RATIONALITY
& COMPETITION
CRC TRR 190

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

Tobias König (HU Berlin)
Sebastian Schweighofer-Kodritsch (HU Berlin and WZB Berlin)
Georg Weizsäcker (HU Berlin)

Discussion Paper No. 123

October 25, 2018

Collaborative Research Center Transregio 190 | www.rationality-and-competition.de
Ludwig-Maximilians-Universität München | Humboldt-Universität zu Berlin
Spokesperson: Prof. Dr. Klaus M. Schmidt, University of Munich, 80539 Munich, Germany
+49 (89) 2180 3405 | info@rationality-and-competition.de

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

Tobias König, Sebastian Schweighofer-Kodritsch and Georg Weizsäcker*

October 25, 2018

Abstract

We repeatedly elicit beliefs about the returns to study effort in a panel survey of students of a large university course. A behavioral model of quasi-hyperbolic discounting and malleable beliefs yields the prediction that the dynamics of return beliefs mirrors the importance of exerting self-control, such that return expectations first increase as the exam approaches, and then sharply drop post-exam. Exploiting variation in exam timing to control for common information shocks, we find this prediction confirmed: average subjective expectations of returns increase by about 20% over the period before the exam, and drop by about the same amount afterwards.

*König and Schweighofer-Kodritsch: School of Economics and Business, Humboldt-Universität zu Berlin & WZB Berlin Social Science Center; Weizsäcker: School of Economics and Business, Humboldt-Universität zu Berlin & DIW Berlin. Mail: Humboldt-Universität zu Berlin, Unter den Linden 6, D-10099 Berlin, Germany. Phone: +49 (0)30 2093 5998. Email: sebastian.kodritsch@gmail.com. We thank Yves Breitmoser, Anna Gumpert, Andreas Haufler, Peter Schwardmann, Felix Weinhardt and Florian Zimmermann for insightful discussions, our various seminar and conference audiences for helpful comments, and Felix Bönisch for excellent research assistance. Special thanks go to Martina Albers who was involved in early stages of this project. Financial support by the German Science Foundation (CRC TRR 190) and the ERC (Starting Grant 263412) is gratefully acknowledged.

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; Brunnermeier et al., 2017) and they may thereby improve material outcomes for a behaviorally biased decision maker. In this paper, we provide the first field evidence that beliefs 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 returns to their study effort on 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. 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 relevance 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.

To guide our empirical design and analysis, we first formalize this intuition with a simple behavioral model of (β, δ)-discounting and malleable beliefs. The model indeed yields the prediction that the decision maker's subjectively expected return to effort (i) is most upward-biased in the final study period before the exam, when self-control is most valuable, and (ii) sharply drops in hindsight, when the exam is over and the instrumental motive is gone.

We then design a dynamic student survey to test the model's predictions. Our main variable of interest is a student's belief about an unknown entity: the difference between (i) her performance (in % of maximal achievable point score) 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. 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 mean-zero and iid across students, then the beliefs should not change over time if they were rational (the martingale property of expectations). In this case, any theory predicting a non-trivial pattern in average expectations could easily be tested against the null hypothesis of rational expectations. However, it is likely that students' information shocks are non-mean-zero and correlated, arising from the students' common experiences in lectures, class tutorials, and from other

[1] There is a large literature in both psychology and economics demonstrating overconfidence, see the survey by DellaVigna, 2009. For a quick introduction to the psychology literature on unrealistic optimism see Shepperd et al., 2015.

common observations that may inform them about the returns to study. We therefore rely on a particular feature in the university's exam organization: we compare expectations 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' perspectives, they share the same lectures and class tutorials, and at any given point in time they have access to the same information about the course. But the two groups vary in the importance of self-control because one group's exam comes several weeks earlier.

Empirically, we find that both groups' average return expectations follow the model's predicted pattern over time *relative to their respective exam*: return expectations increase towards the final study period before one's own exam and sharply drop post-exam. No reaction occurs around the time that the other group has their exam. When we combine both groups into a single pool, average expectations are essentially constant. Quantitatively, our main finding is that average beliefs show an average upward-bias of return expectations of around 20% in the period before the exam, and this push-up belief disappears post-exam. Moreover, we also find empirical support for our maintained assumption that students experience a self-control problem, which underlies our interpretation of what causes biased beliefs.

