Beliefs as a Means of Self-Control? Evidence from a Dynamic Student Survey

Felix Bönisch, Tobias König, Sebastian Schweighofer-Kodritsch, and Georg Weizsäcker*

May 28, 2025

Abstract

We repeatedly elicit beliefs about the returns to study effort, in a large university course. A behavioral model of quasi-hyperbolic discounting and malleable beliefs predicts that the dynamics of beliefs mirrors the importance of exerting self-control, such that believed returns increase as the exam approaches, and drop post-exam. Exploiting variation in exam timing to control for common information shocks, we find this prediction confirmed: average believed study returns increase by about 20% over the period before the exam, and drop by about the same afterwards. Additional analyses further support the hypothesized mechanism that beliefs serve as a means of self-control.

Keywords: beliefs, present bias, self-control, effort, survey

JEL classification: C81, D81, D91

*Bönisch: WZB Berlin Social Science Center; König: School of Business and Economics, Linnaeus University; Schweighofer-Kodritsch and Weizsäcker: School of Business and Economics, Humboldt-Universität zu Berlin. Mail: Humboldt-Universität zu Berlin, Unter den Linden 6, D-10099 Berlin, Germany. Email: sebastian.kodritsch@gmail.com. We thank Yves Breitmoser, Anna Gumpert, Andreas Haufler, Peter Schwardmann, Felix Weinhardt and Florian Zimmermann for insightful discussions, and our various seminar and conference audiences for helpful comments. Special thanks go to Martina Albers who was involved in early stages of this project. Financial support by the *Deutsche Forschungsgemeinschaft* through CRC TRR 190 (project number 280092119) and by the European Research Council (ERC Starting Grant 263412) is gratefully acknowledged. Declarations of interest: none.

1 Introduction

People exhibit systematically biased beliefs in a variety of domains.[1] To a classical decision maker, these biases are often costly, and neutral at best. Economists therefore assess evidence of belief biases mostly from a "mistakes" perspective. But biased beliefs can also serve to overcome a self-control problem (see, e.g., Bénabou and Tirole, 2002), and they may thereby improve material outcomes for a behaviorally biased decision maker. In this paper, we provide direct field evidence that beliefs about the *returns* to effort with delayed benefits indeed systematically respond to this instrumental motive, i.e., that beliefs serve as a means of self-control.

We investigate the dynamics of students' beliefs about the effectiveness of study effort for exam performance. Studying for an exam has immediate costs and delayed rewards, which are the typical features of a self-control problem that arises due to present bias (see, e.g., Ariely and Wertenbroch, 2002; Augenblick, Niederle, and Sprenger, 2015; Clark, Gill, Prowse, and Rush, 2020; Cheung, Tymula, and Wang, 2022, for related evidence). To examine whether beliefs may be used to overcome this self-control problem, we exploit a time pattern: the returns to studying for an exam increase as the exam comes closer in time, implying that the importance of the self-control problem grows, too. Under instrumental belief distortion, the students' return beliefs should therefore be upward-biased most when the exam is imminent.[2]

To guide our empirical design and analysis, we first formalize this intuition with a simple behavioral model of (β, δ)-discounting and malleable beliefs (i.e., beliefs that are purposefully—if subconsciously—managed and may thus be distorted, within limits, away

[1] There is a large literature in both psychology and economics demonstrating overconfidence and other belief biases, see the survey by DellaVigna (2009). For a quick introduction to the psychology literature on unrealistic optimism, following the seminal work by Weinstein (1980), see Shepperd, Waters, Weinstein, and Klein (2015). For evidence on belief distortions that feature in applied economic analyses, see the contributions in Bachmann, Topa, and van der Klaauw (2023).

[2] A similar pattern of belief distortions may arise in any setting where an agent has to exert preparation effort over time to improve their performance at a fixed future date/event, e.g., a sports competition or a "stage" performance, including sales or job talks. We note that the proposed mechanism keeps constant the (utility-) value of any given outcome, to emphasize beliefs regarding the value-differences as they depend on effort. Our theoretical model in Section 3 would, however, also apply to agents that alter their overall perception of the values of various outcomes, scaling the *difference* in values obtained under lower vs. higher effort, to motivate themselves. For instance, before a sports competition they may tell themselves about important knock-on effects of being successful.

from a Bayesian benchmark). The model indeed yields the prediction that the decision maker's subjectively expected return to effort is most upward-biased in the final study period before the exam, when self-control is most valuable, and also that it drops in hindsight, when the exam is over and the instrumental motive gone.

We then design a dynamic student survey to test the model's predictions on a sample of students in a large university course (first-year Bachelor's microeconomics). Our main variable of interest is a student's belief about an unknown entity: the difference between (i) her performance (measured in point-score percentage) if she were to study for 40 hours during the last two weeks before the exam, and (ii) her performance if she were to study for 20 hours during the same period. Both subjective expectations are elicited at multiple points in time, keeping the target—the return to studying in the last two weeks prior to the exam—constant.

Notice that if information shocks about this target were i.i.d. across students, then their average belief, if formed rationally, should not change over time (the martingale property). In this case, any theory predicting non-trivial belief dynamics could easily be tested against the null hypothesis of rational expectations. However, the students' information shocks are correlated due to their common experiences in lectures, class tutorials, and other observations that inform them about the returns to studying. We therefore rely on a particular feature of the university's exam organization (see Section 4.1 for details): we compare beliefs between two groups of students that take the course's final exam at different points in time. The two groups are indistinguishable from the instructors' perspective, they share the same lectures and class tutorials, and have access to the same information throughout the data collection period. But the two groups vary in the importance of self-control because one group's exam comes several weeks earlier. Indeed, various demographic variables or even beliefs elicited in the initial survey do not allow to predict group membership, supporting the use of one group as control group for the other.

We find that, for each of the two groups, average beliefs follow the model's predicted pattern over time, *relative to their respective exam*: return beliefs increase towards the final study period before one's own exam and sharply drop post-exam. No such reaction occurs around the time that the other group has their exam. Quantitatively, believed returns show an average increase of around 20% in the period before the exam, which largely disappears post-exam and is unaffected by various control variables. When combining both groups into a pooled data set, we find that initial and final average return beliefs coincide, suggesting unbiased baseline priors in combination with a build-up of beliefs in the effectiveness of studying during the exam preparation.

Although 214 students responded in the survey's first wave, our main analysis is based on a subsample of only 84 students that completed all waves and could be unambiguously assigned to one exam date based on additional administrative data. All regression coefficients concerning our hypotheses show the predicted sign, but we obtain statistical significance only when we pool the two exam dates. It is worthwhile noting, however, that selective attrition would likely leave us with students that suffer less-than-average from present bias (for related evidence, see, e.g., Kim, 2020; Harrison, Lau, and Yoo, 2023), whereby effects may be underestimated. Additional analyses further support the hypothesized mechanism. First, we find evidence for the assumption that students experience a self-control problem, in the sense that they systematically over-predict their future study effort.[3] In light of the model, this provides a possible motive for biased beliefs about study returns. Second, we confirm our main results when restricting the non-exam-writing control students to those who, after the respective other group's exam, report knowing the exam questions (which are publicly available immediately after each exam). The effects are therefore not driven by surprising exam content. Third, additional data relating to another course taken by a subset of our sample (introductory mathematics) delivers qualitatively similar results. These data also show that the students' tendencies to report beliefs that change in the predicted manner are correlated between the two courses, as one would expect if present bias is a stable individual characteristic and driver of these beliefs. Finally, other correlations are also in line with the model: students with a greater upward bias before the exam show a larger drop in their return beliefs right after the exam, and these belief movements are positively correlated with the extent of effort over-predictions.

Evidence that self-control problems cause belief bias is hard to come by, and accordingly

[3] While this evidence for present bias is only indirect, the assumption that studying is negatively affected by present bias is as standard as it is empirically well-founded in the literature.

scarce. This is especially true for evidence from the field, with "naturally occurring" tasks. Our approach exploits the dynamic variation in the importance of self-control, similar to the most closely related work to ours, Ma (2020), who considers the dynamics of confidence as measured by students' predictions of their individual grade. Ma elicits economics students' confidence concerning their grades on an essay three times over several weeks – first, well before essay submission, second, shortly after submission, and finally, shortly before grade release. He finds that (average) confidence drops following submission and then stays constant. He interprets this finding through the lens of a model where a single working period (essay writing) is followed by a waiting period, which starts following work completion (essay submission) and ends with the realization (grade release). The model simultaneously allows for both emotional and instrumental benefits of overconfidence, with the former stemming from anticipatory utility and the latter stemming from present bias. Thus, Ma concludes that his evidence supports the instrumental motive (self-control) while rejecting the emotional motive (anticipatory utility), which would predict monotonically declining (over-) confidence.

While the basic setting and focus on dynamics of belief distortion of Ma (2020) are similar to ours, the main differences are that (i) our model has two effort provision periods and focuses solely on the instrumental value of belief distortion due to present bias, abstracting from emotional benefits, (ii) our elicitation concerns itself with the beliefs about the *returns* to studying, which is our main innovation in line with our modeling focus, and (iii) our identification relies on a control group that receives the same information as the treatment group but is in a different period relative to performance, also in line with our multi-period effort model. Overall, the two studies are highly complementary in their approaches, but their findings are consistent and mutually support each other's conclusions; in particular, both find evidence that beliefs are managed to serve the instrumental motive of self-control.

The rest of the paper is organized as follows. After a discussion of more related literature in Section 2, Section 3 presents the simple theoretical model whose predictions inform our empirical approach; Section 4 then describes the survey and explains the empirical strategy to test the theoretical predictions; Section 5 presents our results—tests of the predictions, results on their robustness, and findings from further related analyses—and Section 6 has concluding remarks. The Appendix has various details omitted from the main text.

2 Related Literature

Our point of departure is the general idea that belief distortions may be instrumental in overcoming a self-control problem. In influential work, Bénabou and Tirole (2002) make this point by combining (β, δ)-discounting (e.g., Laibson, 1997, or O'Donoghue and Rabin, 1999) with imperfect self-knowledge, where belief distortions concern *intrinsic* personal characteristics (ability or preferences) and arise from sophisticated self-persuasion of the kind that bad news may be optimally forgotten.[4] By contrast, we study belief distortions about the largely extrinsic return to effort and propose a reduced-form model of belief manipulation. Our model differs in terms of its predictions in two important ways: first, it yields optimal belief distortions also under naïveté about the self-control problem, and second, it generates systematic belief distortions that violate Bayesian updating.[5]

We extend the multiple-selves model with (β, δ)-discounting by an over-arching unbiased "planner" who can directly (and sub-consciously, for the decision-making "doer") distort the return beliefs and thereby achieve self-control. While the model is related to various dual-systems theories (e.g., Thaler and Shefrin, 1981; Fudenberg and Levine, 2006; Brocas and Carrillo, 2008), none of these feature malleable beliefs. Our model shares the feature of optimal (sub-conscious) belief choice with Akerlof and Dickens (1982), Brunnermeier and Parker (2005), Gollier and Muermann (2010) and Bracha and Brown (2012), but with a different motive for belief distortion: instrumental self-control rather than non-instrumental anticipatory pleasure. Indeed, for the case of dynamic consistency ($\beta = 1$), belief distortion is never optimal in our model. It is worthwhile noting that "feel-good" belief bias—regardless of whether it stems from an anticipatory utility motive (Loewenstein, 1987) or an ego utility motive (Kőszegi, 2006)—would concern the exam score for any given study effort, whereas to

[4] Relatedly, Brocas and Carrillo (2000), and Carrillo and Mariotti (2000) show that information avoidance can be optimal under dynamic inconsistency. Compte and Postlewaite (2004) study a model where confidence enhances the probability of success on a task that is performed repeatedly and show that a positively biased perception of one's chance of success is then optimal and can be maintained even in the long run (for related causal evidence from sports, see Rosenqvist and Skans, 2015; Ahammer, Lackner, and Voigt, 2019).

[5] Bénabou and Tirole (2002) maintain full sophistication about how memory is manipulated, so Bayesian updating delivers that the average posterior belief equals the prior. From the perspective of Kamenica and Gentzkow (2011), their model is therefore one of Bayesian (self-) persuasion (relatedly, see also Hagenbach and Koessler, 2022; Mariotti, Schweizer, Szech, and von Wangenheim, 2023), while Bénabou and Tirole (2006) generalize this to various degrees of naïveté about memory manipulation concerning the "belief in a just world."

be instrumentally effective, belief bias has to concern the *returns* to effort, i.e., the effect on the exam score of changing study effort—which involves comparisons of counterfactuals. Put simply, none of these non-instrumental motives make a clear prediction for our elicitation of the returns to effort, whereas the instrumental/self-control motive does. Our abstraction from non-instrumental motives is thus not to deny their relevance, but rather to focus on the mechanism of interest arising from present bias.[6] Indeed, policy research highlights, and confirms, that overcoming present bias is key to students' success (e.g., Levitt, List, Neckermann, and Sadoff, 2016; Himmler, Jäckle, and Weinschenk, 2019).

The repeated elicitation of beliefs around an exam also relates our work to research in psychology on the temporal evolution of confidence—e.g., as a category of "unrealistic optimism" (see Windschitl and Stuart, 2015)—which is measured by beliefs about how well one performs in a self-relevant task (typically, students' predictions of their grade on an exam). A robust insight from this research is that such confidence decreases in the temporal proximity to the time of performance (see seminal work by Nisan, 1972; Gilovich, Kerr, and Medvec, 1993).[7] The study by Ma (2020), discussed in the introduction, can be considered a follow-up on this paradigm.

Two experiments complement our study and support our conclusions in parts where our field evidence is naturally less cleanly controlled.[8] First, Huck, Szech, and Wenner (2018) study costly effort provision under varying information about the piece rate. They find that not knowing whether the piece rate is high or low increases participants' effort, to roughly the same level as the one by participants who know that they face the high one;

[6] Indeed, see Breig, Gibson, and Shrader (2020) and Cordes, Friedrichsen, and Schudy (2023), for experimental evidence supporting the relevance of (also) anticipatory utility for procrastination in completing an exogenously given number of tasks over multiple time periods (see Brunnermeier, Papakonstantinou, and Parker, 2017, for a purely belief-based theory of procrastination). Relatedly, Engelmann, Lebreton, Salem-Garcia, Schwardmann, and van der Weele (2024) provide direct causal evidence of wishful thinking to alleviate anxiety about adverse future outcomes.

[7] Grimes (2002) shows this for economics students, and Sweeny and Krizan (2013) offer a meta-analysis as well as review of the psychology literature. There is also evidence of defensive pessimism developing shortly before the self-relevant feedback among individuals with low self-esteem (Shepperd, Ouellette, and Fernandez, 1996). Relatedly, Chen and Schildberg-Hörisch (2019) experimentally provide the first causal evidence that (absolute) confidence has a motivational value, and Banerjee, Gupta, and Villeval (2020) further show confidence spillovers between unrelated tasks in a field experiment in India.

[8] Different from our study, they come with caveats about using artificial tasks and the short time horizon of a typical laboratory session, where the assumption of a self-control problem due to present-biased utility discounting is somewhat questionable (for an alternative psychological account, see Muraven and Baumeister, 2000; Baumeister, Vohs, and Tice, 2007).

moreover, a third of participants choose to avoid freely available wage information. These findings can be explained by a simple (one-period) effort provision version of our model that assumes low mental costs of belief distortion, and they indicate potentially large instrumental benefits of distorting one's return beliefs.[9] Second, while Huck et al.'s experiment gives the participants either perfect or partial information (beliefs are not elicited), Lobeck (2022) elicits *belief updating* from a noisy signal about whether the participant will be paid randomly or according to effort on a tedious task. While all participants will ultimately have to perform the task a second time in the same "payoff state," Lobeck's treatment varies whether beliefs about returns are elicited before or after they learn about this second task. He finds that participants' updated beliefs that effort pays are greater for those who already know they will have to perform again, and this appears driven by differences in processing discouraging information.[10] Though in a binary setting (effort either pays or does not at all) and without our temporal dimension, Lobeck's findings support the hypothesis that return beliefs respond to the instrumental motive of managing self-control, and they also indicate that this may be driven by positively biased processing of discouraging information.[11]

From a purely empirical perspective our work is also related to tests of rational expectations in the field, where the researcher does not observe (or control) all information that agents receive (e.g., Bernheim, 1990; Benítez-Silva and Dwyer, 2005). Whereas this literature essentially relies on the assumption of i.i.d. forecast errors in their tests of the rational expectations hypothesis, our empirical strategy is able to control also for correlated information.

[9] The authors themselves pitch "plain" present bias against motivated beliefs, but not their combination.

[10] This supports the good-news-bad-news asymmetry in updating put forward by Sharot, Korn, and Dolan (2011), for which evidence is otherwise rather mixed, however (e.g., see the handbook chapter discussion by Benjamin, 2019).

