu, G., Heath, C. and Larrick, R. (2008). A prospect theory model of goal behavior. Unpublished manuscript.

Figures and Tables

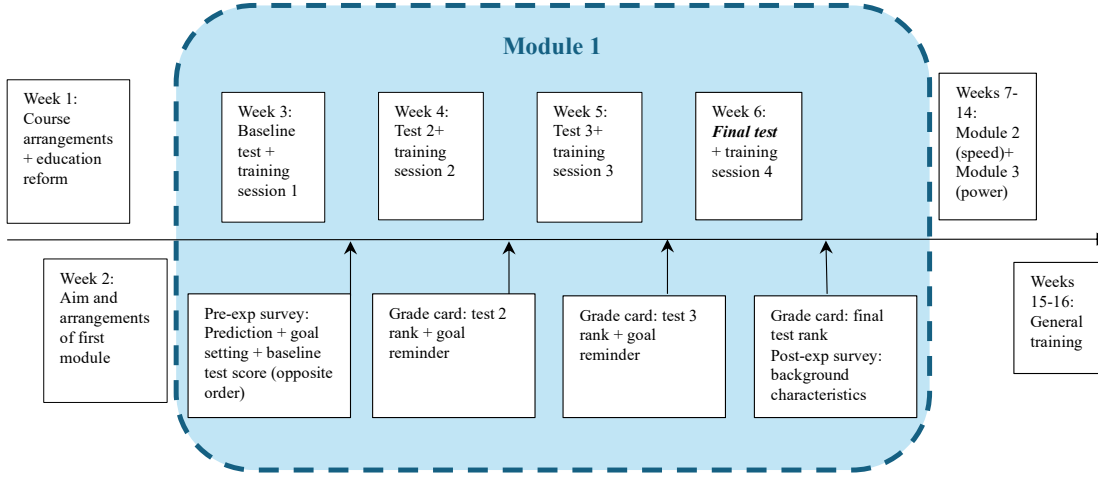


Figure 1: Timeline of the field experiment

Notes: This figure illustrates the timeline of the field experiment. The baseline test was conducted before the pre-experimental survey. In the pre-experimental survey, subjects predicted their baseline relative performance and, depending on treatment assignment, set goals either before or after receiving information about their actual baseline performance. Subsequent reminder messages reported updated scores and repeated subjects' goals. The post-experimental survey was administered after the final training session.

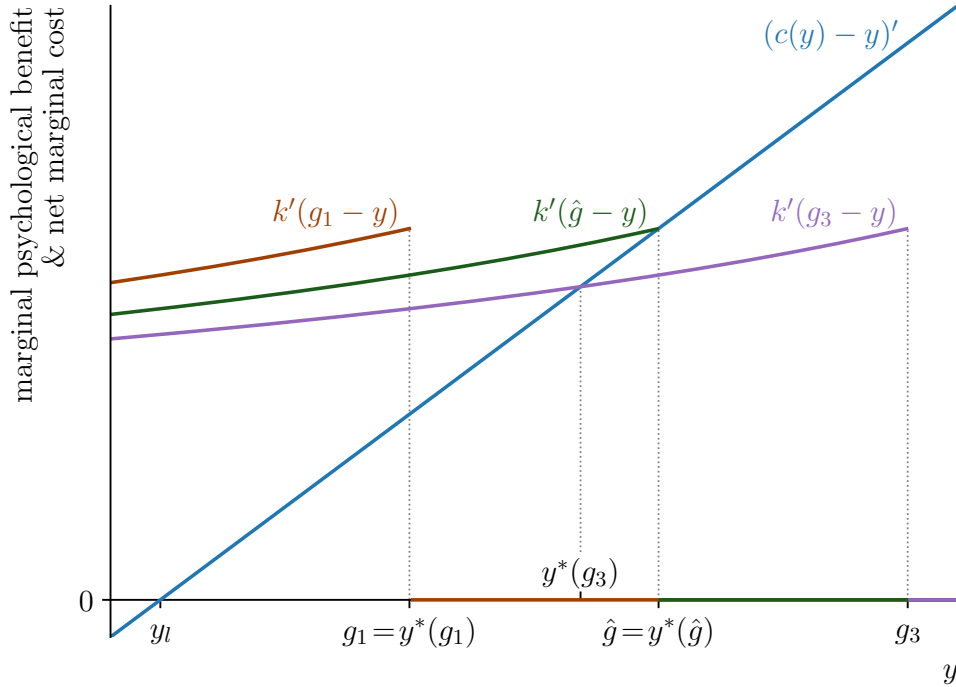
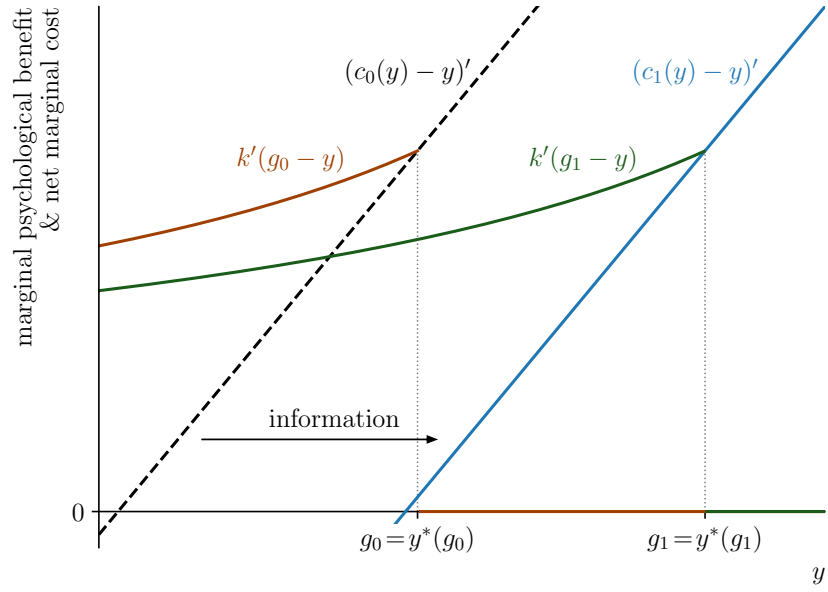
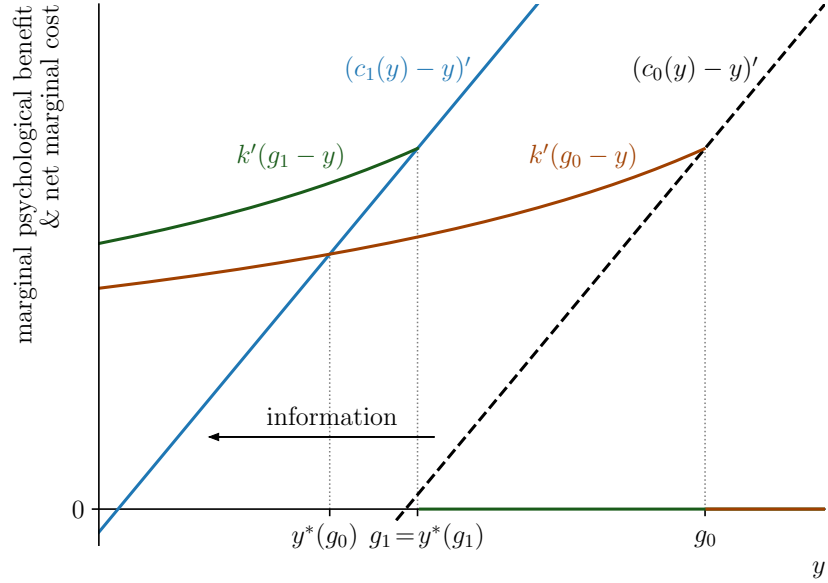


Figure 2: Performance for different goals

Notes: The figure shows optimal performance $y^*(g)$ for three goals $g_1 < \hat{g} < g_3$. The curve $(c(y) - y)'$ graphs the net marginal cost while the other three curves depict the marginal psychological benefits $k'(g - y)$ for the respective goals. Optimal performance for each goal is: $y^*(g_1) = g_1$, $y^*(\hat{g}) = \hat{g}$, $y^*(g_3) < g_3$. While a higher goal may increase performance, an excessively ambitious goal need not be attained. Performance is maximized by the challenging but attainable goal $\hat{g}$ rather than by the highest goal g_3 .



(a) Underestimators



(b) Overestimators

Figure 3: Effect of information about relative performance

Notes: This figure illustrates how information about relative baseline performance affects goal setting and performance when students hold incorrect beliefs about their relative ability. In both panels, the curve $(c_0(y) - y)'$ represents the perceived net marginal cost under incorrect beliefs, while $(c_1(y) - y)'$ represents the true net marginal cost. The curves $k'(g_0 - y)$ and $k'(g_1 - y)$ denote the marginal psychological benefits associated with goals chosen before and after information, respectively. $y^*(g_0)$ and $y^*(g_1)$ represent the performance level chosen by the doer based on the true net costs for either goal. In panel (a), the student underestimates her relative ability, so information shifts the net marginal cost curve to the right and leads to a higher goal and higher performance. In panel (b), the student overestimates her relative ability, so information shifts the net marginal cost curve to the left and leads to a lower but more realistic goal that results in higher performance.

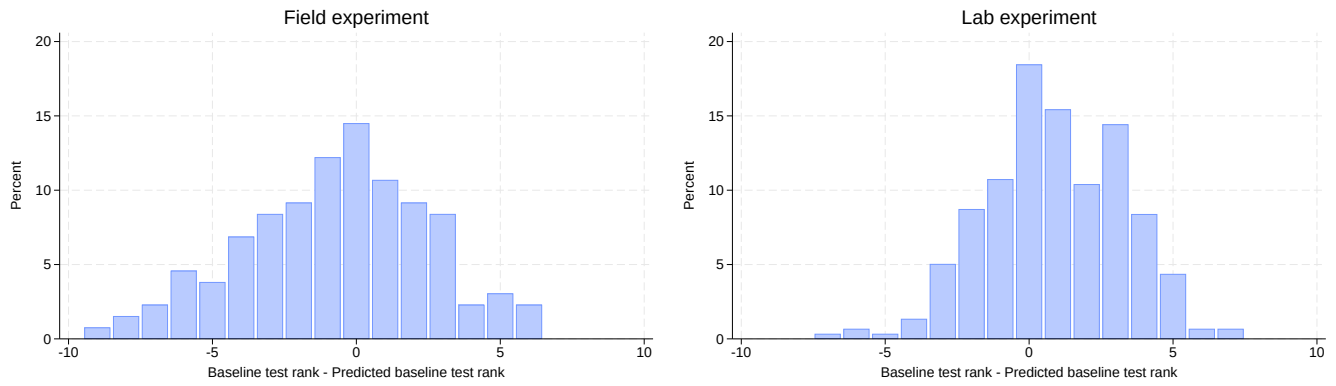


Figure 4: Distribution of belief (in)accuracy.

Notes: This figure plots the distribution of belief errors, measured as actual baseline rank minus predicted baseline rank. Positive values indicate underestimation of one's relative performance, negative values indicate overestimation, and zero indicates accurate beliefs. In both experiments, ranks range from 1 to 10, with higher ranks indicating better relative performance.

Table 1: Treatment effect on goals by accuracy

	Field		Lab		Pooled		Lab, goal reset	
Dep. Var.: Goal	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Underestimate	-1.120*	-1.831**	-2.051***	-1.923***	-1.800***	-1.789***	-1.501**	-1.400**
	(0.601)	(0.733)	(0.368)	(0.321)	(0.296)	(0.262)	(0.601)	(0.528)
Overestimate	1.639**	1.306**	1.337**	1.371***	1.441***	1.400***	1.208**	1.155**
	(0.578)	(0.572)	(0.473)	(0.411)	(0.363)	(0.316)	(0.505)	(0.439)
Info	0.117	-0.367	-0.594	-0.384	-0.412	-0.342	-0.846	-0.699
	(0.444)	(0.421)	(0.651)	(0.638)	(0.487)	(0.488)	(0.722)	(0.686)
Info $\times$ Underestimate	1.314*	2.005**	1.801***	1.647**	1.713***	1.667***	1.254	1.095
	(0.624)	(0.764)	(0.587)	(0.584)	(0.445)	(0.466)	(0.798)	(0.711)
Info $\times$ Overestimate	-1.452*	-1.151	-0.915	-1.042	-1.041*	-1.081**	-0.819	-0.861
	(0.684)	(0.684)	(0.684)	(0.620)	(0.526)	(0.478)	(0.704)	(0.597)
<i>Estimated Differences: Info - NoInfo</i>								
Underestimate	1.430***	1.638***	1.207***	1.263***	1.301***	1.325***	0.408	0.395
	(0.436)	(0.531)	(0.205)	(0.190)	(0.180)	(0.172)	(0.430)	(0.404)
Accurate	0.117	-0.367	-0.594	-0.384	-0.412	-0.342	-0.846	-0.699
	(0.444)	(0.421)	(0.651)	(0.638)	(0.487)	(0.488)	(0.722)	(0.686)
Overestimate	-1.335***	-1.518***	-1.510***	-1.426***	-1.452***	-1.423***	-1.665***	-1.560**
	(0.403)	(0.453)	(0.283)	(0.296)	(0.222)	(0.230)	(0.406)	(0.553)
Baseline test performance	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Group fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Personal characteristics	No	Yes	No	Yes	No	Yes	No	Yes
Observations	131	128	298	298	429	427	298	298

Notes: this table reports the estimates from OLS models (top panel) and the implied differential treatment effects (bottom panel) for the field experiment, the laboratory experiment, the pooled data, and the laboratory experiment where subjects in the *NoInfo* treatment, who revised their goal, were reclassified to the *Info* treatment. For the reclassified subjects in column (7) and (8), the revised goal is used as the outcome variable. Indicators for “Baseline test performance” refer to the nine dummy variables for the respective baseline test ranks. Indicators for “Group fixed effects” refers to the indicator variables for the semester-class-gender level in the field experiment and the individual experimental sessions in the laboratory experiment. For the field experiment, individual characteristics include gender, monthly income (asinh), no. of siblings, birth order, GPA, age, height, weight, fat rate, and metabolic rate, Han, affiliated, urban Hukou, boarding school, healthy; the female indicator is absorbed by the semester-class-gender fixed effects. For the laboratory experiment, individual characteristics include a female and other-gender indicators, monthly income (asinh), no. of siblings, birth order, GPA, an indicator for missing recent GPA, an indicator for undergraduate status and for master’s degree or above, and an indicator for number of past experiment participations. For the pooled data, individual characteristics include female indicator, monthly income (asinh), no. of siblings, birth order. Although GPA data are available for both experiments, they are excluded from the pooled regressions, because the grading scales differ substantially across universities and are therefore not directly comparable. Since the pooled specification uses fewer individual controls than the field and laboratory specifications, column (6) has more observations than columns (2) and (4) combined. One subject in the field experiment is excluded from the analysis because of non-attendance at the baseline test, and one subject in the laboratory experiment is excluded because of a protocol violation (cheating on the baseline test). Both nevertheless remain part of the original ranking sets used to generate the relative-performance information: the field subject was assigned the lowest rank under the original procedure, whereas the laboratory subject retained the rank implied by actual performance. Estimates for control variables are not reported. Robust standard errors allowing for heteroskedasticity and clustered at the group level (as defined above) are reported in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Table 2: Determinants of goals

Dep. Var.: Goal	Field		Lab		Pooled		Lab, goal reset	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Info	-0.103 (0.555)	-0.414 (0.427)	-1.167 (0.939)	-0.935 (0.847)	-0.929 (0.713)	-0.829 (0.663)	-1.822* (0.990)	-1.789* (0.942)
Baseline test rank	0.0965 (0.0803)	0.0937 (0.0819)	-0.0292 (0.0590)	-0.00523 (0.0562)	0.0366 (0.0503)	0.0507 (0.0484)	-0.0299 (0.206)	0.0174 (0.211)
Baseline test rank $\times$ Info	0.427*** (0.0802)	0.447*** (0.0994)	0.597*** (0.0610)	0.569*** (0.0661)	0.517*** (0.0525)	0.515*** (0.0538)	0.649*** (0.196)	0.614*** (0.203)
Predicted baseline test rank	0.518*** (0.0868)	0.504*** (0.0957)	0.665*** (0.0573)	0.636*** (0.0495)	0.576*** (0.0583)	0.564*** (0.0540)	0.638*** (0.168)	0.583*** (0.169)
Predicted baseline test rank $\times$ Info	-0.376*** (0.0887)	-0.345*** (0.0869)	-0.413*** (0.115)	-0.413*** (0.0907)	-0.350*** (0.0902)	-0.357*** (0.0765)	-0.445** (0.181)	-0.404** (0.168)
<i>Linear Combination:</i>								
Info: Baseline test rank	0.523*** (0.0506)	0.540*** (0.0386)	0.568*** (0.0504)	0.564*** (0.0641)	0.554*** (0.0378)	0.566*** (0.0439)	0.620*** (0.0365)	0.631*** (0.0478)
Info: Predicted baseline test rank	0.142** (0.0613)	0.160*** (0.0475)	0.252*** (0.0767)	0.223*** (0.0739)	0.226*** (0.0555)	0.207*** (0.0523)	0.194*** (0.0449)	0.178*** (0.0448)
Group fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Personal characteristics	No	Yes	No	Yes	No	Yes	No	Yes
Observations	131	128	298	298	429	427	298	298

Notes: this table reports the estimates from OLS models (top panel) and the linear combination of the baseline test and predicted baseline test with their respective info-treatment interaction term (bottom panel). Estimates for control variables are not reported. See Table 1 notes for a list of the control variables and details on sample construction. Robust standard errors allowing for heteroskedasticity and clustered at the group level are reported in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Table 3: Determinants of whether goal is re-set

