Biased Memory and Perceptions of Self-Control

Afras Sial, Justin Sydnor, and Dmitry Taubinsky*

December 2022

Abstract

Using data from a field experiment on exercise, we analyze the relationship between imperfect memory and people's awareness of their limited self-control. We find that people overestimate past gym attendance, and that larger overestimation of past attendance is associated with (i) more overestimation of future attendance, (ii) a lower willingness to pay to motivate higher future gym attendance, and (iii) a smaller gap between goal and forecasted attendance. We organize these facts with a structural model of quasi-hyperbolic discounting and naivete, estimating that people with more biased memories are more naive about their time inconsistency, but not more time-inconsistent.

*Sial: UC Berkeley. afras@berkeley.edu. Sydnor: University of Wisconsin–Madison and NBER. jsydnor@bus.wisc.edu. Taubinsky: UC Berkeley and NBER. dmitry.taubinsky@berkeley.edu. We thank seminar and conference participants, as well as Mariana Carrera, Heather Royer, Mark Stehr, Collin Raymond, Tristan Gagnon-Bartsch, and Daniel Gottlieb for early comments on this project.

A large and growing literature, spanning many economically-consequential domains, shows that people appear to not be fully aware of their self-control problems (e.g., Acland and Levy, 2015; Chaloupka et al., 2019; Bai et al., 2021; Allcott et al., 2022a; Allcott et al., 2022b; Carrera et al., 2022)—a phenomenon the literature refers to as naivete (O'Donoghue and Rabin, 2001). This is consistent with a broader body of work on overconfidence and misprediction of own future behavior (e.g., Barber and Odean, 2001; Malmendier and Tate, 2005, 2015; DellaVigna and Malmendier, 2006; Kőszegi, 2006; Oster et al., 2013; Cheng et al., 2014; Gottlieb, 2014; Hoffman and Burks, 2020; Huffman et al., 2022).

A fundamental question is whether and how naivete and overconfidence can persist in settings where people receive repeated feedback (Ali, 2011; Heidhues et al., 2018). One leading explanation is that imperfect memory (see, e.g., Zimmermann, 2020; Enke et al., 2022, for recent economic evidence) can lead to the development of persistent biases in beliefs (e.g., Mullainathan, 2002; Gottlieb, 2014). Recent theoretical work has built on this insight to develop models in which imperfect memory facilitates overconfidence and naivete about self-control problems (e.g., Bénabou and Tirole, 2002; Gagnon-Bartsch et al., 2021; Gottlieb, 2021).

This paper provides new empirical evidence on the link between biased memory and awareness of self-control problems. Using data from a field experiment on gym attendance, we investigate how gym members' bias in recall of their past gym attendance relates to (i) their beliefs about future attendance, (ii) their willingness to pay to alter their future selves' behavior via incentives, and (iii) structural estimates of their time inconsistency and their awareness of it. The gym attendance setting is a natural one for studying the link between memory and naivete, because most gym members have significant past experience to draw on when forming beliefs about their future attendance. Carrera et al. (2022) previously used these data to study gym members' self-control and naivete, but did not investigate the relationship with biased memory.

We begin by documenting that people overestimate both past and future gym attendance, and that there is a strong, positive association between the two. This replicates, in a very different field setting, the core finding in Huffman et al. (2022) and Müller (2022) about the link between biased memories and misprediction of future behavior. The key new reduced-form and structural results in this paper link imperfect memory to naivete about limited self-control. First, we find that while goal attendance is generally higher than forecasted attendance, consistent with some awareness of imperfect self-control, those who overestimate their past attendance more believe that their future attendance will be closer to their goal. Second, those with more upwardly-biased memories display less desire to use incentives to change their future behavior. We establish this finding by utilizing the *behavior change premium* measure of Carrera et al. (2022) and Allcott et al. (2022b). This measure captures awareness of self-control problems

through the gap between a person's willingness to pay for a future incentive for gym attendance and their subjective expected earnings from this incentive. A larger behavior change premium is indicative of more awareness of self-control problems because it indicates that the person values the expected behavior change from the incentive more. We find that participants with above-median memory bias are on average willing to pay $0.98 to increase their future gym attendance by one visit, while those with below-median memory bias are willing to pay $2.51. This suggests that those with inflated memories of their past attendance perceive themselves to be less time-inconsistent and thus in less need of incentives to motivate future behavior. Third, we examine take-up of commitment contracts with no financial upside that require attending the gym a minimum number of times during a four-week period. We find that those with more upwardly biased memories are *more* likely to take up commitment contracts. This is consistent with Carrera et al.'s (2022) theoretical and empirical results that take-up of commitment contracts is *positively* associated with naivete in this gym setting.

Using these reduced-form moments, we estimate a structural model of quasi-hyperbolic preferences and naivete, allowing the parameters to vary by memory bias. We find that individuals with above- and below-median memory bias have essentially identical levels of time inconsistency.[1] However, individuals with above-median memory bias are much less aware of their time inconsistency. Specifically, those with above-median memory bias are aware of only about twenty-five percent of their degree of time inconsistency, while those with below-median memory bias are aware of approximately fifty percent of their degree of time inconsistency. We also show that other forms of misperceptions—such as over-optimism about one's future availability and hassle costs of attendance—cannot account for all of the patterns in the data.

In the rest of this paper, Sections 1-4 contain the experimental design, reduced-form results, structural estimates, and conclusion, respectively.

1 Experimental Design

This paper reports on a field experiment conducted by a research group that included Sydnor and Taubinsky, a subset of whose data was first reported in Carrera et al. (2022). The main new data introduced in this paper are people's memories of their past gym attendance. Our description of the experimental design mirrors Carrera et al. (2022). 1,292 participants were recruited from a gym associated with a private university in the Midwestern U.S. In addition to regular membership available to the general public, the gym offers subsidized memberships to

[1] Interestingly, though not directly related to our findings, Chew et al. (2020) conduct a lab experiment and find that participants with positive false memories about their past performance on a cognitive test are more likely to exhibit present focus in monetary time-discounting tasks.

graduate-student, faculty, and staff affiliates of the university and members of a health insurance company's wellness program. Participation in the study was limited to those over the age of 18 with membership lasting at least eight weeks prior to the start of the online survey component of the study. The study consisted of three waves of recruitment via email invitations and flyers between October 2015 and March 2016, avoiding long breaks in the academic calendar.

In each wave, participants first completed an online component that included questions about their next four weeks of gym attendance (starting the Monday following the online survey). Some participants were then randomly assigned an experimental incentive for attendance for those following four weeks. In order to enter the gym, members were required to swipe their membership ID cards, creating a record of their visit. Participants provided consent for us to access the attendance records associated with their membership cards, which is what we use to construct all measures of participants' attendance reported in this paper.[2]

The online component of the study consisted of information about experimental procedures and a series of questions relating to past and future gym attendance, willingness to pay (WTP) for various attendance incentives, numeracy, comprehension, attention, and demographics.[3] After providing consent, participants were first asked the following question about their prior attendance: *Please think back over the past 100 days (about 14 weeks). What is your best guess as to the number of days you went to [the gym]?* For the 83% of gym members who had maintained their membership for at least 100 days prior to the online component of the study, our measure of their memory bias is the difference between their answer and their actual attendance, divided by 100. For the 17% of members who had been members for fewer than 100 days, our measure of their memory bias is the difference between their answer and their actual attendance, divided by the number of days that they were members.[4] Participants did not receive monetary incentives for accurate recall, which was a deliberate methodological decision.[5] If the lack of incentives leads to low effort in reporting that generates noise, this would attenuate our main results since our measure of memory bias is a right-hand-side variable in our analysis. In principle, the lack of incentives could also lead to a systematic upward or downward reporting bias, though there is no obvious reason this would be the case. In our study, participants provided explicit consent to share their past and future attendance records

[2]As reported in Carrera et al. (2022), most visits lasted considerably longer than 10 minutes, suggesting that participants continued to go to the gym to exercise (rather than simply swipe to obtain an experimental financial incentive), as they presumably did prior to the experiment.

[3]See the Study Instructions Appendix for a complete outline of the online component of the study.

[4]See Appendix Figure A2 for histograms of membership durations for all participants in panel (a) and those who had been members for fewer than 100 days in panel (b).

[5]As discussed in Carrera et al. (2022) and Allcott et al. (2022b), it is not possible to incentivize truthful reporting of forecasts of future behavior when people perceive themselves to be time-inconsistent—correspondingly, we did not incentivize forecasts. We did not incentivize the memory question to keep them as comparable as possible to the forecast questions.

with the researchers, actively sharing their membership barcode to enroll in the study (see the consent form in the experimental screenshots in the Study Instructions Appendix). Thus, participants were aware that they could not affect the researchers' beliefs through misreporting. Because our main results are about the relationship between recall and other decisions (which were largely incentivized), confounds can only arise when the reporting bias in recall is also correlated with people's preferences in the other decisions.

Before proceeding to the other parts of the online survey, 50% of participants were randomly assigned to receive an information treatment. In wave 1, participants assigned to the treatment were shown a line graph of their recorded number of gym visits per week over the prior 20 weeks—this is referred to as the "basic" information treatment. In waves 2 and 3, participants received an "enhanced" information treatment: they were (i) shown the same graph as in the basic information treatment, (ii) asked to estimate the average days per week they attended the gym over the prior 20 weeks, and (iii) were told that participants in wave 1 overestimated their future attendance during the four-week experimental period by one day per week on average. In all waves, participants not assigned to the information treatment proceeded without viewing the treatment screens. The main results in this paper pool data from all participants regardless of information-treatment assignment. In Section 2.3, we discuss the information-treatment results.

In the next part of the online component, participants forecasted the number of days they would visit the gym in the four following weeks, as well as their goal attendance during that period. They were then introduced, in random order and on separate screens, to six possible incentive schemes for gym attendance during the four-week experimental period: \$1 per visit, \$2 per visit, \$3 per visit, \$5 per visit, \$7 per visit, and \$12 per visit. On the survey screen for each scheme, participants (i) forecasted the number of days they would visit the gym during the experiment under the relevant incentive, and (ii) stated their WTP for the scheme. Participants revealed their WTP by choosing the smallest fixed payment for which they would trade away the incentive scheme. They used a slider allowing responses from \$0 to 30 times the piece rate; a fill-in-the-blank question allowed them to indicate higher values if they positioned the slider at the maximum value.

Participants were then asked about their willingness to take up commitment contracts both for more and fewer gym visits. Specifically, they were asked whether they preferred an unconditional \$80 fixed payment or \$80 conditional on attending the gym at least, e.g., 12 days over the next four weeks in the case of the more-visits contracts or, e.g., 11 or fewer days in the case of the fewer-visits contracts. In waves 1 and 2, participants made decisions about contracts for 8, 12, and 16 or more visits as well as 7, 11, or 15 or fewer visits. In wave 3, participants only considered two contracts (contracts for $\geq$12 and $\leq$11 visits), and were additionally asked to choose

between $0 and $80 conditional on visiting the gym at least 12 times over the next four weeks. In this paper we focus only on the more-visits contracts; see Carrera et al. (2022) for a detailed analysis and discussion of the importance of the fewer-visits contracts.

Participants' decisions about exercise incentives were incentive-compatible. Each of the piece-rate-incentive and commitment-contract questions was selected for potential assignment to participants with positive probability. Participants for whom a commitment contract question was selected to count received their preferred of the two options. When the selected question involved a piece-rate incentive, we used the Becker-DeGroot-Marschak (BDM) mechanism, where a participant's WTP for that incentive was compared against a randomly-drawn fixed payment. If a participant's WTP was above the randomly chosen fixed payment, they would receive the piece-rate incentive. If their WTP was below the randomly chosen fixed payment, they would receive the randomly chosen fixed payment. For all piece-rate incentives and commitment contracts, participants were informed that all payments would be made after the conclusion of the four-week experimental period.

To generate random assignment of attendance incentives for the majority of participants, fixed payments in the BDM were drawn from a mixture distribution with two components: a uniform distribution from $0-$7 (mixture weight = 0.99), and a uniform distribution from the full range of slider values (mixture weight = 0.01). This guaranteed that incentives were exogenously assigned, with the exception of two rare cases that total to 44 participants that are excluded from our analysis.[6] Finally, to create an exogenously determined "control" group that did not face any incentive to visit the gym, the study also included a choice between a $0 per-visit incentive and a $20 fixed payment, and this question was chosen with 0.33 probability.[7]

The online component also included questions to check for numeracy, comprehension, and attention. Fewer than 5% of participants failed to pass each of these checks, indicating high levels of engagement and understanding.[8] The final set of questions in the online component

[6]The first case is when the fixed payment draw exceeded $7 ($n = 12$). The second case is when a participant indicated a WTP value within the $0-$7 range from which our fixed payments were heavily drawn ($n = 32$).