Related Literature Our point of departure is the general idea that belief distortions may be instrumental in overcoming a self-control problem. In influential earlier 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.[2] By contrast, we study belief distortions regarding the largely extrinsic return to effort and propose a reduced-form model of belief manipulation that is agnostic about the exact psychological mechanism for manipulation (because we could not observe what information students obtain). Importantly, our theoretical model also differs in terms of its predictions: it predicts belief distortions also under naïveté about the self-control problem, and it generates systematic belief distortions that violate Bayesian updating.[3]

We extend the multiple-selves model with (β, δ)-discounting by an over-arching unbiased planner who can directly (and sub-consciously) distort the return beliefs to achieve self-

[2]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, even in the long run.

[3]Bénabou and Tirole (2002) maintain full sophistication about how memory is manipulated, so Bayesian updating delivers that the average posterior belief equals the prior.

control. While the model is related to various versions of dual-systems theories (e.g., Thaler and Shefrin, 1981, Fudenberg and Levine, 2006, or Brocas and Carrillo, 2008), none of these feature malleable beliefs. At the same time, while our model shares the feature of optimal belief choice with Brunnermeier and Parker (2005) and Bracha and Brown (2012), the motive for belief distortions is different: it is 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.

Evidence that self-control problems lead to belief distortions is hard to come by, and accordingly scarce. It requires variation in the nature of the same self-control problem together with control over information. The closest work to ours is the experiment by Schwardmann and van der Weele (2017). They had all subjects perform an intelligence test as in Möbius et al. (2014) and elicited beliefs about this performance. The pre-announced prospect of possible subsequent money earnings—in case they successfully convince an interviewer that they outperformed their peers—induced more favorable beliefs. Whereas their analysis suggests a *strategic* value of self-deception for being better able to deceive others, we study a non-strategic self-control problem. A further important difference is that we investigate a natural field environment.

From a purely empirical perspective our work is also related to tests of rational expectations in the field, where the researcher does not observe (nor control) all information that agents receive (e.g., Bernheim, 1990; Benítez-Silva and Dwyer, 2005). Whereas this literature essentially relies on the assumption of iid forecast errors in their tests of the rational expectations hypothesis, our empirical strategy is able to control also for correlated information.

Finally, our paper contributes to the literature on policy interventions to overcome self-control. This literature has mostly focused on the provision of external commitment devices, where take-up requires sophistication (e.g., Ariely and Wertenbroch, 2002; Ashraf et al., 2002; Kaur et al., 2015). Our findings suggest that even seemingly naïve people might achieve some degree of self-control by distorting how they perceive reality. This has not been considered in the policy literature so far, and it is important to understand how this mechanism could inform policy interventions; e.g., what kind of information to provide, and how (in particular, when).

2 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 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 (present bias) in the form of quasi-hyperbolic discounting, with parameters $(\beta, \delta) \in (0, 1)^2$. Given belief $\hat{R}_t$ in period t and given risk neutrality, she maximizes, respectively in the two periods,

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

where $\hat{e}_2$ denotes her expectation of e_2 as of period 1.

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

$$e_t\left(\hat{R}_t\right) = \epsilon_t \hat{R}_t \text{ for } \epsilon_1 = \beta\delta^2,\ \epsilon_2 = \beta\delta.$$

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 ($\epsilon_2 > \epsilon_1$).[4]

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 acts in $t = 0$ and 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

[4] Even with perfect long-run patience ($\delta = 1$), this would obtain upon assuming that $g = R \cdot (\phi e_1 + e_2)$, with $0 < \phi < 1$, so that early study effort "depreciates". For simplicity, we ignore this realistic aspect here.

assume that belief distortion has some mental cost increasing in the amount of self-delusion (cf. Bracha and Brown, 2012).[5]

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[6]

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

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 $\epsilon_3 \equiv 0$):

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

Before her exam, Sue will come to believe that the returns are excessively high, and the more so the close 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 ($\hat{R}_3^* = \hat{R}_0$).

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.

Prediction:
Sue's expectation of the returns to her study effort (i) is the higher the closer ahead is her exam, and (ii) drops sharply after she took her exam.

As a final note, observe that what matters for this prediction is sophistication of the planner, while the prediction is qualitatively unaffected by naïveté of Sue as doer. Observe also that the assumption of naïveté implies, over and above the made prediction, that Sue

[5] 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.