[11] If one views deception as a self-control problem, then the experiment by Schwardmann and van der Weele (2019) is similarly related. They find that people deceive themselves (more) into confidence about an IQ test performance when facing incentives to subsequently deceive others. See also Schwardmann, Tripodi, and van der Weele (2022) for field evidence. A further related experiment is by Drobner and Orhun (2024), who measure the believed return to effort, but as an outcome variable in a learning environment. They show that misperception about returns to effort can be driven by wrong priors about one's own ability.

3 Theoretical Background

We first model the "study problem" of a present-biased student, whom we call Sue, taking an exam at a fixed date. To formalize the intuition that Sue's beliefs about the returns to effort respond to the instrumental benefits of overcoming her self-control problem, we then introduce a self-regulatory system—Sue's "planner"—that subconsciously chooses her beliefs at some mental cost.

Sue's study problem consists of three periods. In the first two periods, $t = 1, 2$, Sue exerts study effort e_t at cost $c(e_t) = \frac{e_t^2}{2}$ in preparation for her exam, which takes place at the end of period 2. She receives her grade

$$g(e_1, e_2, R) = R \cdot (e_1 + e_2) \tag{1}$$

in period 3, where R is the return to her effort, and she trades off her desire to achieve a higher grade against the cost of higher study effort.

Importantly, Sue faces uncertainty about R, where we denote by $\hat{R}_t$ her expectation as of the beginning of period t, and she faces a self-control problem in the form of quasi-hyperbolic discounting (present bias), with parameters $(\beta, \delta) \in (0,1)^2$. Given belief $\hat{R}_t$ in period t, and assuming risk neutrality, she chooses effort e_t to maximize utility U_t, given by

$$\begin{aligned} U_1\left(e_1, e_2 | \hat{R}_1\right) &= -\frac{e_1^2}{2} - \beta\delta\frac{e_2^2}{2} + \beta\delta^2\hat{R}_1 \cdot (e_1 + e_2) \text{ and} \\ U_2\left(e_1, e_2 | \hat{R}_2\right) &= -\frac{e_2^2}{2} + \beta\delta\hat{R}_2 \cdot (e_1 + e_2), \end{aligned}$$

respectively, for the two periods $t = 1, 2$.

Her optimal effort in t, as a function of her return belief, is therefore

$$e_t\left(\hat{R}_t\right) = \kappa_t \hat{R}_t \text{ for } \kappa_1 = \beta\delta^2,\ \kappa_2 = \beta\delta.$$

Sue under-provides effort due to her present bias, $\beta < 1$: for given return beliefs $\hat{R}_t$, an unbiased Sue would want greater effort in both periods. Moreover, Sue exerts greater effort the closer she finds herself to the exam, since the reward of a better grade weighs more heavily in the later period ($\kappa_2 > \kappa_1$).

This completes the description of Sue as a "doer," for given return beliefs. We now turn to the main focus of our model, the determination of beliefs. They are chosen by Sue's planner, who has the same preferences except that she has no present bias ($\beta = 1$). We think of the planner as a subconscious self-regulatory system with the sole capacity to distort the doer's perception of environmental uncertainty in order to overcome her self-control problem. Since the planner has to somehow suppress what the doer "knows," we also assume that belief distortion has some mental cost, increasing in the intensity of self-delusion (this follows Bracha and Brown, 2012; see Coutts, 2019, for supporting evidence).[12]

Concretely, for each t, the planner chooses belief $\hat{R}_t$ at cost $b_t = \gamma \frac{1}{2}\left(\hat{R}_t - \hat{R}_0\right)^2$, where $\gamma > 0$ is a scaling-parameter and $\hat{R}_0 > 0$ is Sue's planner's belief in period $t = 0$. Under the simplifying assumption that no information arrives during $t = 0, 1, 2$, Sue's planner maximizes[13]

$$\begin{aligned} V\left(\hat{R}_1, \hat{R}_2, \hat{R}_3 | \hat{R}_0\right) = &- \frac{e_1\left(\hat{R}_1\right)^2}{2} - \delta \frac{e_2\left(\hat{R}_2\right)^2}{2} + \delta^2 \hat{R}_0 \cdot \left(e_1\left(\hat{R}_1\right) + e_2\left(\hat{R}_2\right)\right) \\ &- \gamma \frac{1}{2}\left(\hat{R}_1 - \hat{R}_0\right)^2 - \delta\gamma \frac{1}{2}\left(\hat{R}_2 - \hat{R}_0\right)^2 - \delta^2 \gamma \frac{1}{2}\left(\hat{R}_3 - \hat{R}_0\right)^2. \end{aligned} \tag{2}$$

Correctly predicting the doer's effort response, the planner trades off instrumental benefits and mental costs of belief distortion. The solution to this problem is (where we let $\kappa_3 \equiv 0$):

$$\hat{R}_t^* = \hat{R}_0 \cdot \left(1 + \frac{1-\beta}{\beta} \cdot \frac{\kappa_t^2}{\kappa_t^2 + \gamma}\right). \tag{3}$$

Before her exam, Sue will come to believe that the returns are excessively high, and the more so the closer is the exam ($\hat{R}_2^* > \hat{R}_1^* > \hat{R}_0$). After the exam, there is no instrumental value to costly self-delusion, hence her return beliefs will be undistorted ($\hat{R}_3^* = \hat{R}_0$).

[12] While our model presents a reduced form of the underlying psychological mechanism(s), this might for instance involve selective recall or selective attention (cf. Bénabou and Tirole, 2002), with the doer being unaware of the strategic selection, however. See also our related concluding discussion in Section 6.

[13] Allowing for information arrival would complicate the model but can be incorporated. We abstract from it here to focus on the proposed behavioral channel, so any belief change is caused by the motive to overcome the self-control problem. Moreover, together with the fact that Sue's planner has dynamically consistent preferences, it also conveniently implies that it is without loss of generality to have her choose beliefs once-and-for-all at the outset.

Indeed, the only reason to distort beliefs here is instrumental: absent present bias, beliefs would be undistorted also before the exam (for $\beta = 1$, $\hat{R}^*_t = \hat{R}_0$). Note also that as the mental cost to self-delusion becomes arbitrarily small, this self-regulatory mechanism allows Sue to achieve the long-run optimal level of effort (as $\gamma \to 0$, $\hat{R}^*_t \to \hat{R}_0/\beta$ for $t = 1, 2$).

In sum, we obtain the following prediction.

Main Prediction (Return Beliefs): *Sue's expectation of the returns to her study effort*

(1a) is the higher the closer ahead is her exam, and

(1b) drops sharply after she took her exam.

What matters for this prediction is sophistication of the planner, it is qualitatively unaffected by naïveté of Sue as the doer. However, naïveté implies, over and above the made prediction, that Sue mispredicts her future effort, which affords a simple test of whether she indeed suffers from present bias.[14]

We emphasize that the goal of our model is to offer but the simplest concrete formalization of our hypothesized mechanism (as highlighted in the above prediction), before investigating it empirically. However, the main prediction is robust to several model extensions/variations: specifically, in Appendix A, we show robustness to generalized present bias and knowledge decay for Sue as the doer, as well as some present bias also for Sue as the planner.[15] Clearly, other variations of the model specifications may yield additional predictions and suggest different empirical tests. One such additional prediction concerns the relationship between the extent of belief distortion and the extent of present bias. Sue's belief $\hat{R}^*_t$ increases in her present bias (i.e., decreases in β) unless she suffers from overwhelming present bias as the doer.[16] Specifically, this is guaranteed if $\beta > 0.5$, a threshold safely below the available estimates (see Augenblick and Rabin, 2019; Cheung, Tymula, and Wang, 2021).

[14] At the beginning of period 2, when she already has belief $\hat{R}^*_2$, naïve Sue would overpredict her subsequent effort under any "partial" naïveté in the sense of O'Donoghue and Rabin (1999). Earlier on, in period 1, she may however over- or underpredict her period-2 effort: while naïveté, as usual, works towards overprediction, her beliefs will subsequently become upwards-distorted, which works towards underprediction.

[15] While we assume $\gamma > 0$ throughout, as an additional note, the result also obtains with costless belief distortion together with a lexicographic taste against self-delusion of the planner whenever indifferent.

[16] The sign of $\partial \hat{R}^*_t/\partial \beta$, $t \in \{1, 2\}$, is negative if and only if $(0.5 - \beta)2\gamma < \kappa^2_t$, where $\kappa_1 = \beta\delta^2$ and $\kappa_2 = \beta\delta$. To see intuitively why this relation breaks down for small β, take the extreme case of $\beta = 0$: belief distortion then has no instrumental value, because such a myopic doer Sue would exert zero effort anyways.

4 Data Collection, Hypotheses, and Identification

4.1 Data Collection, Sample and Belief Measures

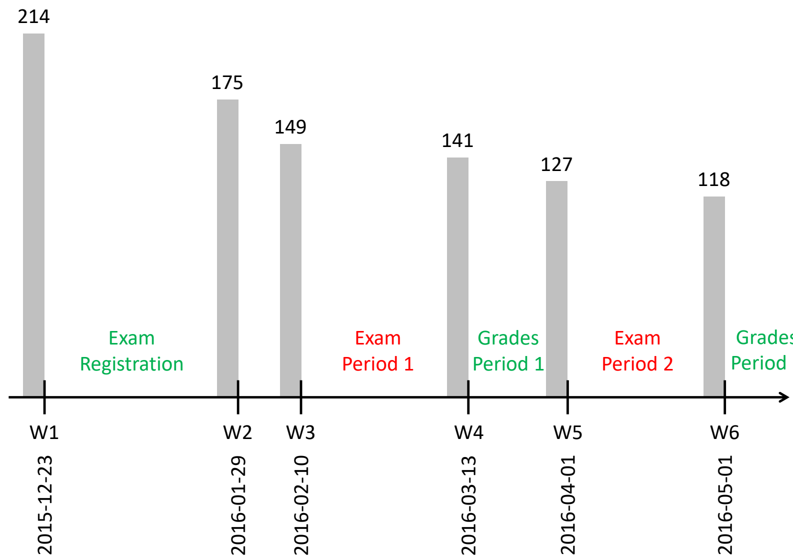
We collected the data in a repeated online survey of the students taking our first-semester microeconomics course at Humboldt-Universität zu Berlin, in the winter term 2015/16.[17] The survey consisted of six waves, eliciting students' beliefs about their study effort, grades and returns to studying at different stages in the semester cycle. We describe here its key features; further details are provided in Appendix C. Participation in the survey was voluntary and incentivized with a €10 completion payment, plus the chance (1:7) of winning an Amazon voucher worth €100. After the final wave of the survey, the data were anonymously matched with data from the university's examination office.

The timeline of our survey is visualized in Figure 1. An important feature for our design is that any given student faces one of two different exam dates that were several weeks apart, namely February 23, 2016 (in exam period 1), and April 15, 2016 (in exam period 2). All students were required to commit themselves to one of the two dates by January 25, 2016 (exam registration) – a decision that students typically make with the goal of balancing their individual schedules of about six final course exams per semester. The students' individual schedules are heterogeneous, as they depend on the program of study and year/semester of entry, as well as individual study progression and course choices. We started the first wave of our survey in mid-December 2015, and the final wave was completed in early May 2016.

In our initial wave we had 214 respondents, which is about one-half of the students who ended up writing the exam. As we are interested in the dynamics of beliefs over time, we include only students who participated in all six waves (118 students). Also crucial for our identification strategy is that students can be unambiguously assigned to one of the two exam dates; otherwise, we cannot maintain consistent treatment/control groups over time. Our sample of interest therefore includes only those "stay-ons" who took the exam on the date for which they had actually registered. This further reduces our sample by 34 students,

[17] One author (Weizsäcker) gave the weekly lectures, two authors (König and Schweighofer-Kodritsch) held weekly class tutorials, and one author (Bönisch) conducted the survey without being involved in teaching. There were three further TAs who were not involved in this research in any way.

Figure 1: Survey Timeline



Notes: Figure 1 visualizes the timeline of the survey comprising of six waves (W1–W6). Grey bars indicate the number of participants in each wave. First-exam takers write the exam between W3 and W4, second-exam takers between W5 and W6.

leaving us with 84 students in the final sample.[18] In Appendix B, we show that this (reduced) sample of 84 students does not significantly differ from other students who completed the first wave in terms of background characteristics elicited there. Moreover, the same is true when comparing the two groups of students that form our final sample.

From wave 2 onwards, we elicited beliefs about the returns to studying. We asked students to give us an estimate of the percentage points they expect to achieve in their microeconomics exam for two hypothetical effort scenarios: (a) if they were to study for the exam for 20 hours in the 14 days prior to their exam date, and (b) if they were to study for 40 hours in the 14 days prior to their exam date. In a wave occuring after a student's exam date, this question was adjusted to refer to the past; i.e., we asked what percentage of points a student thought she would have achieved if in the 14 days prior to her exam she had studied 20 hours, and similarly for the case of 40 hours. The numbers for the hypothetical effort scenarios were chosen based on the students' own effort expectations: in wave 1, 20 and 40 hours are the

[18]Of these 34 students, 31 did not take either of the two exams, and three were registered for the first exam date but ended up taking the exam on the second date.

two tertiles (rounded) of responses to a question about own expected study effort during the 14 days prior to their exam date.[19] Our return belief measure is the difference in the subjectively expected percentage points achieved in the exam between the two hypothetical effort scenarios. We denote this belief by r^i_τ for student i and wave τ.

Our second variable of interest, in addition to *return* expectations, is the students' prediction of their own study *effort*, which allows to investigate our assumption of a self-control problem. Specifically, we also asked students how many hours they expected to study for the exam in the 14 days prior to it (with straightforward adjustment to refer to the past after the exam), corresponding to e_2 in the model. Summarizing our findings, we indeed find evidence of a self-control problem for students, via naïveté about and hence "overprediction" of their future study effort. In wave 3, shortly before the exam of group-1 students, these students on average predict their study effort to be 43.25 hours, while shortly after the exam, in wave 4, they report to have studied only for 38.69 hours. For group-2 students, we observe a similar pattern: in wave 5, they report an average prediction of 42.69 hours, whereas in wave 6 they report an average of 37.35 hours.[20] Confirming this pattern of overprediction, we also find that the students tend to overpredict the total number of course exams they would write: the average difference between their prediction at the initial wave 1, before any exams, and their report in the final wave 6, after all exams, is 0.24 for group 1 and 0.30 for group 2.[21] Effort overprediction and exam overprediction exhibit a statistically significant positive correlation (0.248, $p = 0.025$).

4.2 Empirical Strategy and Hypotheses

As explained in the introduction, we deal with the possibility of correlated information shocks by exploiting the timing of the exam. Group-1 students take the exam seven weeks before

[19]The median response in wave 1 is 30. These aggregate statistics are the only information we received before the end of the survey and finalization of all grading.

[20]The average overprediction in our data is 4.78, which is significantly different from zero at the 5-percent level. When calculating this average, we removed two students' observations who reported more than 200 hours in the final study period (one per group). This implies that they would study more than 14 hours a day in each of their last 14 days before the exam, which we find implausible and constitutes extreme outliers by any standard.

[21]These are the numbers excluding the two students with "unrealistic" effort reports also here, for comparability. (Otherwise the numbers would be 0.25 for group 1 and 0.29 for group 2.)

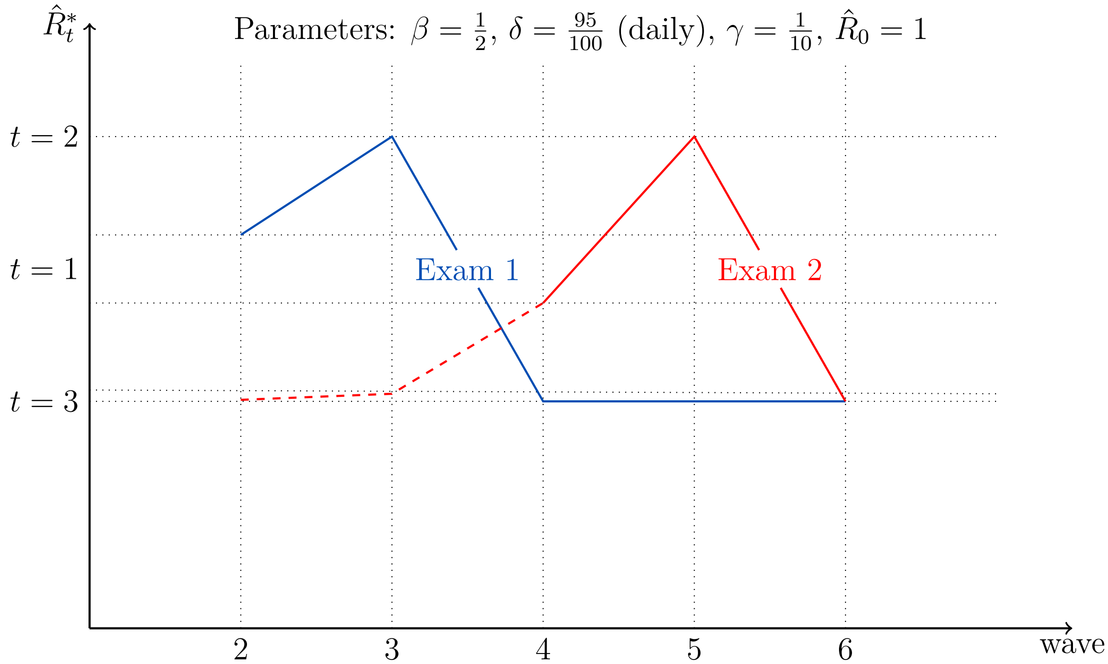
group-2 students, whereby, at any given point in *calendar time*, the two groups are identical in terms of available information but differ in terms of *model time* (i.e., distance to exam), and hence in terms of how important self control is.