[7]Only 1.8% of the people chose the dominated $0 option instead of a $20 fixed payment. The probabilities of questions being selected to count varied across waves. In wave 1, the $0, $2, and $7 per-visit incentive questions were each selected with probability 0.33. In wave 2, the $0 and $2 per-visit incentive questions were again selected with probability 0.33, while the $5 and $7 per-visit incentive questions were each selected with probability 0.165. In wave 3, the $0 and $7 per-visit incentive questions and a question with the choice between $0 and the contract for $80 conditional on at least 12 visits were each selected with probability 0.33. In each wave, the remaining probability 0.01 was equally allocated across all six per-visit incentive questions and all commitment contract questions with the choice between an unconditional $80 fixed payment and $80 conditional on a certain level of gym attendance.

[8]4.8% of participants incorrectly answered at least one of two numeracy questions from Lusardi and Mitchell (2007). 1.8% of participants failed one attention check which involved not choosing a strictly dominated option from a pair of options, and 3.5% failed a second attention check involving clicking to continue to the next survey screen without selecting any option in a multiple-choice question. 4.3% incorrectly answered two questions regarding comprehension of the WTP elicitation procedure.

of the study asked about participant demographic characteristics. Only 6 participants declined to answer at least one of the optional demographic questions; these participants are excluded from the parts of our analyses that use demographic controls. Our final sample consists of 1,247 participants,[9] of which 61% are female, 57% are full-time students, the mean imputed age is 34 years old, and the mean duration of membership is 998 days.

2 Reduced-Form Results

2.1 Graphical Results

While people's recollection of their past gym attendance closely tracks their actual past attendance, there is systematic bias in participants' memories. Figure 1a presents a binned scatter plot that compares participants' actual likelihood of visiting the gym on a given day in the 100 days prior to the study to their reported recollection of that daily visit likelihood. If participants were unbiased, then the relationship between their memory of attendance and actual past attendance would be on the dashed 45-degree line. Instead, while the best-fit line is parallel to the 45-degree line, participants on average overestimate their past attendance. The difference between the best-fit line and the 45-degree line is approximately constant across past attendance, suggesting that bias in memory is independent of past performance in this setting.

Appendix Figure A1 presents a histogram of memory bias, showing that on average, memories of past gym attendance are biased upwards. Slightly over one-third of the participants correctly remember their past visit likelihood within 5 percentage points. Of the remaining participants with larger errors, 90 percent overestimate their past attendance.

In the rest of the reduced-form results section, we investigate how memory bias relates to proxies for naivete about self-control. Because exercise tends to involve immediate costs and delayed health benefits, limited self-control would reduce gym attendance, and naivete about these limitations would lead people to overestimate their future gym attendance. The first proxy of naivete that we study is thus the gap between forecasted and actual attendance. Figure 1b presents a binned scatter plot comparing participants' memory bias with their forecast bias. Our measure of forecast bias is the difference between the forecasted daily likelihood of visiting the gym during the four-week attendance experiment and the actual daily visit likelihood during that period. Because different people were randomly assigned different attendance incentives, to construct the figure we use forecasted attendance at the assigned incentive level. On average, people overestimate their future gym attendance, and there is a strong positive

[9] Our sample consists of one fewer participant than that of Carrera et al. (2022) since we could not reliably match one participant to pre-study attendance records to compare actual and remembered prior attendance.

association between memory bias and forecast bias.

Figure 2 plots forecasted and actual visits, at each incentive level, for participants with above- versus below-median memory bias. Participants in the above-median memory bias group have a more positive forecast of their future attendance than those in the below-median memory bias group, while actual attendance during the experiment is similar across the two groups. This suggests that memory bias is associated with forecast bias rather than with preferences for gym attendance or self control.

2.2 Regression Analysis

Column 1 of Table 1 quantifies the relationship between forecast bias and memory bias. We find that a 10 percentage-point increase in a participant's overestimation of their past daily attendance likelihood is associated with a 3 percentage-point increase in their bias in forecasted future daily attendance likelihood, and this is highly statistically significant.

Columns 2-4 study three other proxies of awareness of present focus. Column 2 studies people's perceptions of how much they will fall short of their goal attendance. This measure is a proxy for the gap between people's desired attendance—which is the attendance that would be attained by a time-consistent future self—and the attendance they expect given their beliefs about the degree of time inconsistency of their future self. The mean gap between goal and forecasted daily attendance likelihood is 12 percentage points, which suggests some awareness of time inconsistency. We find that a 10 percentage-point increase in a participant's memory bias about past attendance is associated with a 0.7 percentage-point decrease in the gap between goal and forecasted future daily attendance likelihood.

In column 3 we introduce a measure of desire to change one's future self's behavior, the *behavior change premium* (BCP), as formulated by Carrera et al. (2022) and Allcott et al. (2022b). The BCP is how much participants are willing to pay for the behavior change induced by a marginal increase in their per-visit incentive, and is a measure of a person's *perceived* time inconsistency. Following Carrera et al. (2022) and Allcott et al. (2022b), the BCP at per-visit incentive p and an increment in the per-visit incentive Δ is defined as

$$BCP(p,\Delta) := \underbrace{\frac{w(p+\Delta)-w(p)}{\Delta}}_{\substack{\text{WTP per dollar}\\ \text{of incentive}}} - \underbrace{\frac{\tilde{\alpha}(p+\Delta)+\tilde{\alpha}(p)}{2}}_{\substack{\text{Forecasted earnings}\\ \text{per dollar of incentive}}} \tag{1}$$

where $w(\cdot)$ is the WTP for a given incentive and $\tilde{\alpha}(\cdot)$ is forecasted attendance under a given incentive. The first term is the increase in WTP for attendance incentives, per dollar increase in the per-visit incentive. The second term is the average of forecasted attendance under the

original per-visit incentive and the slightly higher incentive. Carrera et al. (2022) and Allcott et al. (2022b) show that the BCP is increasing in the degree of perceived time inconsistency, and that for individuals who perceive themselves to be time-consistent, $BCP(p,\Delta) \leq 0$ and $\lim_{\Delta \to 0} BCP(p,\Delta) = 0$. Intuitively, the Envelope Theorem implies that a time-consistent person should be willing to pay $\tilde{\alpha}(p)dp$ for a marginal change dp in incentives. For small but non-marginal changes, a second-order approximation implies a time-consistent person should be willing to pay $\Delta(\tilde{\alpha}(p+\Delta) + \tilde{\alpha}(p))/2$ for a Δ increase in incentives. A WTP above the time-consistent benchmark implies that the person places a premium on the behavior change induced by the incentive. See Appendix B.2 for further details.

Because we can compute the BCP for different levels of incentives p, in column 3 we use each person's average BCP value across different levels of incentives. Specifically, for two adjacent incentive levels p_k and p_{k+1}, we compute $BCP(p_k, p_{k+1} - p_k)$, and then take the average across all such pairs of incentives. Column 3 shows a strong, negative association between the BCP and memory bias. A 10 percentage-point increase in a participant's bias in recalling their past daily attendance likelihood is associated with a 44-cent decrease in their BCP (a 37% reduction). Thus, those with larger positive biases in their memory tend to express less desire for behavior change, which is consistent with a lack of awareness of self-control problems. Appendix Figure A3 provides a binned scatter plot of the BCP by memory bias, Appendix Table A4 additionally controls for the forecasted change in attendance per dollar increase in incentive and fully replicates column 3, and Appendix Table A5 shows that there is no statistically significant association between memory bias and noise in the BCP estimate.

A key advantage of the BCP is that it is a belief-free measure of people's (perceived) time-inconsistency: it mechanically controls for people's perceived earnings by subtracting out beliefs $\tilde{\alpha}$. This is in contrast to typical commitment contract designs, where whether a person wishes to commit to a penalty for falling short of a target depends on the person's beliefs about the likelihood of incurring the penalty. We study how take-up of such commitment contracts relates to memory bias in Column 4. Theoretically, take-up of commitment contracts is not necessarily increasing in perceived time inconsistency because more naive individuals believe that they have a higher likelihood of follow-through and thus a lower chance of incurring the contract penalty. Carrera et al. (2022) show that under a wide range of economic conditions that are plausible in our setting, more naive individuals will in fact have *higher* take-up of commitment contracts because they are more optimistic about avoiding the penalty. Consistent with this, column 4 shows that memory bias is significantly positively associated with commitment contract take-up.

Robustness Appendix Table A1 includes two variations of Table 1 with different controls. Appendix Table A2 includes the maximum number of observations available in each column, instead of using a constant sample. Appendix Table A3 studies an indicator for above-median memory bias as the dependent variable. The results are largely the same as in Table 1.

Appendix A.3 implements the method developed Oster (2019), building on work by Altonji et al. (2005), to quantify potential bias from the omission of unobservable controls. The bias-adjusted coefficient estimates are largely similar to those in Table 1.

2.3 Information treatments

Appendix A.4 studies the interaction between memory bias and the information treatments. One hypothesis is that people want to have maximally accurate memories and information, and thus will update from freely-available information about their past. Under this hypothesis, information about the past would help people make more accurate forecasts about the future, and the effect would be strongest for people with the largest memory bias. A second hypothesis is that imperfect memory is motivated, and that people may attempt to maintain (positively) biased beliefs about themselves by ignoring "debiasing" information. Under this hypothesis, there would be no interaction between the information treatments and past behavior. A third possibility is that our treatments did not present information in a sufficiently easy-to-process manner, in which case there would also be no interaction with memory bias.

We find that the effects of the information treatments on the dependent variables in Table 1 are largely orthogonal to memory bias. In fact, the basic information treatment that simply showed participants a line graph of their past weekly attendance over the prior twenty weeks had no average effect. Only the information treatment that additionally asked people to re-estimate their past attendance from that graph and informed participants that people in a prior wave of the experiment overestimated their future gym attendance had a significant average effect on forecasts and decisions about incentives. The fact that this enhanced information treatment changed forecasts and decisions but did not interact with memory bias suggests that the treatment largely operates through the information that most people overestimated their future attendance, rather than through improving the accuracy of beliefs about past attendance. These results are consistent with motivated beliefs as in Bénabou and Tirole (2002), where people avoid, ignore, or forget unfavorable information about their past behavior, or with Gagnon-Bartsch et al. (2021), where people simply ignore information about their past behavior if they do not believe that it can update their beliefs. The fact that an information treatment affected forecast bias without interacting with memory bias suggests that biased memory is not the only channel that facilitates naivete about self-control.

3 Structural Estimates

3.1 Structural Model and Identification

Building on Carrera et al. (2022), we structurally estimate a model of quasi-hyperbolic discounting with imperfect perception. We assume that participants face immediate stochastic costs c_t of going to the gym on any day t and receive a fixed, delayed health benefit b from each visit at the conclusion of the experimental period. We assume that costs are on net always non-negative and distributed independently and identically according to the exponential distribution with mean $1/\lambda$.[10] We assume quasi-hyperbolic preferences with present focus parameter $\beta \in [0,1]$ applied to future utility flow. As in O'Donoghue and Rabin (2001), people may be partially naive and perceive their present focus in the future to be $\tilde{\beta} \in [\beta, 1]$. Formally, in period t, people evaluate a stream of instantaneous utility flows u_τ by $U^t(u_t, u_{t+1}, ..., u_T, u_{T+1}) = u_t + \beta \sum_{\tau=t+1}^{T+1} u_\tau$, but believe their time $t' > t$ self uses the discount factor $\tilde{\beta}$ to form discounted expected utility $U^{t'}$.[11] Each day in the experiment corresponds to one period in time $t \in \{1, \ldots, T = 28\}$, and all financial incentives and delayed health benefits are received at time $T+1$. The participant visits the gym when the perceived discounted benefits of the gym visit exceed the costs: when $\beta(b+p) > c_t$. Given per-visit incentive p, these assumptions imply that the forecasted and actual attendance curves are given by $\tilde{\alpha}(p) = 28 \cdot \left[1 - e^{-\lambda\tilde{\beta}(b+p)}\right]$ and $\alpha(p) = 28 \cdot \left[1 - e^{-\lambda\beta(b+p)}\right]$, respectively.