[6] Allowing for information arrival would complicate the model but can be incorporated. Ruling out any information arrival is also convenient because with it, the fact that Sue's planner has dynamically consistent preferences implies that it is without loss of generality to have her choose beliefs once-and-for-all at the outset.

mispredicts her future effort, which affords a simple test of whether she indeed suffers from present bias.[7]

3 Data Collection, Hypotheses, and Identification

3.1 Data Collection, Sample and Belief Measures

We collected the data in a repeated online survey of the students of our first-semester microeconomics course at Humboldt Universität zu Berlin, in the winter term 2015/16.[8] 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. Participation in the survey was incentivized by payment for completion, plus the chance of winning an online shopping voucher. After the final wave of the survey, the data were anonymously matched with data from the university's examination office. We committed to not having access to any data until the end of the course, when all grades would be finalized.[9]

The timeline of our survey is visualized in Figure 1. An important feature for our design is that any given students 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). They 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 schedules of about six final course exams per semester. We started the first wave of our survey in mid-December 2015 and the final wave was conducted 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. Over the survey period of six months, 96 students dropped out of the survey.

A crucial part of our identification strategy is to have observations from students who participated in all six waves and who can be unambiguously assigned to one of the two exam dates. Our main sample therefore includes only the "stay-ons" who wrote the exam on the date they had registered for.[10] This leaves us with a total of 84 observations: 60 first-exam

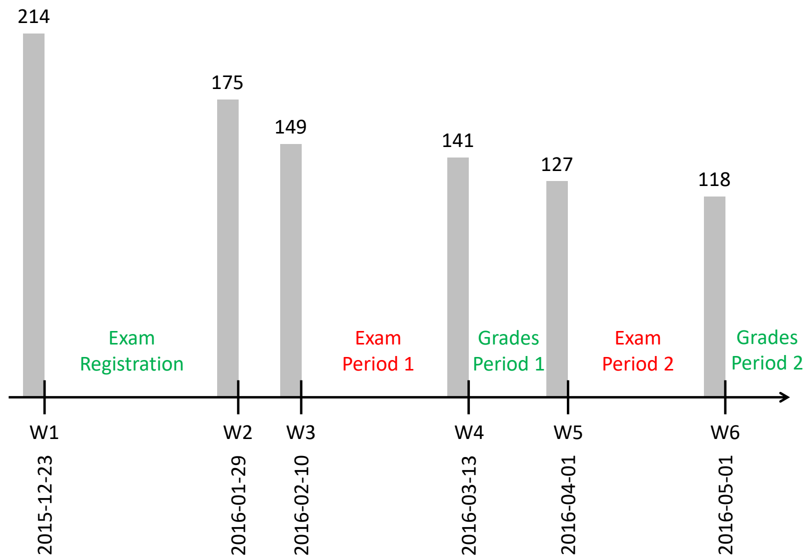
[7] At the beginning of period 2, when she already has belief $\hat{R}_2^*$, naïve Sue would overpredict her subsequent effort under any "partial" naïveté in the sense of O'Donoghue and Rabin (1999). Earlier on, in period 1, she may however over- or underpredict her period-2 effort: While naïveté, as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

[8] One author gave the weekly lectures and the other two authors held weekly class tutorials. There were three further TAs who were not involved in this research.

[9] The details of our data and our data protection concept are provided in appendix A.

[10] Students who failed their exam on the first date could repeat the exam on the second date, and for students who initially planned to write the exam on the first date but then had to take the second date due to proven illness, our return questions could be mis-interpreted, which is why we dropped these observations

Figure 1: Survey Timeline



Notes: Grey bars indicate the number of participants in each wave.

takers (group 1) and 24 second-exam takers (group 2). In appendix A, we show that our main (reduced) sample of 84 students does not significantly differ from other students who completed the first wave in terms of background characteristics. Moreover, there are no significant differences in observables between group 1 and group 2.

From wave 2 onwards, we elicited beliefs about the return to study. We asked students to give us an estimate of the percentage of points they expect to achieve in their microeconomics exam for two hypothetical effort scenarios: (a) if they were to study 20 hours in the 14 days prior to their exam date, and (b) if they were to study 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: 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 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 two tertiles (rounded) of responses to a question about own expected study effort during the 14 days prior to their exam date.[11]

Our return belief measure is the difference in the subjectively expected percentage points achieved in the exam between the two hypothetical effort scenarios. We notate this belief

from our main sample.

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

by r^i_τ for student i and wave date τ.