Define $s_t^i \equiv r_t^i - r_{t-1}^i$ as individual i's belief revision between times $t-1$ and t. Rational expectations imply that this revision has a zero expectation as of $t-1$. Moreover, if information innovations (belief revisions) are i.i.d. within a sample of n people, then the group average $s_t = \sum_{i=1}^{n} s_t^i$ approaches zero with growing n. In our application, however, belief revisions are likely to have a common component due to common information: $s_t^i = \epsilon_t^i + \eta_t$, where only the first term is i.i.d., while the second is a common innovation. Under rational expectations, any such common innovation η_t shifts the sample's average. However, the difference between two groups' averages (at a given time t) has this common component removed, hence $s_t^{G_1} - s_t^{G_2}$ still approaches zero ($s_t^{G_k}$ is group k's average). Comparing our two groups, we can test whether expectations are rational against the alternative hypothesis of a systematic pattern over time, as predicted by our model.

Figure 2 illustrates the "literal" prediction of our theoretical model, which abstracts from information shocks, with calendar time (i.e., waves 2 through 6) on the horizontal axis and the agent's return beliefs at different moments in model time on the vertical axis. The figure illustrates how both groups progressively "build up" their return beliefs as their respective exam date approaches. Group 1 takes the exam at the first date, between waves 3 and 4, whereas group 2 takes the exam at the second date, between waves 5 and 6. Hence, they go through the same dynamic pattern of beliefs, but in a staggered fashion. At wave 3, group 1 is close to their exam, corresponding to $t = 2$ in the model, whereas group 2's exam is still distant, corresponding to model time $t = 1$. (Note here that group 1 starts relatively high already in wave 2, carried out right after the end of the exam registration phase, because of its temporal proximity to wave 3 and hence also group 1's exam.) At wave 4, group 1 is past their exam (without having learned their grades yet), corresponding to model time $t = 3$, whereas group 2's exam is still distant, corresponding to another version of model time $t = 1$ (the figure's notes explain the parameterization). At wave 5, group 2 is close to their exam, corresponding to model time $t = 2$, etc.

Common information may shift both groups' beliefs, however. In our statistical analysis,

Figure 2: Predicted Return Beliefs over Time



Notes: Figure 2 depicts the model's prediction of expected returns $\hat{R}_t^*$ over the five waves (2 through 6) where this belief was elicited, for our two groups (group 1 in blue, and group 2 in red). To derive numerical predictions that take into account the different distances to the exam (e.g., in wave 2 both groups are at model time $t = 1$ but group 1 is much closer to their exam), we extend Section 3's model to T pre-exam periods (days), which is straightforward. We then use actual days of the survey (see Figure 1), together with a discount factor of 0.95 per day.

we control for this by basically considering an upcoming exam as a treatment for which we have a control group. Specifically, we then let each group be the treatment group in waves where, according to our model, belief manipulation incentives are the strongest: wave 3 for group 1 and wave 5 for group 2. We take the respective other group as the control group and predict return beliefs to be higher for treatment than for control. We also predict that once the exam is written and there is no instrumental motive to distort return beliefs, they become undistorted. Hence, within each group we predict a drop in average beliefs in the wave immediately following this group's exam, relative to the group that did not just write the exam.[22] Our main hypotheses, concerning return beliefs, are summarized below.

[22]More generally, students' beliefs may be subject to systematic biases in updating that are independent of the instrumental motive we aim to test for. From this perspective, the control group serves as baseline, against which we compare updating under exam treatment, which intensifies the self-control problem.

Main Hypotheses (Return Beliefs):

(1a) In wave 3, the average return belief of group 1 exceeds that of group 2, and in wave 5, the average return belief of group 2 exceeds that of group 1.

(1b) Between waves 3 and 4, there is a drop in the average return belief of group 1 relative to group 2, and between waves 5 and 6, there is a drop in the average return belief of group 2 relative to group 1.

5 Results

5.1 Main Hypotheses

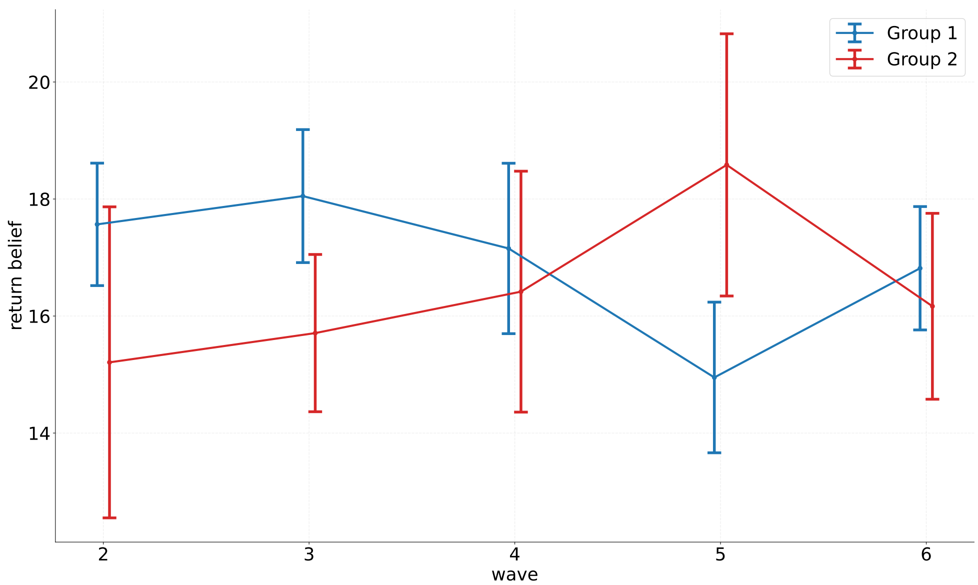
Graphical Illustration We begin the analysis with a graphical illustration of our key variable of interest: Figure 3 shows the dynamic pattern of mean return beliefs for our two groups. The figure covers waves 2 through 6, where the questions about returns were included in the survey. The blue and red lines represent the means of group-1 and group-2 students, respectively.

Even for each group in isolation, we observe that beliefs build up towards the exam, reaching their single peak at the wave immediately preceding it, and then drop right afterwards. Figure 3 quite closely resembles the illustration of the model's literal prediction of Figure 2, remarkably so in fact for group 2.[23] Some more detailed data patterns are not predicted by our simple model. First, in wave 4, group 2's beliefs are slightly (though insignificantly) below those of group 1, where one may have expected the opposite because group 1 has no instrumental motive for distorting beliefs any more, while the motive may be at work for group 2 already, even a month's time before their exam. Second, and relatedly, while group 1's beliefs do drop right after the exam, they subsequently drop further, and even more sharply, before ultimately returning to a level similar to that in wave 4.[24]

[23]Appendix B's Figure 4 shows box plots of return beliefs by exam group and wave period. While interquartile ranges are rather similar between groups and quite stable over time, our main hypotheses are qualitatively confirmed also in terms of the groups' *median* return beliefs (though these do not exhibit an absolute increase over pre-exam waves as observed for *mean* return beliefs).

[24]While this could be due to a common information shock that pushes "true" return beliefs downward, we suspect that, more plausibly, the fact that students in group 1 learn their exam grade shortly before wave 5

Figure 3: Return Beliefs Over Time



Notes: Figure 3 plots mean return beliefs over waves 2-6 by group, blue for the earlier exam-takers of group 1 ($N = 60$), red for the later exam-takers of group 2 ($N = 24$). Error bars represent the standard error of the mean.

However, it is notable that both groups' beliefs reach similar peak levels before their respective exams, and the two groups' beliefs basically converge after each exam. The latter holds true in particular in the final wave 6. Although group 2 then still has not yet learned their grades, neither group has any instrumental distortion motive at this point (both are post-exam), and their beliefs largely coincide. In fact, the *overall* average return belief of the pooled sample in the final wave 6 is very similar to that in wave 2, when this belief was first elicited, albeit with a much reduced variance (means of 16.89 vs. 16.63, and variances of 94.84 vs. 64.28, in wave 2 vs. wave 6, respectively). This initial-final comparison is consistent with an unbiased prior and rational (Bayesian) updating from i.i.d. information

is influential here, in either of two ways. First, the freshly received factual grade may temporarily act as an anchor for the students with intermediate actual effort, which would limit the variation between expected scores in the two hypothetical/counterfactual effort scenarios of 20 vs. 40 hours (Tversky and Kahneman, 1974), thereby reducing the implied expected return to effort. Second, students with actual preparation effort (not too far) below 40 hours may avoid immediate negative feelings of regret (Bell, 1982; Loomes and Sugden, 1982) by telling themselves that had they really studied harder, this would not have made that much of a difference, also moving the two estimates closer together. Neither of these psychological aspects is captured by our model, which effectively focuses on beliefs until only shortly after the exam.

– but the dynamic (diverging) patterns of the two groups in-between are inconsistent with it; instead, they are largely in line with the instrumental belief-distortion channel. Summing up this first graphical investigation, our simple model naturally misses some aspects of the observed patterns but it matches the key qualitative features of each group's return beliefs and supports our main hypotheses regarding the group differences rather well.

Table 1: Return Beliefs

	Date 1		Date 2		Pooled	
	(1)	(2)	(3)	(4)	(5)	(6)
Exam	2.342	1.772	3.633	3.958	2.987**	2.988**
	(1.753)	(1.861)	(2.569)	(2.680)	(1.265)	(1.296)
Date 2					-0.113	-0.113
					(1.265)	(1.296)
Const.	15.708***	6.496	14.950***	6.953	15.247***	6.406
	(1.332)	(4.392)	(1.292)	(4.648)	(1.235)	(4.001)
Controls	no	yes	no	yes	no	yes
Obs.	84	84	84	84	168	168
R^2 / Pseudo	0.017	0.142	0.026	0.071	0.027	0.075
F-statistic	1.784	1.465	2.000	1.303	4.109**	2.489**

S.E. in brackets, for Date 1 and 2 robust (HC1), for pooled OLS clustered at ID level.
*** p<0.01, ** p<0.05, * p<0.1

Regression Analysis: Hypothesis 1a Table 1 contains regression results that allow testing whether students whose exam is imminent have higher return beliefs than students whose exam lies in the more distant future, or in the past (Hypothesis 1a). We carry out every regression with and without control variables elicited in the initial wave 1.[25] Columns 1 and 2 report the corresponding results from regressing wave 3 return beliefs, r_3^i, on an *Exam* dummy indicating that the student takes the exam between waves 3 and 4 (here,

[25]These control variables are gender, semester of study, whether they take the microeconomics course the first time, program of study (economics, business, other), age, their expected result (percentage score on their exam) and their expected effort (study hours during the 14 days before their exam) as of wave 1.

meaning that she is in group 1). The *Exam* coefficient is positive, as predicted, but not significantly so. Running the corresponding regression for the second exam date – here, using r_5^i as the dependent variable and letting *Exam* equal one for group-2 students – also gives the predicted sign without significance (columns 3 and 4). A pooled regression that combines both data sets and includes a *Date 2* dummy (equal to one for group-2 students) to capture time-fixed effects produces a similar estimate on *Exam*, but with much smaller standard errors, such that we obtain statistical significance (columns 5 and 6). These results are robust to including various controls, which we list in Footnote 25. The full regression results can be found in Appendix B's Table 7.

Table 2: Return Beliefs – After vs. Before Exam

	Date 1		Date 2		Pooled	
	(1)	(2)	(3)	(4)	(5)	(6)
Exam	-1.603	-1.448	-4.283*	-4.090	-2.943**	-2.943**
	(2.600)	(2.559)	(2.445)	(2.493)	(1.290)	(1.322)
Date 2					-0.182	-0.182
					(1.290)	(1.322)
Const.	0.708	1.276	1.867*	0.802	1.665	1.596
	(2.194)	(5.620)	(0.986)	(4.433)	(1.297)	(4.211)
Controls	no	yes	no	yes	no	yes
Obs.	84	84	84	84	168	168
R^2 / Pseudo	0.005	0.058	0.048	0.105	0.021	0.067
F-statistic	0.380	0.696	3.068*	0.814	2.647*	1.360

S.E. in brackets, for individual dates robust (HC1), for pooled OLS clustered at ID level.
*** p<0.01, ** p<0.05, * p<0.1

Regression Analysis: Hypothesis 1b To test for relative drops in return beliefs, we difference the beliefs between wave dates 3–4 and 5–6, respectively. For each individual we subtract the return belief in the wave occurring immediately before an exam date from the return belief in the wave occurring immediately after that exam date. We regress the resulting difference, $\Delta_t^i := r_t^i - r_{t-1}^i$ (for $t = 4$ and $t = 6$), on an *Exam* dummy indicating that the student's exam immediately preceded wave t (in $t = 4$ it equals one for group 1, and in $t = 6$ it equals one for group 2). We run this regression again for each exam date separately and also for the pooled sample, in the latter case controlling for time-fixed effects via a *Date 2* dummy, all with and without the aforementioned control variables.

The results are summarized in Table 2, whereas the full results including all controls can be found in Appendix B's Table 8. The *Exam* coefficient is negative throughout, indicating that students who are "treated" in the sense that they recently wrote the exam (*Exam*=1) experience a drop in their return beliefs, relative to "untreated" students. This finding is again robust to including controls, and it is statistically significant for the pooled regression.

5.2 Robustness of Main Results

We here address what we consider the three main concerns regarding our findings: Endogenous selection into groups, group-specific exam information, and how the belief distortion observed in our microeconomics course relates to analogous data from another course.

Selection Due to university regulations, we could not randomly assign students their exam dates exogenously, nor incentivize them toward exam dates in any way. Hence, one might worry about endogenous selection of students into exam dates (corresponding to our two groups), such that it somehow generates the particular pattern observed. Though we could not think of any plausible selection condition that would produce the rather intricate dynamic pattern of return beliefs for *both* groups separately, we cannot rule out that one exists.

However, in terms of the personal characteristics that we elicited in wave 1, we find that the two groups are balanced (see Appendix B's Table 5 (a)). This holds true also in terms of final exam grades and various proxies related to self-control – specifically, their expected and actual exam performance, how much they overpredict their study effort and how much they overpredict the total number of exams they would write that semester (see Appendix B's Table 5 (b)). Probit regressions further confirm that personal characteristics, exam grades and proxies for self-control are not predictive of group membership, i.e., whether a student in our final sample would write (and actually wrote) the exam at the first or second date (see Appendix B's Table 6).

Regarding attrition, we note that the students who dropped out of the survey over time could reasonably be expected to experience *more* severe self-control problems than the students who managed to complete all six waves. This would suggest that the observed belief distortions about the return to study effort, based on the latter subsample, may well be underestimated (for related evidence, see Kim, 2020; Harrison et al., 2023).

Exam Information While the two groups of students wrote exams on the same course, their exams were not identical, of course. To the extent that our empirical confirmation of the predicted drop in return beliefs (Hypothesis 1b) stems from updating from the actual exam, group differences may arise by construction. However, information conditions were equalized across groups as far as possible, as each exam was made available to all course participants directly after it had been written. Moreover, we made all reasonable efforts to keep the contents and levels of difficulty of the exam questions identical.