Estimates of the BCP and the forecasted and actual attendance curves identify the parameters β, $\tilde{\beta}$, b, and λ. Proposition 1 of Carrera et al. (2022) (duplicated in Appendix B.2) shows that up to negligible higher-order terms, the BCP can be approximately expressed as a function of the structural parameters as follows:

$$BCP(p, \Delta) \approx (1 - \tilde{\beta})(b + p + \Delta/2) \frac{\tilde{\alpha}(p + \Delta) - \tilde{\alpha}(p)}{\Delta}. \tag{2}$$

Intuitively, the BCP is increasing in (i) perceived time inconsistency $1 - \tilde{\beta}$, (ii) the average of per-attendance benefits at incentives p and $p + \Delta$, and (iii) the perceived behavior change, $\tilde{\alpha}(p + \Delta) - \tilde{\alpha}(p)$. The intuition for identification is then as follows. Delayed benefits b are identified from the projected intersection of the forecasted and actual attendance curves, $\tilde{\alpha}(p)$ and $\alpha(p)$. This is because $\tilde{\alpha}(p) = \alpha(p)$ at $p = -b$. With b identified, $\tilde{\beta} - \beta$ is identified from the difference between the forecasted and actual attendance curves ($\tilde{\alpha}(p)$ and $\alpha(p)$), and $\tilde{\beta}$ is identified from

[10] This assumption on the cost distribution does not rule out the possibility of stochastic benefits from going to the gym (e.g., for socializing or entertainment). It requires that the sum of costs (e.g., the hassle costs associated with transportation) be larger than those benefits. Carrera et al. (2022) consider alternative assumptions and find that an exponential distribution with a cost floor of zero is most consistent with the data.

[11] More generally, there could also be exponential discounting: $U^t(u_t, u_{t+1}, ..., u_T, u_{T+1}) = \delta^t u_t + \beta \sum_{\tau=t+1}^{T+1} \delta^\tau u_\tau$. For simplicity, we set $\delta = 1$, as the periods in our context are short.

the BCP statistic. With $\tilde{\beta}$ and $\tilde{\beta}-\beta$ identified, β is clearly identified as well. Finally, the rate parameter λ is identified by the slopes of $\tilde{\alpha}(p)$ and $\alpha(p)$. Appendix C.1 formally describes our estimating equations and the generalized method of moments (GMM) approach to obtain the estimates. We cluster standard errors at the participant level.

Including other forms of misprediction While the baseline model assumes that all misprediction of future behavior is due to naivete about β, we can enrich the model to add misforecasting of the future costs (and benefits) of attendance, such as how busy one is in the future. In our framework, this corresponds to individuals misperceiving the cost parameter λ as $\tilde{\lambda}$. This parameter is identified by the slope of the perceived attendance curve $\tilde{\alpha}(p) = 28 \cdot \left[1 - e^{-\tilde{\lambda}\tilde{\beta}(b+p)}\right]$. However, because actual attendance $\alpha(p) = 28 \cdot \left[1 - e^{-\lambda\beta(b+p)}\right]$ is determined only by the product $\lambda\beta$ and not by each of the two parameters separately, these two parameters are not separately identified. In our main results, we fix β at the estimate from the baseline model, and we examine robustness to this assumption in the appendix. See Appendix C.1.1 for further details.

3.2 Results

Table 2a presents parameter estimates of our baseline model, by below- versus above-median overestimation of past attendance. Columns 1 and 2 report estimates of β and $\tilde{\beta}$, the actual and perceived present focus parameters, respectively. Columns 3 and 4 report estimates of b and $1/\lambda$, the perceived health benefit and mean cost of a gym visit, respectively. Column 5 reports a measure of naivete suggested by Augenblick and Rabin (2019): the fraction of present focus $1-\beta$ that individuals are aware of. Consistent with reduced-form results in Section 2, parameters that determine gym attendance—β, b, and λ—do not meaningfully vary with memory bias. However, the perceived present focus parameter $\tilde{\beta}$—which affects forecasted attendance and the BCP—is significantly higher in the above-median memory bias group. Appendix Table A8 reports parameter estimates by quartile of memory bias, with largely the same conclusions.

Table 2b shows that the estimated model matches the empirical moments well. The estimated model almost perfectly matches average actual attendance and average misprediction of actual attendance. Appendix Figure A4 shows the tight in-sample fit of model predictions to forecasted and actual attendance curves. The model does not fully capture the difference in the average BCP between the above- and below-median-memory-bias groups, but the model's prediction is within the confidence interval of the empirically-estimated difference. The BCP is slightly over-estimated for the above-median-memory-bias group and more significantly for the below-median group. This mismatch is potentially due to our baseline model understating the degree of heterogeneity. The more heterogeneous model in Appendix Table A8 produces

different estimates of the BCP that better match the difference in empirical moments, without any changes to predictions about actual and forecasted attendance.

Table 3 clarifies how our reduced-form results about the negative association between the BCP and memory bias imply that misperceptions of β must vary with memory bias. Table 3a reports estimates of a model where $\tilde{\beta}$ is assumed homogeneous across the two memory-bias groups, but cost perceptions $\tilde{\lambda}$ are potentially heterogeneous. Table 3b and Appendix Figure A5 show that while heterogeneous misperception of costs can account for our result that attendance misprediction varies with memory bias, it cannot account for our result that the BCP also varies with memory bias. Intuitively, this is because equation (2) shows that once people's perceived elasticity with respect to incentives is controlled for, the BCP reflects only perceptions of time-inconsistency $\tilde{\beta}$, and not perceptions of future behavior (and Appendix Table A4 shows that controlling for perceived response to incentives does not alter how the BCP varies with memory bias). Thus, our reduced-form results about the association between memory bias and the BCP require perceptions of time inconsistency to vary with memory bias.

Consistent with Table 3, Appendix Table A9 shows that when both $\tilde{\beta}$ and $\tilde{\lambda}$ are allowed to vary by memory bias, all heterogeneity of misperceptions loads on heterogeneity in $\tilde{\beta}$, and $\tilde{\lambda}$ is estimated to approximately equal λ for both memory bias groups. Despite additional parameters, the model fit is not better than in the baseline of Table 2.

Additional results and robustness To achieve identification in a model with misperceptions of costs without restricting assumptions on parameter values, we can estimate the product $\lambda\beta$ in place of each parameter separately. Appendix Table A10 reports parameter estimates and predicted moments from this model, and again shows that allowing for heterogeneity in misperceptions of $\tilde{\lambda}$ does not improve model fit relative to the baseline model in Table 2.

Appendix Table A11 presents a seven-parameter version of the model in Table 2, under the assumption that β is homogenous across memory bias groups. The model fit in Appendix Table A11 remains superior to that of the seven-parameter model in Table 3, further supporting the baseline modeling assumptions.

4 Conclusion

This paper contributes new evidence of a link between memory biases and awareness of self-control limitations. This provides empirical support for recent theoretical models in which biases in learning and memory formation support persistent overconfidence and naivete. At the same time, our results should not be interpreted to suggest that biased memory is the only channel that supports naivete in settings where people have ample opportunities to learn. For

example, as we noted in Section 2.3, an information treatment that informed gym members that participants in our study tend to be over-optimistic seems to have reduced naivete through non-memory-based channels.

An implication of our results is that if naivete is at least partly linked to biases in memory, then it is likely that the degree of naivete is context-dependent. Memory distortions are probably more likely in some environments than others, due to factors such as ego-related motivations, availability of clear and salient feedback, and incentives for maintaining accurate records and beliefs.

References

Acland, Dan, and Matthew R. Levy. 2015. "Naiveté, Projection Bias, and Habit Formation in Gym Attendance." *Management Science* 61 (1): 146–160.

Ali, Nageeb. 2011. "Learning Self-Control." *The Quarterly Journal of Economics* 126 (2): 857–893.

Allcott, Hunt, Matthew Gentzkow, and Lena Song. 2022a. "Digital Addiction." *American Economic Review* 112 (7): 2424–2463.

Allcott, Hunt, Joshua J. Kim, Dmitry Taubinsky, and Joshua Zinman. 2022b. "Are High-Interest Loans Predatory? Theory and Evidence from Payday Lending." *The Review of Economic Studies* 89 (3): 1041–1084.

Altonji, Joseph, Todd Elder, and Christopher Taber. 2005. "An Evaluation of Instrumental Variable Strategies for Estimating the Effects of Catholic Schooling." *Journal of Human Resources* 40 791–821.

Augenblick, Ned, and Matthew Rabin. 2019. "An Experiment on Time Preference and Misprediction in Unpleasant Tasks." *The Review of Economic Studies* 86 (3): 941–975.

Bai, Liang, Benjamin Handel, Edward Miguel, and Gautam Rao. 2021. "Self-Control and Demand for Preventive Health: Evidence from Hypertension in India." *The Review of Economics and Statistics* 103 (5): 835–856.

Barber, Brad M., and Terrance Odean. 2001. "Boys Will Be Boys: Gender, Overconfidence, and Common Stock Investment." *The Quarterly Journal of Economics* 116 (1): 261–292.

Bénabou, Roland, and Jean Tirole. 2002. "Self-Confidence and Personal Motivation." *The Quarterly Journal of Economics* 117 (3): 871–915.

Carrera, Mariana, Heather Royer, Mark Stehr, Justin Sydnor, and Dmitry Taubinsky. 2022. "Who Chooses Commitment? Evidence and Welfare Implications." *The Review of Economic Studies* 89 (3): 1205–1244.

Chaloupka, Frank J., IV, Matthew R. Levy, and Justin S. White. 2019. "Estimating Biases in Smoking Cessation: Evidence from a Field Experiment." NBER Working Paper No. 26522.

Cheng, Ing-Haw, Sahil Raina, and Wei Xiong. 2014. "Wall Street and the Housing Bubble." *American Economic Review* 104 (9): 2797–2829.

Chew, Soo Hong, Wei Huang, and Xiaojian Zhao. 2020. "Motivated False Memory." *Journal of Political Economy* 128 (10): 3913–3939.

DellaVigna, Stefano, and Ulrike Malmendier. 2006. "Paying Not to Go to the Gym." *American Economic Review* 96 (3): 694–719.

Enke, Benjamin, Frederik Schwerter, and Florian Zimmermann. 2022. "Associative Memory and Belief Formation." Working Paper.

Gagnon-Bartsch, Tristan, Matthew Rabin, and Joshua Schwartzstein. 2021. "Channeled Attention and Stable Errors." Working Paper.

Gottlieb, Daniel. 2014. "Imperfect Memory and Choice Under Risk." *Games and Economic Behavior* 85 127–158.

Gottlieb, Daniel. 2021. "Will You Never Learn? Self Deception and Biases in Information Processing." Working Paper.

Hall, Alistair R. 2005. *Generalized Method of Moments*. Oxford University Press.

Hansen, Lars Peter. 1982. "Large Sample Properties of Generalized Method of Moments Estimators." *Econometrica* 50 (4): 1029–1054.

Heidhues, Paul, Botond Kőszegi, and Philipp Strack. 2018. "Unrealistic Expectations and Misguided Learning." *Econometrica* 86 (4): 1159–1214.

Hoffman, Mitchell, and Stephen V. Burks. 2020. "Worker Overconfidence: Field Evidence and Implications for Employee Turnover and Firm Profits." *Quantitative Economics* 11 (1): 315–348.

Huffman, David, Collin Raymond, and Julia Shvets. 2022. "Persistent Overconfidence and Biased Memory: Evidence from Managers." *American Economic Review* 112 (10): 3141–3175.

Kőszegi, Botond. 2006. "Ego Utility, Overconfidence, and Task Choice." *Journal of the European Economic Association* 4 (4): 673–707.

Lusardi, Annamaria, and Olivia S. Mitchell. 2007. "Baby Boomer Retirement Security: The Roles of Planning, Financial Literacy, and Housing Wealth." *Journal of Monetary Economics* 51 (1): 205–224.

Malmendier, Ulrike, and Geoffrey Tate. 2005. "CEO Overconfidence and Corporate Investment." *The Journal of Finance* 60 (6): 2661–2700.

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

Mullainathan, Sendhil. 2002. "A Memory-Based Model of Bounded Rationality." *The Quarterly Journal of Economics* 117 (3): 735–774.

Müller, Maximilian. 2022. "Selective Memory Around Big Life Decisions." Working Paper.

O'Donoghue, Ted, and Matthew Rabin. 2001. "Choice and Procrastination." *The Quarterly Journal of Economics* 116 (1): 121–160.