Dep. Var.: Goal was reset	(1)	(2)	(3)	(4)
Baseline test rank - Predicted baseline test rank	0.110*** (0.0260)	0.106*** (0.0266)	0.119*** (0.0245)	0.113*** (0.0292)
Baseline test rank - Predicted baseline test rank $\times$ Overestimate	-0.0137 (0.0340)	0.0128 (0.0314)	-0.0296 (0.0359)	0.000531 (0.0338)
Final test rank			-0.0681 (0.0790)	-0.0550 (0.0884)
Final test rank, squared			0.00468 (0.00696)	0.00390 (0.00796)
Group fixed effects	Yes	Yes	Yes	Yes
Personal characteristics	No	Yes	No	Yes
Observations	150	150	150	150

Notes: this table reports the estimates from OLS models for the subjects in the *NoInfo* treatment of the laboratory experiment. The outcome variable, goal was reset, equals 1 if subjects stated in the post-experimental survey that they revised their initial goal and set a goal different from the initial one, and 0 otherwise. Estimates for control variables are not reported. See Table 1 notes for a list of the control variables and details on sample construction relevant to the laboratory experiment. Robust standard errors allowing for heteroskedasticity and clustered at the group level are reported in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Table 4: Goal updating

Dep. Var.: Revised goal - original goal	(1)	(2)	(3)	(4)	(5)	(6)
Baseline test rank - Predicted baseline test rank	0.498*** (0.0736)	0.499*** (0.0873)	0.371*** (0.120)	0.380** (0.131)	0.670*** (0.156)	0.711*** (0.133)
(Baseline test rank - Predicted baseline test rank) $\times$ Overestimate			0.338 (0.193)	0.323 (0.191)	0.157 (0.270)	-0.00838 (0.207)
Group fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Personal characteristics	No	Yes	No	Yes	No	Yes
Observations	150	150	150	150	88	88

Notes: this table reports the estimates from OLS models for subjects in the *NoInfo* treatment of the laboratory experiment. The outcome variable is the difference between the revised goal and the original goal. For subjects who reported no revision, the difference is coded as zero. In column (5) and (6), the sample is further restricted to only those who revised their goals. Estimates for control variables are not reported. See Table 1 notes for a list of the control variables and details on sample construction relevant to the laboratory experiment. Robust standard errors allowing for heteroskedasticity and clustered at the group level are reported in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Table 5: Treatment effect on relative performance in the final test by accuracy

Dep. Var.: Final test rank	Field		Lab		Pooled		Lab, goal reset	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Underestimate	-0.922** (0.398)	-1.509** (0.502)	-0.528 (0.814)	-0.468 (0.836)	-0.504 (0.660)	-0.435 (0.685)	-1.012 (0.758)	-0.985 (0.841)
Overestimate	0.316 (0.688)	0.706 (0.802)	-0.504 (0.589)	-0.576 (0.586)	-0.159 (0.506)	-0.229 (0.502)	-0.0672 (0.709)	-0.234 (0.616)
Info	-0.932 (0.734)	-1.133 (0.698)	-0.378 (0.654)	-0.441 (0.760)	-0.322 (0.554)	-0.324 (0.568)	-0.925 (0.532)	-1.042 (0.701)
Info $\times$ Underestimate	1.849** (0.840)	2.207** (0.934)	0.0862 (0.727)	0.0331 (0.922)	0.266 (0.651)	0.178 (0.700)	0.864 (0.495)	0.896 (0.699)
Info $\times$ Overestimate	1.122 (0.912)	1.424 (1.138)	0.136 (0.901)	0.263 (0.902)	0.176 (0.687)	0.264 (0.684)	-0.313 (0.784)	-0.0707 (0.688)
<i>Estimated Differences: Info - NoInfo</i>								
Underestimate	0.917** (0.409)	1.074** (0.496)	-0.292 (0.303)	-0.408 (0.354)	-0.0561 (0.283)	-0.145 (0.304)	-0.0617 (0.423)	-0.145 (0.554)
Accurate	-0.932 (0.734)	-1.133 (0.698)	-0.378 (0.654)	-0.441 (0.760)	-0.322 (0.554)	-0.324 (0.568)	-0.925 (0.532)	-1.042 (0.701)
Overestimate	0.190 (0.879)	0.291 (0.964)	-0.242 (0.649)	-0.179 (0.600)	-0.146 (0.491)	-0.0596 (0.498)	-1.239* (0.652)	-1.112 (0.634)
Baseline test performance	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Group fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Personal characteristics	No	Yes	No	Yes	No	Yes	No	Yes
Observations	131	128	298	298	429	427	298	298

Notes: This table reports the estimates from OLS models (top panel) and the implied differential treatment effects (bottom panel) for the field experiment, the laboratory experiment, the pooled data, and the laboratory experiment where subjects in the *NoInfo* treatment, who revised their goal, were reclassified to the *Info* treatment. Estimates for control variables are not reported. See Table 1 notes for a list of the control variables and details on sample construction. Robust standard errors allowing for heteroskedasticity and clustered at the group level are reported in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Table 6: Treatment effect on absolute performance in the final test by accuracy

	Field		Lab		Lab, goal reset	
Dep. Var.: Final test abs. performance	(1)	(2)	(3)	(4)	(5)	(6)
Underestimate	-9.650 (16.33)	-16.29 (19.80)	-1.396 (2.018)	-1.898 (2.131)	-3.814** (1.514)	-3.887** (1.753)
Overestimate	30.04* (16.65)	27.85 (19.48)	-1.750 (1.803)	-1.787 (1.963)	-0.761 (1.743)	-1.004 (1.787)
Info	6.446 (14.51)	1.666 (14.94)	-1.186 (2.094)	-1.305 (2.123)	-2.408 (2.088)	-2.622 (2.348)
Info $\times$ Underestimate	22.81 (23.78)	21.05 (21.43)	0.584 (2.047)	0.440 (2.326)	3.831** (1.735)	3.305* (1.853)
Info $\times$ Overestimate	-14.51 (20.82)	1.952 (24.21)	-0.0791 (2.732)	0.207 (2.618)	-0.870 (2.126)	-0.412 (1.942)
<i>Estimated Differences: Info - NoInfo</i>						
Underestimate	29.25** (12.93)	22.72* (11.07)	-0.602 (0.866)	-0.865 (1.012)	1.423* (0.786)	0.683 (1.205)
Accurate	6.446 (14.51)	1.666 (14.94)	-1.186 (2.094)	-1.305 (2.123)	-2.408 (2.088)	-2.622 (2.348)
Overestimate	-8.063 (20.99)	3.618 (23.70)	-1.265 (1.572)	-1.098 (1.526)	-3.278** (1.339)	-3.034* (1.655)
Baseline test performance	Yes	Yes	Yes	Yes	Yes	Yes
Group fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Personal characteristics	No	Yes	No	Yes	No	Yes
Observations	131	128	298	298	298	298

Notes: This table reports the estimates from OLS models (top panel) and the implied differential treatment effects (bottom panel) for the field experiment, the laboratory experiment, and the laboratory experiment where subjects in the *NoInfo* treatment, who revised their goal, were reclassified to the *Info* treatment. Estimates for control variables are not reported. Controls for baseline test performance are the time the plank is held in the field experiment and the total correct entries in the lab experiment, each in the baseline test. See Table 1 notes for a list of the remaining control variables and details on sample construction. Robust standard errors allowing for heteroskedasticity and clustered at the group level are reported in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Information, Goal Setting and Performance: Experimental Evidence

Haoran He

haoran.he@bnu.edu.cn

Beijing Normal University

Marcus Roel

marcus.roel@nottingham.edu.cn

University of Nottingham Ningbo China

Ronghai Su

sronghai@163.com

Beijing Normal University

Qian Weng

qian.weng@ruc.edu.cn

Renmin University of China

– Online Appendix –

Contents

A	Additional details about the experiment	3
A.1	First survey goal-setting related questions for the <i>Info</i> treatment	3
A.2	An example of the WeChat message for the goal-setting treatments	4
A.3	Survey questions related to socioeconomic background in the second survey	4
A.4	Earnings by rank and test in laboratory experiment	6
A.5	Laboratory experimental instructions	6
B	Power analysis	20
B.1	Field experiment: realized power	20
B.2	Laboratory experiment	20
B.3	Interpretation for the combined evidence	22
C	Conceptual Framework	23
C.1	Discussion of the model	23
C.2	Assumptions	24
C.3	Proofs	25
C.4	A simple equilibrium model of goalsetting	28
D	Additional figures and tables	33
E	Satisfaction with final test performance	42
F	Analysis incorporating a control group without goal-setting intervention	44
	References	48

List of Figures

D.1	Distribution of the number of pages submitted	33
D.2	Distribution of goals by treatment	37
D.3	Distribution of goals against predicted and actual baseline test ranks by treatment and accuracy	39

List of Tables

D.1	Variable definitions	34
D.2	Summary statistics	35
D.3	Randomization check	36
D.4	Treatment effect on relative performance in the final test by accuracy; Robustness check: broader notion of accurate type	40
D.5	Treatment effect on absolute performance in the final test by accuracy; Robustness check: broader notion of accurate type	41
E.1	Summary statistics satisfaction	42
E.2	Performance compared to goal	42

E.3	Satisfaction regarding final test performance	43
F.1	Summary statistics - NoGoal treatment	44
F.2	Randomization check - field experiment, NoGoal, NoInfo, and Info treatment	45
F.3	Treatment effect on performance in the final test by accuracy - field experiment, NoGoal, NoInfo, and Info treatment	47

A. Additional details about the experiment

A.1. FIRST SURVEY GOAL-SETTING RELATED QUESTIONS FOR THE *INFO* TREATMENT

1. Please predict your score on last week's plank test (test 1) (choose one option in the following table).

Notes: Please note that you will earn 2 out of 3 points by completing all other questions in this survey; the other 1 point depends on how accurate your prediction is in this question. The scoring criteria are as follows: if your prediction ranking is the same as the actual ranking, you will earn 1 point; if your prediction is 1 ranking higher or lower than the actual ranking, you will earn 0.5 points; if your prediction is more than 1 ranking higher or lower than the actual ranking, you will get 0 points.

	Score	Male/Female rank decile interval
☐	100	(0, 10%]
☐	94	(10%, 20%]
☐	88	(20%, 30%]
☐	84	(30%, 40%]
☐	80	(40%, 50%]
☐	76	(50%, 60%]
☐	72	(60%, 70%]
☐	68	(70%, 80%]
☐	64	(80%, 90%]
☐	60	(90%, 100%]

<Next page>

2. Your actual score in last week's plank test (test 1) is shown in the following table:

Task	Weight (Highest possible score)	Test score (points)	Male/Female rank decile interval
Test 1	8	88	(20%, 30%]
Test 2	8		
Test 3	8		
Final test	60		
Class attendance	10		
Survey 1	3		
Survey 2	3		
Total	100		

Please check the box below

☐ I already know my "Test 1" score

<Next page>

3. Please set a goal for your score in the final test (choose one option in the following table).

Note: You will be reminded of your goal when you are informed of your score for each test through WeChat messages.

	Score	Male/Female rank decile interval
☐	100	(0, 10%]
☐	94	(10%, 20%]
☐	88	(20%, 30%]
☐	84	(30%, 40%]
☐	80	(40%, 50%]
☐	76	(50%, 60%]
☐	72	(60%, 70%]
☐	68	(70%, 80%]
☐	64	(80%, 90%]
☐	60	(90%, 100%]

<Next page>

A.2. AN EXAMPLE OF THE WECHAT MESSAGE FOR THE GOAL-SETTING TREATMENTS

Hello, XXX

The goal that you set for your score in the final test of the plank is: 94 – (10%,20%].

Your actual score on last week’s plank test 2 is shown in the table below.

Task	Weight (Highest possible score)	Test score (points)	Male/Female rank decile interval
Test 1	8	88	(20%, 30%]
Test 2	8	84	(30%, 40%]
Test 3	8		
Final test	60		
Class attendance	10		
Survey 1	3		
Survey 2	3		
Total	100		

Please reply when you receive this message. Thank you!

Education reform project assistant

Date

A.3. SURVEY QUESTIONS RELATED TO SOCIOECONOMIC BACKGROUND IN THE SECOND SURVEY

1. What is your ethnicity?

- ☐ Han
- ☐ Minority
2. What is your political affiliation?
- ☐ Not politically affiliated
- ☐ Youth League member
- ☐ Communist Party member
- ☐ Democratic Party member
3. Which kind of Hukou did you hold before enrolling into university?
- ☐ Non-agricultural
- ☐ Agricultural
- ☐ No Hukou
4. Have you been to a boarding elementary, middle or high school?
- ☐ Yes
- ☐ No
5. What is your average total after-tax monthly income from all sources (including from your parents, etc.)? _____ Chinese yuan
6. Do you have any siblings (including half-siblings and siblings adopted by parents)?
- ☐ No
- ☐ Yes
7. [Only shown when choosing “Yes” in question 6] How many siblings do you have excluding yourself (including half-siblings and siblings adopted by parents)? _____
8. [Only shown when choosing “Yes” in question 6] Please specify your order in the siblings. _____
9. What is your health status?
- ☐ Very healthy
- ☐ Rather healthy
- ☐ Fair
- ☐ Rather unhealthy
- ☐ Very unhealthy
10. What was your grade point average (on a 100 points scale) for all courses, including physical education and optional courses, but excluding additional credits earned from extracurricular activities, in the previous academic year (2017–2018)? _____

A.4. EARNINGS BY RANK AND TEST IN LABORATORY EXPERIMENT

Rank	ECUs earned in test	
	Round 1	Round 2
Top 10%	120	360
11%—20%	110	330
21%—30%	100	300
31%—40%	90	270
41%—50%	80	240
51%—60%	70	210
61%—70%	60	180
71%—80%	50	150
81%—90%	40	120
91%—100%	30	90

A.5. LABORATORY EXPERIMENTAL INSTRUCTIONS

Hello everyone, welcome to participate in this economic experiment conducted by the Centre for Decision Research and Experimental Economics (CeDEx). The results will be used for scientific research. We greatly appreciate your participation.

This experiment will last about 85 minutes. During the experiment, you will be asked to complete some tasks. You can earn real money by completing them, the amount of which will depend on your performance in these tasks as well as the other participants' performance. It is therefore very important that you read the following instructions carefully.

Rules of Conduct

We kindly ask you to follow the following rules of conduct in the lab. Before the experiment starts, please switch off your cell phone and other electronic devices and put them into the basket on the upper right of your screen. Throughout the whole experiment, please do not use cell phone and any other electronic devices; please do not communicate with other participants; please do not save any screenshots that are relevant to the content of this experiment; and please do not conduct any other activities that are irrelevant to the experiment. Participants who violate the rules will be immediately excluded from the experiment and will not be paid. If you have any question during the experiment, please press the buzzer at your desk on your right-hand side. The experimenter will approach you to answer your questions in private.

Procedures and Anonymity

This experiment will be conducted on PCs. This experiment follows a no-deception policy. All tasks will be implemented exactly as described in the experimental task instructions, and which are described on the following pages. The instructions are the same for everyone. All participants' personal identities will remain confidential and anonymous. It means no other participant learns about who made which decision and who got how much payment. It also means no personal identities will be disclosed in the research analysis.

Payments

During the experiment, your earnings will be measured in Experimental Currency Units (ECUs). One ECU is worth 0.2 Yuan. When the experiment is over, your ECU earnings will be converted to RMB. In addition, you will be paid an extra 36 Yuan for your attendance.

After the experiment we will ask you to sign a receipt for your total earnings. The experimenter will then pay you immediately via WeChat in private.

If you have any question, please press the buzzer at your desk on your right-hand side. The experimenter will come and answer your question in private. If not, please wait for all remaining participants to be ready.

Experimental Task Instructions

In today's experiment, you will be tasked with memorizing the English names of African countries and their capitals, along with their geographic locations. You will do this in two independent rounds. Each round consists of two stages: **memorization and testing**.

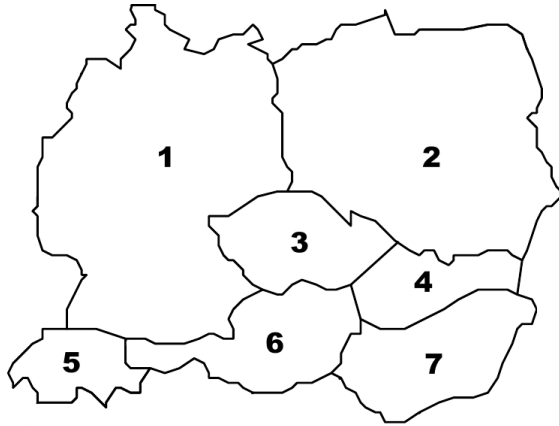
During the memorization stage, a portion of the African continent's map will be displayed on your computer screen. The map will feature the names of the countries in large bold text, with the corresponding capitals in smaller non-bold text underneath. Both elements will be positioned within the national borders of each country, or outside of them with lines pointing to their respective locations when space is limited. It will look similar to the map below, which, however, uses a part of European map instead.



Each round, you will be given a **fixed amount of time** to memorize as many items and as accurately as possible. On the top right corner of your screen, you will be able to see how much time is left.

Once the memorization period concludes, you will proceed to the test stage, where we will test how much you have memorized within a set time limit. In this stage, an empty map, with names of countries and capitals hidden will appear on your screen. Instead, each country location will be marked with a number. The country in the top-left position will be designated as number 1, with the next country to the right as number 2, continuing until the right end of that row. After reaching the right edge of the map, numbering will proceed to the countries in the most left of "row 2," and so on.