In addition, our later survey waves asked the participants whether they knew the contents of the exam of the earlier date. Thus we are able to redo the corresponding regressions using only the subsets of the respective control group that knew the relevant exam (with pooled data). Despite the accordingly smaller number of observations – as a control group, group 2 reduces to 13 out of 24 for the first exam, while group 1 reduces to 16 out of 60 for the second exam – the results are robust to this variation: We obtain statistically significant estimates with the predicted sign (*Exam* coeff. -4.27, s.e. 1.81, without controls, and -4.50, s.e. 1.87, with controls, both significant at the 5%-level; see Table 9 in Appendix B for the full regression results).[26]

Cross-Validation with Additional Course In addition to the microeconomics course, we also elicited students' return beliefs regarding the introductory mathematics course, which most students take simultaneously to microeconomics. There are several reasons for why the responses about the mathematics course may be less reliable: most importantly, the survey waves were timed towards the exam dates in microeconomics, which differ from those in mathematics. Therefore, at the time of wave 3 the students who took the first

[26]We also confirm Hypothesis 1a for this subsample, though the corresponding *Exam* coefficient is only marginally significant.

mathematics exam were already one week into the 14 days period before the exam, and thus the timing relative to the exam is not "symmetric" across groups. Moreover, we have no hard information regarding which mathematics exam date students registered for or actually took, nor any grade information. Nonetheless, we use this data to carry out a similar analysis on the responses about the mathematics exams and investigate whether the belief distortion that we find regarding the microeconomics course correlates with a potential belief distortion regarding the mathematics course.

Performing regressions analogous to those above for our main hypotheses—here with 44 group-1 students and 19 group-2 students, based on self-reports in final wave 6—we find some further support for both hypotheses (for the full regression details, including controls, see Appendix B's Tables 10 and 11). First, students that are just about to write the exam tend to exhibit higher return beliefs than students whose exam lies in the more distant future or in the past (i.e., for both exam dates), though this tendency is rather weak here and not statistically significant (even for the pooled sample). Second, return beliefs of students who have just written the exam decrease relative to their respective control group; this effect is statistically significant for the first exam date (1%-level) as well as for the pooled sample (5%-level). Notably, our controls once again leave estimated treatment effects unaffected.

Table 3: Link Between Micro and Math

	Pre-Exam Belief Change		Post-Exam Belief Change		Effort Overprediction	
	(1)	(2)	(3)	(4)	(5)	(6)
Bel. Change Math (Pre)	0.710**	0.743**				
	(0.301)	(0.354)				
Bel. Change Math (Post)			0.522***	0.577***		
			(0.146)	(0.131)		
Eff. Overpred. Math					0.622**	0.609**
					(0.251)	(0.260)
Const.	1.938	1.555	1.712	4.167*	3.428	0.846
	(1.772)	(1.763)	(1.458)	(2.342)	(2.156)	(2.773)
Controls	No	Yes	No	Yes	No	Yes
Obs.	52	52	52	52	52	52
R^2 / Pseudo	0.297	0.310	0.188	0.219	0.271	0.302
F-statistic	5.580**	4.369***	12.832***	5.786***	6.143**	2.127*

Robust S.E. in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Most importantly, however, we find strong correlations between microeconomics and mathematics in all three key measures related to instrumental return belief distortion to overcome a self-control problem. These are (i) effort overprediction, as predicted effort in the wave immediately preceding the exam minus reported effort in the wave immediately succeeding the exam (wave 3 vs. wave 4 for group 1, and wave 5 vs. wave 6 for group 2), (ii) "post-exam belief deflation," as the drop in return beliefs subsequent to writing the exam ($r_3^i - r_4^i$ for group 1, and $r_5^i - r_6^i$ for group 2), and (iii) "pre-exam belief inflation," as the gradual inflation of return beliefs in the period leading up to the exam ($r_3^i - r_2^i$ for group 1, and $r_5^i - r_4^i$ for group 2). Table 3 presents the results of regressing each of these individual measures for microeconomics on the same measure for mathematics, with and without control variables. Despite the smaller sample—only 52 students out of the 84 students in our main sample took the mathematics exam—the corresponding (partial) correlations are statistically significant, and they are all high, well above 0.5.[27] Again, including control variables hardly affects the estimates (see Appendix B's Table 12 for the full results). This evidence is consistent with the argument that the fundamental reason for belief distortion in our model is a personal trait, present bias, which has similar consequences in both courses.

5.3 Further Results

Return Beliefs and Study Effort Our return belief elicitation concerns the difference between two hypothetical scenarios (40 vs. 20 hours of study effort). Our measure exhibits great variance, as can be seen from Figure 3, raising the question of how strongly it relates to actual behavior. To address this question we regress effort (expected or ex-post reported) on the return belief measure, as well as various controls. The results show a strongly significant positive correlation, supporting the validity of our belief measure as a basis for behavior (see Appendix B's Table 13).

Belief Distortions, Present Bias, and Exam Performance An intuitive but theoretically not entirely straightforward hypothesis is that the amount of upwards distortion in

[27] We have 29 group-1 and 9 group-2 students who took the mathematics exam at the first date. For the second date these numbers are 7 and 7, respectively.

return beliefs increases in the degree of present bias (decreases with β). The students' effort overprediction provides a measure of their present bias, allowing to investigate also this hypothesis. We again consider both pre-exam return inflation and post-exam return deflation for belief distortion here, and we additionally investigate whether the same individuals who inflate return beliefs pre-exam also deflate post-exam.

The correlation coefficients between these three measures for the pooled sample are all positive;[28] in particular, effort overprediction is positively correlated with both pre-exam belief inflation (correlation coeff. 0.342, significant at the 1%-level) and post-exam belief deflation (correlation coeff. 0.198, significant at the 10%-level). Moreover, there is a very high correlation between the two measures of belief distortion (correlation coeff. 0.548, significant at the 1%-level).[29]

Finally, we note that if belief distortions were unrelated to self-control problems, we would expect students with stronger present bias to perform significantly worse on the exam. We therefore use effort overprediction as a measure of present bias and relate it to actual exam scores. While with a negative sign, the plain correlation is close to zero, equal to -0.083, and statistically insignificant. This is well in line with the motivational mechanism suggested here, namely that the students' self-control problem is at least partially resolved by distorting return beliefs. Of course, this is only suggestive, as present bias may be correlated with unobserved "natural" microeconomics skills.

Power Analysis While we have obtained results that consistently align with our hypothesis and the underlying theory that beliefs serve as a means of self-control to overcome present bias, including cross-validation, they are based on a rather small sample. In Appendix B.4.2, we therefore add a basic investigation of the extent to which our main analysis (i.e., just the tests of the main hypotheses on a single course) may be underpowered, to better inform its reliability and also guide related future research in natural settings where sample size is a

[28] We again exclude the two students (one per group) with impossible effort reports of more than 200 hours, though the findings do not hinge on that; see footnote 20.

[29] Splitting the sample by group yields qualitatively similar results, though for group 1 all correlations are weaker than for group 2 (indeed, for group 1, effort overprediction is hardly correlated with the belief distortion measures, while the latter two remain highly and significantly correlated); and a similar picture emerges when running regressions, including controls (see Appendix B's Table 14).

critical aspect.[30] Here, we offer a brief summary.

Because ex-post power analysis based on already observed data is viewed very critically (see, e.g., Hoenig and Heisey, 2001), we conduct an "ex-ante" version of power analysis. We posit different effect sizes (group differences in belief measures) due to belief manipulation at different times relative to an exam and calculate the sample size required for our tests to achieve conventional power levels of 80% or more (two-sided t-tests of Null hypothesis of no difference, $\alpha = 0.05$), for comparison with our actual sample size. Absent information from prior research, we employ our simple model of Section 3 to derive effect size ranges, where we vary the key present-bias parameter β over the 95% confidence interval of $[0.51, 0.85]$ for its meta-analytic estimate of 0.66 reported in the meta-study by Cheung et al. (2021) for non-monetary rewards, while fixing other parameters.

Thus, we find that our cross-sectional analysis of Hypothesis 1a (comparing pre-exam return-belief levels, waves 3 and 5) is underpowered: 80% power would require 1.2 to 12 times our actual sample size, depending inversely on the extent of present bias. By contrast, our difference-in-differences analysis of Hypothesis 1b (comparing pre- to post-exam return belief revisions, waves 3 to 4 and 5 to 6) is generally better powered, easily achieving conventional power of 80% for large present bias and very nearly so for intermediate present bias. In particular, our calculations deliver that a sample size of around 200 observations would ex ante have been sufficient to detect belief manipulation effects with conventional power in our setting for intermediate present bias.

We also illustrate how our actual data may inform future research ex post. We calculate the standardized effect size for the two hypotheses (Cohen's d, see Cohen, 1988) and the implied sample requirement for conventional power in future studies like ours. Thus, we basically confirm the ex-ante result of 200, in line with intermediate present bias. Despite some caveats, we consider the findings quite consistent and encouraging with regards to the feasibility of similar future studies. However, given our own experience, starting from 214 students, we emphasize that achieving a target sample size in settings like ours depends critically on appropriate participation/completion incentives and management of also other potential risks to sample size.

[30] We would like to thank an anonymous reviewer and the associate editor for this excellent suggestion.

6 Concluding Remarks

This paper provides evidence of systematic deviations from rational expectations about the returns to studying. A key feature of the analysis is that the test for rational expectations works without any observation on the *actual* returns to studying. The violation of the martingale property is enough to conclude that rational expectations are rejected. For natural settings such as ours, we show how to augment this basic test by a suitable control group, in order to rule out that the effect is driven by correlated information. Thus, we provide a method that extends its applicability.

Of course, the value of finding such a violation of rational expectations rests on the behavioral mechanism that one desires to test (as does, in this paper, the entire empirical design). Here, the particular dynamic pattern of beliefs we observe is predicted by the motivational incentives that arise with self-control problems: the importance of self-control increases as the exam gets closer in time and vanishes after the exam, and belief deviations follow exactly the same pattern.

The paper thereby contributes to the literature on self-control, which has mostly focused on extrinsic commitment opportunities (for a review see Bryan, Karlan, and Nelson, 2010) or intra-personal equilibrium strategies employing self-punishments and self-rewards (in a non-cooperative game between the multiple selves with conflicting preferences; e.g., see Bernheim, Ray, and Yeltekin, 2015; Kodritsch, 2015). Both of the latter are effective only under a high degree of sophistication, however; by contrast, empirical evidence suggests that people instead tend to be rather naïve (Augenblick and Rabin, 2019; Breig et al., 2020; Le Yaouanq and Schwardmann, 2022). Indeed, we confirm this also in our sample, where students significantly over-estimate their future study effort.

Our simple planner-doer model moves the sophistication to a time consistent planner, as a way of reconciling naïveté with successful self-control. The doer has a present bias and may well be naïve, but she is sub-consciously regulated by the planner who sophisticatedly employs belief manipulation. The model thus also captures the notion of willpower in a very simple and intuitive manner, namely in terms of the mental costs of such belief manipulation (cf. Mischel, Shoda, and Rodriguez, 1989). We therefore consider its further development in terms of (other) distinct implications and broader applicability a promising avenue for future research.

We have empirically tested (and confirmed) the predictions of this simple model only against rational expectations. Despite our robustness and further supporting results, there are rather obvious data limitations, largely due to our study's natural setting and resulting constraints. Alternative explanations can surely be constructed, and indeed, we consider this paper's main contribution a conceptual/methodological one. However, what is remarkable about the mechanism proposed is its parsimony/simplicity in explaining the dynamic belief pattern of *both* groups, while also explaining related experimental findings, by combining two leading insights from behavioral economics: present bias and malleable beliefs. For instance, while the drop in return beliefs post-exam could be ex-post rationalization of low study effort, our model actually explains this part of the dynamic pattern of beliefs the same way as it explains why they are inflated pre-exam, on which ex-post rationalization is silent. The explanations differ then only in whether post-exam return beliefs are pessimistic or realistic, but our study's aim is not at all to dismiss the psychology of ex-post rationalization (or of emotions-management more broadly), but to develop and test a hypothesis about the psychology of self-control. Also, while we cannot rule out selection on unobservables (of course), we could also not think of any alternative theory that is based on equally well-established principles and would explain our findings upon allowing for some such selection.

Predicting and measuring belief dynamics is a novel area of research. An important aspect in it for policy purposes is how people select, process and recall information. Given the purpose of this research as well as its empirical setting, we employed a reduced form model of belief formation, without specifying the psychology of how exactly students arrive at and maintain distorted return beliefs – in particular, whether by motivated processing of information (Sharot, Korn, and Dolan, 2011; Lobeck, 2022; Möbius, Niederle, Niehaus, and Rosenblat, 2022) or by motivated memory (Chew, Huang, and Zhao, 2020; Zimmermann, 2020; Roy-Chowdhury, 2023). In the presence of self-control problems or other behavioral deviations from the standard model, the short- and long-run psychological and behavioral effects of information are far from obvious: for example, in our model, additional information may help by making the planner more informed (she may have a false prior expectation;

relatedly, see Breig et al., 2020, and also Drobner and Orhun, 2024) but it may also increase her costs of distorting beliefs and thereby make self-control harder (relatedly, see Huck et al., 2018, and also Roy-Chowdhury, 2023).

Altogether, we see this paper's contributions in (i) proposing as well as formalizing the idea that beliefs are an instrumental means of self-control, and highlighting the relevance of return beliefs; in (ii) designing a dynamic empirical approach repeatedly eliciting return beliefs via two fixed counterfactuals and exploiting natural variation over time as well as between groups in the importance of self-control for identification; and in (iii) presenting supportive evidence for the hypotheses derived from our model, using this approach on a convenience field sample. Thus, we hope to inspire future research that develops the proposed link and approach further, both theoretically and empirically.

References

Ahammer, A., M. Lackner, and J. Voigt (2019). Does confidence enhance performance? Causal evidence from the field. *Managerial and Decision Economics 40*(6), 704–717. 6

Akerlof, G. A. and W. T. Dickens (1982). The economic consequences of cognitive dissonance. *American Economic Review 72*(3), 307–319. 6

Ariely, D. and K. Wertenbroch (2002). Procrastination, deadlines, and performance: Self-control by precommitment. *Psychological Science 13*(3), 219–224. 2

Augenblick, N., M. Niederle, and C. Sprenger (2015). Working over time: Dynamic inconsistency in real effort tasks. *Quarterly Journal of Economics 130*(3), 1067–1115. 2, B-16

Augenblick, N. and M. Rabin (2019). An experiment on time preference and misprediction in unpleasant tasks. *Review of Economic Studies 86*(3), 941–975. 11, 27, B-16

Bachmann, R., G. Topa, and W. van der Klaauw (2023). *Handbook of Economic Expectations*. Academic Press. 2

Banerjee, R., N. D. Gupta, and M. C. Villeval (2020). Feedback spillovers across tasks, self-confidence and competitiveness. *Games and Economic Behavior 123*, 127–170. 7

Baumeister, R. F., K. D. Vohs, and D. M. Tice (2007). The strength model of self-control. *Current Directions in Psychological Science 16*(6), 351–355. 7

Bell, D. E. (1982). Regret in decision making under uncertainty. *Operations Research 30*(5), 961–981. 18

Bénabou, R. and J. Tirole (2002). Self-confidence and personal motivation. *Quarterly Journal of Economics 117*(3), 871–915. 2, 6, 10

Bénabou, R. and J. Tirole (2006). Belief in a just world and redistributive politics. *Quarterly Journal of Economics 121*(2), 699–746. 6

Benítez-Silva, H. and D. S. Dwyer (2005). The rationality of retirement expectations and the role of new information. *Review of Economics and Statistics 87*(3), 587–592. 8

Benjamin, D. J. (2019). Chapter 2 – errors in probabilistic reasoning and judgment biases. In B. D. Bernheim, S. DellaVigna, and D. Laibson (Eds.), *Handbook of Behavioral Economics – Foundations and Applications 2*, Volume 2 of *Handbook of Behavioral Economics: Applications and Foundations 1*, pp. 69–186. North-Holland. 8

Bernheim, B. D. (1990). How do the elderly form expectations? An analysis of responses to new information. In D. A. Wise (Ed.), *Issues in the Economics of Aging*, pp. 259–286. University of Chicago Press. 8

Bernheim, B. D., D. Ray, and S. Yeltekin (2015). Poverty and self-control. *Econometrica 83*(5), 1877–1911. 27

Bracha, A. and D. J. Brown (2012). Affective decision making: A theory of optimism bias. *Games and Economic Behavior 75*(1), 67–80. 6, 10

Breig, Z., M. Gibson, and J. G. Shrader (2020, March). Why do we procrastinate? Present bias and optimism. IZA DP No. 13060. 7, 27, 29

Brocas, I. and J. D. Carrillo (2000). The value of information when preferences are dynamically inconsistent. *European Economic Review 44*, 1104–1115. 6

Brocas, I. and J. D. Carrillo (2008). The brain as a hierarchical organization. *American Economic Review 98*(4), 1312–1346. 6

Brunnermeier, M. K., F. Papakonstantinou, and J. A. Parker (2017). Optimal time-inconsistent beliefs: Misplanning, procrastination, and commitment. *Management Science 63*(5), 1318–1340. 7

Brunnermeier, M. K. and J. A. Parker (2005). Optimal expectations. *American Economic Review 95*(4), 1092–1118. 6