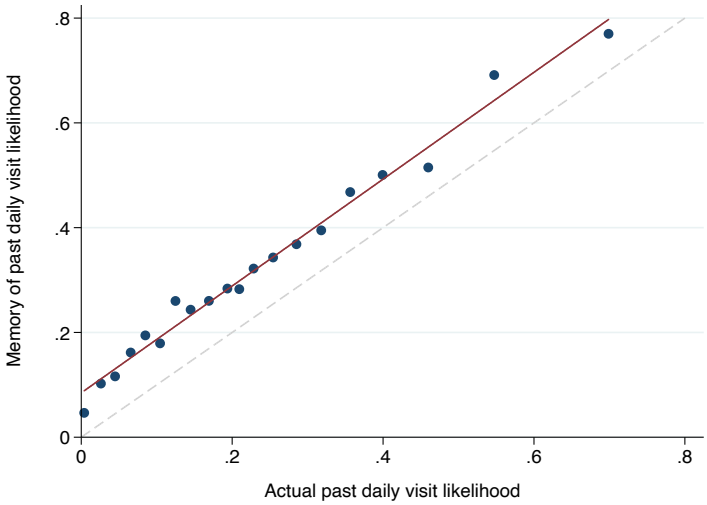
Oster, Emily. 2019. "Unobservable Selection and Coefficient Stability: Theory and Evidence." *Journal of Business and Economic Statistics* 37 (2): 187–204.

Oster, Emily, Ira Shoulson, and E. Ray Dorsey. 2013. "Optimal Expectations and Limited Medical Testing: Evidence from Huntington Disease." *American Economic Review* 103 (2): 804–30.

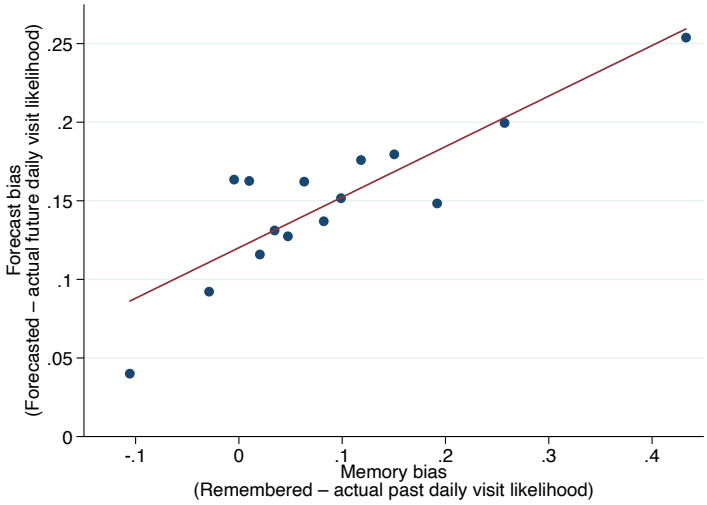
Zimmermann, Florian. 2020. "The Dynamics of Motivated Beliefs." *American Economic Review* 110 (2): 337–361.

Figure 1: Memory of daily likelihood of visiting the gym

(a) Actual vs. remembered past daily visit likelihood

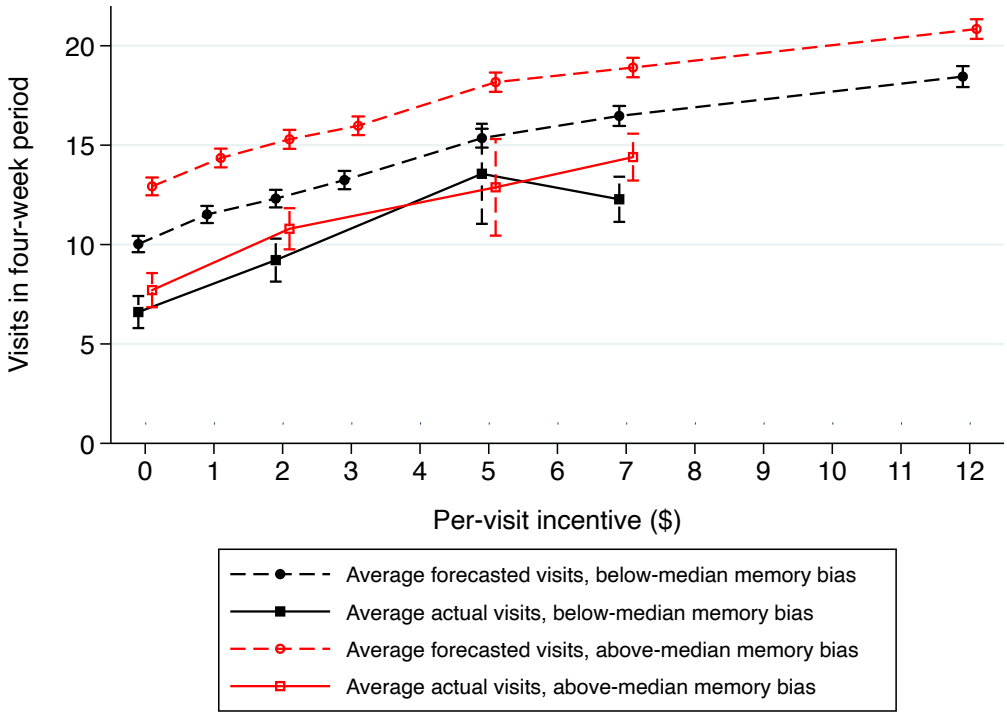


(b) Biases in forecasts vs. memories of daily visit likelihood



Notes: Panel (a) of this figure presents a binned scatterplot comparing participants' actual past daily likelihood of visiting the gym and their recollection of their past daily likelihood of visiting the gym. Actual past daily likelihood is the fraction of days—either out of the past 100 or out of the total membership duration when the duration was lower than 100—on which the participant attended the gym. Remembered past daily likelihood is defined analogously, but using the participant's recollection in the numerator. A dashed 45-degree line is included for reference. Panel (b) of this figure presents a binned scatterplot comparing participants' memory bias and forecast bias. Memory biases are defined on a scale from 0 to 1 as the difference between participants' recalled and actual visit likelihood. Forecast biases are similarly defined as the difference between participants' forecasted daily likelihood of visiting the gym during the four-week experimental period under the incentive randomly assigned to them and their actual daily likelihood of visiting the gym during that period. Panel (a) includes the full sample, while panel (b) excludes 122 participants assigned a commitment contract since forecasted attendance under commitment contracts was not elicited.

Figure 2: Forecasted and actual attendance by memory bias and incentive



Notes: This figure compares the means and 95% confidence intervals of participants' subjective forecasts of gym visits during the four-week experimental period, and their actual attendance under their assigned per-visit incentive for the subsamples with below- and above-median memory bias. As in Figure 1, memory bias is the difference between recalled past daily likelihood of visiting the gym and the actual daily likelihood of visiting the gym. The sample median of this difference is 0.06. Forecasted visits are averaged over all participants in each subsample, while actual visits are averaged over the participants within each subsample who were randomly assigned each per-visit incentive. The incentive levels were probabilistically targeted differently in each wave, so the sample sizes for the average actual visits statistics differ across incentive levels ($0: N= 413; $2: N= 293; $5: N= 75; $7: N= 341).

Table 1: Awareness of present focus by memory bias

	Forecasted – actual attendance	Goal – forecasted attendance	Behavior change premium	Take-up of "more" visits contract
	(1)	(2)	(3)	(4)
Memory bias	0.30***	–0.07**	–4.36***	0.24***
	(0.06)	(0.03)	(1.63)	(0.09)
Dependent var. mean	0.15	0.12	1.17	0.49
	(0.01)	(0.00)	(0.22)	(0.01)
Past attendance control	Yes	Yes	Yes	Yes
Demographic controls	Yes	Yes	Yes	Yes
Information fixed effects	Yes	Yes	Yes	Yes
Wave fixed effects	Yes	Yes	Yes	Yes
Incentive fixed effects	Yes	No	No	No
Contract fixed effects	No	No	No	Yes
N	1,119	1,119	1,119	2,807
Clusters	1,119	1,119	1,119	1,119

Notes: This table reports the association between memory bias and attendance-related proxies for naivete, the estimated behavior change premium, and take-up of commitment contracts. As in Figure 1, memory bias is the difference between recalled past daily likelihood of visiting the gym and the actual daily likelihood of visiting the gym. Each column presents coefficient estimates from OLS regressions and dependent variable means, with standard errors reported in parentheses. In column 1, the dependent variable is the difference between participants' forecasted attendance under their assigned incentive and actual attendance. In column 2, the dependent variable is the difference between participants' goal and forecasted attendance in the absence of incentives. The dependent variables in columns 1-2 are expressed in terms of the daily visit likelihood (i.e., aggregate attendance divided by the 28 days of the experiment). In column 4, observations are pooled across the three types of visit-threshold contracts, with standard errors clustered at the participant level. In all columns, "demographic controls" include gender, student status, age, and the natural log of membership duration. A "past attendance control" is also included as participants' daily visit likelihood in the 100 days immediately preceding the experiment. The sample excludes 122 participants assigned a commitment contract since forecasted attendance under commitment contracts was not elicited, and 6 participants who declined to state their gender or age. *, **, ***: statistically significantly different from 0 at the 10%, 5%, and 1% level, respectively.

Table 2: Model with naivete about present focus

(a) Parameter estimates

	Memory bias	(1) $\hat{\beta}$	(2) $\hat{\tilde{\beta}}$	(3) $\hat{b}$	(4) $1/\hat{\lambda}$	(5) $\frac{(1-\hat{\tilde{\beta}})}{(1-\hat{\beta})}$
1	Below med. (N=563)	0.54 (0.48, 0.59)	0.78 (0.72, 0.85)	9.05 (8.24, 9.86)	15.38 (13.48, 17.27)	0.47 (0.36, 0.57)
2	Above med. (N=562)	0.55 (0.51, 0.59)	0.89 (0.85, 0.93)	10.04 (9.14, 10.94)	14.07 (12.65, 15.48)	0.25 (0.17, 0.33)
3	Difference	−0.01 (−0.08, 0.05)	−0.11 (−0.18, −0.03)	−0.99 (−2.20, 0.22)	1.31 (−1.06, 3.68)	0.22 (0.09, 0.35)

(b) Empirical and model-predicted moments

		Memory bias	(1) Behavior change premium ($)	(2) Actual attendance (likelihood)	(3) Forecasted – actual attend. (likelihood)
1	Empirical	Below med. (N=563)	1.81 (1.11, 2.50)	0.34 (0.32, 0.36)	0.12 (0.10, 0.14)
2		Above med. (N=562)	0.53 (0.06, 1.01)	0.39 (0.37, 0.41)	0.17 (0.16, 0.19)
3		Difference	1.27 (0.43, 2.11)	−0.05 (−0.08, −0.02)	−0.05 (−0.08, −0.02)
4	Predicted	Below med. (N=563)	2.07 (1.45, 2.69)	0.34 (0.32, 0.36)	0.11 (0.10, 0.13)
5		Above med. (N=562)	1.12 (0.70, 1.54)	0.39 (0.37, 0.41)	0.16 (0.14, 0.17)
6		Difference	0.95 (0.20, 1.70)	−0.05 (−0.08, −0.03)	−0.04 (−0.07, −0.02)

Notes: Panel (a) of this table reports parameter estimates and 95% confidence intervals for two subsamples, split at the median memory bias. Memory biases are defined as in Table 1. The present focus parameter is denoted by β, the perceived present focus parameter by $\tilde{\beta}$, the perceived health benefits of a gym visit by b, and the mean costs of a gym visit by $1/\lambda$. Standard errors are clustered at the participant level. Inference for the statistics in columns 4-5 and the last row is conducted using the Delta method. See Appendix C.1 for details about the GMM estimation procedure. Panel (b) reports empirical and model-predicted means, differences in means, and 95% confidence intervals for three moments in the same subsamples as in panel (a). In columns 1, 2, and 3, the moments of interest are the average behavior change premium, actual attendance during the experiment, and the difference between forecasted attendance under assigned incentives and actual attendance, respectively. Rows 1-3 report the empirical means and 95% confidence intervals, while rows 4-6 report the model-predicted moments using the parameter estimates in panel (a). Inference for the statistics in rows 4-6 is conducted using the Delta method. The sample excludes 122 participants assigned a commitment contract since forecasted attendance under commitment contracts was not elicited.