Using the aforementioned European example, your computer screen will look like the following:



Time left (sec.): 235

Country location	1
Country name	
Capital name	

< Confirm/Next >

On the screen of the test stage, you will be asked to enter the **English** names of the country and the capital corresponding to the location of a number on the map, one country at a time and in increasing order of country numbers, starting with country 1, then country 2, and so on.

Once you have entered your answers, or if you wish to skip any entry, please press the “Confirm/Next” button to move on to the next country. **Please be aware that once you have moved to the next entry, you will not be able to return to previous entries to make any changes.**

Each correct answer – country or capital name – is worth 1 point. This means that if you get both the country and capital names correct for a location, you will get 2 points. If you only get one of them correct, you will get 1 point. For an answer to be correct, the spelling must be 100% correct (but whether any letter is capitalized does not matter). *Any* wrong entry is scored at 0 point, and unfilled blanks are scored at 0 point too.

Your earnings for each test will depend on **how your total points rank among all participants in this room**. In essence, your **relative performance** will determine how many ECUs you earn in that specific test. The better your ranking – categorized in 10% increments such as the top 10%, 11%–20%, 21%–30%, and so forth – the greater your earnings.

As previously mentioned, this experiment consists of two independent rounds. This means that the portion of the African continent map you will encounter in round 1 will be different from that in round 2. This also means that your memorization performance in round 1 will not directly affect your performance in round 2. Specifically, round 1 will consist of 12 countries, while round 2 will feature 36 different countries. The time allotted for the memorization and test stages will reflect this: in round 1, you will have 8 minutes to memorize these items and 6 minutes for completing the test; whereas in round 2, you will be given 24 minutes for memorization and 18 minutes to complete the test.

Your total earnings from this experiment will be the sum of the earnings from both tests (plus the show-up fee). In consideration of the time allocated to each round, **the earnings in round 2 for each corresponding rank will be three times those of round 1.**

The following table summarizes the ECUs you will earn by your ranking position in each test:

Rank	ECU earned in round	
	Test 1	Test 2
Top 10%	120	360
11%—20%	110	330
21%—30%	100	300
31%—40%	90	270
41%—50%	80	240
51%—60%	70	210
61%—70%	60	180
71%—80%	50	150
81%—90%	40	120
91%—100%	30	90

If two or more participants tie with respect to their points, they will always be ranked in the same *and* more favorable ranking category. E.g., suppose two out of 10 participants get everything correct. In this case, both will be ranked in the top 10% instead of both being placed in the 11%–20% rank. The 3rd best participant will be ranked at their natural position, i.e., 21%–30%.

If you have any question, please press the buzzer at your desk on your right-hand side. The experimenter will come and answer your question in private. Otherwise, if you have fully understood the task, please press <Next page>. After pressing <Next page>, please remain quiet, and wait for all other participants to be ready so that we can start the memorization stage of round 1.

<Next page>

[Round 1. Memorization]

Round 1. Please memorize the names of the countries and capitals, and their location.

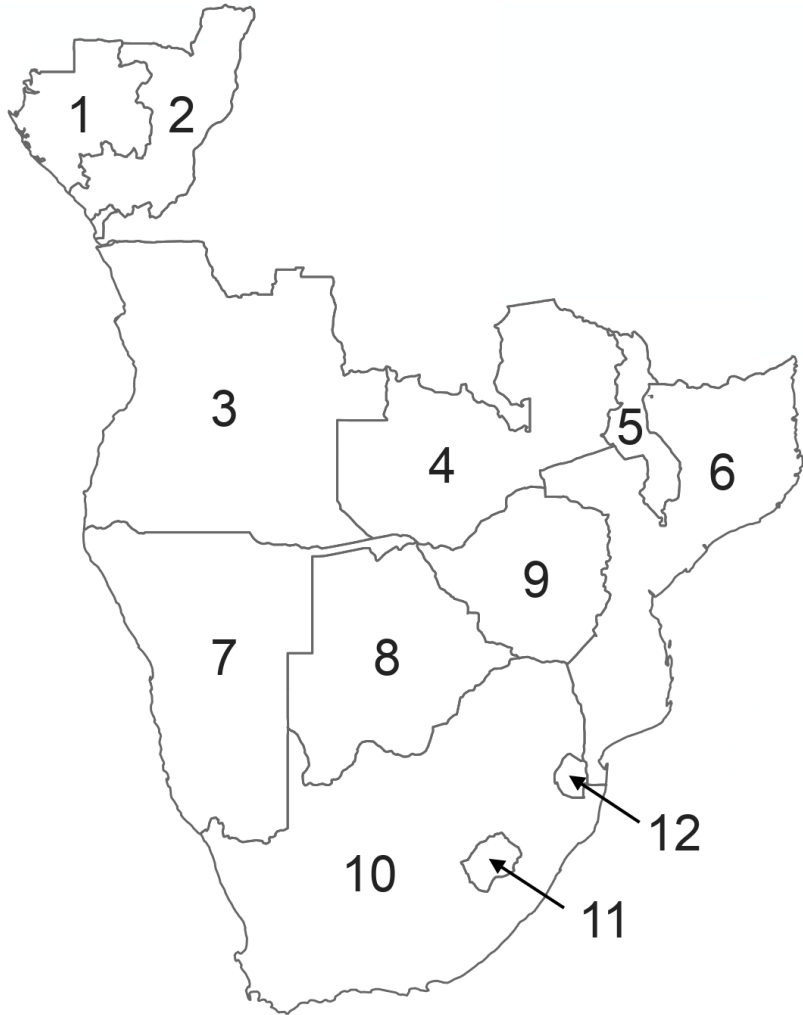
Time left (sec.): 174



[Round 1. Test]

Round 1. Test

Time left (sec.): 253



Country location	4
Country name	
Capital name	

<Confirm/Next>

Note: You cannot return to previous entries.

[Round 1 Questionnaire]

After this initial round, we would like you to fill in a short questionnaire. If you have any questions while filling in the questionnaire, please press the buzzer at your desk on your right-hand side. The experimenter will come and answer your question in private.

1. Please predict your ranking in the test for **the previous round 1** by choosing one of the options below.

[Single-choice question]

Notes: If the ranking you predicted is consistent with your actual ranking category, you will earn 40 ECUs as an additional reward; if your prediction is within 1 rank category higher or lower than your actual ranking category, you will earn 20 ECUs as an additional reward; in all other cases (i.e., more than 1 rank category), no reward will be earned.

Predicted rank	Please select
Top 10%	☐
11%—20%	☐
21%—30%	☐
31%—40%	☐
41%—50%	☐
51%—60%	☐
61%—70%	☐
71%—80%	☐
81%—90%	☐
91%—100%	☐

Please click “Next page” to continue.

<Next page>

[Goalsetting page; shown before (after) the Information provision page for the NoInfo (Info) treatment]

We now ask you to set a goal for the ranking that you would like to achieve in round 2.

2. Please set a goal for your ranking in **round 2** by choosing one of the options below. [Single-choice question]

Predicted rank	Please select
Top 10%	☐
11%—20%	☐
21%—30%	☐
31%—40%	☐
41%—50%	☐
51%—60%	☐
61%—70%	☐
71%—80%	☐
81%—90%	☐
91%—100%	☐

Please click “Next page” to continue.

<Next page>

[Information provision page; shown after (before) the Goalsetting page for the NoInfo (Info) treatment]

Performance in the test for Round 1

Your rank in the test of round 1: [actual ranking in round 1]

Your predicted ranking for the test of round 1 was: [predicted ranking in round 1]

Your actual ranking is [XX] category(ies) [better/worse] than your predicted ranking / Your actual ranking is the same as your predicted ranking.

Please click “Next page” to continue.

<Next page>

[Round 2. Memorization]

Round 2. Please memorize the names of the countries and capitals, and their location.

Your goal for test round 2: [goal of ranking category]

Time left (sec.): **513**

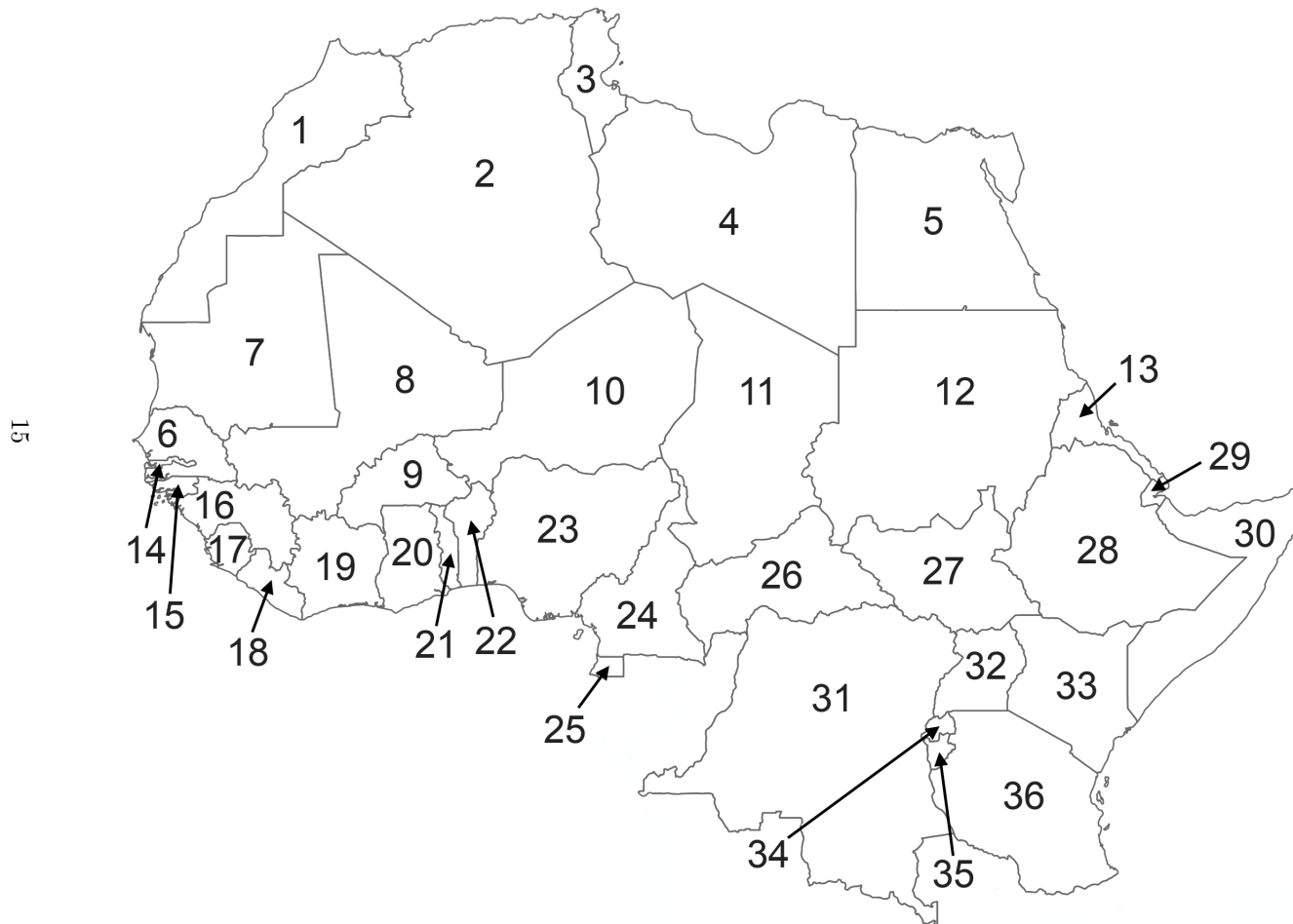


[Round 2. Test]

Round 2. Test

Your goal for test round 2: [goal of ranking category]

Time left (sec.): 832



Country location	12
Country name	
Capital name	

<Confirm/Next>

Note: You cannot return to previous entries.

Performance in the test for Round 2

Your rank in the test of round 2: [actual ranking in round 2]

Your goal for test round 2 was: [goal of ranking category]

You have completed this experiment. Well done!

While we calculate your earnings from the experiment, please complete a post-experiment survey.

<Next page>

[Post-experiment questionnaire survey, separated across multiple pages]

1. What is the full name of your major? _____ [Insert text box]
2. In which year of study are you? [Single choice]
 - ☐ Year 1 Undergraduate
 - ☐ Year 2 Undergraduate
 - ☐ Year 3 Undergraduate
 - ☐ Year 4 Undergraduate
 - ☐ Year 5 Undergraduate
 - ☐ Year 1 Master student
 - ☐ Year 2 Master student
 - ☐ Year 3 Master student
 - ☐ Year 1 PhD student
 - ☐ Year 2 PhD student
 - ☐ Year 3 PhD student
 - ☐ Year 4 PhD student
 - ☐ Year 5 PhD student
 - ☐ Other, please specify _____ [Insert text box]
3. What is your monthly disposable income (including all possible sources)? _____ Yuan
4. What is your average score during your last academic semester? [Single choice]
 - ☐ 90–100
 - ☐ 80–89
 - ☐ 70–79
 - ☐ 60–69
 - ☐ 50–59
 - ☐ 40–49
 - ☐ 30–39

- ☐ 20–29
- ☐ 10–19
- ☐ 0–9
- ☐ Not relevant. I don't need to take exams

5. How many times have you participated in incentivized scientific experiment as a subject before? [Single choice]

- ☐ Never
- ☐ 1–3
- ☐ 4–6
- ☐ more than 6

6. What is your gender? [Single choice]

- ☐ Female
- ☐ Male
- ☐ Other

7. How many siblings do you have (including half-siblings and siblings adopted by parents) excluding yourself? _____ [Insert text box]

8. [Only shown when answer in Q7 > 0] What birth order are you among all children in your family (including half-siblings and siblings adopted by parents)? _____ [Insert text box]

9. Was the time given for the test stages sufficient for you to enter all names of countries and capitals that you had memorized?

Time in the test for round 1 was sufficient

- ☐ Yes
- ☐ No

[If answered “No”] Why insufficient? _____ [Insert text box]

Time in the test for round 2 was sufficient

- ☐ Yes
- ☐ No

[If answered “No”] Why insufficient? _____ [Insert text box]

10. During the memorization or testing stage in round 2, did you generally feel that the effort required to memorize names was greater or less than what you initially anticipated when setting your goals? [Single-choice question]

- ☐ Much greater than anticipated
- ☐ Somewhat greater than anticipated

- ☐ About what I anticipated
- ☐ Somewhat less than anticipated
- ☐ Much less than anticipated

11. [Control Only] Did learning about your rank in the test for round 1 cause you to change your goal for the test for round 2 in your mind after you had already set your goal? [Single-choice]

- ☐ Yes
- ☐ No

[If answered “Yes”] Please recall what your updated goal was by choosing one of the options below. [Single-choice]

Ranking goal	Please select
Top 10%	☐
11%–20%	☐
21%–30%	☐
31%–40%	☐
41%–50%	☐
51%–60%	☐
61%–70%	☐
71%–80%	☐
81%–90%	☐
91%–100%	☐

12. How satisfied are you with your performance in the test for round 2? [Single-choice]

- ☐ Very satisfied
- ☐ Somewhat satisfied
- ☐ Neither satisfied nor unsatisfied
- ☐ Somewhat unsatisfied
- ☐ Very unsatisfied

13. How satisfied are you with your effort in memorization and test in round 2? [Single-choice]

- ☐ Very satisfied
- ☐ Somewhat satisfied
- ☐ Neither satisfied nor unsatisfied
- ☐ Somewhat unsatisfied
- ☐ Very unsatisfied

<Next page>

[End of Post-experiment questionnaire survey]

[Payoff Summary Page]

You will now see on your screen a summary of your ranking and earnings in each round.

Round	Your ranking	Payoff in round (ECUs)
1	Actual number	Actual number
2	Actual number	Actual number

Your total earnings from the tests plus the [prediction payments] ECUs from your prediction in round 1 are: [earnings task, in ECU] ECUs, which equals to [earnings task, in RMB] Yuan.

Show-up fee: 36 Yuan.

Your total earnings from the experiment: [total earnings] Yuan.

Please click confirm to acknowledge and continue

<Confirm>

The experiment is over. Thank you for your participation!

To ensure that other participants can participate in other related experiments of the School of Economics under the same information conditions as you, we kindly ask you not to publicly discuss or disclose the content and details of this experiment.

Please remain seated. We will now distribute a receipt form. Please write down your total earnings stated above, fill in the details and sign. The experimenter will then come to your seat and pay out your earnings via WeChat. Until then, please remain seated and stay on this page.

Your total earnings: [total earnings] Yuan.

Once you are paid, you are free to leave. Please leave both the instructions and the receipt form on your table. Do not take anything with you but your belongings.

B. Power analysis

B.1. FIELD EXPERIMENT: REALIZED POWER

This section reports the realized statistical power of the field experiment for subgroup-specific performance effects. Since the main analyses distinguish subjects by the accuracy of their beliefs about relative baseline performance, statistical precision depends not only on the overall sample size but also on the realized number of observations within each belief-accuracy subgroup.

In our field experiment, the number of observations by group was as follows: $N_{\text{NoInfo}} = 23$ & $N_{\text{Info}} = 24$ for underestimators, ($N_{\text{NoInfo}} = 9$ & $N_{\text{Info}} = 10$ for the accurate group), and $N_{\text{NoInfo}} = 33$ & $N_{\text{Info}} = 32$ for overestimators. It is clear that the power for such a small number of observations is somewhat limited. This limitation motivates the inclusion of a follow-up laboratory experiment that studies the same mechanism in a more controlled environment, since the alternative of adding sufficient observations in the field would have been too time and resource demanding, if not impossible.