Bryan, G., D. Karlan, and S. Nelson (2010). Commitment devices. *Annual Review of Economics 2*(1), 671–698. 27

Carrillo, J. D. and T. Mariotti (2000). Strategic ignorance as a self-disciplining device. *Review of Economic Studies 67*, 529–544. 6

Chen, S. and H. Schildberg-Hörisch (2019). Looking at the bright side: The motivational value of confidence. *European Economic Review 120*, 103302. 7

Cheung, S. L., A. Tymula, and X. Wang (2021, July). Quasi-hyperbolic present bias: A meta-analysis. IZA Discussion Paper No. 14625. 11, 26, B-16

Cheung, S. L., A. Tymula, and X. Wang (2022). Present bias for monetary and dietary rewards. *Experimental Economics 25*(4), 1202–1233. 2, B-16

Chew, S. H., W. Huang, and X. Zhao (2020). Motivated false memory. *Journal of Political Economy 128*(10), 3913–3939. 28

Clark, D., D. Gill, V. Prowse, and M. Rush (2020). Using goals to motivate college students: Theory and evidence from field experiments. *Review of Economics and Statistics 102*(4), 648–663. 2

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2 ed.). Lawrence Erlbaum Associates. 26, B-16, B-19, B-20

Compte, O. and A. Postlewaite (2004). Confidence-enhanced performance. *American Economic Review 94*(5), 1536–1556. 6

Cordes, C., J. Friedrichsen, and S. Schudy (2023, December). Motivated procrastination. Collaborative Research Center Transregio 190 Discussion Paper No. 471. 7

Coutts, A. (2019). Testing models of belief bias: An experiment. *Games and Economic Behavior 113*, 549–565. 10

DellaVigna, S. (2009). Psychology and economics: Evidence from the field. *Journal of Economic Literature 47*(2), 315–372. 2

Drobner, C. and A. Y. Orhun (2024, August). Misguided effort. Mimeo. 8, 29

Engelmann, J. B., M. Lebreton, N. A. Salem-Garcia, P. Schwardmann, and J. J. van der Weele (2024). Anticipatory anxiety and wishful thinking. *American Economic Review 114*(4), 926–960. 7

Falk, A., A. Becker, T. Dohmen, D. Huffman, and U. Sunde (2016, January). The preference survey module: A validated instrument for measuring risk, time, and social preferences. IZA DP No. 9674. C-2

Fudenberg, D. and D. K. Levine (2006). A dual-self model of impulse control. *American Economic Review 5*(5), 1449–1476. 6

Gilovich, T., M. Kerr, and V. H. Medvec (1993). Effect of temporal perspective on subjective confidence. *Journal of Personality and Social Psychology 64*(4), 552–560. 7

Gollier, C. and A. Muermann (2010). Optimal choice and beliefs with ex ante savoring and ex post disappointment. *Management Science 56*(8), 1272–1284. 6

Grimes, P. W. (2002). The overconfident principles of economics student: An examination of a metacognitive skill. *Journal of Economic Education 33*(1), 15–30. 7

Hagenbach, J. and F. Koessler (2022). Selective memory of a psychological agent. *European Economic Review 142*, 104012. 6

Harrison, G. W., M. I. Lau, and H. I. Yoo (2023, May). Constant discounting, temporal instability and dynamic inconsistency in Denmark: A longitudinal field experiment. Mimeo. 4, 22

Himmler, O., R. Jäckle, and P. Weinschenk (2019). Soft commitments, reminders, and academic performance. *American Economic Journal: Applied Economics 11*(2), 114–42. 7

Hoenig, J. M. and D. M. Heisey (2001). The abuse of power. *The American Statistician 55*(1), 19–24. 26, B-16

Huck, S., N. Szech, and L. M. Wenner (2018, March). More effort with less pay: On information avoidance, optimistic beliefs and performance. Mimeo. 7, 8, 29

Kamenica, E. and M. Gentzkow (2011). Bayesian persuasion. *American Economic Review 101*(6), 2590–2615. 6

Kim, J. (2020, May). Discounting, dynamic consistency, and cooperation in an infinitely repeated game experiment. Mimeo. 4, 22

Kodritsch, S. (2015, February). A note on the welfare of a sophisticated time-inconsistent decision-maker. WZB Discussion Paper SP II 2015-201. 27

Köszegi, B. (2006). Ego utility, overconfidence, and task choice. *Journal of the European Economic Association 4*(4), 673–707. 6

Laibson, D. (1997). Golden eggs and hyperbolic discounting. *Quarterly Journal of Economics 112*(2), 443–478. 6

Le Yaouanq, Y. and P. Schwardmann (2022). Learning about one's self. *Journal of the European Economic Association 20*(5), 1791–1828. 27

Levitt, S. D., J. A. List, S. Neckermann, and S. Sadoff (2016). The behavioralist goes to school: Leveraging behavioral economics to improve educational performance. *American Economic Journal: Economic Policy 8*(4), 183–219. 7

Lobeck, M. (2022, September). Motivating beliefs in a just world. Mimeo. 8, 28

Loewenstein, G. (1987). Anticipation and the valuation of delayed consumption. *Economic Journal 97*(387), 666–684. 6

Loewenstein, G. and D. Prelec (1992). Anomalies in intertemporal choice: Evidence and an interpretation. *Quarterly Journal of Economics 107*(2), 573–597. A-1

Loomes, G. and R. Sugden (1982). Regret theory: An alternative theory of rational choice under uncertainty. *Economic Journal 92*(368), 805–824. 18

Ma, M. (2020, June). Motivated overconfidence: Theory and evidence. Mimeo. 5, 7

Mariotti, T., N. Schweizer, N. Szech, and J. von Wangenheim (2023). Information nudges and self-control. *Management Science 69*(4), 2182–2197. 6

Mischel, W., Y. Shoda, and M. L. Rodriguez (1989). Delay of gratification in children. *Science 244*(4907), 933–938. 27

Möbius, M. M., M. Niederle, P. Niehaus, and T. S. Rosenblat (2022). Managing self-confidence: Theory and experimental evidence. *Management Science 68*(11), 7793–7817. 28

Muraven, M. and R. F. Baumeister (2000). Self-regulation and depletion of limited resources: Does self-control resemble a muscle? *Psychological Bulletin 126*(2), 247–259. 7

Nisan, M. (1972). Dimension of time in relation to choice behavior and achievement orientation. *Journal of Personality and Social Psychology 21*(2), 175–182. 7

O'Donoghue, T. and M. Rabin (1999). Doing it now or later. *American Economic Review 89*(1), 103–124. 6, 11

Rosenqvist, O. and O. N. Skans (2015). Confidence enhanced performance? – The causal effects of success on future performance in professional golf tournaments. *Journal of Economic Behavior & Organization 117*, 281–295. 6

Roy-Chowdhury, V. (2023, September). Biased recall and the dynamics of beliefs: Evidence from schools. Mimeo. 28, 29

Schwardmann, P., E. Tripodi, and J. J. van der Weele (2022). Self-persuasion: Evidence from field experiments at international debating competitions. *American Economic Review 112*(4), 1118–1146. 8

Schwardmann, P. and J. van der Weele (2019). Deception and self-deception. *Nature Human Behavior 3*, 1055–1061. 8

Schweighofer-Kodritsch, S. (2018). Time preferences and bargaining. *Econometrica 86*(1), 173–217. A-1

Sharot, T., C. W. Korn, and R. J. Dolan (2011). How unrealistic optimism is maintained in the face of reality. *Nature Neuroscience 14*(11), 1475–1479. 8, 28

Shepperd, J. A., J. A. Ouellette, and J. K. Fernandez (1996). Abandoning unrealistic optimism: Performance estimates and the temporal proximity of self-relevant feedback. *Journal of Personality and Social Psychology 70*(4), 844–855. 7

Shepperd, J. A., E. A. Waters, N. D. Weinstein, and W. M. P. Klein (2015). A primer on unrealistic optimism. *Current Directions in Psychological Science 24*(3), 232–237. 2

Sweeny, K. and Z. Krizan (2013). Sobering up: A quantitative review of temporal declines in expectations. *Psychological Bulletin 139*(3), 702–724. 7

Thaler, R. H. and H. M. Shefrin (1981). An economic theory of self-control. *Journal of Political Economy 89*(2), 392–406. 6

Tversky, A. and D. Kahneman (1974). Judgment under uncertainty: Heuristics and biases – Biases in judgments reveal some heuristics of thinking under uncertainty. *Science 185*(4157), 1124–1131. 18

Weinstein, N. D. (1980). Unrealistic optimism about future life events. *Journal of Personality and Social Psychology 39*(5), 806–820. 2

Windschitl, P. D. and J. O. Stuart (2015). *Optimism Biases*, Chapter 15, pp. 431–455. John Wiley & Sons, Ltd. 7

Zimmermann, F. (2020). The dynamics of motivated beliefs. *American Economic Review 110*(2), 337–61. 28

Appendix

A Theoretical Robustness

We sketch here three extensions of our model simultaneously: Generalized present bias and knowledge decay, on the part of Sue as the doer, and some present bias also on the part of Sue as the planner.

First, we consider the following generalization beyond (β, δ)-discounting for Sue as the doer: Denoting by $d(t)$ the discount factor on any utility delayed by $t \in \{0, 1, 2\}$ periods, we assume $d(0) = 1 > d(1) = \beta_1\delta > d(2) = \beta_1\beta_2\delta^2$ with $(\beta_1, \delta) \in (0, 1)^2$ and $\beta_2 \in (\beta_1, 1]$. While the model in Section 3 corresponds to the special case where $\beta_2 = 1$ (and $\beta_1 = \beta$), this extension allows for general increasing patience, as in hyperbolic discounting (e.g., Loewenstein and Prelec, 1992).[31]

Second, let Sue's grade now be determined as

$$\tilde{g}(e_1, e_2, R) = R \cdot (\phi e_1 + e_2), \tag{4}$$

where $\phi \in (0, 1]$ measures the extent to which early study effort depreciates until the exam due to knowledge decay. The model in Section 3 corresponds to the special case where $\phi = 1$, see (1) there.

It is straightforward to solve for Sue's (doer's) effort in any period $t \in \{1, 2\}$, for given return beliefs, also under the above extension(s). Recalling from Section 3 that $e_t\left(\hat{R}_t\right) = \kappa_t \hat{R}_t$ for $\kappa_1 = \beta\delta^2$ and $\kappa_2 = \beta\delta$, we now have the generalization

$$\tilde{e}_t\left(\hat{R}_t\right) = \tilde{\kappa}_t \hat{R}_t \text{ for } \tilde{\kappa}_1 = \phi\beta_1\beta_2\delta^2,\ \tilde{\kappa}_2 = \beta_1\delta.$$

The only difference worth noting here is that when $\phi < 1$, given constant beliefs, Sue's study effort would then increase over time toward the exam even with perfect long-run patience, due to early study effort's "depreciation" ($\tilde{\kappa}_2 > \tilde{\kappa}_1$ then holds true even for $\beta_2 = 1$ and

[31] Patience about the first period of delay is governed by $d(1)/d(0) = \beta_1\delta$ whereas that about the second is governed by $d(2)/d(1) = \beta_2\delta$, so that $d(2)/d(1) > d(1)/d(0)$ holds true if and only if $\beta_2 > \beta_1$. See also the general notion of "weak present bias" in Schweighofer-Kodritsch (2018, pp. 191–192).

$\delta \to 1$).

For the third extension, we now allow the planner to be present-biased as well, just like the doer, but to a lesser extent. Specifically, we assume discounting by the planner according to function $d(t)$ such that $d(0) = 1 > d(1) = \tilde{\beta}_1\delta > d(2) = \tilde{\beta}_1\tilde{\beta}_2\delta^2$ with $\tilde{\beta}_1 \in (\beta_1, 1]$ and $\tilde{\beta}_2 \in [\beta_2, 1]$, while maintaining the key assumption of Sue's sophistication as planner.[32] The model in Section 3 corresponds to the special case where $\tilde{\beta}_1 = \tilde{\beta}_2 = 1$. Any present bias on the part of the planner means that the early and late planner's choice of (late) beliefs, determining late effort, diverge. The planner now non-cooperatively chooses just $\hat{R}_t$ in every period $t \in \{1, 2, 3\}$, taking as given her other belief choices (and resulting effort), rather than choosing all beliefs once and for all at the outset; for any t, she maximizes the corresponding $\tilde{V}_t$ as given below:

$$
\begin{aligned}
\tilde{V}_1\left(\hat{R}_1, \hat{R}_2, \hat{R}_3|\hat{R}_0\right) =& -\frac{\tilde{e}_1\left(\hat{R}_1\right)^2}{2} - \tilde{\beta}_1\delta\frac{\tilde{e}_2\left(\hat{R}_2\right)^2}{2} + \tilde{\beta}_1\tilde{\beta}_2\delta^2\hat{R}_0 \cdot \left(\phi\tilde{e}_1\left(\hat{R}_1\right) + \tilde{e}_2\left(\hat{R}_2\right)\right) \\
& -\gamma\frac{1}{2}\left(\hat{R}_1 - \hat{R}_0\right)^2 - \tilde{\beta}_1\delta\gamma\frac{1}{2}\left(\hat{R}_2 - \hat{R}_0\right)^2 - \tilde{\beta}_1\tilde{\beta}_2\delta^2\gamma\frac{1}{2}\left(\hat{R}_3 - \hat{R}_0\right)^2; \\
\tilde{V}_2\left(\hat{R}_1, \hat{R}_2, \hat{R}_3|\hat{R}_0\right) =& -\frac{\tilde{e}_2\left(\hat{R}_2\right)^2}{2} + \tilde{\beta}_1\delta\hat{R}_0 \cdot \left(\phi\tilde{e}_1\left(\hat{R}_1\right) + \tilde{e}_2\left(\hat{R}_2\right)\right) \\
& -\gamma\frac{1}{2}\left(\hat{R}_2 - \hat{R}_0\right)^2 - \tilde{\beta}_1\delta\gamma\frac{1}{2}\left(\hat{R}_3 - \hat{R}_0\right)^2; \\
\tilde{V}_3\left(\hat{R}_1, \hat{R}_2, \hat{R}_3|\hat{R}_0\right) =& \hat{R}_0 \cdot \left(\phi\tilde{e}_1\left(\hat{R}_1\right) + \tilde{e}_2\left(\hat{R}_2\right)\right) - \gamma\frac{1}{2}\left(\hat{R}_3 - \hat{R}_0\right)^2.
\end{aligned}
$$

In the resulting equilibrium, we clearly still have $\tilde{\hat{R}}_3^* = \hat{R}_0$, as any period-3 distortion away from $\hat{R}_0$ only has costs. Concerning return beliefs in periods $t \in \{1, 2\}$, that period's self of the planner chooses these as

$$
\begin{aligned}
\tilde{\hat{R}}_2^* &= \hat{R}_0 \cdot \left(1 + \frac{\tilde{\beta}_1 - \beta_1}{\beta_1} \cdot \frac{\tilde{\kappa}_2^2}{\tilde{\kappa}_2^2 + \gamma}\right), \text{ and} \\
\tilde{\hat{R}}_1^* &= \hat{R}_0 \cdot \left(1 + \frac{\tilde{\beta}_1\tilde{\beta}_2 - \beta_1\beta_2}{\beta_1\beta_2} \cdot \frac{\tilde{\kappa}_1^2}{\tilde{\kappa}_1^2 + \gamma}\right),
\end{aligned}
$$

which is a straightforward generalization of Section 3's (3). Consequently, given our basic assumptions $\tilde{\beta}_1 > \beta_1$ and $\tilde{\beta}_2 \geq \beta_2$ (implying also $\tilde{\beta}_1\tilde{\beta}_2 > \beta_1\beta_2$), we obtain the same main

[32] Whether the doer is naïve or sophisticated about her present bias does not affect our main prediction also under the extension(s) considered here.

prediction concerning the dynamics of return beliefs. Besides establishing robustness, an additional observation from this extension is that if the planner's preferences regarding study effort for the exam coincide with those of the doer, there is generally no (motive for) belief distortion here, i.e., we have $\tilde{\hat{R}}_t^* = \hat{R}_0$ for all periods t not only when both doer and planner have no present bias ($\tilde{\beta}_2 = \beta_2 = 1$ and $\tilde{\beta}_1 = \beta_1 \to 1$) but more generally whenever they have the same discounting ($\tilde{\beta}_t = \beta_t$ for both $t \in \{1, 2\}$).

In particular, knowledge decay as such then entails no belief distortion. While we have assumed full "sophistication" about the decay parameter ϕ, note that if the doer were naïve about it—i.e., act as if $\phi = 1$ while really $\phi < 1$—she would *over*-provide study effort early on, in period 1. In the absence of other frictions—i.e., when doer and planner discount utility the same way—the planner would then distort return beliefs in period 1 *down*-ward and select accurate beliefs afterwards. The predicted dynamic pattern of return beliefs would then be that beliefs move up before the exam as well, but in contrast to our theory based on present bias and also to what we observe, they would then remain constant.