Our second variable of interest is the students' predictions of their own study *effort* (not return to effort) and how it develops over time. Specifically, we repeatedly asked students how many hours they expect to study for the exam in the 14 days prior to it. We will refer to this variable as $\hat{e}^i_{\tau,2}$. To summarize it briefly, we indeed find evidence of lack of self-control in our data. In wave 3, shortly before the exam of group-1 students, they on average predict their study effort to be 43.72 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 [6] they report an average of 44.47 [37.35] hours.[12]

3.2 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 group-2 students and we can therefore use the fact that 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). They therefore also differ in how important self control is.

Define $s^i_t \equiv r^i_t - r^i_{t-1}$ as individual i's belief revision between times $t-1$ and t. Rational expectations imply that it has a zero expectation as of $t-1$, and if information innovations (belief revisions) are iid within a sample of n people, then the group average $s_t = \sum_{i=1}^n s^i_t$ is approximately zero. In our application, however, belief revisions are likely to have a common component due to common information: $s^i_t = \epsilon^i_t + \eta_t$, where only the first term is iid, while the second is a common innovation. Under rational expectations, any such common innovation η_t still shifts the sample's average. However, the difference between two groups' averages has this common component removed, hence $s^{G_1}_t - s^{G_2}_t$ still approaches zero ($s^{G_k}_t$ 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.

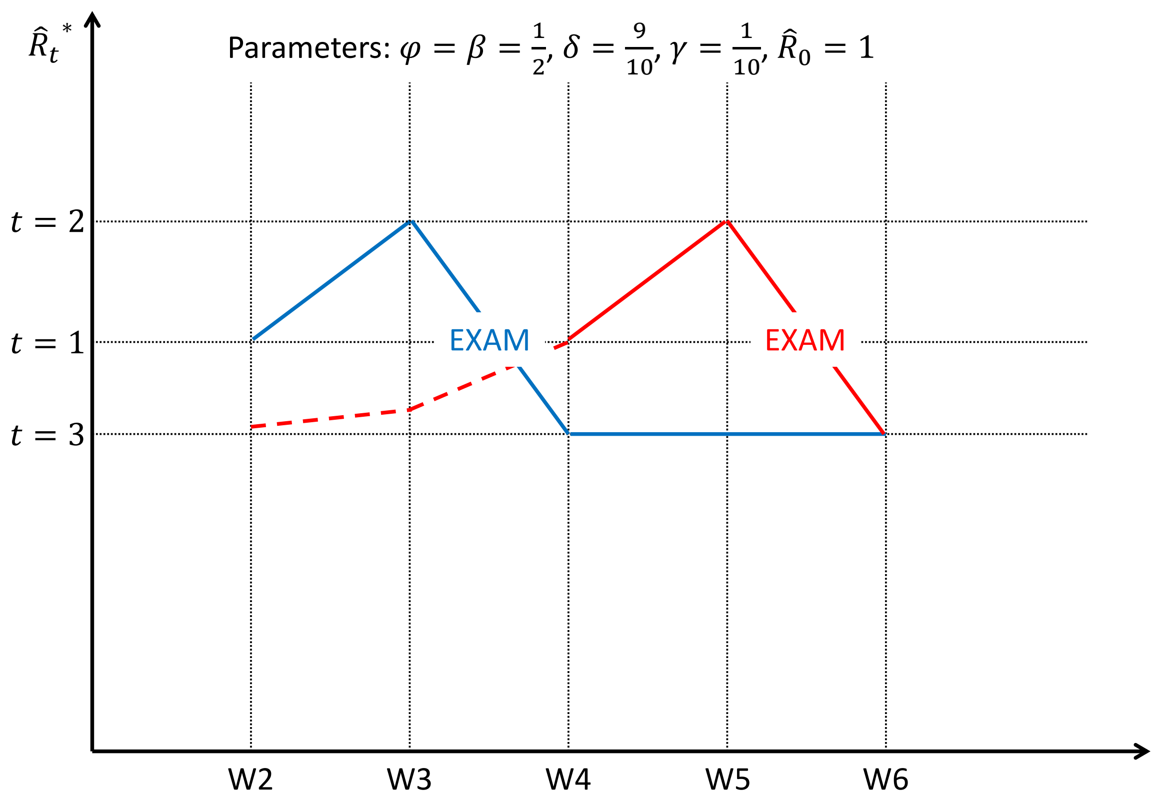
Figure 2 illustrates our empirical strategy based on the theoretical model, with calendar time (waves 2 through 6) on the horizontal axis and the agent's return beliefs at different moments in model time on the vertical axis.[13] The figure illustrates how both groups to progressively "build up" their 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

[12]The average over-prediction 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 this final study period (one per group). This would imply that they would study more than 14 hours a day in each of their last 14 days before the exam, which we find implausible. These are extreme outliers, by any standard.