Given a residual standard deviation of 2.068 (after controlling for baseline performance)¹, the realized power at a two-tailed 5% (10%) significance level at minimum detectable mean difference of 1 rank is 38.13% (50.55%) for underestimators and 49.58% (61.42%) for overestimators. The corresponding minimum detectable mean differences at 80% power are 1.69 (1.50) ranks for underestimators and 1.44 (1.28) for overestimators.

These calculations indicate that, while the field experiment is informative about behavior in a natural setting, statistical precision is limited for subgroup-specific performance effects. This motivates the use of a follow-up laboratory experiment to obtain more precise evidence on the same mechanism in a more controlled environment.

B.2. LABORATORY EXPERIMENT

Ex-ante power analysis

To complement the field evidence, the laboratory experiment was designed to provide better-powered evidence on the same treatment contrast and mechanism. This section presents the ex-ante power calculations underlying the target sample size.

The key outcome variable in our experiment is the subject's final test rank, which equals 10 if a subject ranks in the (0, 10%] bracket, 9 if in (10%, 20%], ..., down to 1 if in (90%, 100%]. Disregarding potential ties in ranks and assuming an equal distribution of subjects across ranks, the standard deviation of final rank is in theory $\sum_{rank=1}^{10} \frac{(rank-5.5)^2}{10} = 2.872$.

We use 1.1 ranks as the benchmark minimum detectable mean difference. This value is guided by the magnitude observed for underestimators in the field experiment that the information treatment improved underestimators' ranks by around 1 (Table 5 columns (1)–(2) on estimated differences for underestimators). Because the field setting introduces external motivational factors that might attenuate the effect of goal setting, we expect the improvement to be larger in a tightly controlled laboratory experiment, possibly up to 1.5 ranks. To remain reasonable in our *ex-ante* power analysis, we use an expected minimum detectable mean difference of 1.1 ranks (which represents a performance increase by only 11 percentage points in the overall distribution).

¹Specifically, this is the standard deviation of the residuals from a regression of the final test rank on the dummies for the baseline test rank.

Our experimental design features a baseline test prior to the experimental intervention. Subjects’ performance in this baseline test is highly predictive of their performance in the final test. From the field experiment, we observed that controlling for the baseline test performance reduces the unexplained standard deviation of the final test rank by 28%. For this power analysis, we conservatively assume a 25% reduction from the theoretical standard deviation of 2.872, yielding an adjusted standard deviation of 2.154.

We base our calculation of expected mean difference and standard deviation on our own field experiment rather than on previous laboratory studies of goal setting for a specific reason: our laboratory experiment was explicitly designed to mirror the institutional features of the field experiment, including a real-effort task, relative grading scheme, incentive structures, and informational timing of goal setting. Because previous laboratory experiments on goal setting often differ substantially in task nature, the presence of relative performance evaluation, or the specific belief-elicitation mechanisms, we considered our own field data to be the most contextually relevant prior for calibrating the expected effect size.

As a result, to detect the minimum mean difference of 1.1 ranks at a power of 80% and a two-tailed significance level of 5%, it requires $[(1.96 + 0.84) \cdot \sqrt{2 \cdot 2.154^2} / 1.1]^2 = 60.124$ subjects per treatment for each accuracy type (underestimate, accurate, overestimate). Therefore, 120 subjects in total for both treatments for a belief accuracy type are needed to be adequately powered.

As observed in our field experiment, 38% of the students underestimated their relative baseline performance, 48% overestimated, and only 13% held accurate beliefs. Powering the laboratory experiment for the smallest subgroup would therefore have required a substantially larger total sample. We consequently treat analyses for the accurate-belief group as exploratory and base the sample-size target on the smallest substantively relevant subgroup among underestimators and overestimators. Since accurate subjects are the least likely to benefit from the provision of information they already know (albeit it may increase their confidence thereof), we deemed this tradeoff to be acceptable.

Instead, we suppose that the smallest relevant subgroup (either underestimators or overestimators) would constitute 40% of the sample. This leads us to the pre-registered total targeted number of observations of $300 = 120/40\%$ to meet our power target. We stated in the pre-registry that in case of large deviations from our assumptions, on e.g., group size for subgroups, or explanatory power of baseline performance, the sample size would be adjusted accordingly to maintain our power target of 80%.

Realized power

We recruited 298 subjects for the laboratory experiment, which is very close to the target sample size of 300. The predictive power of the baseline rank for the final rank was also aligned well with our expectations – after controlling for baseline performance, the standard deviation of the residuals was 2.186, compared with the assumed value of 2.154. However, the realized distribution of belief types deviated noticeably from our initial conjectures: 54.4% of subjects underestimated their performance ($N_{\text{NoInfo}} = 82$ & $N_{\text{Info}} = 80$), 18.5% were accurate ($N_{\text{NoInfo}} = 28$ & $N_{\text{Info}} = 27$), and 27.2% overestimated their performance ($N_{\text{NoInfo}} = 40$ & $N_{\text{Info}} = 41$). Given these subgroup sizes, the realized power for overestimators dropped to 61.96% at a two-tailed 5% significance level (and 73.22% at 10%) for the minimum detectable mean difference of 1.1; the minimum detectable mean difference at 80% power would be 1.361 ranks. By contrast, the power for underestimators improved to 89.29%.

A back-of-the-envelope calculation suggests that to reach the target of 120 subjects who overestimate their baseline performance, we would need a total of $441 = 120/27.2\%$ subjects. This would require recruiting an additional 143 subjects. We decided against running further sessions for two primary reasons. First, our

laboratory experiment was intentionally designed so that the data could be pooled with that of the field experiment for the main analyses. Indeed, the pooled data provided sufficient observations to meet our power targets by belief accuracy type: $N_{\text{NoInfo}} = 105$ & $N_{\text{Info}} = 104$ for underestimators, ($N_{\text{NoInfo}} = 37$ & $N_{\text{Info}} = 37$ for the accurate group), and $N_{\text{NoInfo}} = 73$ & $N_{\text{Info}} = 73$ for overestimators, each of which is well in excess of the targeted 60. Second, the estimated treatment effects for overestimators in the laboratory experiment turned out to be very small (see Table 5, 0.19 to 0.29 ranks, equivalent to $0.07(= 0.19/2.872)$ to $0.10(= 0.29/2.872)$ standard deviations). This suggested that the true effect is likely small, and our initially expected minimum detectable mean difference was unlikely to be reached in this subgroup, making further data collection practically unjustified.

Relative to the field experiment, the laboratory data substantially improve statistical precision, especially for underestimators. At the same time, realized subgroup shares imply that precision remains heterogeneous across belief types.

B.3. INTERPRETATION FOR THE COMBINED EVIDENCE

Taken together, the power calculations clarify the evidentiary role of the two experimental settings. The field experiment provides evidence from a natural setting but has limited precision for subgroup-specific performance effects. The laboratory experiment substantially improves precision in a more controlled environment, although precision in a single sample remains sensitive to ex-post deviations in subgroup shares from the ex-ante assumptions.

Because the two experiments share highly similar institutional features and experimental designs, it is conceptually and methodologically permissible to pool the data. We find it reassuring that the behavioral patterns and results from the laboratory experiment and the field experiment are broadly consistent, especially on goal setting (see section 4 of the paper for more details). This consistency not only lays a solid empirical foundation for pooling the data, but also brings greater confidence to the initial findings from the field.

Crucially, pooling the data from both the field and the laboratory experiments provides ample observations to comfortably meet our ex-ante power targets (i.e., at least 60 subjects per treatment-subgroup). As a result, the pooled sample provides sufficient statistical power to detect realistic effect sizes across all belief-accuracy subgroups. These considerations motivate interpreting the single-experiment estimates with appropriate caution, and placing particular weight on the pooled analyses when evaluating the subgroup-specific performance effects in the main text.

C. Conceptual Framework

This section of the appendix provides further details on the model presented in Section 3. Section C.1 discusses the modeling choices and simplifications relative to the literature. Section C.2 formalizes the cost function over relative performance, the relevant domains, and the doer’s tie-breaking rule. Section C.3 proves Proposition 1 and the predictions stated in Section 3.4. Finally, Section C.4 develops a simple equilibrium extension that examines how a player’s goal and performance depend on her competitor’s.

C.1. DISCUSSION OF THE MODEL

The main text presents a deliberately stripped-down model, aiming for a simple framework to guide the subsequent empirical analysis. Three simplifications deserve comment.

1. Falling short of goals leads to losses, but exceeding does not result in gains.

The conventional reference-point model based on Prospect Theory’s value function (Kahneman and Tversky, 1979) features both losses from falling short of the goal and gains from exceeding it (see, e.g., Heath et al., 1999; Wu et al., 2008). We retain only the loss side, as in the present-bias model of Clark et al. (2020). The motivation is twofold. First, omitting gains keeps the analysis simple, eliminating a concave reference-dependent utility term over relative performance.² Second, in a model of endogenous goal setting, gains from exceeding the goal could create implausible incentives for the goalsetter to set a goal below y_t – possibly much lower – generating “free utility” at no effort cost, which strikes us as empirically implausible. In our reduced form the goalsetter maximizes performance and does not directly value psychological gains or losses, so this incentive does not literally arise; we note it, however, because it motivates the losses-only assumption in endogenous-goal models more generally.

Our model differs from Clark et al. (2020) in two fundamental ways (beyond the obvious absence of an explicit β - δ structure). First, Clark et al. assume a linear loss function, while we allow diminishing sensitivity to losses. This has the crucial implication that goals that are too high can be demotivating. Second, Clark et al. focus on absolute performance, while we model relative performance and examine how miscalibrated beliefs about relative ability affect goal setting.

2. The goalsetter’s ignorance of psychological costs.

In the model, the goalsetter chooses g to maximize $y^*(g)$, meaning she disregards any potential psychological costs arising from $k(g - y)$. In our deterministic, full-information setting, this is without loss of generality: by Proposition 1, the goal that maximizes performance is also met exactly, $y^*(\hat{g}) = \hat{g}$. Any goal $g > \hat{g}$ generates strictly positive psychological costs and strictly lowers performance, and thus would never be chosen. Note that this argument relies on the absence of randomness in the doer’s environment; in a stochastic extension (e.g., Koch and Nafziger, 2011) the goalsetter would trade off expected performance against the expected psychological cost of falling short due to random chance.

3. The goalsetter’s ignorance of material costs.

In a fully specified present-bias model, the goalsetter’s utility function would include material costs $c(y)$,

²Like Köszegi and Rabin (2006) and many other subsequent papers, our decision maker has both consumption utility (material benefits and cost) and a reference dependent loss utility, but no gain utility. In Heath et al. (1999) and Wu et al. (2008), utility is linearly separable in costs and benefits, with outcomes relative to the goal (the benefits) being evaluated through a Prospect Theory value function.

which together with the material benefits y pin down a preferred performance level $y_h > y_l$ that the goalsetter would like to implement. Any performance above y_h would be inefficient from her perspective. Our simplified model drops $c(y)$ from the goalsetter’s objective entirely. This corresponds to a setting where goals are “weak”: the doer’s response $y^*(g)$ never reaches y_h for any goal the goalsetter could set, so that the effective constraint $y < y_h$ that the full model would impose is never binding. In an education or learning setting, where the marginal cost of effort rises steeply and the motivational impact of goals appears limited, this seems empirically reasonable as a first approximation.

Consider instead a setting where goals are “strong” – where the goalsetter could in principle implement a performance above y_h . Here, the optimal goal g^* implements y_h exactly, with $k'(g^* - y)$ lying strictly above the net marginal cost curve at $y = y_h$.³ This “corner case” changes how information affects the performance of someone who slightly overestimates her relative ability – i.e., whose true net marginal cost curve lies marginally to the left of the perceived one. For such a person, reaching a given performance is harder when others perform better, so the goalsetter’s preferred performance also shifts leftward, from y_h^0 (under the perceived curve) to y_h^1 (under the true curve), with $y_h^1 < y_h^0$. In the *NoInfo* treatment, the goalsetter sets g_0 to implement y_h^0 under the perceived curve. By continuity, $k'(g_0 - y)$ still lies above true net marginal costs at $y = g_0$, so the doer still reaches her goal g_0 , performing above y_h^1 . In the *Info treatment*, the goalsetter targets the revised y_h^1 and sets a lower goal g_1 that implements it exactly. *NoInfo* performance therefore exceeds *Info* performance for sufficiently small degrees of overestimation.⁴

A more complex model can therefore generate richer but less sharp predictions, but only for overestimators and only in some cases. In any event, our experimental design and data would not allow us to distinguish these cases, so we adopt the simpler, more parsimonious model.

C.2. ASSUMPTIONS

Cost of relative performance. Although the main text works with cost as a function of relative performance, the natural primitive is the cost of absolute performance. Let $x \in \mathbb{R}^+$ denote absolute performance, with an associated cost function $\hat{c}(x)$, which is continuous, twice differentiable, and satisfies $\hat{c}(0) = 0, \hat{c}'(0) = 0, \hat{c}'(x) > 0$ for $x > 0$, $\hat{c}''(x) > 0$ for $x \geq 0$, and $\lim_{x \rightarrow \infty} \hat{c}'(x) = \infty$. Relative performance is the difference between the decision maker’s absolute performance and the average absolute performance of others, $\bar{x}$, i.e., $y = x - \bar{x}$. The relative cost function is therefore $c(y, \bar{x}) = \hat{c}(y + \bar{x})$. Where $\bar{x}$ is held fixed, we suppress it and write $c(y)$.⁵

Domains. The construction of $c(y, \bar{x})$ also pins down the domains that the main text left implicit. Relative performance $y = x - \bar{x}$ takes values in $[-\bar{x}, \infty)$. Goals are defined on the same domain, $g \in [-\bar{x}, \infty)$.

The doer’s maximization problem. We make the technical assumption that if the solution to the doer’s problem, $\max_y y - k(g - y) - c(y)$, is non-unique, she will choose the highest performance level.

³Note, if $k'(g - y)$ intersected the net marginal cost curve at or below y_h , the goalsetter could not push performance above y_h .

⁴The case above is specific to overestimators with small belief errors. For larger degrees of overestimation and for underestimators, information affects goals and performance as in our reduced form.

⁵The construction $c(y, \bar{x}) = \hat{c}(y + \bar{x})$ justifies the shifts in the marginal net cost curves in Figures 2 and 3. Since students can observe their own performance reasonably accurately, information about relative baseline test performance primarily teaches them about the performance of others. A student who overestimates the performance of others (i.e., underestimates her own relative ability) updates $\bar{x}$ downward upon receiving information, shifting $((c(y) - y))'$ to the right, as depicted in Figure 3a. The opposite holds for students who overestimate their own relative ability.

C.3. PROOFS

Proposition 1: *There exists a performance-maximizing goal $\hat{g} > y_l$. For $g \leq y_l$, $y^*(g) = y_l$; for $g \in (y_l, \hat{g}]$, $y^*(g) = g$; for $g > \hat{g}$, $y^*(g) \in (y_l, \hat{g})$ and is strictly decreasing in g .*

To prove the proposition, it is useful to first define some helpful notation. Denote the material utility by $a(y) := y - c(y)$. As $c(y)$ is strictly convex, together with its regularity conditions, the solution y_l that maximizes $a(y)$ exists, is unique, and the first order condition is both necessary and sufficient; it satisfies $a'(y_l) = 0$.

Next, we express the psychological utility by $b(y, g) := -k(g - y)$. Following our assumptions on $k(\cdot)$, $b(y, g)$ is strictly increasing and strictly convex in y for $y \leq g$, and equals 0 otherwise. Whenever we refer to the derivative of $b(y, g)$ with respect to y at $y = g$, we mean the right derivative of $k(z)$ at $z = 0$, denoted by $k'(0^+)$. Intuitively, this corresponds to the doer reaching her goal from below. By assumption, $k'(0^+)$ is bounded. Finally, by the concavity of k , an increase in g lowers $b_y(y, g)$ at every $y < g$: a higher goal means less marginal motivation for any unit increase in y .

Using these terms, we can express the doer's problem as $\max_{y \in [-\bar{x}, \infty)} a(y) + b(y, g)$.

Proof of Proposition 1. We first establish existence. The doer's objective is continuous in y on $[-\bar{x}, \infty)$ and tends to $-\infty$ as $y \rightarrow \infty$, so $y^*(g)$ exists. We characterize $y^*(g)$ by considering the two cases: $g \leq y_l$ (trivial) and $g > y_l$. The performance-maximizing goal $\hat{g}$ – the threshold at which $y^*(g)$ turns from rising to falling – emerges in the second case.⁶

Case 1: $g \leq y_l$. As $a(y)$ is maximized at y_l , and $b(y, g) \leq 0$ everywhere with $b(y_l, g) = 0$ when $g \leq y_l$, the doer's objective $a(y) + b(y, g)$ attains its maximum at $y = y_l$, giving $y^*(g) = y_l$.

Case 2: $g > y_l$. Since $a'(y_l) = 0$ and $a'(y) < 0$ for $y > y_l$, increasing y above y_l reduces the material payoff; this loss must be compensated by a sufficiently large marginal psychological benefit $b_y(\cdot)$. But since $a'(y_l) = 0$, it must be that when g is sufficiently close to y_l , $b_y(y, g) + a'(y) > 0$ for any $y \in [y_l, g]$. For such goals, the optimal y must hence be *at the corner*, i.e., $y = g$, so $y^*(g) = g$.