At the same time, if the doer suffers from present bias, there is also no discontinuity of beliefs at the point where the planner has none, which we assumed in Section 3 (here, more generally, there is no discontinuity at $\tilde{\beta}_1 = \tilde{\beta}_2 = 1$). The terminology "drops sharply" in our main prediction refers to the fact that return beliefs are already upward-distorted in period 1, then rise further in period 2, and subsequently—after having taken the exam—fall below their period-1 level. The above shows when this is qualitatively true, but as $\tilde{\beta}_t \to \beta_t$ for both $t \in \{1, 2\}$, all of these differences converge to zero.

B Additional Analyses, Tables and Figures

B.1 Summary Statistics and Selection Analysis

Table 4 compares the students from our main restricted sample of 84 observations with all others who participated in wave 1. The latter group consists of students who participated in wave 1 but are not included in our main sample (N=130). This group includes those whom we did not lose due to attrition (N=96) and those who do not meet our inclusion criteria of being unambiguously assigned to a treatment group (N=34), as explained in Section 4.1. There are no significant differences between these two groups in terms of gender, study program, first-time takers of the course, number of semesters of study, or age (all except semester and year of birth are either one or zero).

Table 4: Summary Statistics – Main Sample versus Others

	main sample		others in wave 1			
	mean	sd	mean	sd	diff	t-stat
male	0.43	0.50	0.51	0.50	0.08	(1.13)
economics	0.45	0.50	0.34	0.48	-0.11	(-1.68)
business	0.44	0.50	0.52	0.50	0.07	(1.07)
other program	0.11	0.31	0.15	0.35	0.04	(0.82)
first time micro	0.79	0.41	0.75	0.43	-0.03	(-0.54)
semester	1.90	1.48	2.22	2.11	0.31	(1.18)
year of birth	1993.99	3.18	1993.14	4.13	-0.85	(-1.60)
Observations	84		130		214	

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

Table 5 (a) shows that there are also no significant differences in these background characteristics between the treatment groups in our final sample, that is, between those who write the exam on date 1 (group 1) and those who take the exam on date 2 (group 2).The table's second panel (b) compares the two groups by exam score (percentage of points achieved), expected exam score at wave 1 as well as at the respective pre-exam wave, expected effort pre-exam minus reported effort post-exam (effort "overprediction"; pre-exam is wave 3 for group 1 and wave 5 for group 2, post-exam is the wave immediately thereafter in each case), and initially (as of wave 1) expected total number of course exams that would be writ-

Table 5: Summary Statistics by Treatment Group

(a)

	group 1		group 2			
	mean	sd	mean	sd	diff	t-stat
male	0.45	0.50	0.38	0.49	-0.07	(-0.63)
economics	0.43	0.50	0.50	0.51	0.07	(0.54)
business	0.45	0.50	0.42	0.50	-0.03	(-0.27)
other program	0.12	0.32	0.08	0.28	-0.03	(-0.47)
first time micro	0.77	0.43	0.83	0.38	0.07	(0.70)
semester	1.98	1.47	1.71	1.52	-0.27	(-0.76)
year of birth	1993.75	3.11	1994.58	3.36	0.83	(1.09)
Observations	60		24		84	

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

(b)

	group 1		group 2			
	mean	sd	mean	sd	diff	t-stat
achieved points (in %)	58.94	18.55	57.63	18.51	-1.31	(-0.29)
exp. result (wave 1)	72.03	11.53	69.96	11.56	-2.08	(-0.73)
exp. result (pre-exam)	67.12	14.63	67.22	14.01	0.10	(0.03)
effort overprediction	4.56	17.35	5.34	21.05	0.79	(0.17)
exam overprediction	0.24	1.28	0.30	1.15	0.07	(0.22)
Observations	59		23		82	

Removed one observation per group with unrealistic effort reports (> 200 hours). Exam overprediction corresponds to the difference in the number of planned (wave 1) and written (wave 6) exams. Effort overprediction is defined as the difference between expected effort in the pre-exam and post-exam wave.

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

ten that semester minus ultimately reported number (as of wave 6). Again, groups appear indistinguishable, including even in terms of actual exam scores.[33]

We use Probit regression analysis to further examine the selection into the first or second exam date based on exam scores and proxies for self-control (see Table 6). Exam scores represent a joint measure of ability and effort; therefore, they may capture unobservable characteristics such as skills, intrinsic work motivation, and self-control. It is reassuring that students in the first and second exam groups do not systematically differ in exam grades, picking up these unobservable factors.

[33] Grading exams was done by class tutors who were not involved in this research.

In addition, individuals' expectations about future exam performance do not significantly predict selection into the exam groups. Nor do the differences between ex-ante expected and ex-post reported effort or between the ex-ante expected and ex-post reported total number of course exams written—both of which may serve as proxies for self-control—predict students' choice of exam date.[34] There is also no joint significance of these measures, as inferred from the Likelihood-ratio test. Thus, we do not find evidence of self-selection into exam groups based on the observables we have.[35]

Table 6: Probit Regressions

	Probit (Dep. Var.: Group 2)				Marginal Effects			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Exp. Result W1	-0.019		-0.018	-0.026	-0.006		-0.006	-0.008
	(0.016)		(0.016)	(0.018)	(0.005)		(0.005)	(0.005)
Exp. Res. Pre-Exam	0.010		0.010	0.010	0.004		0.003	0.003
	(0.013)		(0.013)	(0.013)	(0.004)		(0.004)	(0.004)
Points	0.000		-0.001	-0.004	0.000		-0.000	-0.001
	(0.009)		(0.009)	(0.009)	(0.003)		(0.003)	(0.003)
Effort Overpred.		-0.003	-0.003	-0.003		-0.001	-0.001	-0.001
		(0.005)	(0.005)	(0.005)		(0.002)	(0.002)	(0.002)
Exam Overpred.		0.028	0.012	0.061		0.009	0.004	0.020
		(0.117)	(0.119)	(0.132)		(0.040)	(0.040)	(0.043)
Const.	0.053	-0.572***	0.040	1.665				
	(0.895)	(0.149)	(0.911)	(2.229)				
Controls	no	no	no	yes	no	no	no	yes
Obs.	84	84	84	84	84	84	84	84
R^2 / Pseudo	0.013	0.005	0.015	0.046				
LLR (p-value)	0.7378	0.7889	0.9071	0.9488				

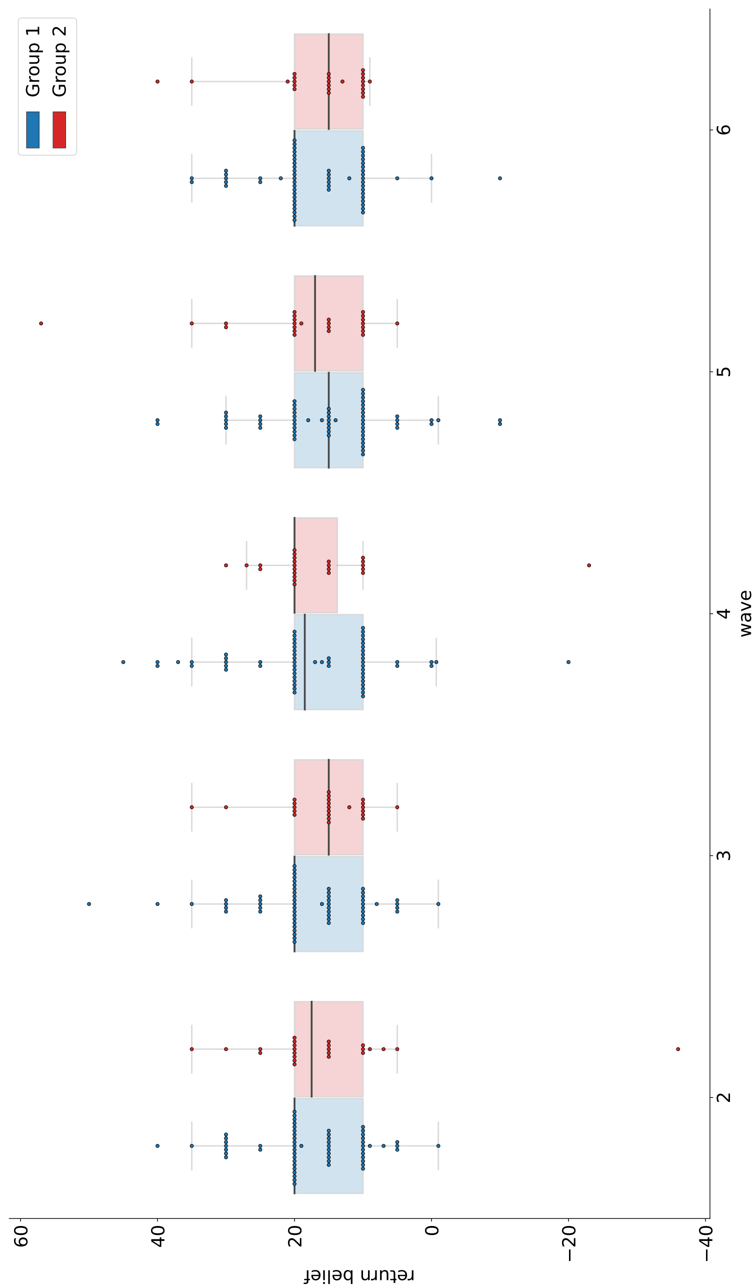
Robust S.E. in parentheses.
Marginal Effects are calculated as the average of the marginal effects at each observation.
*** p<0.01, ** p<0.05, * p<0.1

[34] The Probit results are robust to excluding the two observations with unrealistic effort reports.
[35] Of course, we cannot rule out selection into exam grades based on unobservables.

B.2 Main Hypotheses

Figure 4 shows box plots of return beliefs over exam groups and waves 2-6. The upper hinge of each box represents the 75th percentile, the lower hinge the 25th percentile, and the line the median.

Figure 4: Return Beliefs Over Time



Here we show the full regression tables, including various control variables, underlying Tables 1 and 2 in the main text, in Tables 7 and 8, respectively.

Table 7: Return Beliefs

	Date 1		Date 2		Pooled	
	(1)	(2)	(3)	(4)	(5)	(6)
Exam	2.342	1.772	3.633	3.958	2.987**	2.988**
	(1.753)	(1.861)	(2.569)	(2.680)	(1.265)	(1.296)
Date 2					-0.113	-0.113
					(1.265)	(1.296)
Female		3.639		-0.537		1.438
		(2.517)		(2.280)		(1.877)
Semester		1.366*		0.932		1.155
		(0.765)		(1.132)		(0.870)
First Micro		2.956		1.829		2.424
		(2.924)		(2.885)		(2.626)
Econ		2.800		6.055		4.562
		(3.292)		(3.927)		(3.198)
Business		2.866		5.373*		4.218
		(3.134)		(2.958)		(2.581)
Age		-0.244		0.425		0.073
		(0.356)		(0.453)		(0.374)
Exp. Result		0.168		0.013		0.083
		(0.123)		(0.105)		(0.083)
Exp. Effort		0.013		0.021		0.017
		(0.030)		(0.033)		(0.027)
Const.	15.708***	6.496	14.950***	6.953	15.247***	6.406
	(1.332)	(4.392)	(1.292)	(4.648)	(1.235)	(4.001)
Obs.	84	84	84	84	168	168
R^2 / Pseudo	0.017	0.142	0.026	0.071	0.027	0.075
F-statistic	1.784	1.465	2.000	1.303	4.109**	2.489**

Age, Expected Effort (wave 1) and Expected Result (wave 1) are centered around their means.
S.E. in brackets, for Date 1 and 2 robust (HC1), for pooled OLS clustered at ID level.
*** p<0.01, ** p<0.05, * p<0.1

Table 8: Return Beliefs – After vs. Before Exam

	Date 1		Date 2		Pooled	
	(1)	(2)	(3)	(4)	(5)	(6)
Exam	-1.603	-1.448	-4.283*	-4.090	-2.943**	-2.943**
	(2.600)	(2.559)	(2.445)	(2.493)	(1.290)	(1.322)
Date 2					-0.182	-0.182
					(1.290)	(1.322)
Female		0.596		3.858*		2.364
		(2.877)		(2.181)		(1.724)
Semester		-0.644		-0.182		-0.420
		(0.906)		(0.931)		(0.789)
First Micro		1.946		2.427		2.148
		(4.027)		(2.671)		(2.361)
Econ		-1.516		-2.197		-2.019
		(3.483)		(3.290)		(2.770)
Business		-0.474		-2.646		-1.679
		(3.373)		(2.322)		(2.109)
Age		0.014		-0.009		0.024
		(0.340)		(0.265)		(0.216)
Exp. Result		-0.112		-0.016		-0.054
		(0.131)		(0.091)		(0.069)
Exp. Effort		0.077		0.030		0.054*
		(0.053)		(0.032)		(0.029)
Const.	0.708	1.276	1.867*	0.802	1.665	1.596
	(2.194)	(5.620)	(0.986)	(4.433)	(1.297)	(4.211)
Obs.	84	84	84	84	168	168
R^2 / Pseudo	0.005	0.058	0.048	0.105	0.021	0.067
F-statistic	0.380	0.696	3.068*	0.814	2.647*	1.360

Age, Expected Effort (wave 1) and Expected Result (wave 1) are centered around their means.
S.E. in brackets, for individual dates robust (HC1), for pooled OLS clustered at ID level.
*** p<0.01, ** p<0.05, * p<0.1

B.3 Robustness of Main Results

B.3.1 Exam Information

Table 9 shows the full regression results for the robustness check concerning exam information. The table summarizes pooled regression results for our two main hypotheses, restricting the sample to students in the respective control group, who would/did not write the exam at the given date, but subsequently reported knowing its content (recall that the exam became publicly available right after its date). Specifically, at wave 4, 13 out of the 24 students in group 2 reported knowing group 1's exam (taking place between waves 3 and 4). Conversely, at wave 6, 16 out of the 60 students in group 1 reported knowing group 2's exam (taking place between waves 5 and 6). By pooling the data – comparing 60 students from group 1 with 13 from group 2 around exam date 1, and 16 from group 1 with 24 from group 2 around exam date 2 – we arrive at a total sample size of 113 observations for the pooled regressions.

Table 9: Regressions for Information Sample

	Return Beliefs		After vs. Before Exam	
	(1)	(2)	(3)	(4)
Exam	2.788*	2.642	-4.276**	-4.502**
	(1.579)	(1.755)	(1.805)	(1.874)
Date 2	-0.262	-0.172	-1.093	-1.228
	(1.717)	(1.822)	(2.019)	(2.101)
Female		2.113		2.172
		(2.315)		(2.325)
Semester		0.778		0.034
		(0.789)		(0.822)
First Micro		2.022		2.187
		(3.174)		(2.920)
Econ		4.008		-0.550
		(3.086)		(2.586)
Business		3.713		-0.948
		(2.973)		(2.593)
Age		-0.021		-0.033
		(0.399)		(0.254)
Exp. Result		0.151		-0.080
		(0.108)		(0.106)
Exp. Effort		0.017		0.068*
		(0.032)		(0.041)
Const.	15.489***	7.743*	3.258**	1.701
	(1.646)	(4.520)	(1.645)	(4.575)
Obs.	113	113	113	113
R^2 / Pseudo	0.020	0.096	0.034	0.084
F-statistic	1.992	1.763*	2.852*	1.421

Age, Expected Effort (wave 1) and Expected Result (wave 1)
are centered around their means.
S.E. in brackets are clustered at ID level.
*** p<0.01, ** p<0.05, * p<0.1

B.3.2 Cross-Validation with Additional Course

Tables 10 and 11 present the full regression results regarding our two main hypotheses for the data on the introductory mathematics course. Table 12 presents the full regression results linking belief distortion in microeconomics and in mathematics, extending Table 3 from the main body of the paper. The sample size for the mathematics course is comparatively smaller because not all students enrolled in the microeconomics course are also registered for mathematics. However, in the comparisons focused purely on mathematics, some students who are excluded from the microeconomics sample are included, leading to a larger sample size (63 students) compared to the analysis linking belief distortions in microeconomics and mathematics, for which we only consider the subset of students that meet the microeconomics inclusion criteria (52 students). Due to the particular smaller subsample, we had to exclude even some controls that are not course-specific (all 52 students in the sample are in economics or business, 51 of them take the microeconomics exam for the first time, and 50 of them are in their first semester of study).