Table 3: Model with naivete about present focus and misperceptions of costs, homogeneous perceived present focus

(a) Parameter estimates

	Memory bias	(1) $\hat{\beta}$	(2) $\hat{\tilde{\beta}}$	(3) $\hat{b}$	(4) $1/\hat{\lambda}$	(5) $1/\hat{\tilde{\lambda}}$
1	Below med. (N=563)	0.54 By assump.	0.85 (0.82, 0.89)	9.46 (8.61, 10.31)	15.82 (14.34, 17.29)	17.14 (15.52, 18.76)
2	Above med. (N=562)	0.55 By assump.	0.85 (0.82, 0.89)	9.68 (8.83, 10.52)	13.79 (12.54, 15.03)	13.24 (11.98, 14.50)
3	Difference	−0.01 By assump.	0 By assump.	−0.22 (−1.40, 0.97)	2.03 (0.11, 3.95)	3.91 (2.15, 5.66)

(b) Empirical and model-predicted moments

		Memory bias	(1) Behavior change premium ($)	(2) Actual attendance (likelihood)	(3) Forecasted – actual attend. (likelihood)
1	Empirical	Below med. (N=563)	1.81 (1.11, 2.50)	0.34 (0.32, 0.36)	0.12 (0.10, 0.14)
2		Above med. (N=562)	0.53 (0.06, 1.01)	0.39 (0.37, 0.41)	0.17 (0.16, 0.19)
3		Difference	1.27 (0.43, 2.11)	−0.05 (−0.08, −0.02)	−0.05 (−0.08, −0.02)
4	Predicted	Below med. (N=563)	1.41 (1.07, 1.75)	0.34 (0.32, 0.36)	0.11 (0.10, 0.13)
5		Above med. (N=562)	1.49 (1.13, 1.84)	0.39 (0.37, 0.41)	0.16 (0.14, 0.17)
6		Difference	−0.07 (−0.10, −0.05)	−0.05 (−0.08, −0.02)	−0.04 (−0.07, −0.02)

Notes: Panel (a) of this table modifies panel (a) of Table 2 by allowing the actual mean costs of a gym visit to differ from the perceived mean costs of a gym visit. The present focus parameter β is set equal to the values estimated in Table 2. The perceived present focus parameter $\tilde{\beta}$ is restricted to be constant across the two memory bias groups. Panel (b) of this table is analogous to panel (b) of Table 2.

Online Appendix

Biased Memory and Perceptions of Self-Control

Afras Sial, Justin Sydnor, and Dmitry Taubinsky

Table of Contents

A Reduced-Form Results **2**

A.1 Distributions of Memory Bias and Membership Duration 2

A.2 Additional Results on Awareness of Present Focus by Memory Bias 4

A.3 Bias-Adjusted Estimates of the Effects of Memory Bias 8

A.4 Further Results on Information Treatments 13

B Theoretical Results from Carrera et al. (2022) **15**

B.1 Predictions on Commitment Contract Take-up 15

B.2 Behavior Change Premium Derivation 18

C Structural Results **20**

C.1 Details on GMM Estimation of Parameters 20

C.2 Additional Structural Estimates of Baseline Present Focus Model 22

C.3 Additional Structural Estimates with Misperceptions of Costs 24

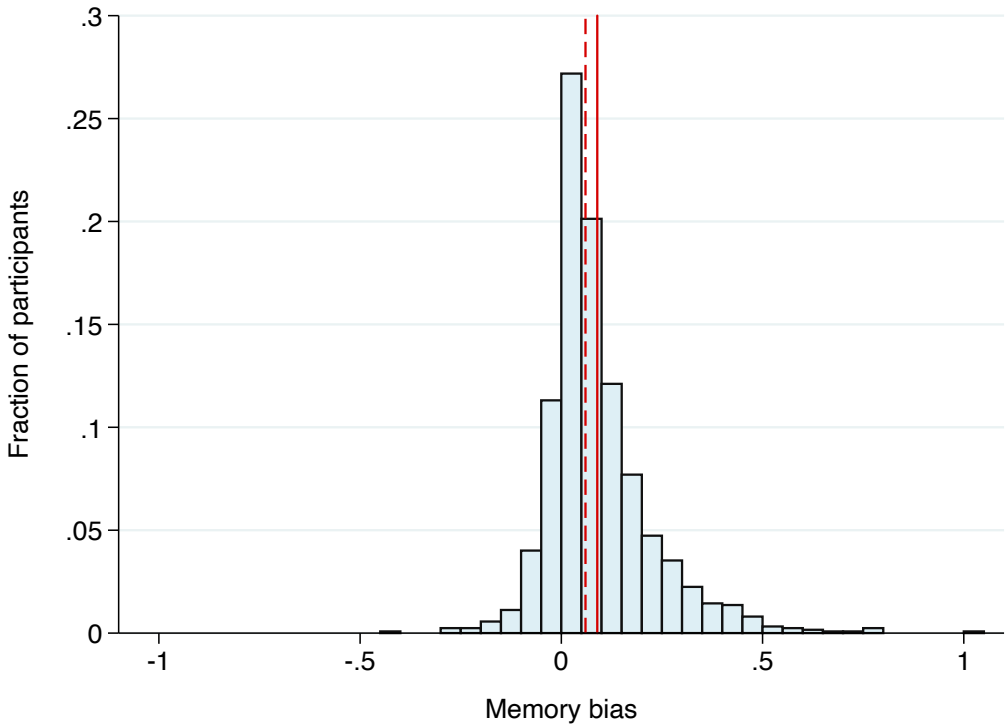
C.4 Additional Structural Estimates Under a Homogeneity Assumption 26

C.5 In-Sample Fit of Structural Models 29

A Reduced-Form Results

A.1 Distributions of Memory Bias and Membership Duration

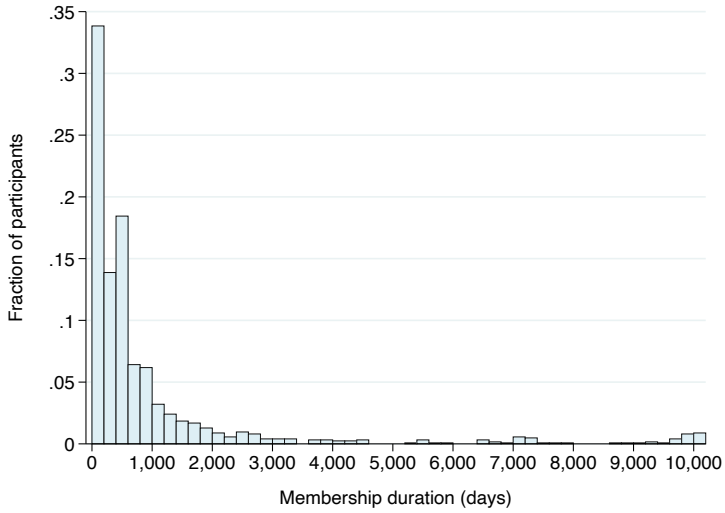
Figure A1: Histogram of biases in memory of attendance likelihood



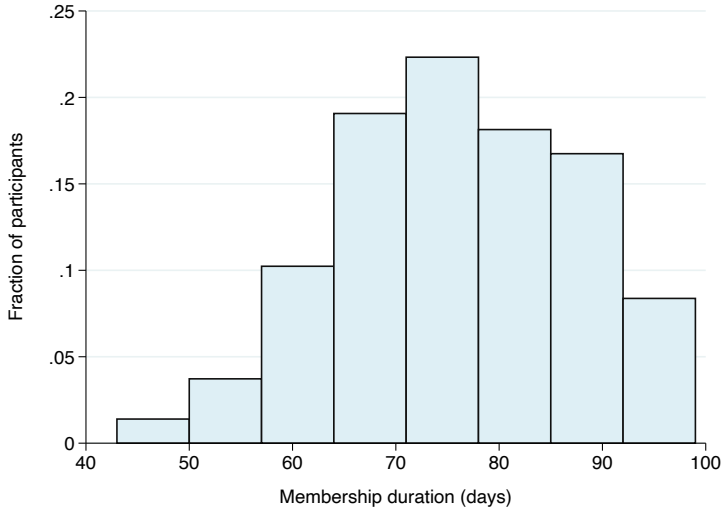
Notes: This figure presents a histogram showing the distribution of memory biases in our sample. As in Figure 1, memory bias is the difference between recalled past daily likelihood of visiting the gym and the actual daily likelihood of visiting the gym. The dashed and solid red lines indicate the median and mean memory biases, respectively. Memory bias exceeds one when a participant remembers attending the gym for a greater number of days than their recorded membership duration.

Figure A2: Histograms of membership duration

(a) All participants



(b) Membership duration <100 days



Notes: This figure presents histograms showing the distribution of membership duration in our sample. Panel (a) includes all participants, while panel (b) is restricted to participants who had been members for fewer than 100 days prior to completing the online survey component of the study.

A.2 Additional Results on Awareness of Present Focus by Memory Bias

Table A1: Awareness of present focus by memory bias, alternative controls

(a) Fixed effects only

	Forecasted – actual attendance	Goal – forecasted attendance	Behavior change premium	Take-up of "more" visits contract
	(1)	(2)	(3)	(4)
Memory bias	0.31***	–0.04	–4.14**	0.30***
	(0.06)	(0.03)	(1.65)	(0.09)
Dependent var. mean	0.15	0.12	1.17	0.49
	(0.01)	(0.00)	(0.22)	(0.01)
Past attendance control	No	No	No	No
Demographic controls	No	No	No	No
Information fixed effects	Yes	Yes	Yes	Yes
Wave fixed effects	Yes	Yes	Yes	Yes
Incentive fixed effects	Yes	No	No	No
Contract fixed effects	No	No	No	Yes
N	1,119	1,119	1,119	2,807
Clusters	1,119	1,119	1,119	1,119

(b) No controls

	Forecasted – actual attendance	Goal – forecasted attendance	Behavior change premium	Take-up of "more" visits contract
	(1)	(2)	(3)	(4)
Memory bias	0.32***	–0.04	–4.47***	0.29***
	(0.06)	(0.03)	(1.68)	(0.09)
Dependent var. mean	0.15	0.12	1.17	0.49
	(0.01)	(0.00)	(0.22)	(0.01)
Past attendance control	No	No	No	No
Demographic controls	No	No	No	No
Information fixed effects	No	No	No	No
Wave fixed effects	No	No	No	No
Incentive fixed effects	No	No	No	No
Contract fixed effects	No	No	No	No
N	1,119	1,119	1,119	2,807
Clusters	1,119	1,119	1,119	1,119

Notes: This table modifies the controls in Table 1 by including only fixed effects—omitting the past attendance and demographic controls—in panel (a) and omitting all controls in panel (b).

Table A2: Awareness of present focus by memory bias, full samples

	Forecasted − actual attendance	Goal − forecasted attendance	Behavior change premium	Take-up of "more" visits contract
	(1)	(2)	(3)	(4)
Memory bias	0.30***	–0.08***	–3.89**	0.21**
	(0.06)	(0.03)	(1.52)	(0.09)
Dependent var. mean	0.15	0.12	1.19	0.49
	(0.01)	(0.00)	(0.20)	(0.01)
Past attendance control	Yes	Yes	Yes	Yes
Demographic controls	Yes	Yes	Yes	Yes
Information fixed effects	Yes	Yes	Yes	Yes
Wave fixed effects	Yes	Yes	Yes	Yes
Incentive fixed effects	Yes	No	No	No
Contract fixed effects	No	No	No	Yes
N	1,119	1,241	1,241	2,929
Clusters	1,119	1,241	1,241	1,241

Notes: This table replicates Table 1 with the maximum number of participants available in each column, instead of using a constant sample across all columns. The sample in each column excludes the 6 participants who declined to state their gender or age. The sample in column 1 also excludes the 122 participants assigned a commitment contract since forecasted attendance under commitment contracts was not elicited.

Table A3: Awareness of present focus, above- vs. below-median memory bias

	Forecasted – actual attendance	Goal – forecasted attendance	Behavior change premium	Take-up of "more" visits contract
	(1)	(2)	(3)	(4)
Above-med. memory bias	0.05***	–0.01	–1.22***	0.07***
	(0.01)	(0.01)	(0.40)	(0.03)
Dependent var. mean	0.15	0.12	1.17	0.49
	(0.01)	(0.00)	(0.22)	(0.01)
Past attendance control	Yes	Yes	Yes	Yes
Demographic controls	Yes	Yes	Yes	Yes
Information fixed effects	Yes	Yes	Yes	Yes
Wave fixed effects	Yes	Yes	Yes	Yes
Incentive fixed effects	Yes	No	No	No
Contract fixed effects	No	No	No	Yes
N	1,119	1,119	1,119	2,807
Clusters	1,119	1,119	1,119	1,119

Notes: This table modifies Table 1 by binarizing the measure of memory bias along its median. The sample median memory bias is 0.06, where memory bias is defined as in Table 1 as the difference between recalled past daily likelihood of visiting the gym and the actual daily likelihood of visiting the gym.