Since goals only affect $b_y(\cdot)$ but not $a'(y)$, and since $b_y(y, g)$ is bounded, increasing in y (up to g) but decreasing in g , and $c'(y) \rightarrow \infty$ as $y \rightarrow \infty$, there must exist some $\tilde{g}$ for which $b_y(\cdot, \tilde{g})$ first meets $-a'(\cdot)$ (from above). Since $b_y(\cdot)$ and $-a'(\cdot)$ do not have kinks, this first meeting is either (a) a crossing at the corner $y = \tilde{g}$, or (b) a tangency at a point $y^t \leq \tilde{g}$. Thus for $g \in (y_l, \tilde{g}]$, the solution is at the corner and performance rises one-for-one with the goal: $y^*(g) = g$. Cases 2a and 2b below characterize what happens for $g > \tilde{g}$ under (a) and (b), respectively.

Case 2a. Further increasing g above $\tilde{g}$ lowers $b_y(\cdot)$ without affecting $a'(y)$, so the curves intersect at some $y < \tilde{g}$, the new doer's optimum, which falls short of the goal. It follows that the performance-maximizing goal is $\hat{g} = \tilde{g}$, and $y^*(g)$ is decreasing in g for $g > \hat{g}$ (assuming a single intersection).

Case 2b. If $y^t = \tilde{g}$ (the tangency occurs at the corner), the argument from Case 2a applies directly. Otherwise, $y^t < \tilde{g}$, and so the curves touch at an interior point, with $b_y(\cdot, \tilde{g})$ lying above $-a'(\cdot)$ for all $y \leq \tilde{g}$ except at y^t , which is a saddle point. Increasing g slightly above $\tilde{g}$ separates the saddle point into a local maximum and a local minimum, with the former – where $b_y(\cdot)$ crosses $-a'(\cdot)$ from above – being the focus of our analysis. Hence, we need to check whether the corner $y^c = g$ or the local maximum, y^{loc} , is globally optimal.

⁶We reserve g^* for the goalsetter's optimal goal, which coincides with $\hat{g}$ since the goalsetter maximizes $y^*(g)$.

At $g = \tilde{g}$, the local maximum y^{loc} is just emerging from the saddle point and is thus strictly worse than the corner y^c . As g rises above $\tilde{g}$, however, the utility difference $U(y^c) - U(y^{loc})$ is strictly decreasing in g , so the corner's advantage shrinks. To see this, write:

$$U(y^c) - U(y^{loc}) = b(y^c, g) + a(y^c) - b(y^{loc}, g) - a(y^{loc})$$

Differentiating with respect to g using $b_y(y^{loc}, g) + a'(y^{loc}) = 0$ yields

$$\frac{d[U(y^c) - U(y^{loc})]}{dg} = [b_y(y^c, g) + a'(y^c)] \cdot \frac{dy^c}{dg} + b_g(y^c, g) - b_g(y^{loc}, g)$$

Using the fact that at the corner $y^c = g$, so $dy^c/dg = 1$ and that $b_y(y^c, g) = -b_g(y^c, g)$, the derivative of the utility difference simplifies to

$$a'(y^c) + b_y(y^{loc}, g).$$

Adding and subtracting $a'(y^{loc})$ and making use once again of the observation that the derivative is zero at the local maximum, we get

$$a'(y^c) + b_y(y^{loc}, g) + a'(y^{loc}) - a'(y^{loc}) = a'(y^c) - a'(y^{loc}).$$

Since $y^c > y^{loc} > y_l$, we have $a'(y^c) < a'(y^{loc}) < 0$. Hence:

$$\frac{d[U(y^c) - U(y^{loc})]}{dg} = a'(y^c) - a'(y^{loc}) < 0 \quad (\text{C.1})$$

Moreover, since $b_y(\cdot)$ is bounded while $c'(y) \rightarrow \infty$ as $y \rightarrow \infty$, the corner cannot remain optimal indefinitely: there exists a goal at which the doer is indifferent between the two, and the doer prefers the local maximum thereafter. By the assumed tie-breaking rule, the doer chooses the corner at the indifference goal. Performance there equals the goal, so this is the performance-maximizing goal $\hat{g}$.⁷

A marginal increase in g beyond this point leads to a discontinuous decrease in y^* from the corner to the local maximum. For further increases in g , $b_y(\cdot)$ shifts down pointwise while $-a'(\cdot)$ is unchanged, so the intersection moves leftward: y^{loc} is strictly decreasing in g . Altogether, $y^*(g) = g$ on $(y_l, \hat{g}]$, and $y^*(g) \in (y_l, \hat{g})$ is strictly decreasing for $g > \hat{g}$ (assuming a single local maximum, and noting that if $b_y(\cdot)$ crosses $-a'(\cdot)$ at the corner before the indifference point is reached, the analysis of Case 2a applies).

Robustness. In Case 2a, we assumed b_y and $-a'$ cross only once (so the corner gives way to a single interior optimum). In Case 2b, we assumed that a single interior local maximum competes with the corner. Because we impose relatively few assumptions on the shape of c' and k' , it could be possible that $b_y(\cdot, g)$ and $-a'(\cdot)$ intersect multiple times for a given g , and that the number of local maxima may change as g rises. We now show that the characterization of $y^*(g)$ is unaffected by such configurations.

First, suppose that for a given g , the doer's problem has two local maxima y^i and y^j with $y^i < y^j$. Following the same steps as in Case 2b, the lower maximum becomes relatively more attractive as g rises:

$$U(y^j) - U(y^i) = b(y^j, g) + a(y^j) - b(y^i, g) - a(y^i)$$

⁷Without the tie-breaking rule, if the doer took y^{loc} at the indifference point, the supremum of $y^*(g)$ would then be approached but never attained, and no performance-maximizing goal would exist. Crucially, this would imply that the *goalsetter's* problem would not have a maximum.

$$\frac{d[U(y^j) - U(y^i)]}{dg} = b_g(y^j, g) - b_g(y^i, g) = b_y(y^i, g) - b_y(y^j, g) < 0 \quad (\text{C.2})$$

where the inequality follows from b_y being increasing in y and $y^i < y^j$. In other words, the doer would never want to switch from the lower local maximum y^i to the higher y^j as g rises. Combined with the fact that each individual local maximum is strictly decreasing in g (as shown in Cases 2a and 2b), $y^*(g)$ is strictly decreasing in g as long as the set of local maxima remains unchanged.

Next, suppose an increase in g to g^n introduces a new local maximum y^n . Following the argument of Case 2b, y^n must have been a saddle point at $g^n - \epsilon$, with a higher local maximum or the corner case y^k existing above it (so $y^n < y^k$). By continuity, $U(y^k) > U(y^n)$ at g^n . Note that the appearance of y^n and any downward jump from y^k to y^n at a higher goal is consistent with our characterization. The only concern is an upward jump. Hence, suppose there exists a lower local maximum $y^i < y^n$ that the doer was choosing at $g^n - \epsilon$. By equation (C.1) or (C.2), y^i remains preferred to y^k at g^n . Combined with y^k preferred to y^n at g^n , y^i is also preferred to y^n at g^n , and remains so as g rises (by equation (C.2)). Hence the introduction of y^n cannot cause an upward jump in $y^*(g)$.

Lastly, suppose a local maximum y^j is eliminated when g rises to g^e – i.e., y^j becomes a tangency point at g^e , and for $g > g^e$, b_y drops below $-a'$ in the neighborhood of y^j . For our characterization, the only concern is whether this elimination causes y^* to jump upward. A local maximum below y^j always exists (b_y lies above $-a'$ at y_l). Let y^i be the lower local maximum just before y^j . If the doer was at y^i just before g^e , the elimination of y^j does not affect her choice. If the doer was at y^j itself and she drops to the lower local maximum y^i , the downward jump is consistent with our characterization. Let $y^k > y^j$ denote a higher local maximum or the corner. An upward jump to y^k cannot occur: by construction, $U(y^j) > U(y^k)$ at $g^e - \epsilon$ and $U(y^i) > U(y^j)$ at g^e (as y^j becomes tangent from below). By continuity, $U(y^i) > U(y^j)$ also at $g^e - \epsilon'$, for some $\epsilon' < \epsilon$, where equation (C.1) or (C.2) gives $U(y^j) > U(y^k)$ as well. Hence $U(y^i) > U(y^k)$ at $g^e - \epsilon'$, and – since equation (C.1) or (C.2) preserves the inequality as g rises to g^e – at g^e as well. The doer therefore falls to y^i , not up to y^k .

This completes the proof of Proposition 1. $\square$

Proof of Predictions. The goalsetter’s problem is to set a goal that results in the highest performance. Let $\bar{x}_1$ denote the true average absolute performance of others. Since all students learn their baseline performance before the final test, the doer responds to the true cost $c(y, \bar{x}_1) = \hat{c}(y + \bar{x}_1)$ regardless of treatment. Treatments differ only at the goal-setting stage: an *Info* goalsetter knows $\bar{x}_1$ and sets the performance-maximizing goal g_1 ; by Proposition 1, this goal is met exactly, so realized performance is $y_1 = g_1$. A *NoInfo* goalsetter instead holds a possibly mistaken belief $\bar{x}_0$ and sets the goal g_0 that would maximize performance if $\bar{x}_0$ were correct. The realized performance $y^*(g_0)$, however, is the doer’s response to the true cost, which may differ from the performance the goalsetter anticipates.

Case 1. Overestimating others’ absolute performance: $\bar{x}_0 > \bar{x}_1$.

Overestimating others’ absolute performance is equivalent to *underestimating one’s own relative ability* (and, in the context of the experiment, underestimating one’s baseline performance). Since the goalsetter overestimated others’ performance, the truth is lower: $\bar{x}_1 = \bar{x}_0 - \Delta$, $\Delta > 0$. This decrease shifts the cost curve, and hence the net marginal cost $-a'(\cdot)$, as a function of y to the right: $c(y + \Delta, \bar{x}_1) = c(y, \bar{x}_0)$ since

$$c(y, \bar{x}_0) = \hat{c}(y + \bar{x}_0) = \hat{c}(y + (\bar{x}_1 + \Delta)) = \hat{c}((y + \Delta) + \bar{x}_1) = c(y + \Delta, \bar{x}_1).$$

In other words, the true net marginal cost curve (under $\bar{x}_1$) is the perceived one (under $\bar{x}_0$) shifted to the right by Δ . Raising the goal by Δ shifts the marginal psychological benefit $b_y(\cdot)$ to the right by the same amount. Hence if g_0 is the optimal goal under $\bar{x}_0$, then $g_1 = g_0 + \Delta$ is optimal under $\bar{x}_1$.

Under the true cost $c(\cdot, \bar{x}_1)$, the *Info* goal g_1 is performance-maximizing, so $y^*(g_1) = g_1$. The *NoInfo* goal lies below it, $g_0 = g_1 - \Delta < g_1 = \hat{g}$, so by Proposition 1, it induces strictly lower performance ($y^*(g_0) = g_0$ if $g_0 > y_l$ and $y^*(g_0) = y_l$ otherwise). The *Info* performance therefore exceeds *NoInfo* performance.

Case 2. Underestimating others' absolute performance: $\bar{x}_0 < \bar{x}_1$.

Underestimating others' absolute performance is equivalent to *overestimating one's own relative ability*. Since the goalsetter underestimated others' performance, the truth is higher: $\bar{x}_1 = \bar{x}_0 + \Delta$, $\Delta > 0$. This increase shifts the cost curve, and hence the net marginal cost $-a'(\cdot)$, as a function of y by Δ to the left. Following the same argument as in Case 1, it must be that if g_1 is the optimal goal under $\bar{x}_1$, then $g_0 = g_1 + \Delta$ is optimal under $\bar{x}_0$.

Under the true cost $c(\cdot, \bar{x}_1)$, the *Info* goal g_1 is performance-maximizing, so $y^*(g_1) = g_1$. The *NoInfo* goal lies above it, $g_0 = g_1 + \Delta > g_1 = \hat{g}$, so by Proposition 1, it induces strictly lower performance ($y^*(g_0) \in (y_l, \hat{g})$). The *Info* performance therefore exceeds *NoInfo* performance.

Case 3. Accurate beliefs: $\bar{x}_0 = \bar{x}_1$.

As students in the *Info* and *NoInfo* treatment hold the same beliefs, they solve the same problem at all stages. The prediction follows immediately. □

C.4. A SIMPLE EQUILIBRIUM MODEL OF GOALSETTING

Lastly, we formalize a simple equilibrium model of goal setting. As in our main model, the goalsetter has an inherent preference for more effort than the doer. The model differs from the reduced form in the main text in two respects: it adopts explicit β - δ preferences, and the action space is binary (high vs. low effort) to keep the strategic problem tractable.

We will show that the optimal goal may depend on the competitor's goal. Informing a player that her competitor's cost of effort is higher (lower) – e.g., that her relative performance in the baseline test of the experiment is better (worse) than expected – can lead her to increase (decrease) her goal and her effort, thereby affecting her chance of winning.

The Setting. Suppose two present-biased players, $i \in \{1, 2\}$, compete against each other. At time $t = 2$, each player decides on an effort level $e_i \in \{0, 1\}$, which results in an immediate cost of $c_i > 0$ (for $e_i = 1$). Winning the game results in a delayed benefit of 1 at time $t = 3$. Let $p_i(e_i, e_j)$ denote player i 's probability of winning. Winning probabilities are symmetric among players, meaning that $p_i = 1/2$ if $e_i = e_j$, $p_i = p_H$ if $e_i > e_j$ and $p_i = p_L$ if $e_i < e_j$ for any i , with $0 < p_L < 1/2 < p_H < 1$ and $p_L + p_H = 1$. We denote the increase in the likelihood of winning arising from high effort by $\Delta \equiv p_H - 1/2 (= 1/2 - p_L)$.

At time $t = 1$, the player can set a goal to win, $g_i = 1$, or “to lose”, $g_i = 0$. As a convention, we speak of setting “a goal” when the goal is to win and setting “no goal” otherwise. Neither achieving nor exceeding the goal has any value. It is only when the decision maker misses her goal that she suffers a psychological cost of failure of $k > 0$ at time $t = 3$.

The goal's only benefit is the ‘stick’, which the $t = 1$ self may use to motivate the future version of herself to exert more effort. Let $\beta_i \in (0, 1)$ denote her present-bias discount factor (with exponential time

discounting normalized to $\delta = 1$ for simplicity throughout).

The Effort Problem. At time $t = 2$, the doer faces the following maximization problem:

$$\max_{e_i \in \{0,1\}} \beta_i \cdot [p_i(e_i, e_j) - (1 - p_i(e_i, e_j)) \cdot g_i \cdot k] - c_i \cdot e_i \quad (\text{C.3})$$

Notice that we can rewrite the discounted utility from the perspective of $t = 2$ as $U_i^2(e_i, e_j, g_i) = \beta_i \cdot [(1 + g_i \cdot k) \cdot p_i(e_i, e_j) - g_i \cdot k] - c_i \cdot e_i$. This highlights the key idea, namely that a goal of winning increases the relative incentives to exert effort while imposing a “fixed cost” of k .

Solving the $t = 2$ problem, we see that high effort ($e_i = 1$) is preferred over low effort ($e_i = 0$) whenever

$$\beta_i \cdot (1 + g_i \cdot k) \cdot \Delta \geq c_i. \quad (\text{C.4})$$

What may be surprising at first is that i 's effort is independent of the other player's effort. This is due to the binary nature of effort and winning probabilities, which imply a symmetry in how effort translates to winning. While it may appear as if this structure creates a meaningless game, i.e., where i 's choices do not depend on j 's choices, we will shortly see that this is not at all the case when we consider the goal-setting problem.

Denote the highest costs at which high effort is still exerted, i.e., the cutoff, as $\bar{c}_i \equiv \beta_i \cdot (1 + g_i \cdot k) \cdot \Delta$.⁸ It highlights that (the cutoff cost of) effort is decreasing in the degree of present bias ($\beta_i \downarrow$), increasing in how strongly effort affects the likelihood of winning ($\Delta \uparrow$), and increasing in having a goal. Moreover, the higher the cost of falling short of the set goal ($k \uparrow$), the easier it becomes to incentivize high effort.

The Goalsetting Problem. At time $t = 1$, each player takes the Subgame Perfect Nash Equilibrium outcome of the effort provision stage as given. Denoting i 's equilibrium effort following some goal g_i as $e_i^*(g_i)$ (noting that i 's effort is independent of j 's effort and therefore also unrelated to j 's goal), player i solves

$$\max_{g_i \in \{0,1\}} \beta_i \cdot [(1 + g_i \cdot k) \cdot p_i(e_i^*(g_i), e_j^*(g_j)) - g_i \cdot k - c_i \cdot e_i^*(g_i)] \quad (\text{C.5})$$

The time $t = 1$ self decides whether it is worthwhile to impose higher incentives for effort at $t = 2$. Due to present-bias, she may find this to be worthwhile if higher effort outweighs the psychological cost from non-attainment, which occurs with a strictly positive probability since $p_H < 1$.

As we have not yet imposed any constraints on any of our model's parameters, it comes as no surprise that the goal setting problem can be entirely uninteresting in some cases. These cases immediately follow from inequality (C.4), which determines effort. In particular, if $\beta_i \cdot \Delta \geq c_i$, the $t = 2$ self prefers to exert high effort regardless of whether she has a goal or not. Here, there is no disagreement between her and her previous $t = 1$ self, who prefers high effort as long as $\Delta > c_i$. Alternatively, if $\beta_i \cdot (1 + k) \cdot \Delta < c_i$, the $t = 2$ self is uninterested in pursuing high effort even with a goal of winning as the cost of effort is simply too large. Since goals impose a strictly positive psychological cost (on both the doer and the goal-setter) yet do not fundamentally change the effort problem, they would never be set. For the remaining analysis we will therefore restrict ourselves to:

⁸As a convention, we break indifference always in favor of high effort and setting a goal.