Table 10: Return Beliefs – Math

	Date 1		Date 2		Pooled	
	(1)	(2)	(3)	(4)	(5)	(6)
Exam	1.853	2.294	1.193	0.334	1.523	1.523
	(2.321)	(2.119)	(2.296)	(1.927)	(1.153)	(1.182)
Date 2					0.477	0.477
					(1.153)	(1.182)
Female		3.799		1.686		2.647*
		(2.404)		(2.005)		(1.591)
Semester		2.797**		1.027		1.833***
		(1.348)		(1.143)		(0.669)
Econ		-1.035		-0.347		-0.776
		(2.124)		(2.092)		(1.724)
Age		0.099		0.084		0.081
		(0.547)		(0.361)		(0.370)
Exp. Result		0.105		-0.102		0.007
		(0.102)		(0.097)		(0.072)
Exp. Effort		0.122**		0.158***		0.141***
		(0.050)		(0.043)		(0.034)
Const.	14.579***	9.650***	15.386***	14.188***	14.810***	11.914***
	(1.717)	(2.985)	(1.640)	(2.224)	(1.391)	(1.943)
Obs.	63	63	63	63	126	126
R^2 / Pseudo	0.008	0.171	0.003	0.193	0.005	0.161
F-statistic	0.637	2.727**	0.270	5.394***	0.910	5.710***

Age, Expected Effort (wave 1) and Expected Result (wave 1) are centered around their means.
Due to the different and smaller sample, some controls had to be excluded due to lack of variation, i.e., only three (out of 63) students do not take the mathematics course for the first time and only one student does not major in economics or business.
S.E. in brackets, for individual dates robust (HC1), for pooled OLS clustered at ID level.
*** p<0.01, ** p<0.05, * p<0.1

Table 11: Return Beliefs – After vs. Before Exam – Math

	Date 1		Date 2		Pooled	
	(1)	(2)	(3)	(4)	(5)	(6)
Exam	-5.316**	-5.459**	-1.279	-1.581	-3.297**	-3.297*
	(2.281)	(2.302)	(2.055)	(2.055)	(1.672)	(1.714)
Date 2					-0.229	-0.229
					(1.672)	(1.714)
Female		-5.223**		-1.311		-3.078**
		(2.318)		(2.100)		(1.369)
Semester		-2.369		2.861*		0.401
		(1.575)		(1.592)		(1.095)
Econ		0.501		-2.017		-0.589
		(2.327)		(1.958)		(1.489)
Age		0.255		0.207		0.252
		(0.468)		(0.361)		(0.238)
Exp. Result		-0.161		-0.096		-0.140***
		(0.108)		(0.080)		(0.050)
Exp. Effort		-0.011		-0.093**		-0.053*
		(0.043)		(0.039)		(0.028)
Const.	5.316***	11.008***	3.068**	1.907	3.906**	5.792***
	(1.761)	(3.213)	(1.266)	(2.648)	(1.538)	(2.105)
Obs.	63	63	63	63	126	126
R^2 / Pseudo	0.069	0.169	0.006	0.115	0.035	0.093
F-statistic	5.430**	2.608**	0.387	1.562	2.028	2.514**

Age, Expected Effort (wave 1) and Expected Result (wave 1) are centered around their means.
Due to the different and smaller sample, some controls had to be excluded due to lack of variation, i.e., only three (out of 63) students do not take the mathematics course for the first time and only one student does not major in economics or business.
S.E. in brackets, for individual dates robust (HC1), for pooled OLS clustered at ID level.
*** p<0.01, ** p<0.05, * p<0.1

Table 12: Link Between Micro and Math

	Pre-Exam Belief Change		Post-Exam Belief Change		Effort Overprediction	
	(1)	(2)	(3)	(4)	(5)	(6)
Bel. Change Math (Pre)	0.710**	0.743**				
	(0.301)	(0.354)				
Bel. Change Math (Post)			0.522***	0.577***		
			(0.146)	(0.131)		
Eff. Overpred. Math					0.622**	0.609**
					(0.251)	(0.260)
Female		-1.811		-4.177		-1.622
		(2.534)		(3.085)		(4.014)
Econ		1.814		-0.792		3.890
		(3.978)		(2.904)		(3.891)
Age		-0.113		0.146		-0.746
		(0.333)		(0.492)		(0.631)
Const.	1.938	1.555	1.712	4.167*	3.428	0.846
	(1.772)	(1.763)	(1.458)	(2.342)	(2.156)	(2.773)
Obs.	52	52	52	52	52	52
R^2 / Pseudo	0.297	0.310	0.188	0.219	0.271	0.302
F-statistic	5.580**	4.369***	12.832***	5.786***	6.143**	2.127*

Age is centered around its mean.
Due to the different and smaller sample, some controls had to be excluded due to lack of variation, i.e., only two (out of 52) students are not in the first semester, 51 out of 52 students take microeconomics for the first time and all 52 students major either in economics or business.
Robust S.E. in parentheses
*** p<0.01, ** p<0.05, * p<0.1

B.4 Further Results

B.4.1 Beliefs, Effort, Present Bias, and Exam Performance

In Table 13, columns 1–3, we report regressions of expected effort (resp., ex-post reported effort) during the 14 days before the exam on return beliefs, where we pool all five waves for which we have data. We find our measure of expected returns economically validated: Students who believe the returns to studying to be higher, according to the hypothetical scenario of our survey question, also expect to study more (resp., report to have studied more).

The first column's regression includes only a group dummy indicating when the student wrote the exam and our standard set of controls, the second column's regression adds a variable indicating self-reported importance of the grade for one's career (ranging from 1 for unimportant to 6 for very important), and the third column's regression adds wave dummies (omitting wave 2).[36] In column 4, we drop all observations from waves 2 to 5: The positive statistical relation between effort and expected returns holds when using ex-post exam measures only (wave 6); though it is not statistically significant with this smaller set of observations, this turns out to be due to including control variables;[37] in a regression without controls, it remains statistically significant (coeff. 0.416, S.E. 0.198, $p<0.035$).

Table 14 shows full regression results for the correlations between effort overprediction, as a measure of present bias (assuming similar naïveté), pre-exam return belief inflation and post-exam return belief deflation. It shows these are all positively correlated, while only the positive correlation between the latter two measures of belief distortion is statistically significant, and it is so at the 1%-level.

[36] We exclude two students' observations here, which were extreme outliers, as explained in Footnote 20. Including them has essentially no effect, except for marginally increasing the standard errors on the estimated return coefficient.

[37] The only significant control variable here is gender: Women expect to study and report to have studied almost 8 hours more than men (coeff. 7.78, S.E. 3.851, $p<0.05$).

Table 13: Return Belief Measure and Behavior

Dep. Var.:	*Expected or Ex-Post Reported Effort*			
	Ordinary Least Squares			
	Waves 2-6 (pooled)			Wave 6 only
Exp. Return	0.354***	0.343***	0.337**	0.268
	(0.131)	(0.132)	(0.132)	(0.209)
Group 2	3.075	2.832	2.824	0.949
	(2.849)	(2.793)	(2.808)	(3.741)
Importance		0.603	0.619	1.507
		(1.098)	(1.117)	(1.509)
Wave 3			3.999***	
			(1.372)	
Wave 4			0.858	
			(2.218)	
Wave 5			0.719	
			(2.123)	
Wave 6			-1.910	
			(2.010)	
Female	6.711**	6.736**	6.754**	7.779**
	(3.163)	(3.163)	(3.179)	(3.851)
Semester	1.189	1.316	1.324	1.742
	(1.550)	(1.618)	(1.627)	(2.237)
First Micro	-0.355	-0.388	-0.369	4.721
	(6.037)	(6.075)	(6.107)	(7.065)
Age	0.456	0.521	0.523	0.723
	(0.664)	(0.666)	(0.669)	(0.848)
Const.	28.232***	26.311***	25.595**	17.367
	(8.278)	(9.539)	(9.972)	(11.086)
Obs.	410	410	410	82
R^2	0.114	0.116	0.128	0.127
F-statistic	2.927**	2.468**	3.600***	1.724

Age is centered around its mean.

Removed one observation per group with unrealistic effort reports (> 200 hours).

Robust/ID-clustered S.E. in brackets. *** p<0.01, ** p<0.05, * p<0.1

Table 14: Return Belief Changes and Effort Overprediction

Dep. Variable:	Post-Exam Belief Change			
	(1)	(2)	(3)	(4)
Effort Overprediction	0.117	0.100		
	(0.101)	(0.100)		
Pre-Exam Belief Change			0.563***	0.552***
			(0.062)	(0.057)
Group 2	1.560	1.437	0.288	0.002
	(2.537)	(2.633)	(1.820)	(1.963)
Female		-1.517		-1.689
		(3.168)		(3.008)
Semester		-0.110		-0.211
		(0.948)		(0.719)
First Micro		-1.340		-0.581
		(4.434)		(4.226)
Econ		1.079		1.443
		(3.387)		(2.946)
Business		0.853		0.054
		(3.477)		(3.162)
Age		0.116		0.101
		(0.394)		(0.316)
Exp. Result		0.057		0.012
		(0.131)		(0.117)
Exp. Effort		-0.076*		-0.055
		(0.046)		(0.043)
Const.	0.291	1.296	0.501	1.460
	(1.541)	(5.737)	(1.342)	(5.067)
Obs.	82	82	82	82
R^2 / Pseudo	0.043	0.083	0.300	0.330
F-statistic	0.674	0.606	41.859***	14.395***

Age, Expected Effort (wave 1) and Expected Result (wave 1) are centered around their means.
Removed one observation per group with unrealistic effort reports (> 200 hours).
Robust S.E. in parentheses.
*** p<0.01, ** p<0.05, * p<0.1

B.4.2 Power Analysis

Given ex-post power analysis using realized effect sizes is viewed very critically (see, e.g., Hoenig and Heisey, 2001), we provide here a basic "ex-ante" power analysis concerning the tests of our main hypothesis (i.e., Hypotheses 1a and 1b, applied to each of the two exam dates). Absent prior studies to inform ours, we posit effect sizes derived from our simple (β, δ)-discounting model with malleable beliefs from Section 3, which we combine with empirical estimates (wherever possible), in particular of the present-bias parameter β, which governs the hypothesized motive of belief distortion as a means of self-control. We note that the analysis is also basic in that it applies to each of the (four) tests in isolation and ignores any additional checks (such as our cross-validation with another course). We will end this section drawing lessons from this ex-ante power analysis as well as from our ex-post realized data—in particular, the corresponding standardized effect sizes according to Cohen's d (Cohen, 1988)—for future studies' sample considerations.

Effect Sizes We rely on the meta-study by Cheung et al. (2021) and use their mean estimate of $\beta = 0.66$ for non-monetary outcomes as defining an intermediate level of present bias. Additionally, we use the bounds of their 95% confidence interval, $[0.51, 0.85]$, to define large and small levels of present bias, respectively. We consider these different levels of present bias to derive the range of ex-ante predicted effect sizes (group differences) at different waves, while holding other parameters of our model fixed (as we will show, varying just this key parameter over the above interval already covers a wide range of effect sizes). In particular, we set the ("long-run" daily) discount factor to $\delta = 0.95$.[38] Furthermore, we assume $\gamma = 0.05$, meaning that relative to the costs of study effort those of belief distortion are low, and we normalize the model's baseline belief, $\hat{R}_0$, to 1, to obtain belief distortions directly in percentage terms (see Footnote 40 for further explanation).

We detail our procedure here for waves 3 and 4 around the first exam date, corresponding to results for Date 1 in Tables 1 and 2, for the second it is analogous. Using (3), under the

[38]Other related literature (Augenblick et al., 2015; Augenblick and Rabin, 2019; Cheung et al., 2022) estimates long-run discount factors δ close to one, even weekly ones, whereby this may appear too low. However, this parameter effectively captures also any knowledge decay in our model, see Appendix A, which the main exposition abstracts from for simplicity.

parameters above, we obtain group 1's return beliefs at wave 3 as follows for various values of β:[39]

$$r_3^{G_1} = \begin{cases} 1.5555, & \text{if } \beta = 0.51, \\ 1.3589, & \text{if } \beta = 0.66, \\ 1.1398, & \text{if } \beta = 0.85. \end{cases} \tag{5}$$

This means that students in group 1 upwardly bias their return beliefs by 14% to 55% two weeks before their exam, relative to their baseline belief. In wave 4, return beliefs for group 1 drop to baseline $\hat{R}_0$, so $r_4^{G_1} = 1$ independently of present bias. Analogously, group 2's wave-3 and wave-4 return beliefs are:

$$(r_3^{G_2}, r_4^{G_2}) = \begin{cases} (1.0063, 1.1439), & \text{if } \beta = 0.51, \\ (1.0056, 1.1174), & \text{if } \beta = 0.66, \\ (1.0032, 1.0580), & \text{if } \beta = 0.85. \end{cases} \tag{6}$$

From (5) and (6), for any given β, we can calculate $\Delta_3^r \equiv r_3^{G_1} - r_3^{G_2}$, i.e., the group difference in return beliefs between group 1 and group 2 at wave 3, as the corresponding ex-ante predicted effect size in percentage points for Hypothesis 1a (wave 3), tested in Table 1 (Date 1's EXAM coefficient).[40]

Likewise, again for any given β, we can calculate the belief revision between waves 3 and 4 for group k using $s_4^{G_k} \equiv r_4^{G_k} - r_3^{G_k}$. The group difference $\Delta_4^s \equiv s_4^{G_1} - s_4^{G_2}$ in these revisions is then the ex-ante predicted effect size in percentage points for Hypothesis 1b, tested in Table 2 (Date 1's EXAM coefficient). For instance, for small present bias, we expect an effect size of $\Delta_4^s = -0.1946$, whereas $\Delta_4^s = -0.6931$ for large present bias.

[39] In using (3), we substitute κ_t by the discount factor $\beta\delta^t$, so—unlike the "model-time" notation in Section 3—t denotes the delay to the exam. Following the timeline of our study, as given in Figure 1, e.g., at wave 3 this equals $t = 13$ for group 1 and $t = 65$ for group 2. Figure 6 shows the Mathematica code we used for the calculations (exemplified for $\beta = 0.66$).

[40] To obtain an ex-ante prediction for the EXAM coefficient in Table 1, we need to assume a "true" return, $\hat{R}_0$. Upon calibrating standard deviations of beliefs to fixed proportions of mean beliefs, however, only percentages as relative differences matter and the sample sizes we compute are independent of $\hat{R}_0$, whereby $\hat{R}_0 = 1$ is a mere normalization indeed.

Standard Deviations and Group Ratio Given the novelty of our return belief elicitation to the literature, we derive an ex-ante expectation for the variability of return beliefs by assuming that they vary similarly to other belief data elicited in our own study's wave 1, which preceded the elicitation of return beliefs. Specifically, we use the responses to the wave-1 question about how much time the students believe they will spend on studying for the exam in the two weeks preceding it. The mean response is 35.67 hours, and this comes with a standard deviation of 23.53, hence the ratio of standard deviation to mean equals approximately 2/3.[41] Assuming the same such ratio, the standard deviation of group k's return beliefs at a given wave time τ is given by:

$$\mathrm{SD}_{r_\tau^{G_k}} \equiv \frac{2}{3} \cdot r_\tau^{G_k}. \tag{7}$$

The corresponding standard deviation of group k's return belief revisions between waves 3 and 4 follows as

$$\mathrm{SD}_{s_4^{G_k}} = \sqrt{\left(\mathrm{SD}_{r_3^{G_k}}\right)^2 + \left(\mathrm{SD}_{r_4^{G_k}}\right)^2 - 2 \cdot \rho \cdot \mathrm{SD}_{r_3^{G_k}} \mathrm{SD}_{r_4^{G_k}}}, \tag{8}$$

where ρ is the correlation coefficient of return beliefs between waves 3 and 4. We assume a (moderately) positive correlation of $\rho = 0.5$ for both groups.