[13]The illustration employs another parameter ϕ of the return function g, as given in footnote 4.

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 $t = 1$. 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 $t = 1$. At wave 5, group 2 is close to their exam, corresponding to model time $t = 2$, etc.

Figure 2: Predicted Return Beliefs over Time



In our statistical analysis we 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 are 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 are undistorted.

Hence, within each group we predict a drop in averages beliefs, relative to the group that did not just write the exam. Our hypotheses about retrun beliefs are summarized below.

Main Hypothesis (Return beliefs):

(1a) In wave 3 [5] the average return belief of group 1 [2] exceeds that of group 2 [1].

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

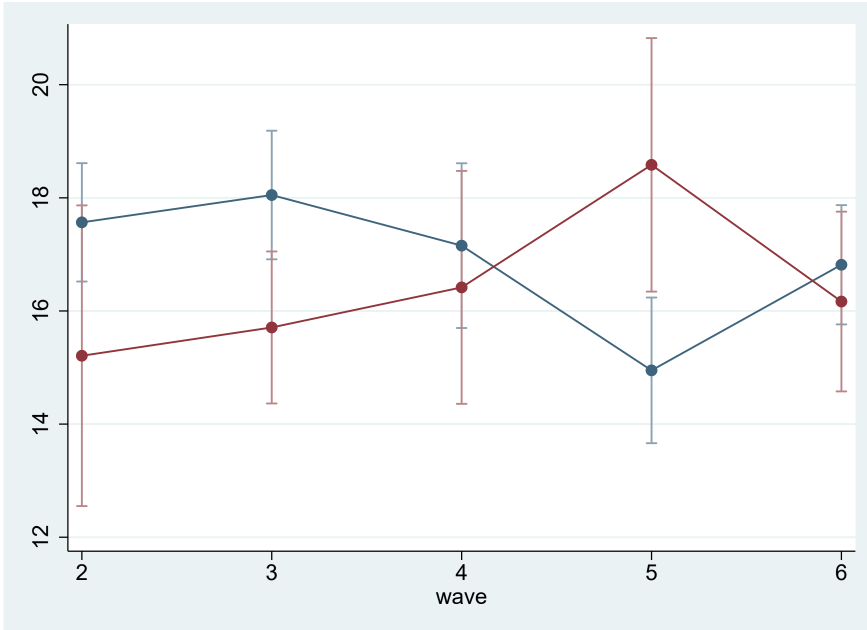
4 Results

Graphical Illustration. Figure 3 (a) shows the dynamic pattern of average 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 [red] line represents the mean of group 1 [2] students. For each exam group, we observe that beliefs build up leading up to the exam, peak at the wave date that occurs just before the respective exam, and drop afterwards. The humped-shaped pattern predicted by our model is supported by the data, in both groups.

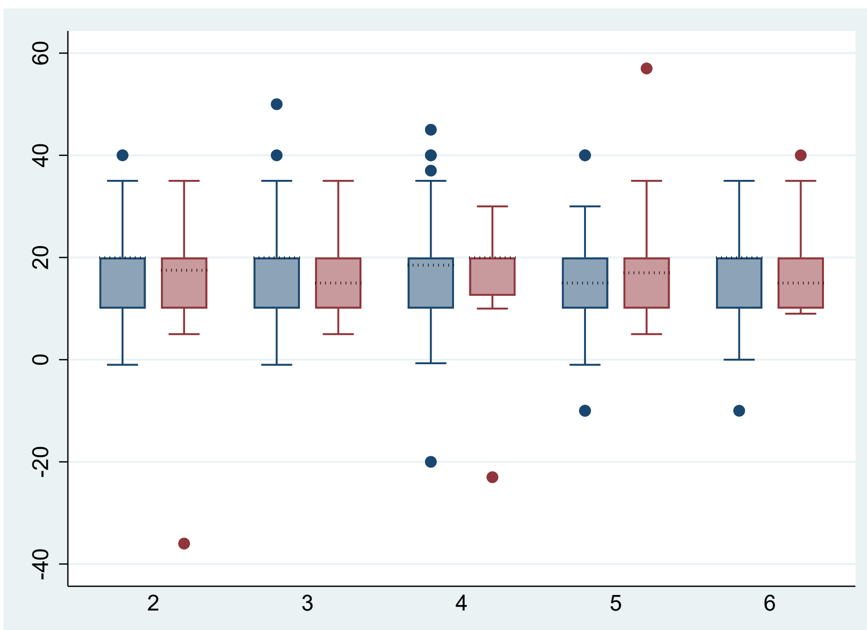
While the overall picture is also consistent with our main statistical hypothesis of the previous section—Figure 3 (a) closely resembles Figure 2—some features were not predicted. First, in wave 4, group 2's beliefs are slightly (but insignificantly) below those of group 1, where one may have expected the opposite because group 1 has no motive for distorting beliefs any more. Second, while group 1's beliefs drop after the exam, they return to the level of wave 4, which is not predicted by the model. A further relevant observation is the two groups' beliefs are very similar in wave 6. Although group 2 still has not yet learned their grades, neither group has a distortion motive (both are post-exam), and their beliefs largely coincide. Overall, we regard the first graphical investigation as confirming both parts of our main hypothesis.