Assumption 1: Assume $\beta_i \cdot \Delta < c_i \leq \beta_i \cdot (1 + k) \cdot \Delta$ for all $i \in \{1, 2\}$.

The implication of this assumption is that *goals matter*, meaning that any player i at time $t = 2$ exerts effort if and only if $g_i = 1$, or, in short, $e_i^*(g_i) = g_i$. But while having a goal will induce high effort, the question remains whether the player is interested in setting a goal at time $t = 1$.

Setting a goal and not setting a goal result in the following payoffs (with i 's future effort pinned down by her respective goal). For notational convenience, we fixed j 's effort and omitted g_j in the following equations.

$$\begin{aligned} U_i^1(g_i = 1, e_i^*(g_i), e_j = 0) / \beta_i &= (1 + k) \cdot p_H - k - c_i \\ U_i^1(g_i = 0, e_i^*(g_i), e_j = 0) / \beta_i &= 0.5 \\ U_i^1(g_i = 1, e_i^*(g_i), e_j = 1) / \beta_i &= (1 + k) \cdot 0.5 - k - c_i \\ U_i^1(g_i = 0, e_i^*(g_i), e_j = 1) / \beta_i &= p_L \end{aligned}$$

Player i thus finds it optimal to set a goal if the other player exerts low effort whenever $c_i \leq \Delta - p_L k := \bar{c}^0$, and if the other player exerts high effort whenever $c_i \leq \Delta - 0.5k := \bar{c}^1$. The difference between these two conditions arises from the fact that goals become relatively more costly to set when the other person is exerting high effort, for the likelihood of not meeting one's goal is higher when the other person exerts high effort.

In general, there will be two types of equilibria, one where behavior on equilibrium path is identical among players, which occurs when goals (or no goals) are the dominant choice, and one where behavior differs between the two players. It is in the second type of equilibrium where players' preferences over goals are typically shaped by their competitor's goals. For ease of exposition, we separate our discussion of the equilibria along these two types, starting with the first. Throughout, let $\mathbf{g} \equiv (g_1, g_2)$.

Proposition: *The symmetric Subgame Perfect Nash Equilibria are:*

1. If $c_i > \Delta - p_L k$ for all $i \in \{1, 2\}$, then $\mathbf{g}^* = (0, 0)$.
 2. If $c_i \leq \Delta - 0.5k$ for all $i \in \{1, 2\}$, then $\mathbf{g}^* = (1, 1)$.
- Effort is characterized by $e_i^*(g_i) = g_i$ for all $i \in \{1, 2\}$ and all $g_i \in \{0, 1\}$.*

The first equilibrium describes an outcome of the game where both players set no goals and subsequently exert no effort. The increase in the likelihood to win (Δ) that arises from inducing high effort via setting a goal does not outweigh the extra cost of effort (c_i) and the chance of suffering the psychological cost after losing (k), which occurs with probability p_L if the other opponent only exerts low effort. For both players, no goal is the dominant choice when their cost from non-attainment or their cost of effort is relatively high. The inequality that pins down the equilibrium also emphasizes the double-edged nature of k : while a higher psychological cost-parameter k helps induce effort (equation C.4), it also makes goals more expensive, possibly resulting in no goals being set in equilibrium.

The second equilibrium describes the opposite situation. Here, setting a goal is the dominant choice for both. The cutoff highlights that even when the psychological cost arises in 50% of all interactions due to one's opponent exerting high effort, the improvement in the chance of winning still outweighs the cost.

Both SPNEs are similar; players face a dominant choice when deciding their goal. At the same time, the discussion also highlights the key aspect of this game, namely that goals become relatively more costly to set when the other person is exerting high effort. This is indicated by the difference between the cost cutoff value when the other player exerts high effort ($\bar{c}^1$) compared to when the other player exerts low effort ($\bar{c}^0$).

Proposition: *The non-symmetric pure strategy Subgame Perfect Nash Equilibria are:*

3. *If $c_i > \Delta - p_L k$ for i yet $c_j \leq \Delta - p_L k$ for $j \neq i$, then $g_i = 0, g_j = 1$.*
 4. *If $\Delta - 0.5k < c_i \leq \Delta - p_L k$ yet $c_j \leq \Delta - 0.5k$ for $j \neq i$, then $g_i = 0, g_j = 1$.*
 5. *If $\Delta - 0.5k < c_i \leq \Delta - p_L k$ for all $i \in \{1, 2\}$, then $g_i = 1, g_j = 0$ where $j \neq i$.*
- Effort is characterized by $e_i^*(g_i) = g_i$ for all $i \in \{1, 2\}$ and all $g_i \in \{0, 1\}$.*

The third type of equilibrium occurs when one player never finds it useful to set a goal due to relatively high cost, while the other has relatively low effort cost (at least below $\bar{c}^0$). Given the lack of goal from her competitor, she herself sets a winning goal. This is our first equilibrium where differential effort costs result in the person with the lower effort cost to set a goal of winning, to exert more effort than her opponent in the effort stage, and thus to win more often.⁹

The fourth equilibrium is similar to the third in that the player with the lower cost is the one setting the goal. Unlike before, however, the player with the relatively higher cost would set a goal as long as the other does not. But since setting a goal is a dominant choice for the lower-cost player, it is perfectly determined who sets the goal.

The fifth equilibrium removes any dominant choice. Effort costs are such that each player finds it optimal to set a goal if and only if the other person has not. Consequently, we have multiple equilibria where only one player setting a goal is the equilibrium outcome.

These equilibria are most relevant to our experimental setting. In equilibria 4 and 5, at least one player's equilibrium choice is determined by that of the other. The same holds for equilibrium 3 when j lacks a dominant choice, i.e., $c_j \in (\bar{c}^1, \bar{c}^0]$. These equilibria predict that a player, who underestimates (overestimates) their opponent's cost, and hence their opponent's goal and effort, will adjust their own goal upwards (downwards). Thus, participants in our experiments, who learn that their performance in a baseline test is better (worse) than initially expected (with relative performance being inversely related to the opponent's effort costs) will revise their goals upwards (downwards).

There are multiple ways to arrive at these predictions. For those who underestimate their opponent's cost, we could take equilibrium 4 as a starting point. Suppose player i thinks j 's dominant choice is to set a goal due to having very low effort cost. Upon learning this is not the case, they "play" equilibrium 5, where in some (but not necessarily all) equilibria, it will be i who is setting a goal and not j . On average, this predicts an upward adjustment in i 's goal setting behavior.

Equilibrium 3, can be used in a similar fashion for someone who overestimate their opponent's cost, yet does not have a dominant choice for their goal ($c_j \in (\bar{c}^1, \bar{c}^0]$). Upon learning that their opponent's cost is not as high as initially thought, they "play" equilibrium 5. Again in some equilibria, it will now be j that doesn't set a goal, which, on average, predicts a downward adjustment of j 's goal setting behavior.

The same prediction can be made using only equilibrium 5, together with an equilibrium refinement that selects which player sets a goal. The simplest criterion here is the lowest effort cost (when such costs are public knowledge). This cost ordering is natural: players with lower effort costs have more to gain and less to lose from setting a goal.¹⁰

⁹A sub-case of this equilibrium occurs when $c_j \leq \bar{c}^1$. In this case, setting a goal would also be j 's dominant action.

¹⁰The presence of multiple equilibria imply that there must also exist a mixed strategy equilibrium (at the goal setting stage). It is easy to show that the likelihood of setting a goal equals $Pr(g_i = 1) = \frac{\Delta - p_L k - c_j}{\Delta k}$. As common in many mixed strategy solutions, the comparative statics are both counter-intuitive, and, more importantly, do not necessarily match the actual data. In particular, the equilibrium suggests i 's likelihood of setting a goal is decreasing in c_j . This is at odds with what we observe when subjects are allowed to adjust their goals in response to information as discussed above. Mechanically speaking, the reason behind this mixed-strategy result is that a higher cost c_j makes it more costly for j to set a goal. In order for j to be indifferent between setting a goal and not, i 's likelihood of having a goal (and thus exerting high effort) must fall in equilibrium as this

Going beyond goals, we observe that the equilibrium model makes the same prediction in terms of goals *and* performance as our non-equilibrium model for those who underestimated their performance. In the equilibrium model, a higher goal is accompanied by higher effort and higher performance (i.e., likelihood of winning). It differs, however, for those who overestimated their performance. While the non-equilibrium model predicted lower goals but higher performance, the equilibrium model predicts lower goals and lower performance. Finally, for those who accurately estimate their performance, the equilibrium model has nothing new to add, for there is no reason to expect any other equilibrium to be played.

Apart from these specifics, what does the equilibrium model of goal setting teach us relative to a non-equilibrium model? First, that the presence of other people who set ambitious goals may cause some people to be less ambitious themselves – unlike in a pure decision problem. This introduces, to a certain degree, some unpredictability, i.e., multiple equilibria. However, once goals are set, effort essentially follows goals in predictable ways. After all, goals are useful only when they are costly to meet - or rather, too costly to fall short of. How effort (and thus performance) follows goals can be analyzed more easily in a non-equilibrium setting, allowing us to go beyond a simple binary choice problem towards a richer continuous setting where incentives from goals can be convex.

improves j 's utility of setting a goal relatively more than not setting one.

D. Additional figures and tables

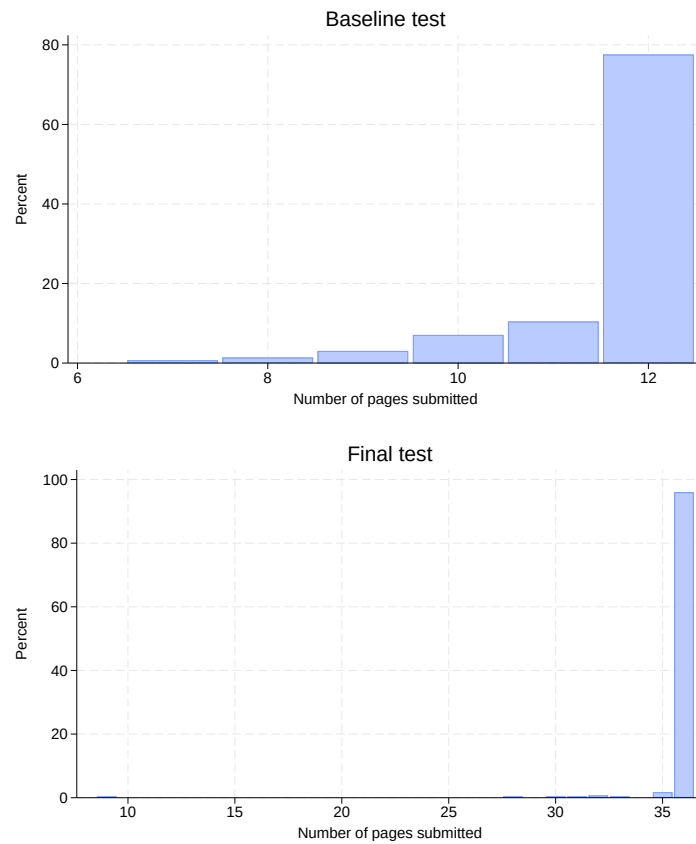


Figure D.1: Distribution of the number of pages submitted

Notes: This figure shows the percentage of participants who submitted a given number of total pages in the baseline test (12 countries total) and the final test (36 countries). One page equals one country.

Table D.1: Variable definitions

Variable Name	Definition
Info	=1 if in <i>Info</i> treatment; 0 otherwise
Baseline test rank	=10 if student ranks in the (0, 10%] bracket, 9 if in (10%, 20%] bracket, etc., for the baseline test.
Absolute performance baseline test	<i>Field experiment</i> : time (in seconds) of holding the plank in the baseline test; <i>Lab experiment</i> : total number of correct entries in baseline test
Predicted baseline test rank	Predicted rank for the baseline test rank, =10 if (0, 10%]; =9 if (10%, 20%], etc.
Goal	Goal for the final test rank, =10 if (0, 10%]; =9 if (10%, 20%], etc.
Underestimate	=1 if actual test rank is higher than the predicted baseline test rank; 0 otherwise
Accurate	=1 if actual test rank equals the predicted baseline test rank; 0 otherwise
Overestimate	=1 if actual test rank is lower than the predicted baseline test rank; 0 otherwise
Final test rank	=10 if student ranks in the (0, 10%] bracket, 9 if in (10%, 20%] bracket, etc., for the final test
Absolute performance final test	<i>Field experiment</i> : time (in seconds) of holding the plank in the final test; <i>Lab experiment</i> : total number of correct entries in final test
Gender: female	=1 if the student is female; 0 otherwise
Monthly income	Monthly disposable income (in CNY)
No. of siblings	Number of siblings
Birth order	Order of birth in all siblings
<i>Specific to field experiment:</i>	
GPA	=GPA score in previous academic year (out of 100 points)
Age	Age
Height	Height (in centimeters)
Weight	Weight (in kilograms)
Fat rate	Body fat rate
Metabolic rate	Basic metabolic rate
Han	=1 if the student's ethnicity is Han; 0 otherwise
Affiliated	=1 if the student is a Communist Party member or a Youth League member; 0 otherwise
Urban Hukou	=1 if the student has non-agricultural Hukou before enrolling in university; 0 otherwise
Boarding school	=1 if the student has ever attended boarding school before enrolling in university; 0 otherwise
Healthy	=1 if the student is very healthy or healthy in a 5-level scale; =0 otherwise
<i>Specific to lab experiment:</i>	
Gender: other	=1 if the student selected the "Gender: other" option; 0 otherwise
GPA	=GPA score in previous academic year (out of 100 points); -1 if no GPA in the previous academic year.
No GPA	=1 if student does not have a GPA from the previous academic year
Student: undergraduate	=1 if student is an undergraduate student; 0 otherwise
Student: master or PhD	=1 if student is either a master or PhD student; 0 otherwise
Participation in experiment	=0 if never participated in an experiment before, =1 if participated 1-3 times before, =2 if participated more than 3 times before
Reset goals	=1 if student stated they revised their goal after learning about their baseline performance and set a goal different from their initial one; 0 otherwise. NoInfo treatment only

Table D.2: Summary statistics

	Median	Mean	Std. Dev.	Min.	Max.	Obs.
Data: Field experiment						
Info	1.0	0.504	0.502	0.0	1.0	131
Predicted baseline test rank	6.0	6.023	2.522	1.0	10.0	131
Goal	8.0	7.771	1.940	3.0	10.0	131
Baseline test rank	5.0	5.298	2.841	1.0	10.0	131
Final test rank	5.0	5.389	2.868	1.0	10.0	131
Absolute performance baseline test	130.1	137.725	58.338	41.3	388.6	131
Absolute performance final test	193.2	203.724	77.843	59.7	489.8	131
Underestimate	0.0	0.359	0.481	0.0	1.0	131
Accurate	0.0	0.145	0.353	0.0	1.0	131
Overestimate	0.0	0.496	0.502	0.0	1.0	131
Gender: female	1.0	0.687	0.465	0.0	1.0	131
Monthly income	2000.0	2276.870	2656.839	0.0	30000.0	131
No. of siblings	0.0	0.558	0.856	0.0	5.0	129
Birth order	1.0	1.198	0.547	1.0	4.0	131
GPA	86.0	84.286	9.279	16.0	94.2	131
Age	19.6	19.567	1.003	15.5	22.0	131
Height	164.7	166.092	8.302	151.1	192.1	130
Weight	56.0	58.198	10.574	37.4	93.9	130
Fat rate	22.7	22.891	5.922	5.8	49.8	130
Metabolic rate	1417.0	1487.223	198.008	1198.0	2181.0	130
Han	1.0	0.916	0.278	0.0	1.0	131
Affiliated	1.0	0.939	0.240	0.0	1.0	131
Urban Hukou	1.0	0.817	0.388	0.0	1.0	131
Boarding school	1.0	0.603	0.491	0.0	1.0	131
Healthy	1.0	0.794	0.406	0.0	1.0	131
Data: Lab experiment						
Info	0.0	0.497	0.501	0.0	1.0	298
Predicted baseline test rank	5.0	4.758	2.269	1.0	10.0	298
Goal	6.0	6.379	2.244	1.0	10.0	298
Baseline test rank	6.0	5.628	2.842	1.0	10.0	298
Final test rank	6.0	5.453	2.869	1.0	10.0	298
Absolute performance baseline test	8.0	8.188	4.120	0.0	23.0	298
Absolute performance final test	22.0	22.493	9.023	0.0	53.0	298
Underestimate	1.0	0.544	0.499	0.0	1.0	298
Accurate	0.0	0.185	0.389	0.0	1.0	298
Overestimate	0.0	0.272	0.446	0.0	1.0	298
Gender: female	1.0	0.705	0.457	0.0	1.0	298
Gender: other	0.0	0.007	0.082	0.0	1.0	298
Monthly income	3000.0	3588.601	1357.681	0.0	10000.0	298
Number of siblings	0.0	0.758	1.319	0.0	10.0	298
Birth order	1.0	1.141	0.841	0.0	8.0	298
GPA	65.0	61.315	24.072	-1.0	85.0	298
No GPA	0.0	0.111	0.314	0.0	1.0	298
Student: undergraduate	1.0	0.913	0.283	0.0	1.0	298
Student: master or PhD	0.0	0.087	0.283	0.0	1.0	298
Participation in experiments	0.0	0.554	0.619	0.0	2.0	298
Earnings total (RMB)	98.0	98.497	21.079	60.0	140.0	298
Data: Lab experiment - goal reset						
Reset goals	1.0	0.587	0.494	0.0	1.0	150
Goal	7.0	6.453	2.399	1.0	10.0	298
Info	1.0	0.792	0.407	0.0	1.0	298