A.2.1 The Behavior Change Premium and Memory Bias

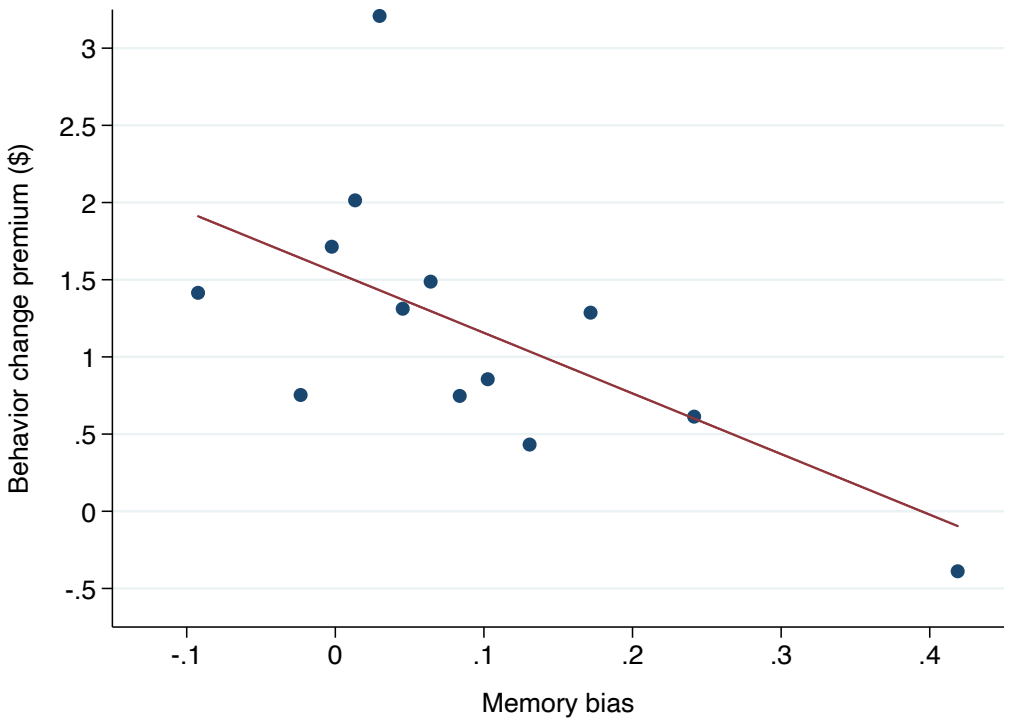
Appendix Figure A3 presents a binned scatterplot comparing participants' average behavior change premium and their memory bias. It shows that there is an approximately linear relationship between the two variables, providing justification for the regression specification in column 3 of Table 1. In addition, it shows that the negative association between the behavior change premium and upward memory bias holds across a large range of memory bias values.

Appendix Table A4 reproduces the results from regressions of the average behavior change premium on memory bias with various controls from the columns 3 of Table 1, Appendix Table A1a, and Appendix Table A1b and compares them to results from an alternative specification. In the alternative specification, we include an additional control for the forecasted change in attendance per dollar increase in incentive (i.e., the "forecasted attendance elasticity"). This variable corresponds to the term $\frac{\tilde{\alpha}(p+\Delta)-\tilde{\alpha}(p)}{\Delta}$ in equation (2), so controlling for it aids in isolating the effect of memory bias on perceived present focus, $1-\tilde{\beta}$. Confirming the robustness of our main result in Table 1, the coefficients on memory bias are not statistically distinguishable across the columns with different controls.

Given the relatively large within-participant variation in our measure of the behavior change premium across incentives, one might be concerned about omitted variable bias stemming from this noise. For example, participants with noisier measures of the behavior change pre-

mium might have lower levels of comprehension and numeracy that causes less informative reporting within the survey, biasing down our calculated approximation of their behavior change premium towards zero. These same participants might also have more biased memories of their past attendance, driving the negative association between the behavior change premium and memory bias. Appendix Table A5 addresses this potential concern by presenting results from regressions of the within-participant standard deviation of the behavior change premium on memory bias. It fails to find any statistically significant associations, indicating that the aforementioned potential source of omitted variable bias is not of concern.

Figure A3: Memory bias and the behavior change premium



Notes: This figure presents a binned scatterplot comparing participants' estimated behavior change premium and their memory bias. As in Figure 1, memory bias is the difference between recalled past daily likelihood of visiting the gym and the actual daily likelihood of visiting the gym. See Sections 2.2 and 3.1 and Appendix B.2 for additional information about the behavior change premium.

Table A4: Control for forecasted attendance elasticity in BCP regressions

	Behavior change premium					
	(1)	(2)	(3)	(4)	(5)	(6)
Memory bias	–4.36***	–3.98**	–4.14**	–3.91**	–4.47***	–4.27**
	(1.63)	(1.65)	(1.65)	(1.66)	(1.68)	(1.69)
$\frac{\Delta \text{ forecasted attendance}}{\Delta \text{ incentive}}$		1.24***		1.51***		1.47***
		(0.25)		(0.25)		(0.25)
Dependent var. mean	1.17	1.17	1.17	1.17	1.17	1.17
	(0.22)	(0.22)	(0.22)	(0.22)	(0.22)	(0.22)
Past attendance control	Yes	Yes	No	No	No	No
Demographic controls	Yes	Yes	No	No	No	No
Information fixed effects	Yes	Yes	Yes	Yes	No	No
Wave fixed effects	Yes	Yes	Yes	Yes	No	No
N	1,119	1,119	1,119	1,119	1,119	1,119

Notes: This table reproduces the columns 3 from Table 1, Appendix Table A1a, and Appendix Table A1b in columns 1, 3, and 5, respectively. It also modifies the columns 3 from each of those tables by controlling for the average forecasted change in attendance per dollar increase in incentives in columns 2, 4, and 6, respectively.

Table A5: Noise in behavior change premium measure and memory bias

	Standard deviation of behavior change premium		
	(1)	(2)	(3)
Memory bias	2.37	2.20	2.17
	(1.79)	(1.78)	(1.83)
Dependent var. mean	7.88	7.88	7.88
	(0.24)	(0.24)	(0.24)
Past attendance control	Yes	No	No
Demographic controls	Yes	No	No
Information fixed effects	Yes	Yes	No
Wave fixed effects	Yes	Yes	No
N	1,119	1,119	1,119

Notes: This table modifies the columns 3 of Table 1, Appendix Table A1a, and Appendix Table A1b in columns 1, 2, and 3, respectively, by replacing the average behavior change premium across incentive levels with the standard deviation of the behavior change premium.

A.3 Bias-Adjusted Estimates of the Effects of Memory Bias

We implement the method developed by Oster (2019) to obtain consistent estimates of the effect of memory bias—our treatment of interest—adjusted for bias from the omission of unob-

served confounders.[12] Intuitively, Oster's method leverages movements in both the R^2 and the coefficient estimate on the treatment as controls are added to the OLS regression of the dependent variable on the treatment. The method uses these movements to estimate how the coefficient on the treatment would change if all relevant unobservables were added to the regression such that the maximum possible R^2 (henceforth R^2_{max}) was achieved. In that case, the resulting coefficient would not suffer from omitted variable bias. We call an estimate of this coefficient the "bias-adjusted estimate" of the treatment effect.

Oster's method requires the assumption of a proportional relationship between selection on observables and unobservables. This framework involves the following regression model:

$$Y = \beta X + \psi\omega^o + W_2 + \epsilon, \tag{3}$$

where Y is the dependent variable of interest, X is the treatment of interest, and ω^o is a vector of observables. In our setting, X is the variable *Memory bias*. Define $W_1 := \psi\omega^o$; W_2 is the unobserved analogue of W_1 and is orthogonal to it. ϵ is an error uncorrelated with X, W_1, or W_2. Define δ as the coefficient of proportionality. The assumed proportional selection relationship is:

$$\delta\frac{\text{cov}(W_1, X)}{\text{var}(W_1)} = \frac{\text{cov}(W_2, X)}{\text{var}(W_2)}. \tag{4}$$

Oster explains that $\delta = 1$, which indicates that "the unobservable and observables are equally related to the treatment," is likely to be an appropriately conservative upper bound. Since researchers tend to focus on collecting data on what they perceive as the most relevant control variables and W_2 is residual of those controls, we expect observables to be more related to the treatment than unobservables (i.e., $\delta \leq 1$). We adopt the assumption that $\delta = 1$ throughout our implementation of Oster's method.

To implement Oster's method, we must obtain coefficient estimates and R^2 values from "uncontrolled" and "controlled" regressions of the dependent variable on the treatment of interest. The uncontrolled regression is a regression of Y on X alone, while the controlled regression additionally controls for ω^o. Define $\mathring{\beta}$ and $\mathring{R}^2$ as the coefficient estimate on X—*Memory bias* in our setting—and the R^2 from the uncontrolled regression, respectively. Furthermore, define $\tilde{\beta}$ and $\tilde{R}^2$ as the coefficient estimate on X and the R^2 from the controlled regression, respectively.

Oster's method also requires an assumption about the value of R^2_{max}. One could follow the assumption made in prior related work, like of Altonji et al. (2005), that if all of the unobservables were observed and included in the controlled regression, the R^2 would equal 1.[13] How-

[12] The method does not consider bias from other sources, such as misspecification.

[13] Altonji et al. (2005) propose a formal notion of robustness. They use a consistent estimator for the ratio of (i) the covariance between the treatment and unobservables to (ii) the covariance between treatment and observables that would result in a treatment effect of zero. They suggest that one consider a result with a ratio above 1 to be

ever, as Oster explains and tests using results from randomized studies, this assumption may often be overly conservative and produce misleading bias-adjusted coefficient estimates. Instead, she proposes using the value $R^2_{max} = 1.3\tilde{R}^2$. At this value, 90% of the bias-adjusted coefficient estimates from the randomized studies she considers have the same sign as $\tilde{\beta}$ and are within the 99.5% confidence interval of $\tilde{\beta}$; only 45% of the estimates from the nonrandomized studies she considers meet these criteria.[14]

In addition to adopting the benchmark value $1.3\tilde{R}^2$ for R^2_{max} in our analysis for all of our dependent variables Y, we consider a more conservative alternative for our attendance-related dependent variables, *Forecasted – actual attendance* and *Goal – forecasted attendance*. We obtain the R^2 from a regression of the actual daily likelihood of attendance during the 4-week experiment on a cubic function of the daily likelihood of attendance in the preceding 100 days; demographic controls comprising of gender, student status, age, and the natural log of membership duration in days; information treatment, wave, and per-visit incentive fixed effects; and the interactions of the demographic controls and fixed effects with the daily likelihood of attendance in the preceding 100 days. This regression produces an R^2 of 0.50, which we consider to be a conservative upper bound on R^2_{max}, as the differences between goal, forecasted, and actual attendance are likely to be more difficult to predict than actual attendance itself.

Under the assumption $\delta = 1$, Oster's Corollary 1 defines a set of two elements $\beta^\star$, one which converges in probability to β, the true treatment effect. Specifically, $\beta^\star = \{\tilde{\beta} - \nu_1, \tilde{\beta} - \nu_2\}$, where

$$\nu_1 = \frac{-\Theta - \sqrt{\Theta^2 + 4\big((R^2_{max} - \tilde{R}^2)\text{var}(Y)\big)\big(\mathring{\beta} - \tilde{\beta}\big)^2\text{var}(X)^2\tau_X}}{-2\tau_X\big(\mathring{\beta} - \tilde{\beta}\big)\text{var}(X)}, \tag{5}$$

$$\nu_2 = \frac{-\Theta + \sqrt{\Theta^2 + 4\big((R^2_{max} - \tilde{R}^2)\text{var}(Y)\big)\big(\mathring{\beta} - \tilde{\beta}\big)^2\text{var}(X)^2\tau_X}}{-2\tau_X\big(\mathring{\beta} - \tilde{\beta}\big)\text{var}(X)}, \tag{6}$$

and where $\Theta = (((R^2_{max} - \tilde{R}^2)\text{var}(Y))(\text{var}(X) - \tau_X) - ((\tilde{R}^2 - \mathring{R}^2)\text{var}(Y))\tau_X - \text{var}(X)\tau_X(\mathring{\beta} - \tilde{\beta})^2)$ and τ_X is the variance of the residual from a regression of X on ω^o.[15]

Under one additional assumption, we can select a unique value from $\beta^\star$ as the bias-adjusted treatment effect. Define $\hat{W}_1$ as the analogue of W_1 from a regression of Y on X and ω^o alone. We assume that $\text{Sign}(\text{cov}(X, W_1)) = \text{Sign}(\text{cov}(X, \hat{W}_1))$. In other words, the bias from unobservables is small enough that controlling for them does not switch the direction of covariance between the observable index and X.

robust.