Figure 3 (b) shows box plots of return beliefs by exam groups and wave periods. The figure enables further observations. First, outliers may play a role for some of the averages in Figure 3 (a). Second, while interquartile ranges are quite stable over time, the medians show a relative drop of return beliefs after the exam, both for group 1 and group 2. But the hump-shaped pattern appears in medians only for group 2 (from wave 3 onwards).

Figure 3: Return Beliefs Over Time



(a)



(b)

Notes: Panel (a) plots mean return beliefs over waves 2-6 by groups, 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. Panel (b) 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 dashed line the median.

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. Specifically, we test Hypothesis 1(a). Column 1 reports the results from regressing wave 3 return beliefs, r_3^i, on a dummy *Exam* indicating that the student took the exam between waves 3 and 4 (i.e., is in group 1). The coefficient on *Exam* 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* equals one for the members of group 2—also gives the predicted sign without significance (column 2). A pooled regression that combines both data sets produces a similar estimate on *Exam*, but with much smaller standard errors, such that we obtain statistical significance (column 3). In this pooled regression the coefficient on the *Date 2* dummy, which is one for the second exam date, is essentially zero.

In columns 4 through 6, we report the results from running median regressions instead of OLS, which are sensitive to outliers. The coefficient for *Exam* is larger in then median regressions than in OLS and it is significant also in the regression for the first exam alone (but not for the second exam).

Table 1: Return Beliefs

	Ordinary Least Squares			Median Regression		
	Date 1	Date 2	Pooled	Date 1	Date 2	Pooled
Exam	2.342	3.633	2.9875**	5***	4	5**
	[1.991]	[2.479]	[1.264]	[1.750]	[2.721]	[1.819]
Date 2			-.112			0
			[1.264]			[1.819]
Const.	15.708***	14.95***	15.247***	15***	15***	15***
	[2.203]	[1.325]	[1.235]	[1.480]	[1.454]	[1.742]
Obs.	84	84	168	84	84	168
R^2 / Pseudo	0.0166	0.0255	0.0265	0.0625	0.0000	0.0263

S.E. in brackets, for pooled OLS clustered at ID level.
*** p<0.01, ** p<0.05, * p<0.1

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 a dummy indicating that the student's exam immediately preceded wave t (the dummy is on in $t = 4$ for group 1, and in $t = 6$ for group 2). We run this regression for each of the two exam dates separately and for the pooled sample, in the latter case controlling for date fixed effects. The results are in columns 1-3 of Table 2, which only has OLS regressions because median regression would not produce meaningful output after differencing due to lack of variation.

Table 2: Return Beliefs – After vs. Before Exam

	Ordinary Least Squares		
	Date 1	Date 2	Pooled
Exam	-1.603	-4.283**	-2.943**
	[2.606]	[2.107]	[1.300]
Date 2			-0.182
			[1.290]
Const.	0.708	1.866	1.665
	[2.203]	[1.126]	[1.293]
Obs.	84	84	168
R^2	0.005	0.048	0.021

S.E. in brackets, for pooled OLS clustered at ID level.
*** p<0.01, ** p<0.05, * p<0.1

The sign of the coefficient on the exam dummy is negative throughout, indicating that students who are "treated" in the sense that they have written the exam (*Exam*=1) experience a drop in their return beliefs, controlling for common belief shocks. Only for the date 1 regression, the coefficient on *Exam* is not statistically significantly different from zero.