Table D.3: Randomization check

	NoInfo Treatment (1)			Info Treatment (2)		
	Obs	Mean	Std. Dev.	Obs	Mean	Std. Dev.
Data: Field experiment						
Baseline test rank	65	5.631*	2.950	66	4.970	2.712
Absolute performance baseline test	65	139.794	55.558	66	135.688	61.310
Predicted baseline test rank	65	6.308	2.610	66	5.742	2.420
Underestimate	65	0.354	0.482	66	0.364	0.485
Accurate	65	0.138	0.348	66	0.152	0.361
Overestimate	65	0.508	0.504	66	0.485	0.504
Gender: female	65	0.692	0.465	66	0.682	0.469
Monthly income	65	2,050.308	1,116.638	66	2,500.000	3,576.031
No. of siblings	63	0.413	0.557	66	0.697	1.052
Birth order	65	1.108	0.437	66	1.288	0.627
GPA	65	84.756	7.648	66	83.823	10.685
Age	65	19.639	0.904	66	19.498	1.095
Height	65	166.258	8.080	65	165.925	8.579
Weight	65	59.046	10.331	65	57.349	10.824
Fat rate	65	23.018	5.232	65	22.763	6.578
Metabolic rate	65	1,497.600	201.456	65	1,476.846	195.509
Han	65	0.908	0.292	66	0.924	0.267
Affiliated	65	0.938	0.242	66	0.939	0.240
Urban Hukou	65	0.815	0.391	66	0.818	0.389
Boarding school	65	0.569	0.499	66	0.636	0.485
Healthy	65	0.785	0.414	66	0.803	0.401
Data: Lab experiment						
Baseline test rank	150	5.620	2.905	148	5.635	2.788
Absolute performance baseline test	150	8.173	4.286	148	8.203	3.959
Predicted baseline test rank	150	4.827	2.322	148	4.689	2.219
Underestimate	150	0.547	0.499	148	0.541	0.500
Accurate	150	0.187	0.391	148	0.182	0.388
Overestimate	150	0.267	0.444	148	0.277	0.449
Gender: female	150	0.700	0.460	148	0.709	0.456
Gender: other	150	0.000	0.000	148	0.014	0.116
Monthly income	150	3,519.353	1,249.625	148	3,658.784	1,459.985
Number of siblings	150	0.627	1.059	148	0.892	1.530
Birth order	150	1.113	0.832	148	1.169	0.852
GPA	150	60.307	25.139	148	62.338	22.981
No GPA	150	0.127	0.334	148	0.095	0.294
Student: undergraduate	150	0.913	0.282	148	0.912	0.284
Student: master or PhD	150	0.087	0.282	148	0.088	0.284
Participation in experiments	150	1.147	1.547	148	1.243	1.403

Notes: This table reports the Kolmogorov-Smirnov Pairwise randomization test results between treatments. Significance levels of the difference between treatments are highlighted next to the mean values in column (1). *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

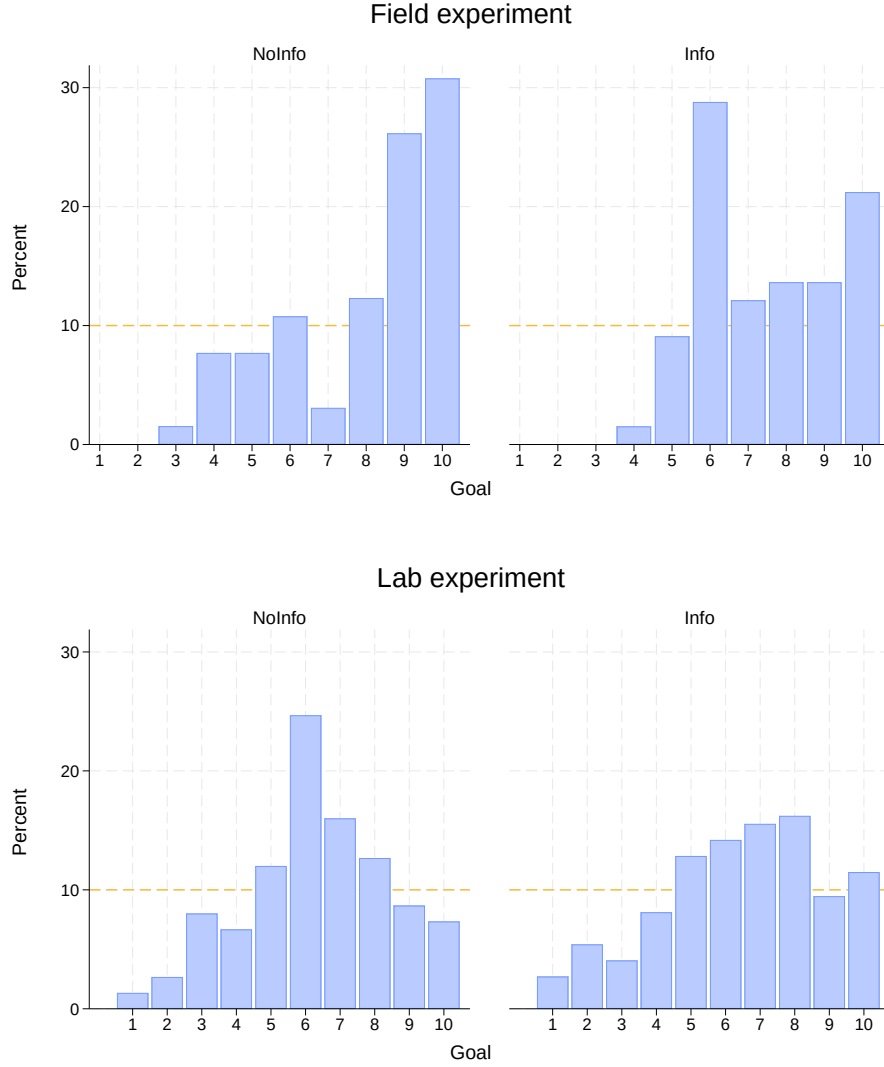


Figure D.2: Distribution of goals by treatment

Notes: This figure shows goals by treatment status. “NoInfo” refers to the treatment in which subjects set goals before receiving information about baseline relative performance, whereas “Info” refers to the treatment in which this information was revealed before goal setting.

Figure D.2 illustrates the distribution of goals by treatment in the field experiment and the laboratory experiment. The dashed orange line represents the theoretical distribution implied by our relative grading rule, under which exactly 10% of students are assigned to each rank based on their final performance (in the absence of ties).

Comparing goals against the 10% line in the field experiment (top panel) reveals that goals are set too high, implying that many students will fail to achieve them. This is particularly evident for students in the *NoInfo* treatment, who set goals in the top two ranks much more often than those in the *Info* treatment. Consequently, the distributions of goals in the two treatments are significantly different (p -value = 0.016, Fisher’s exact test). Moreover, both distributions differ from a distribution in which each goal rank occurs

with 10% likelihood (p -value < 0.001 , χ^2 test).

In contrast, the goal distributions in the laboratory experiment (bottom panel) are noticeably less top-heavy, with the frequency for the top two ranks falling below the 10% threshold in 3 of the 4 rank-by-treatment cells. The distributions for the treatments are visually much more similar, with the difference not being statistically significant (p -value = 0.314, Fisher’s exact test). Again, both distributions differ from a distribution in which each goal rank occurs with 10% likelihood (p -value < 0.001 , χ^2 test).

Figure D.3 plots the distribution of goals against the predicted baseline test ranks (top panel) and actual baseline test ranks (bottom panel) by treatment and belief accuracy in the field experiment (left panel), the laboratory experiment (middle panel), and the laboratory experiment with goal resetting (right panel). Larger bubbles represent a higher frequency of observations, and the 45-degree line indicates the exact equality between the goal set and the respective performance metric.

We find that many students set goals higher than their predicted ranks in the *NoInfo* treatment (top panel, first row) or higher than their actual ranks in the *Info* treatment (bottom panel, second row). Recall that when setting goals, students in the *NoInfo* treatment were uninformed of their relative ability, whereas those in the *Info* treatment were informed. This result thus suggests that students use the best available information as a reference point (as shown in Table 2) to set more challenging goals. In the laboratory experiment, however, we do see a subset of goals set below their predictions in the *NoInfo* treatment or actual baseline performance in the *Info* treatment, especially among those who underestimate their relative baseline performance.

Comparing the distributions that account for goal resetting (right panel) with the laboratory baseline (middle panel) allows us to identify exactly who chooses to revise their targets. As subjects who reset goals in the *NoInfo* treatment are reclassified to the *Info* treatment, the frequency of observations in the *NoInfo* treatment naturally decreases in both the top and bottom panels, making the bubbles fewer and smaller. More importantly, comparing the bottom panels across these two laboratory settings for the same accuracy type reveals a visible further reduction in data points located far from the 45-degree line in either direction. This thinning out implies that subjects whose initial goals deviated most significantly from their baseline performance are the ones most likely to reset them (as shown in Table 3).

The comparison between the *NoInfo* and *Info* treatments in the bottom panel for the same accuracy type highlights that relative performance feedback helps students calibrate goals more accurately against actual baseline performance. In the absence of information, a significant proportion of underestimators set goals below their actual baseline test ranks, which makes such goals too easy to be achieved and renders them ineffective in motivating better performance. Conversely, some overestimators set too ambitious goals in view of their actual baseline performance.

In contrast, the top panel illustrates how information prompts students to move away from their predictions. Compared to their counterparts in the *NoInfo* treatment, subjects in the *Info* treatment who underestimated (overestimated) their performance set higher (lower) goals than their predictions. This demonstrates that information provision corrects prior belief distortions, inducing goal setting that reflect objective reality rather than subjective predictions (as shown in Table 2).

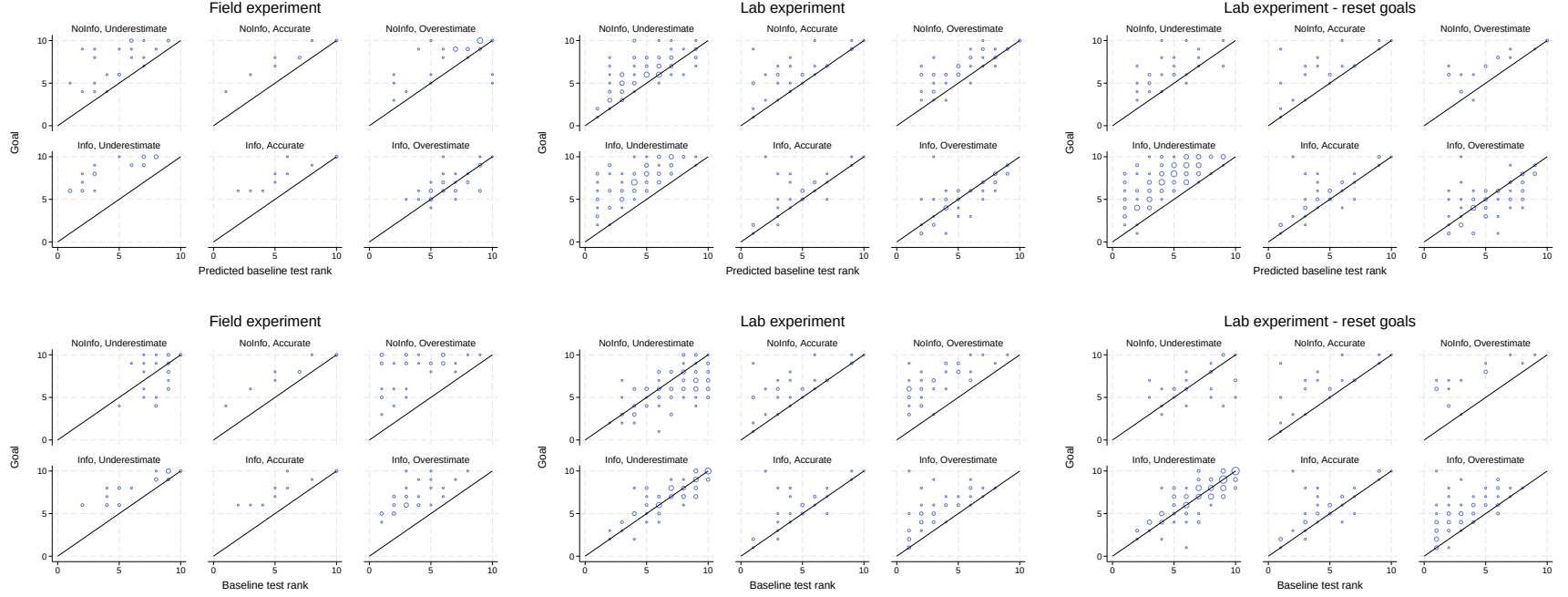


Figure D.3: Distribution of goals against predicted and actual baseline test ranks by treatment and accuracy

Notes: This figure shows goals by treatment status and belief-accuracy type. “NoInfo” refers to the treatment in which subjects set goals before receiving information about baseline relative performance, whereas “Info” refers to the treatment in which this information was revealed before goal setting. “Underestimate”, “accurate”, and “overestimate” are defined based on the difference between actual and predicted baseline rank. Higher ranks indicate better relative performance. Larger bubbles imply higher frequencies.

Table D.4: Treatment effect on relative performance in the final test by accuracy; Robustness check: broader notion of accurate type

	Field		Lab		Pooled		Lab, goal reset	
Dep. Var.: Final test rank	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Underestimate	0.119 (0.656)	-0.855 (0.663)	-0.373 (0.537)	-0.485 (0.567)	-0.188 (0.473)	-0.174 (0.472)	-0.596 (0.713)	-0.801 (0.812)
Overestimate	0.878 (1.005)	0.325 (1.004)	-0.316 (0.688)	-0.429 (0.625)	0.216 (0.498)	0.105 (0.478)	2.047* (1.055)	1.706 (0.977)
Info	0.371 (0.781)	-0.537 (0.859)	-0.240 (0.413)	-0.210 (0.458)	0.00419 (0.383)	-0.0307 (0.380)	-0.396 (0.342)	-0.424 (0.456)
Info $\times$ Underestimate	0.193 (0.997)	1.304 (1.061)	-0.124 (0.621)	-0.239 (0.790)	-0.222 (0.561)	-0.245 (0.615)	0.235 (0.733)	0.297 (0.879)
Info $\times$ Overestimate	-0.535 (1.171)	0.972 (1.158)	0.250 (0.836)	0.169 (0.715)	-0.191 (0.638)	-0.0532 (0.606)	-2.512** (0.928)	-2.248** (0.791)
<i>Estimated Differences: Info - NoInfo</i>								
Underestimate	0.564 (0.544)	0.768 (0.694)	-0.364 (0.319)	-0.449 (0.435)	-0.218 (0.294)	-0.276 (0.339)	-0.161 (0.723)	-0.127 (0.843)
Accurate	0.371 (0.781)	-0.537 (0.859)	-0.240 (0.413)	-0.210 (0.458)	0.00419 (0.383)	-0.0307 (0.380)	-0.396 (0.342)	-0.424 (0.456)
Overestimate	-0.164 (0.951)	0.436 (0.940)	0.00954 (0.752)	-0.0405 (0.631)	-0.187 (0.536)	-0.0839 (0.530)	-2.908*** (0.891)	-2.672*** (0.798)
Baseline test performance	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Group fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Personal characteristics	No	Yes	No	Yes	No	Yes	No	Yes
Observations	131	128	298	298	429	427	298	298

Notes: See notes to Table 5. In this table, students are classified as accurate both when they correctly predicted their rank and when they mispredicted their rank within only one rank (from above or below).

Table D.5: Treatment effect on absolute performance in the final test by accuracy; Robustness check: broader notion of accurate type

	Field		Lab		Lab, goal reset	
Dep. Var.: Final test abs. performance	(1)	(2)	(3)	(4)	(5)	(6)
Underestimate	-15.17 (13.72)	-15.91 (16.22)	-0.285 (1.206)	-0.987 (1.314)	-2.157 (1.820)	-2.665 (2.431)
Overestimate	22.53 (18.27)	18.86 (20.15)	-1.741 (1.694)	-1.313 (1.755)	4.520** (2.052)	4.741** (1.810)
Info	5.827 (13.00)	4.061 (17.37)	-1.356 (1.332)	-1.154 (1.415)	-0.954 (1.227)	-1.172 (1.494)
Info $\times$ Underestimate	27.98 (19.53)	19.47 (27.53)	0.708 (1.654)	0.366 (2.181)	2.582 (2.417)	2.275 (3.067)
Info $\times$ Overestimate	-17.97 (22.79)	-2.229 (21.47)	1.201 (2.183)	0.735 (1.932)	-6.349** (2.263)	-6.280*** (1.637)
<i>Estimated Differences: Info - NoInfo</i>						
Underestimate	33.81** (14.39)	23.53 (16.05)	-0.648 (0.899)	-0.788 (1.263)	1.628 (2.108)	1.103 (2.687)
Accurate	5.827 (13.00)	4.061 (17.37)	-1.356 (1.332)	-1.154 (1.415)	-0.954 (1.227)	-1.172 (1.494)
Overestimate	-12.14 (22.15)	1.833 (23.06)	-0.155 (1.793)	-0.420 (1.597)	-7.303*** (2.090)	-7.452*** (1.542)
Baseline test performance	Yes	Yes	Yes	Yes	Yes	Yes
Group fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Personal characteristics	No	Yes	No	Yes	No	Yes
Observations	131	128	298	298	298	298

Notes: See notes to Table 6. In this table, students are classified as accurate both when they correctly predicted their rank and when they mispredicted their rank within only one rank (from above or below).