[14] For comparison, only 42% and 37% of estimates from the randomized studies meet the sign change and confidence interval criteria, respectively, under the assumption $R^2_{max} = 1$.

[15] See Oster's Proposition 2 for the set of three possible values for the bias-adjusted treatment effect when $\delta \neq 1$.

Oster extends the method to consider an additional set of "unrelated controls" m in the regression model:

$$Y = \beta X + \psi \omega^o + m + W_2 + \epsilon, \tag{7}$$

where m is orthogonal to ω^o, W_2, and ϵ, and the covariance between m and X is unrelated to the covariance between X and ω^o and X and W_2. Regressing all variables on m and using the residuals returns us to the previously described set-up. As a result, we can include m in both the uncontrolled and controlled regressions and residualize X on m before computing var(X) and τ_X to implement the previously described procedure in the presence of m.

Appendix Table A6 reports uncontrolled, controlled, and bias-adjusted coefficient estimates and the relevant R^2 values in columns 1, 2, and 3, respectively, using Oster's method under the aforementioned assumptions. We include as unrelated controls fixed effects for the information treatments and, when relevant, per-visit incentives and commitment contract attendance thresholds. The set of observed controls in the controlled regressions includes all other control variables included in the regressions in Table 1. In rows 1, 3, 5, and 6, $R^2_{max} = 1.3\tilde{R}^2$, and the bias-adjusted coefficient estimates are relatively close to the controlled regression coefficient estimates. In all rows—including rows 2 and 4 where we use our conservative upper bound on R^2_{max}—the sign on the bias-adjusted coefficient estimate is the same as on the controlled regression coefficient estimate, affirming the robustness of our controlled regression results.

Table A6: Bias-adjusted estimates of the effect of memory bias

	Dependent variable	(1) Coeff. [$\mathring{R}^2$], uncontrolled	(2) Coeff. [$\tilde{R}^2$], controlled	(3) Coeff. [R^2_{max}], bias-adjusted
1	Forecasted – actual attendance (N=1,119)	0.31 [0.04]	0.30 [0.10]	0.29 [0.13]
2	Forecasted – actual attendance (N=1,119)	0.31 [0.04]	0.30 [0.10]	0.19 [0.50]
3	Goal – forecasted attendance (N=1,119)	−0.04 [0.01]	−0.07 [0.06]	−0.08 [0.08]
4	Goal – forecasted attendance (N=1,119)	−0.04 [0.01]	−0.07 [0.06]	−0.30 [0.50]
5	Behavior change premium (N=1,119)	−4.13 [0.01]	−4.36 [0.04]	−4.47 [0.05]
6	Take-up of "more" visits contract (N=2,807)	0.29 [0.07]	0.24 [0.08]	0.16 [0.11]

Notes: This table reports coefficient estimates on memory bias and the corresponding R^2 values in brackets from the regressions noted in each column. Column 1 reports values from "uncontrolled" regressions: OLS regressions of the dependent variable noted in each row on memory bias, with fixed effects for information treatments, assigned per-visit incentives in rows 1 and 2 only, and commitment contract visit thresholds in row 6 only. Column 2 reports values from "controlled" regressions: OLS regressions of the dependent variable noted in each row on memory bias, with the same controls as in the corresponding regressions in Table 1. Column 3 reports bias-adjusted coefficient estimates obtained using Oster's (2019) method under the assumption that R^2_{max}—the maximum attainable R^2—takes the value reported in brackets. Oster's method is implemented under the assumption that the proportional selection parameter δ is 1, and that the bias from unobservables is small enough that controlling for them does not switch the sign on the covariance between the observable index and memory bias. In rows 1,3, 5, and 6, R^2_{max} equals 1.3 times the R^2 from column 2, a heuristic value proposed by Oster. In rows 2 and 4, R^2_{max} equals the R^2 from a regression of participants' actual daily likelihood of visiting the gym during the experiment on a cubic function of (i) their actual daily likelihood of visiting the gym in the preceding 100 days; (ii) "demographic controls," which include gender, student status, age, and the natural log of membership duration; (iii) information treatment, wave, and incentive fixed effects; and (iv) the interactions of the demographic controls and fixed effects with the actual daily likelihood of visiting the gym in the preceding 100 days. In row 6, observations are pooled across the three types of visit-threshold contracts. The sample excludes 122 participants assigned a commitment contract since forecasted attendance under commitment contracts was not elicited, and 6 participants who declined to state their gender or age.

A.4 Further Results on Information Treatments

In the online survey component of the study, immediately after participants reported their best guess of their attendance in the past 100 days, and before they reported forecasts of their future attendance, they had a 50% chance of receiving an information treatment. In wave 1, participants who received the information treatment were shown a line graph of their actual number of gym visits in each week over the prior 20 weeks. We refer to this treatment as the *basic* information treatment. In waves 2 and 3, participants received an *enhanced* information treatment. First, they were shown the same graph as in the basic information treatment. Then, they were asked to report the average days per week they attended the gym over the prior 20 weeks using the information from the graph. This question was designed to make participants process the information on the graph. Lastly, they were told that, on average, participants in wave 1 overestimated their future attendance during the four-week experimental period by one day per week. In all waves, participants not assigned to the information treatment proceeded directly from the screen with the question about memory of attendance in the past 100 days to the screen with questions regarding forecasts of future attendance.

In Appendix Table A7, we examine the relationship between proxies for awareness of present focus and memory bias, receipt of information, and their interaction. We study how the different types of engagement with information about past attendance in the basic and enhanced information treatment groups impact participants' beliefs and whether the size of these effects varies with memory bias. For a successful information treatment that causes participants to fully correct their beliefs in accordance with reality, we would expect larger effects on those with greater memory bias, in light of the results in Table 1. If instead participants largely ignore the graphical information provided in the treatments and are primarily affected by the statement that participants in wave 1 overestimated their future attendance by one day per week, we would expect only the enhanced information treatment to affect beliefs. Furthermore, we might expect the effect of the enhanced information treatment to be orthogonal to memory bias since the statement does not relate overestimation of future attendance to any participant characteristics associated with past attendance and is the same for all participants.

To simplify our analysis, we first discretize our measure of memory bias into an indicator for above-median vs. below-median memory bias. Appendix Table A3 presents a version of the results in Table 1 with the indicator for above-median memory bias. In all columns, the results are qualitatively similar to those in Table 1. Forecast bias and the take-up of commitment contracts for more visits increase in memory bias, while the gap between desired and forecasted attendance and the behavior change premium decrease in memory bias. In column 2, however, the coefficient on our measure of memory bias is no longer statistically significant, while the coefficients remain statistically significant at the 1% level in the other three columns.

Appendix Table A7 modifies the regressions in each column of Appendix Table A3 by adding as regressors indicators for receipt of each information treatment and their interaction with the indicator for above-median memory bias. In each column, the estimated coefficient on *above-median memory bias* is similar to the estimate in Appendix Table A3, although in column 2, the coefficient estimate is now zero rather than slightly negative. The estimated coefficients on the indicator for receipt of the basic information treatment and its interaction with *above-median memory bias* are both not statistically distinguishable from zero in all columns. Thus, we do not find a significant effect of simply providing graphical information about past attendance on awareness of present focus.

In columns 1 and 3, however, consistent with the results reported in Carrera et al. (2022), we find that the enhanced information treatment significantly decreased participants' forecast bias and increased the behavior change premium, respectively. Both of these results are consistent with the enhanced information treatment increasing awareness of present focus. Nevertheless, in each column, the effect of the interaction between the enhanced information treatment and above-median memory bias is not statistically distinguishable from zero. Together, these results suggest that the effect of the information treatments on participant beliefs is largely orthogonal to memory bias. In addition, these results suggest that the effect of the enhanced information treatment is likely attributable in large part to its statement that participants in a previous wave of the experiment overestimated their future attendance by one visit per week on average.

As a further look into the effect of the enhanced information treatment, we analyze how the participants who were randomly selected to receive that treatment revised their recall of prior attendance. We use participants' estimates of their average weekly attendance over the past 20 weeks from the enhanced information treatment in waves 2 and 3 to compute their *revised* memory bias. We define a participant's revised memory bias as the difference between their reported daily likelihood of visiting the gym and their actual daily likelihood of visiting the gym in the 20 weeks prior to the experiment. We compare participants' baseline memory bias—computed using their best guess of attendance in the past 100 days, as reported prior to receiving the information treatment—to their revised memory bias. We find that participants in the enhanced information treatment reduced their memory bias by 1.3 percentage points (SE 0.6 percentage points) on average, from a baseline average of 7.7 percentage points.

The fact that participants only had a relatively modest reduction in memory bias after observing a graph of their past attendance could indicate that they struggled to process the graphical information and update their beliefs, similarly to how subjects initially struggled to form correct beliefs from their memories. It is also consistent with motivated beliefs as in Bénabou and Tirole (2002); the same motivations that lead participants to incorrectly process their memories initially could lead them to largely disregard the new information. Channeled attention

as in Gagnon-Bartsch et al. (2021) could also explain this result if participants simply do not pay attention to the information for the purpose of updating their beliefs. Overall, these results provide additional support for the hypothesis that the main effect of the enhanced information treatment came through the statement that prior participants overestimated their future attendance and not through a correction of memory bias.

Table A7: Awareness of present focus, interaction between memory bias and information treatments

	Forecasted − actual attendance	Goal − forecasted attendance	Behavior change premium	Take-up of "more" visits contract
	(1)	(2)	(3)	(4)
Above-med. memory bias	0.05***	0.00	–0.96**	0.09**
	(0.02)	(0.01)	(0.48)	(0.04)
Basic info. treatment	0.00	–0.02	–0.30	–0.01
	(0.03)	(0.02)	(0.68)	(0.05)
Basic info. treatment	–0.03	–0.02	1.10	–0.00
× above-med. memory bias	(0.04)	(0.02)	(0.93)	(0.07)
Enhanced info. treatment	–0.06**	0.00	1.93**	–0.06
	(0.02)	(0.01)	(0.96)	(0.04)
Enhanced info. treatment	0.01	–0.02	–1.24	–0.06
× above-med. memory bias	(0.03)	(0.02)	(1.09)	(0.06)
Dependent var. mean	0.15	0.12	1.17	0.49
	(0.01)	(0.00)	(0.22)	(0.01)
Past attendance control	Yes	Yes	Yes	Yes
Demographic controls	Yes	Yes	Yes	Yes
Information fixed effects	Yes	Yes	Yes	Yes
Wave fixed effects	Yes	No	No	No
Contract fixed effects	No	No	No	Yes
N	1,119	1,119	1,119	2,807
Clusters	1,119	1,119	1,119	1,119

Notes: This table modifies Table 1 by binarizing the measure of memory bias along its median and interacting the resulting indicator for above-median memory bias with indicators for receiving each information treatment. The sample median memory bias is 0.06, where memory bias is defined as in Table 1 as the difference between recalled past daily likelihood of visiting the gym and the actual daily likelihood of visiting the gym. See Section 1 for a description of the basic and enhanced information treatments.

B Theoretical Results from Carrera et al. (2022)

B.1 Predictions on Commitment Contract Take-up

The following discussion on theoretical predictions related to naivete about self-control problems and commitment contract take-up is largely reproduced from Section 2 of Carrera et al.

(2022).

In Appendix A.2.2, Carrera et al. (2022) derive two general results about the demand for commitment contracts when costs are uncertain. First, they show that for a broad class of stochastic cost distributions, the quasi-hyperbolic model predicts that there should not be demand for *any* commitment contract when there is at least a moderate chance that costs exceed delayed benefits. Second, when there is enough uncertainty to make commitment contracts unattractive, the perceived harms of a commitment contract are *increasing* in perceived present focus $1-\tilde{\beta}$. That is, people who perceive themselves to be more present-focused will find commitment contracts less attractive (i.e., more harmful).

Carrera et al. (2022) also show in Appendix A.2.2 that there are two key conditions on the distribution of cost draws under which the value of commitment contracts is eroded. First, the chances of getting a cost draw under which it is suboptimal to take the action ($c_t > b$) must be at least as high as the chances of getting a cost draw under which the time $t = 0$ individual thinks she should visit the gym but thinks that her time $t = 1$ self will not do so. Second, the cost draws exceeding b must not be concentrated in a "small" neighborhood of b.