Discussion. If information was correlated in a way that depends on the students' exam dates, this could potentially produce the pattern of return beliefs (even though that pattern is quite intricate). A particular concern in this respect is an unexpectedly hard first exam, which could induce the return beliefs of group 1 to decrease between waves 3 and 4, but not for members of group 2 if they did not obtain any information about the exam. Anticipating this concern, we decided to equalize information conditions across groups as far as possible. We made each exam available for all course participants directly after it had been written and we advertised this availability in an email to all course participants. We also included a question in the survey asking participants whether they also knew the contents of the exam of the other date. Thus we can redo the above regressions with pooled data but using only the subsets of the respective control group that knew the relevant exam. The results are robust to this variation: we obtain significant estimates with the predicted sign, both for OLS (Hypothesis 1a: *Exam* coefficient 2.79, s.e. 1.58; Hypothesis 1b: coeff. -2.94, s.e. 1.30) and median regressions (Hypothesis 1a: coeff. 5.00, s.e. 2.44).

It could still be argued that informational spillovers to non-writers are limited (e.g., writing an exam is different than just hearing about or reading it). Note however that in order to produce the decrease of return beliefs for both exam dates, we would have had to succeed in "tricking" the students twice, which is unlikely. More generally, the data context is not prone to surprises. All instructors, course contents and materials were essentially identical to past years, of which students were aware. Both exams were identical in format and of similar difficulty compared to past exams, which were also publicly available and are a major part of students' preparation for the exam. The grade results were similar for both groups, and this is true also across years.[14]

While we find evidence for belief manipulation, the question whether belief manipulation pays is difficult to answer with our data set. For this, one would need exogenous assignment of students to belief manipulation, which we do not have. One noteworthy observation is that students who believe returns to studying to be high also expect to study more (and indeed studied more), see Appendix Table 5. This is consistent with the view that the subjectively expected returns are behaviorally relevant.

5 Conclusion

The 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* return to studying. The violation of the martingale property is enough to conclude that rational expections are rejected. However, the test requires a suitable control group, in order to rule out that the effect is driven by correlated information. Moreover, the value of finding such a violation rests on the behavioral mecha-

[14]Students with strictly positive points received, on average, 50.21 points in the first and 47.73 points in the second exam. In the year before, the means were 42.26 and 46.67.

nism that one desires to test (as does, in this paper, the entire empirical design). Here, the particular dynamic pattern of beliefs 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 also contributes to the literature on self control, which has mostly focused on extrinsic commitment opportunities or intra-personal equilibrium strategies employing self-punishments and self-rewards (in a non-cooperative game between the multiple selves with conflicting preferences). Both of the latter are effective only under a high degree of sophistication, however, whereas empirical evidence suggests that people instead tend to be rather naïve. This is also consistent with our sample: students significantly over-estimate the amount of later studying.

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.

Predicting and measuring belief dynamics is a novel area of research. An important aspect in it is how people select, process and recall information. In the presence of self-control problems or other behavioral deviations from the standard model, the effect of information gathering is 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) but it may also increase her costs of distorting beliefs and thereby make self-control harder.

References

Ariely, D. and K. Wertenbroch (2002). Procrastination, deadlines, and performance: Self-control by precommitment. *Psychological Science 13*(3), 219–224. 4

Ashraf, N., D. Karlan, and Y. Wesley (2002). Tying odysseus to the mast: Evidence from a commitment savings product in the philippines. *Quarterly Journal of Economics 121*(2), 635–672. 4

Bénabou, R. and J. Tirole (2002). Self-confidence and personal motivation. *Quarterly Journal of Economics 117*(3), 871–915. 2, 3, 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. 4

Bernheim, B. D. (1990, January). 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. 4

Bracha, A. and D. J. Brown (2012). Affective decision making: A theory of optimism bias. *Games and Economic Behavior 75*(1), 67–80. 4, 6

Brocas, I. and J. D. Carrillo (2000). The value of information when preferences are dynamically inconsistent. *European Economic Review 44*, 1104–1115. 3

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

Brunnermeier, M. K., F. Papakonstantinou, and J. A. Parker (2017). Optimal time-inconsistent beliefs: Misplanning, procrastination, and commitment. *Management Science 63*(5), 1318–1340. 2

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

Carrillo, J. D. and T. Mariotti (2000). Strategic ignorance as a self-disciplining device. *Review of Economic Studies 67*, 529–544. 3

Compte, O. and A. Postlewaite (2004). Confidence-enhanced performance. *American Economic Review 94*(5), 1536–1556. 3

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

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. 22

Fudenberg, D. and D. K. Levine (2006). A dual-self model of impulse control. *American Economic Review 5*(5), 1449–1476. 4