E. Satisfaction with final test performance

To further examine whether goals serve as reference points in the evaluation of outcomes, we analyze how subjects’ satisfaction with their final test performance relates to both their realized final rank and their self-set goal. In this context, satisfaction serves as an empirical proxy for the utility the doer maximizes when aiming for a certain relative performance. The psychological cost $k(g - y)$ that is incurred when relative performance falls short of the goal can be extended to capture psychological benefit when the goal is met or exceeded, although utility derivation from the gain domain is intentionally omitted from our conceptual framework to prevent the acquisition of “free utility” from strategically setting goals too low.

We collected satisfaction data in the post-experimental survey during the spring semester of the field experiment and in the laboratory experiment. Subjects were asked how satisfied they were with their final test performance on a five-point Likert scale, ranging from “very unsatisfied” to “very satisfied”. Table E.1 presents the distribution of satisfaction across both experiments, without and with goal resetting accounted for in the laboratory experiment.

We also obtained a measure of effort cost in the post-experimental survey of the laboratory experiment. Subjects were asked to evaluate the actual effort required for the memorization task compared to their initial anticipations at the time of setting goals. Responses were elicited again on a five-point Likert scale, ranging from “much less” to “much greater”. Table E.1 presents the distribution of effort cost.

Table E.1: Summary statistics satisfaction

	Median	Mean	Std. Dev.	Min.	Max.	Obs.
Data: Field experiment						
Satisfaction with final test performance	2.0	2.579	1.146	1.0	5.0	76
Data: Lab experiment						
Satisfaction with final test performance	3.0	3.114	1.061	1.0	5.0	298
Required vs. anticipated effort in final test	3.0	3.413	0.985	1.0	5.0	298
Data: Lab experiment - goal reset						
Satisfaction with final test performance	3.0	3.114	1.061	1.0	5.0	298
Required vs. anticipated effort in final test	3.0	3.413	0.985	1.0	5.0	298

Table E.2 provides the proportion of students who missed, exactly hit, or exceeded their goals in the final test. We see that only one quarter of the subjects met or exceeded their goals in the field experiment, whereas the proportion doubles in the laboratory experiment. This finding is consistent with the observation that subjects in the field experiment set more ambitious goals, resulting in a larger average goal to final performance discrepancy in the field experiment (7.771 vs. 5.389 ranks) than those in the laboratory experiment (original goal: 6.379 vs. 5.453 ranks; reset goal: 6.453 vs. 5.453 ranks) as documented in Table D.2.

Table E.2: Performance compared to goal

	Missed goal	Hit goal	Exceeded Goal
Field experiment	75.57%	7.63%	16.79%
Lab experiment	53.02%	18.79%	28.19%
Lab experiment, goal reset	53.69%	26.17%	20.13%

Table E.3 reports the regression results for the utility function. The “min” and “max” terms capture the (reduction of) psychological cost from falling short of and benefit from exceeding the goal, incorporating both

linear and quadratic terms of the performance-goal gap to illustrate the curvature assumed in the conceptual framework. We predict that satisfaction increases with final relative performance and decreases with effort cost. Moreover, its relationship with the performance-goal discrepancy is non-linear – convex in the loss domain and concave in the gain domain. That is, satisfaction should increase at an increasing rate when performance approaches the goal from below, and increase at a decreasing rate when performance exceeds the goal.¹¹

Table E.3: Satisfaction regarding final test performance

Dep. Var.: Satisfaction final test performance	Field		Lab		Lab, goal reset	
	(1)	(2)	(3)	(4)	(5)	(6)
Final test rank	0.108 (0.0975)	0.127 (0.137)	0.207*** (0.0292)	0.200*** (0.0303)	0.190*** (0.0406)	0.188*** (0.0410)
$\min\{\text{Final test rank} - \text{goal}, 0\}$	0.381 (0.252)	0.299 (0.285)	0.307*** (0.0871)	0.305*** (0.0997)	0.249* (0.119)	0.234* (0.126)
$\min\{\text{Final test rank} - \text{goal}, 0\}^2$	0.0354 (0.0224)	0.0293 (0.0252)	0.0459** (0.0176)	0.0448** (0.0206)	0.0334* (0.0170)	0.0307 (0.0183)
$\max\{\text{Final test rank} - \text{goal}, 0\}$	0.137 (0.471)	0.197 (0.551)	0.0974 (0.0627)	0.126* (0.0613)	0.187** (0.0716)	0.181** (0.0663)
$\max\{\text{Final test rank} - \text{goal}, 0\}^2$	-0.0203 (0.0966)	-0.0423 (0.106)	-0.0256* (0.0143)	-0.0269* (0.0148)	-0.0369* (0.0175)	-0.0287* (0.0161)
Required vs. anticipated effort in final test			-0.171** (0.0589)	-0.174** (0.0644)	-0.180*** (0.0559)	-0.189*** (0.0610)
Baseline test performance	Yes	Yes	Yes	Yes	Yes	Yes
Group fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Personal characteristics	No	Yes	No	Yes	No	Yes
Observations	76	73	298	298	298	298

Notes: This table reports the estimates from OLS models for the field experiment, the laboratory experiment, and the laboratory experiment where subjects in the *NoInfo* treatment, who revised their goal, were reclassified to the *Info* treatment. Estimates for control variables are not reported. See Table 1 notes for a list of the control variables, with the exception that a continuous variable instead of the usual dummies is used to control for the baseline test rank for the field experiment, and details on sample construction. Robust standard errors allowing for heteroskedasticity and clustered at the group level are reported in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

In the laboratory experiment, the results shown in columns (3)–(6) fully support our predictions. In addition, when testing the equality of the linear and quadratic coefficients across the loss and gain domains, we find that the quadratic terms are significantly larger in the loss domain (Wald test, p -value < 0.05, in columns (3)–(6)). This asymmetry lends support to the shape of the value function of Prospect Theory (Kahneman and Tversky, 1979). In the field experiment, although the effects shown in columns (1)–(2) are imprecisely estimated due to the small sample size, yielding no statistically significant results, the signs of the effects are as expected, and their magnitudes are comparable to those obtained in the laboratory experiment.

¹¹The predictions on convexity and concavity of the psychological cost and benefit with respect to the performance-goal discrepancy in these regressions do not contradict with the assumptions outlined in our conceptual framework. This inversion is simply because the argument of the function here is reversed, measuring final test relative performance minus the goal, instead of the goal minus performance.

F. Analysis incorporating a control group without goal-setting intervention

In the experimental design section, we noted the inclusion of a control group in the field experiment that received no goal-setting intervention (*NoGoal*). We now incorporate this group into the analysis to disentangle the effects of goal setting and informed goal setting on performance. Summary statistics for this group and balance checks across treatments are reported in Tables F.1 and F.2.

Table F.1: Summary statistics - NoGoal treatment

	Median	Mean	Std. Dev.	Min.	Max.	Obs.
Predicted baseline test rank	6.0	5.766	2.629	1.0	10.0	64
Baseline test rank	6.0	5.203	2.918	1.0	10.0	64
Final test rank	5.0	5.000	2.867	1.0	10.0	64
Absolute performance baseline test	121.5	131.945	62.125	20.8	348.4	64
Absolute performance final test	183.0	193.932	88.790	49.1	451.7	64
Underestimate	0.0	0.438	0.500	0.0	1.0	64
Accurate	0.0	0.109	0.315	0.0	1.0	64
Overestimate	0.0	0.453	0.502	0.0	1.0	64
Gender: female	1.0	0.703	0.460	0.0	1.0	64
Monthly income	2000.0	2057.344	782.329	1000.0	6000.0	64
No. of siblings	0.0	0.719	1.253	0.0	7.0	64
Birth order	1.0	1.266	0.877	1.0	7.0	64
GPA	85.0	84.476	7.435	60.0	95.0	64
Age	19.6	19.608	0.964	17.5	22.1	64
Height	164.9	165.895	6.944	151.3	185.9	62
Weight	58.8	58.887	9.001	45.6	90.1	62
Fat rate	24.4	23.287	4.712	13.8	32.0	62
Metabolic rate	1431.0	1485.081	176.361	1279.0	2099.0	62
Han	1.0	0.875	0.333	0.0	1.0	64
Affiliated	1.0	0.891	0.315	0.0	1.0	64
Urban Hukou	1.0	0.766	0.427	0.0	1.0	64
Boarding school	1.0	0.578	0.498	0.0	1.0	64
Healthy	1.0	0.672	0.473	0.0	1.0	64

Table F.2: Randomization check - field experiment, NoGoal, NoInfo, and Info treatment

	NoGoal Treatment (1)			NoInfo Treatment (2)			Info Treatment (3)		
	Obs	Mean	Std. Dev.	Obs	Mean	Std. Dev.	Obs	Mean	Std. Dev.
Baseline test rank	64	5.20	2.92	65	5.63 ^{(3)*}	2.95	66	4.97	2.71
Absolute performance baseline test	64	131.95	62.12	65	139.79	55.56	66	135.69	61.31
Predicted baseline test rank	64	5.77	2.63	65	6.31	2.61	66	5.74	2.42
Underestimate	64	0.44	0.50	65	0.35	0.48	66	0.36	0.48
Accurate	64	0.11	0.31	65	0.14	0.35	66	0.15	0.36
Overestimate	64	0.45	0.50	65	0.51	0.50	66	0.48	0.50
Gender: female	64	0.70	0.46	65	0.69	0.47	66	0.68	0.47
Monthly income	64	2,057.34	782.33	65	2,050.31	1,116.64	66	2,500.00	3,576.03
No. of siblings	64	0.72	1.25	63	0.41	0.56	66	0.70	1.05
Birth order	64	1.27	0.88	65	1.11	0.44	66	1.29	0.63
GPA	64	84.48	7.44	65	84.76	7.65	66	83.82	10.69
Age	64	19.61	0.96	65	19.64	0.90	66	19.50	1.09
Height	62	165.90	6.94	65	166.26	8.08	65	165.92	8.58
Weight	62	58.89	9.00	65	59.05	10.33	65	57.35	10.82
Fat rate	62	23.29	4.71	65	23.02	5.23	65	22.76	6.58
Metabolic rate	62	1,485.08	176.36	65	1,497.60	201.46	65	1,476.85	195.51
Han	64	0.88	0.33	65	0.91	0.29	66	0.92	0.27
Affiliated	64	0.89	0.31	65	0.94	0.24	66	0.94	0.24
Urban Hukou	64	0.77	0.43	65	0.82	0.39	66	0.82	0.39
Boarding school	64	0.58	0.50	65	0.57	0.50	66	0.64	0.48
Healthy	64	0.67	0.47	65	0.78	0.41	66	0.80	0.40

Notes: This table reports the Kolmogorov-Smirnov Pairwise randomization test results between the three treatments. The superscript next to the mean of each treatment shows the column number to which treatment (column) is compared, and the asterisks mark the significance levels of the difference. If, for a given variable, two treatments are not significantly different at conventional levels, no superscript is added. This comparison is only conducted to the “right” to avoid double counting, i.e., column (1) is compared to columns (2)-(3), and column (2) is compared to column (3). Significance level of the difference between treatments are highlighted next to the mean values in column (1). *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Table F.3 presents the treatment effects on final test performance, measured by rank (columns (1)–(2)) and absolute performance (columns (3)–(4)), across various belief accuracy groups, with the accurate type in the *NoInfo* treatment as the reference category.

For underestimators, the estimated treatment effect differences suggest that final test performance in *Info* is better than that in *NoGoal*, which, in turn, is better than in *NoInfo* (though only the difference between *Info* and *NoInfo* is statistically significant). This suggests that mandating goal setting without providing relative ability feedback (*NoInfo*) may actually hinder performance compared to not asking them to set a goal at all (*NoGoal*). Forced to anchor on their pessimistic beliefs, underestimators in the *NoInfo* treatment likely set targets that are too easily attainable, creating a demotivating effect. Providing accurate information (*Info*) corrects this pessimism and yields the best performance.

In contrast, overestimators exhibit a different response pattern: final test performance in *Info* is similar to that in *NoInfo* and better than that in *NoGoal* (with *Info* being significantly better than *NoGoal* in one specification). That is, both explicit goal-setting treatments yield better final performance than the control group. How can we explain this divergence? Survey data from the second wave of the field experiment reveals that 12 out of 37 subjects in the *NoGoal* treatment voluntarily set personal targets. These voluntary goals were highly ambitious (mean = 7.75, s.d. = 1.82) and statistically indistinguishable from those set under the explicit mandate in the *Info* treatment (mean = 7.53, s.d. = 1.77, p -value = 0.695 for t-test, p -value = 0.614 for Fisher’s exact test). Furthermore, these 12 self-motivated individuals performed substantially better than their peers who did not set goals (a 53-second longer plank duration (241.86 vs. 188.77) and a 1.23 higher rank (5.75 vs. 4.52), though statistically insignificant due to the small sample size). Crucially, the vast majority of these voluntary goal setters (8 out of 12) were overestimators.

However, this highly motivated subgroup represents less than half of all overestimators (8 of 21) in the control group (in the second wave of the field experiment). For these individuals, who are naturally prone to overconfidence but lack intrinsic goal-setting discipline, the absence of an explicit mandate proved detrimental. Upon receiving their baseline feedback, they operated without any concrete targets or commitment devices, resulting in a severe drop in motivation and the lowest overall performance. In contrast, the explicit goal-setting mandate in both the *Info* and *NoInfo* treatments forced all overestimators to formally commit to a goal. Even if these mandated goals were subsequently adjusted downward (as discussed in Section 4.3), the mere presence of a formal commitment structure prevented their effort from collapsing completely. Therefore, this analysis sheds light on the potential risks and benefits of uninformed goal setting versus the absence of any goal-setting mandate.

Table F.3: Treatment effect on performance in the final test by accuracy - field experiment, NoGoal, NoInfo, and Info treatment

Dep. Var: Final test	rank		abs. performance	
	(1)	(2)	(3)	(4)
Underestimate	-1.015** (0.398)	-1.386** (0.516)	-8.350 (18.18)	-16.86 (21.30)
Overestimate	0.349 (0.608)	0.344 (0.763)	24.37 (18.02)	24.63 (19.66)
Info	-1.132 (0.735)	-1.376 (0.864)	-3.731 (16.91)	-3.235 (20.84)
Info $\times$ Underestimate	2.132** (0.901)	2.276** (0.903)	28.04 (26.68)	25.34 (30.44)
Info $\times$ Overestimate	1.192 (0.826)	1.538 (0.944)	1.439 (21.31)	1.995 (26.45)
NoGoal	0.187 (0.644)	-0.00326 (1.040)	21.81 (26.32)	25.51 (25.99)
NoGoal $\times$ Underestimate	0.538 (0.891)	0.686 (1.127)	-0.340 (30.81)	-3.300 (31.14)
NoGoal $\times$ Overestimate	-1.105 (0.793)	-0.888 (1.174)	-43.59 (28.21)	-45.58* (24.70)
<i>Estimated Differences:</i>				
<i>Info - NoInfo</i>				
Underestimate	1.001* (0.483)	0.900* (0.451)	24.31* (13.37)	22.11 (13.42)
Accurate	-1.132 (0.735)	-1.376 (0.864)	-3.731 (16.91)	-3.235 (20.84)
Overestimate	0.0600 (0.823)	0.162 (0.790)	-2.292 (21.34)	-1.240 (20.70)
<i>NoGoal - NoInfo</i>				
Underestimate	0.725 (0.675)	0.682 (0.499)	21.47 (15.72)	22.21 (12.73)
Accurate	0.187 (0.644)	-0.00326 (1.040)	21.81 (26.32)	25.51 (25.99)
Overestimate	-0.917 (0.762)	-0.891 (0.817)	-21.79 (17.79)	-20.06 (16.24)
<i>Info - NoGoal</i>				
Underestimate	0.276 (0.717)	0.218 (0.644)	2.848 (16.65)	-0.105 (15.99)
Accurate	-1.319 (0.882)	-1.372 (1.126)	-25.54 (24.83)	-28.75 (21.54)
Overestimate	0.977* (0.552)	1.053 (0.636)	19.49 (13.43)	18.82 (14.49)
Baseline test performance	Yes	Yes	Yes	Yes
Group fixed effects	Yes	Yes	Yes	Yes
Personal characteristics	No	Yes	No	Yes
Observations	195	190	195	190

Notes: This table reports the estimates from OLS models (top panel) and the implied differential treatment effects (bottom panel) for the field experiment for all three treatments. Estimates for control variables are not reported. See Table 1 and 6 notes for a list of the control variables and details on sample construction. Robust standard errors allowing for heteroskedasticity and clustered at the group level are reported in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

References

- Clark, D., Gill, D., Prowse, V., and Rush, M. (2020). Using goals to motivate college students: Theory and evidence from field experiments. *Review of Economics and Statistics*, 102(4):648–663.
- Heath, C., Larrick, R. P., and Wu, G. (1999). Goals as reference points. *Cognitive Psychology*, 38(1):79–109.
- Kahneman, D. and Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2):263–291.
- Koch, A. K. and Nafziger, J. (2011). Self-regulation through goal setting. *The Scandinavian Journal of Economics*, 113(1):212–227.
- Köszegi, B. and Rabin, M. (2006). A model of reference-dependent preferences. *The Quarterly Journal of Economics*, 121(4):1133–1165.
- Wu, G., Heath, C., and Larrick, R. (2008). A prospect theory model of goal behavior. Mimeo.