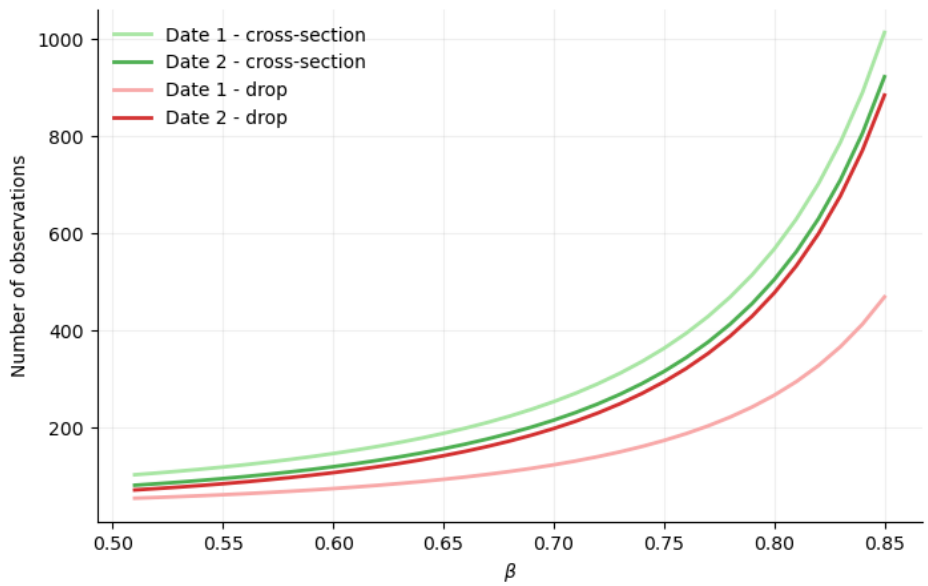
In addition, given our ex-ante perspective, we assume a ratio of group-1 to group-2 students based on previous years. Two years before the study cohort, there were twice as many enrollments for the first exam date as for the second, whereas in the year just before, enrollments were distributed almost equally. We accordingly assume a ratio of students in group 1 to group 2 of 2:1. This is conservative from an ex-ante perspective, at least for our power analysis regarding the first exam date (but slightly optimistic from an ex-post perspective).

Results Figure 5 summarizes the results of our power analysis. For each value of $\beta \in [0.51, 0.85]$, we computed the corresponding effect sizes as described above, for both Hy-

[41] In line with the ex-ante perspective this uses all 214 respondents of wave 1. The numbers are very similar for our final sample at wave 1, however: mean 35.5 and standard deviation 25.21.

potheses 1a ("cross-section") and 1b ("drop"), and for both exams ("Date 1" and "Date 2"). Along with our assumptions regarding standard deviations and the group ratio, we could then determine the required sample size for a power of 80% in testing the Null hypothesis of no effect/difference with a two-sided Student's t-test and significance level $\alpha = 0.05$. The vertical axis of the figure plots this sample size as a (continuous) function of β, for all four tests (exam date $\times$ hypothesis).

Figure 5: Sample size required for a power of 80%



Starting with exam date 1, consider the sample size needed to reliably detect the cross-sectional/level difference Δ_3^r between the two groups' return beliefs at wave 3 (light green line). For intermediate present bias of $\beta = 0.66$ this would be 200 observations, which is 2.38 times more observations than our actual sample size of 84. For large present bias (a small β of 0.51), the required sample size would be 103, and for small present bias (a large β of 0.85), it would be $1,014$.[42] We can thus cap the desired sample size between approximately 1.2 and 12 times the actual sample size.

By contrast, to reliably detect the drop/difference-in-difference Δ_4^s around exam date 1 between waves 3 and 4 (light red line), we would need a sample size of only 55 if the true

[42]To put the magnitude of our posited effect sizes into perspective, we can, assuming the required sample size for 80% power, calculate the associated standardized effect sizes in terms of Cohen's d, where we apply his rule of thumb for interpretation (Cohen, 1988): These values are 0.59 for $\beta = 0.51$ (above medium-size effect), 0.42 for $\beta = 0.66$ (below medium-size effect), and 0.19 for $\beta = 0.85$ (small effect).

β were 0.51, i.e., for large present bias. For intermediate $\beta = 0.66$, the required sample size increases to 99, which is 1.18 times our actual sample size. For small present bias (large $\beta = 0.85$), we would need 470 observations, which is 5.6 times our actual sample size. Hence, for intermediate to larger present bias, our difference-in-differences test at exam date 1 can, from an ex-ante perspective, be regarded as sufficiently powered or close to sufficiently powered, while underpowered for lower levels of present bias.

Figure 5 analogously visualizes the required sample sizes for exam date 2 (dark lines). Here, to detect the cross-sectional effect at wave 5 (dark green), we would need 0.97 to 11 times as many observations, and for the drop effect between waves 5 and 6 (dark red), we would need 0.86 to 10.5 times as many observations.

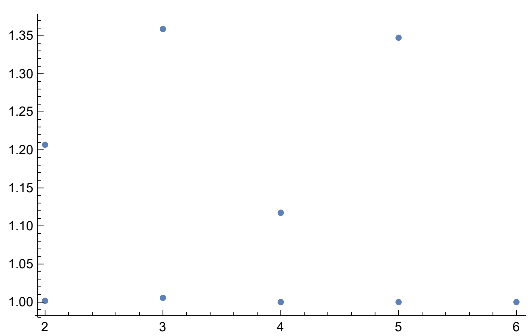
Taken together, and with the caveats pointed out initially, we conclude from this analysis that our study appears generally underpowered for the cross-sectional tests of Hypothesis 1a (see Table 1). By contrast, for the difference-in-differences tests of Hypothesis 1b (see Table 2), the power of our study appears significantly better, achieving better-than-adequate levels for large present bias and close-to-adequate levels for intermediate present bias.

Regarding future studies, we note that our power analysis yields desired sample sizes generally below 200 for intermediate levels of present bias. Thus, our approach to testing for the hypothesized belief distortion on an introductory course's university students, with staggered examination dates, can be encouraged as feasible for future research. Relatedly, with some caution, we can additionally explore how the desired sample sizes would look when using our data as if it came from a pilot. One would then typically calculate the standardized effect sizes in terms of Cohen's d from this data to derive the future (full study's) sample size required to achieve a pre-specified power target (Cohen, 1988). To be clear, this is then not to assess the power of the current study but rather to further guide future studies on sample size considerations. We briefly illustrate this here. For instance, concerning exam date 1, our realized return belief differences between groups 1 and group 2 at wave 3, together with our realized group-1:group-2 ratio of 60:24, yield a relatively small standardized effect size of $d = 0.2841$ for the cross-sectional Hypothesis 1a at this wave. Assuming this was the true effect size and group ratio, to achieve a power level of 80% in a future study would require 478 observations. We can also easily carry out the calculation for pooled tests, however. Taking the belief drop according to Hypothesis 1b around the exam date, pooling the data of both groups (in our study this meant $2 \cdot 84 = 168$ observations), we obtain a value of $d = 0.2933$, implying a sample requirement of 367 observations in future studies. Since in a design like ours this would mean 184 students (367/2), the result of this exercise sits well with the above ex-ante conclusion of around 200 students for intermediate present bias, confirming this suggested (minimum) sample size for future work. As a final lesson, we add that our analysis—subject to the aforementioned caveats—generally highlights the critical importance of incentives that optimize participation/completion subject to the setting's constraints as well as management of further sample-size risk due to background uncertainties beyond researchers' control.

Figure 6: Mathematica Code

```
In[72]:= del = 0.95; (*long-run daily discount factor delta*)
bet = 0.66; (*present bias parameter beta*)
gam = 0.05;(*cost parameter belief distortion*)
r0 = 1;(*baseline return belief*)
kap[t_] := bet del^t;
(*discount factor for delay of t days, kappa(t) - but,
 unlike in paper, uses argument t as delay to exam)*)
ret[x_] := r0 (1 + (1 - bet)/bet (x^2/(x^2 + gam)));
(*optimal return belief as function of discount factor*)
y2 = ret[kap[77]]; (*optimal return beliefs of group 2 over waves yw,
the numbers are days remaining to exam at various wave dates*)
y3 = ret[kap[65]];
y4 = ret[kap[33]];
y5 = ret[kap[14]];
y6 = ret[0];
x2 = ret[kap[25]]; (*optimal return beliefs of group 1 over waves xw,
the numbers are days remaining to exam at various wave dates*)
x3 = ret[kap[13]];
x4 = ret[0];
x5 = ret[0];
x6 = ret[0];
ListPlot[
 {{2, y2}, {2, x2}, {3, y3}, {3, x3}, {4, y4}, {4, x4}, {5, y5}, {5, x5}, {6, y6}, {6, x6}}]
```

Out[83]=




```
In[86]:= {x2, x3, x4, x5, x6}
Out[86]= {1.20674, 1.35885, 1., 1., 1.}

In[87]:= {y2, y3, y4, y5, y6}
Out[87]= {1.00166, 1.00564, 1.11736, 1.34746, 1.}
```

C Panel and Sample Description

This appendix describes the dataset in detail – first its main component, the dynamic online survey, then the additional administrative data. Based on this description, we make explicit how we define our main sample.

C.1 Survey Panel

We conducted our survey in six waves, starting in December 2015 and ending in May 2016. Participants were informed (and also reminded) of each wave in advance through the online course portal, and sent an individual link via email for each wave. Upon opening the link they could respond online to the survey questions (e.g., using a smartphone). Failure to respond to a given wave meant dropping out from the survey (i.e., from all subsequent waves). All responses remained and still remain anonymous. Moreover, we had no access to any of this data until all grades had been finalized (except for the tertiles of expected study effort in wave 1, so we could construct our hypothetical scenarios). Table 15 gives an overview of what information we gathered when, and from how many participants. We invited participants to collect their payment for completion of the survey on May 3 and May 4, 2016. This payment consisted of 10 Euros in cash, plus a 1:7 chance of winning an Amazon voucher worth 100 Euros. This lottery was resolved on May 2, 2016.

Figure 7 provides a screenshot of our main question (in its original German version) regarding return beliefs, described in Section 4.1.

Figure 7: Screenshot of Main Question

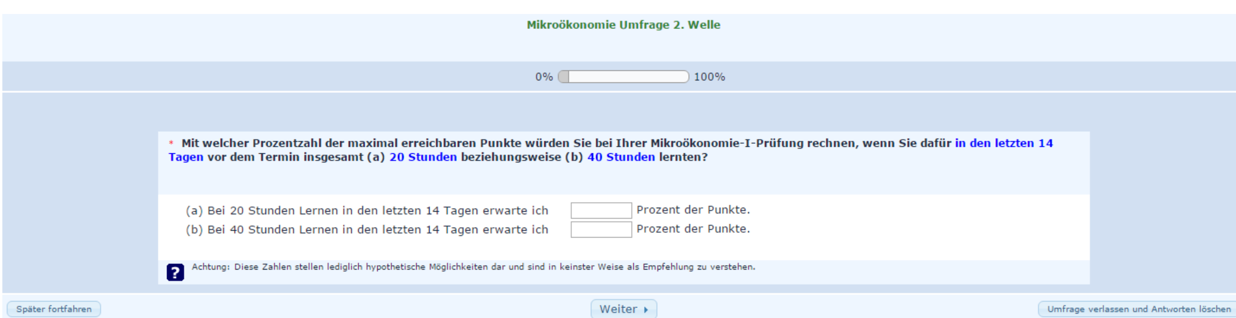


Table 15: Survey Panel

	Wave 1	Wave 2	Wave 3	Wave 4	Wave 5	Wave 6
Wave Start Date	Dec. 9	Jan. 26	Feb. 7	Mar. 8	Mar. 27	Apr. 26
Wave End Date	Dec. 23	Jan. 29	Feb. 10	Mar. 13	Apr. 1	May. 1
# Observations	214	175	149	141	127	118
Response Times	X	X	X	X	X	X
Gender (f/m)	X					
Age (y. & m. of birth)	X					
Survey Info (lec., TA, online, stud.)	X					
Study Prog. (econ., bus., edu., other)	X					
# Semesters of Study	X					
First Timer (y/n)**	X					
Take Maths (y/n)	X					
Exam Date (1, 2, neither)**	X	X	X	X	X	X
Confident in Exam Date (y/n)**	X					
# Exams This Semester	X	X	X	X	X	X
# First Exams This Semester	X	X	X	X	X	X
Career Importance of Grade (1-6)**	X	X	X	X	X	X
Exp. Effort (hs 14 days prior)**	X	X	X	X	X	X
Exp. Performance (% score)**	X	X	X	X	X	X
Exp. Performance with 20 hs Effort**		X	X	X	X	X
Exp. Performance with 40 hs Effort**		X	X	X	X	X
Know First Exam (y/n)				X	X	X
Know Second Exam (y/n)						X
Harder Exam (1st, 2nd, neither, no op.)						X
Patience* (1-10)						X
Risk Tolerance* (1-10)						X
Time-Consistency* (1-10)						X

* We adopted the patience and risk tolerance measures from the preference module of Falk, Becker, Dohmen, Huffman, and Sunde (2016), and we added our own similarly formulated item on time-consistency, namely "Do you generally keep your resolutions?"
** This was also asked about a parallel mathematics course for which the two exam dates were February 16 and April 13; we have no hard data on this course whatsoever, in particular no registrations from the examination office or exam scores.

C.2 Course and Exam Organization, and Administrative Data

Below we provide information regarding course and exam organization:

- Examination periods and dates: Following the end of classroom teaching, there are two examination periods, and every course is examined once in each period. Each examination period lasts for two weeks. In our case the first ended on February 26, and the second on April 15.

- Exam registration and regulations: Students could register from January 1 through January 25. Any student not registered for either of the exam dates after this period could not take the course's exam. Registered students could withdraw from their exam until three working days prior to it, and then could not take the course's exam (at either date). Students registered for the first exam who either supplied a sick note to the examination office for this date or failed the exam could register and then take the second exam.

- Microeconomics exam: The exam dates for our Microeconomics I course were February 23 (between waves 3 and 4) and April 15 (between waves 5 and 6). We published the first exam in the online course portal on March 4 (between the exam and wave 4) and the second on April 21 (between the exam and wave 6). Solutions were never provided. We released the grades for the first exam on March 17 (between waves 4 and 5) and offered exam inspection—a requirement for every exam—on April 6 (between waves 5 and 6). Release of the grades for the second exam and exam inspection took place only after the end of the survey.

The following data from the examination office were anonymously matched with our survey data. (Only for the microeconomics course; we had no access to official data for the mathematics course.)

- Exam registrations as of Feb. 9 (1 or 2 or missing).
- Exam registration lists for each exam date, as of a few days prior (in or out).
- Point scores for each exam date (0-90).

C.3 Sample Definition

Our two groups are defined as follows.

- Group 1:
  - Completed the entire survey (all six waves).
  - Took the exam at the first date, and not at the second date.

- Registered for the first exam with the examination office as of Feb. 9, or else were one of the two students not registered for either exam date as of Feb. 9, but nonetheless on the registration list for the first exam.

- Group 2:
  - Completed the entire survey (all six waves).
  - Took the exam at the second date, and not at the first date.
  - Registered for the second exam with the examination office as of Feb. 9, or else were one of the two students registered for the first exam date as of Feb. 9, but already in wave 3, which took place before Feb. 9, reported in our survey that they would take the exam at the second date.

C.4 Data Protection Concepts

C.4.1 Data protection letter sent to the university administration

We would like to carry out a study of studying behavior with the students in Microeconomics I. We are interested in the students' responses regarding their planned study effort and their expectations regarding grades (please find attached the survey questions). Additionally, we will collect background information as is usual in such surveys (e.g., gender and age).

Participation in the study is voluntary. We commit ourselves to not having any access to the data until all grading is officially finalized. Data will be externally collected and stored by our student assistant Felix Bönisch, and he will make them accessible to us after the final grading. The survey responses can therefore neither affect the design nor the grading of exams.

After grades have been officially awarded, we plan to match the survey data with the respondents' grades in an anonymized manner. To this end, we will use an encryption method that, at no point in time, allows to identify grades and survey responses with the corresponding students—neither by Mr Bönisch nor by us. The encrypted matching of the data works as follows: we will send Felix Bönisch a list assigning each student number a key, where different student numbers are assigned the same key—based on "equivalent" grades—, and Mr Bönisch will then send the list back to us such that student numbers are replaced with the corresponding survey answers. Hence, these responses cannot be linked to student numbers.

Finally, we would like to stress that the careful handling of data protection issues is in our own best interest. Only by doing so can we obtain a large number of participants and credible survey answers, which is essential for the quality of our research.

Prof. Georg Weizsäcker, Sebastian Schweighofer-Kodritsch and Tobias König; Berlin, December 4, 2015

C.4.2 Data protection announcement to participants

Dear students,

Thank you very much for your willingness to participate in this survey. Your sincere responses will make an important contribution to human behavioral research. This is to assure you that your responses are anonymous. Our procedure has been approved by the School of Business and Economics of Humboldt-Universität zu Berlin, and, moreover, your anonymity is also in our best interest for the quality of our research.

1. During the entire survey period, i.e., until all grades have been finalized, none of the researchers involved will have access to the data collected. Until then, these will be externally collected and stored. This rules out any influence of participants' survey responses on exam design and grading.
2. The questions that we will ask you are innocuous. Apart from background information commonly asked in surveys, they relate to your expectations and your study behavior.

3. The use of an encryption mechanism in the data transmission (as mentioned in item 1, this will take place only after grades have been finalized) guarantees that grades and survey responses can never be identified with a survey participant.

In case you have any questions about the details of the study, you can reach us at any time via email to mikro.umfrage@gmail.com, where we are happy to respond as quickly as possible.

Kind regards,

Prof. Georg Weizsäcker.