Kaur, S., M. Kremer, and S. Mullainathan (2015). Self-control at work. *Journal of Political Economy 123*(6), 1127–1277. 4

Laibson, D. (1997). Golden eggs and hyperbolic discounting. *Quarterly Journal of Economics 112*(2), 443–478. 3

Möbius, M. M., M. Niederle, P. Niehaus, and T. S. Rosenblat (2014, August). Managing self-confidence. working paper. 4

O'Donoghue, T. and M. Rabin (1999). Doing it now or later. *American Economic Review 89*(1), 103–124. 3, 7

Schwardmann, P. and J. van der Weele (2017, March). Deception and self-deception. CRC TRR 190 Discussion Paper No. 25. 4

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

Thaler, R. H. and H. M. Shefrin (1981). An economic theory of self-control. *Journal of Political Economy 89*(2), 392–406. 4

Appendix

A Additional Tables and Figures

Table 3: 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

Summary Statistics. Table 3 compares the students from our main restricted sample with all others who participated in wave 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 by Treatment Group

	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)
achieved points	53.13	16.57	53.17	17.48	0.03	(0.01)
Observations	60		24		84	

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

Table 4 shows that there are also no significant differences in observables between groups 1 and 2, suggesting that selection "into treatment" is not a major concern.[15] Remarkably, the mean exam points achieved ex post are almost identical between the two groups.[16]

Decision Relevance. In Table 5, columns 1 to 3, we report regressions of expected (or 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 returns to studying to be higher also expect to study (or reported to have studied) more.

The first column's regression includes only a group dummy indicating when the student wrote the exam, 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 (excluding that for wave 2).[17]

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 (wave 6) measures only.

B Panel and Sample Description

This appendix describes the data set in detail—first its main component, the dynamic online survey, then additional administrative data. Based on this description we make explicit how we define our main sample of participants who remain in the panel throughout the entire study, and how we define our additional sample that is used for additional results concerning only the first exam date.

B.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 page, and sent an individual link via email for each wave. Upon opening the link they could respond online (e.g., using a smartphone) to the survey questions. Failure to respond

[15] Regressing a group-1 dummy on background characteristics results in failure of significance of the F-statistic, inidcating that our observable background characteristics cannot jointly explain selection into treatment.

[16] Grading the students' exams was done by class tutors who were not involved in this research.

[17] In line with our earlier investigation of effort predictions, we also removed two students' observations here. Including them has essentially no effect, except for marginally increasing the standard errors on the estimated return coefficient.

Table 5: Robustness Check: Decision Relevance

Dep. Var.:	*Exp. or Ex-Post Reported Effort*			
	Ordinary Least Squares			
	Waves 2-6 (pooled)			Wave 6 only
Exp. Return	0.408***	0.414***	0.409***	0.416**
	[0.128]	[0.127]	[0.127]	[.198]
Group 2	2.108	2.307	2.302	0.379
	[3.254]	[3.289]	[3.308]	[3.898]
Importance		-0.384	-0.379	0.533
		[1.094]	[1.105]	[1.426]
Wave 3			3.656***	
			[1.344]	
Wave 4			0.543	
			[2.160]	
Wave 5			0.451	
			[2.048]	
Wave 6			-2.246	
			[1.948]	
Const.	32.257***	33.359***	32.959***	28.207***
	[3.158]	[4.846]	[5.225]	[5.372]
Obs.	410	410	410	82
R^2	0.0499	0.0508	0.0620	0.0452

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

to a given wave meant dropping out from also 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 6 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 consistented 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.

Table 6: 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

* This was also asked about a parallel mathematics course for which we only know that: the two exam dates were February 16 and April 13, grades for the first exam were released on March 15 and students could inspect their exams on March 18. We have no data on this course from the examination office.
** We adopted the patience and risk tolerance measures from the preference module of Falk et al. (2016), and we added our own similarly formulated item on time-consistency, namely "Do you generally keep your resolutions?"

B.2 Background Information and Administrative Data

Below is some background 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 cannot take the course's exam. Registered students can withdraw from their exam until three working days prior to it, and then cannot take the course's exam (at either date). Students registered for the first exam that either supply a sick note to the examination office for this date or failed the exam may register and then take the second exam.

- Microeconomics exam: The exam dates for microeconomics 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). 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 (betweem 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 was anonymously matched with our survey data. (No access for mathematics.)

- 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).

B.3 Sample Definitions

Main Sample (only participants that completed the full survey):

- 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.