As a simple numerical illustration for the case of $T = 1$, suppose that c_t is uniformly distributed on $[0,1]$. Then, it can be shown that no individuals with $\tilde{\beta} \geq 0.8$ desire any kind of commitment contract when the costs of attendance exceed the benefits at least 20% of the time—an arguably modest degree of uncertainty. Appendix A.2.2 of Carrera et al. (2022) presents additional examples.

In light of these results, a natural question is why we see *so much* take-up of commitment contracts in behavioral economics experiments. One possible reason is that because evaluating incentive schemes may be complicated, individuals may do so imperfectly. This is in line with a long intellectual history of measuring and modeling stochastic valuation errors in individuals' decisions. Carrera et al. (2022) refer to this mechanism as imperfect perception. Another reason is that some individuals simply like to say "yes" to offers, feel pressure to do so, or falsely assume that the authority offering the contracts must be offering something valuable. Carrera et al. (2022) incorporate such social pressure effects into their model in their Appendix A.2.3, and derive results under more general assumptions that allow for these effects.

They formalize this with a reduced-form econometric model that supposes that for a given choice-set j, individual i behaves as if her forecasted utility under contract (y,P)—where y is a fixed transfer and P is a contingent reward for certain levels of gym attendance—is

$$\widehat{V}(y,P) = V(y,P) + \sigma(P)\varepsilon_{ij} \tag{8}$$

where $V(y,P)$ denotes an individual's subjective expectation of utility under contract (y,P) given

beliefs $\tilde{\beta}$, ε_{ij} has unbounded support, and $\sigma(P) > \sigma(0)$ when $P \neq 0$—i.e., the presence of contingent incentives amplifies complexity and thus stochastic errors. They allow (but do not require) $\sigma(0) = 0$, meaning that individuals have no problems assessing sure incentives. The assumption that P affects the error term only through the variance guarantees that the error term is mean-zero; this is a key assumption of this model, and is typical in standard "random utility" models. For short, they refer to this framework as the *imperfect perception model*.

The take-up of commitment contracts is a particularly problematic measure in the presence of imperfect perception because binary take-up decisions are biased by even mean-zero valuation errors. Even if the errors are symmetric—say 10% of the individuals always choose the wrong option—binary choice data will typically introduce bias. For example, if 10% of choices are mistakes, and only 5% of people actually want a given option, 14% will still end up choosing that given option.

As Carrera et al. (2022) show formally in Appendix A.2.3, the imperfect perception model generates three predictions for penalty-based commitment contracts:

1. Individuals will demand commitment contracts to both exercise more and to exercise less.

2. As long as the average $\tilde{\beta}$ is not too far below 1, there will be a positive correlation between take-up of commitment contracts to exercise more and take-up of commitment contracts to exercise less.

3. In the presence of moderate to high uncertainty about costs, increasing individuals' sophistication about their present focus will decrease their demand for commitment contracts to exercise more.[16]

The intuition for the first prediction is that an extreme enough draw of ε can lead individuals to mistakenly choose undesirable contracts. The intuition for the second prediction is that if commitment contracts would generally look unappealing to individuals in the absence of valuation errors, then individuals with the highest variance in the stochastic valuation term ε will be the most likely to take up both types of contracts. The intuition for the third prediction is that under moderate to large uncertainty, the perceived harms of a commitment contract are decreasing in $\tilde{\beta}$ in the standard quasi-hyperbolic model (see Appendix A.2.2 of Carrera et al., 2022). Although in the standard quasi-hyperbolic model these conditions would lead individuals to never choose a commitment contract, in the imperfect perception model individuals

[16] Interestingly, the converse does not hold for the "less" contracts. Intuitively, this is because a lower $\tilde{\beta}$ dampens the impact of financial incentives in both cases, and thus makes penalty-based contracts potentially more harmful in both cases.

still choose the contract, but with a propensity that is decreasing in the expected harms in the standard model.

The third prediction above from the imperfect perception model justifies our interpretation of the result in column 4 of Table 1. We interpret the statistically significant, positive association between the take-up of commitment contracts for more gym visits and upward bias in memories of past attendance as evidence for decreasing awareness of present focus with upward bias in memory.

B.2 Behavior Change Premium Derivation

To allow this paper to be self-contained, this appendix contains the behavior change premium derivation from Carrera et al. (2022). We largely utilize the same text as in Carrera et al. (2022).

We consider individuals who in periods $t = 1,\ldots,T$ have the option to take an action $a_t \in \{0,1\}$. Choosing $a_t = 1$ generates immediate stochastic costs c_t realized in period t as well as deterministic delayed benefits b realized in period $T+1$. We assume that $c_t > 0$ with positive probability, but don't preclude the possibility of draws $c_t < 0$. For concreteness, we will often refer to $a_t = 1$ as attending the gym and $a_t = 0$ as not attending the gym, with the understanding that our results apply to the general model presented here and not just gym attendance.

For $\bar{a} = \sum_{t=1}^{T} a_t$, we consider incentive contracts that pay out in $T+1$, denoted as $(y, P(\bar{a}))$, that consist of a fixed transfer y (which could be negative), and a contingent reward $P(\bar{a})$ for certain levels of gym attendance. The contingent component $P(\bar{a})$ is non-negative, with $\min_{\bar{a}\in[0,T]} P(\bar{a}) = 0$. We assume for simplicity that utility is quasilinear in money, given the relatively modest incentives involved in our experiment. A piece-rate incentive contract with per-visit incentive p has $y = 0$ and $P(\bar{a}) = p\bar{a}$.

Individuals have quasi-hyperbolic preferences given by $U^t(u_t, u_{t+1}, \ldots, u_T, u_{T+1}) = \delta^t u_t + \beta\sum_{\tau=t+1}^{T+1}\delta^\tau u_\tau$, where u_t is the period t utility flow. By construction, $u_t = -a_t \cdot c_t$ for $1 \le t \le T$ and $u_{T+1} = y + b\bar{a} + P(\bar{a})$. We allow individuals to mispredict their preferences: in period t, they believe that their period $t+1$ self will have a short-run discount factor $\tilde{\beta} \in [\beta, 1]$. For simplicity, we set $\delta = 1$ given the short time horizons involved in our experiment.

Formally, consider a piece-rate contract that pays the agent p every time she chooses $a_t = 1$, and define an individual's willingness to pay for the contract, $w(p)$, to be the smallest y such that she prefers a sure payment of y over this contract. Then:

Proposition 1. *Assume that the costs in each period t are distributed according to smooth density functions, and that terms of order Δ^3 and $\Delta^2\tilde{\alpha}''(p)$ are negligible. If $\tilde{\beta} = 1$, then*

$$\frac{w(p+\Delta) - w(p)}{\Delta} \approx \frac{\tilde{\alpha}(p+\Delta) + \tilde{\alpha}(p)}{2} \tag{9}$$

If $\tilde{\beta} < 1$ and the costs are distributed independently, then

$$\frac{w(p+\Delta)-w(p)}{\Delta} \approx \underbrace{\frac{\tilde{\alpha}(p+\Delta)+\tilde{\alpha}(p)}{2}}_{\text{Surplus if time-consistent}} + \underbrace{(1-\tilde{\beta})(b+p+\Delta/2)\frac{\tilde{\alpha}(p+\Delta)-\tilde{\alpha}(p)}{\Delta}}_{\text{Behavior change premium}} \tag{10}$$

Both approximations are exact in the limit of $\Delta \to 0$, so that (i) $w'(p) = \tilde{\alpha}(p)$ when $\tilde{\beta} = 1$, and (ii) $w'(p) = \tilde{\alpha}(p) + (1-\tilde{\beta})(b+p)\tilde{\alpha}'(p)$ when costs are distributed independently.

The proposition formally shows that the WTP for an increase in incentives consists of two terms. The first term is the surplus, per dollar of incentive change, that an individual would obtain if she were time-consistent and behaved according to her forecasts. This characterization is a corollary of the Envelope Theorem, and analogues of this expression hold in any stochastic dynamic optimization problem, as shown in extensions by Allcott et al. (2022b). Thus, deviations from this expression, which we label

$$BCP(p,\Delta) := \frac{w(p+\Delta)-w(p)}{\Delta} - \frac{\tilde{\alpha}(p+\Delta)+\tilde{\alpha}(p)}{2}, \tag{11}$$

indicate that $\tilde{\beta} \neq 1$. In particular, $BCP > 0$ implies that $\tilde{\beta} < 1$. We call this reduced-form measure the *behavior change premium per dollar of financial incentives*, as it corresponds to individuals' valuation of the behavior change induced by a $\Delta = \$1$ increase in piece-rate incentives.[17]

The assumption about negligible terms is essentially the same as those in the canonical Harberger formula of the dead-weight loss of taxation: the change in incentives is not too large, particularly relative to the degree of curvature in the region of the incentive change. The assumptions are reasonable in our data, where we find that both the actual and forecasted attendance curves are approximately linear.

B.2.1 Extension to mean-zero noise

Carrera et al. (2022) extend the above results to the case where there is mean-zero noise/errors in people's stated beliefs or elicited WTP. In this case, the result about the BCP holds in the aggregate. Specifically, Carrera et al. (2022) show that equation (10) becomes

$$\mathbb{E}\left[\frac{w_i(p+\Delta)-w_i(p)}{\Delta}\right] = \mathbb{E}\left[\frac{\tilde{\alpha}_i(p+\Delta)+\tilde{\alpha}_i(p)}{2} + (1-\tilde{\beta}_i)(b_i+p+\Delta/2)\frac{\tilde{\alpha}_i(p+\Delta)-\tilde{\alpha}_i(p)}{\Delta}\right]. \tag{12}$$

[17] Assuming quasilinearity in money is not without loss, but is plausible for the relatively modest incentive sizes that are offered in field experiments such as ours. If participants are non-negligibly risk-averse over small amounts of money, then the statistic in equation (11) underestimates the WTP for behavior change, and leads to overestimates of $\tilde{\beta}$ (see Allcott et al., 2022b, for further details). Empirically, Carrera et al. (2022) do not find associations between the behavior change premium and their measure of small-stakes risk aversion. This is suggestive evidence that relative to other sources of variation in the behavior change premium, risk aversion doesn't appear to be an important determinant of the BCP.

See Section 2.3.2 of Carrera et al. (2022) for further details.

C Structural Results

C.1 Details on GMM Estimation of Parameters

The following discussion on generalized method of moments estimation of parameters is reproduced from Appendix D.1 of Carrera et al. (2022).

Let $\xi = (\beta, \tilde{\beta}, b, \lambda)$ denote the vector of parameters that we are seeking to estimate. Let $\tilde{\alpha}_i(p)$ denote an individual i's forecasted visits as a function of piece-rate incentive p, and let a_i denote actual visits. Let p_i denote the piece-rate incentive assigned to individual i. We have three sets of moment conditions.

The first set of moment conditions corresponds to forecasted attendance:

$$\mathbb{E}\left[\left(28\left(1-e^{-\lambda(\tilde{\beta}(b+p))}\right)-\tilde{\alpha}_i(p)\right)p^n\right]=0 \tag{13}$$

for all $p \in \mathscr{P} = \{0,1,2,3,5,7,12\}$, and all $n \in \{0,1,2\}$. The set $\mathscr{P}$ is the set of all incentives for which we elicited forecasts. We use 1, p, and p^2 as the instruments for the forecasted attendance equation, and our results are virtually unchanged for smaller and higher n.

The second set of moment conditions corresponds to actual attendance:

$$\mathbb{E}\left[\left(28\left(1-e^{-\lambda(\beta(b+p_i))}\right)-a_i\right)p_i^n\right]=0 \tag{14}$$

for all $n \in \{0,1,2\}$.

The third set of moment conditions corresponds to the behavior change premium:

$$\mathbb{E}\left[(1-\tilde{\beta})(b+(p_k+p_{k+1})/2)\frac{\tilde{\alpha}_i(p+\Delta_k)-\tilde{\alpha}_i(p)}{\Delta_k}-\left(\frac{w_i(p+\Delta_k)-w_i(p)}{\Delta_k}-\frac{\tilde{\alpha}_i(p+\Delta_k)+\tilde{\alpha}_i(p)}{2}\right)\right]=0 \tag{15}$$

where p_k and p_{k+1} are one of five pairs of adjacent incentives from the set $\mathscr{P} \setminus \{0\}$, and $\Delta_k := p_{k+1} - p_k$.

Letting $\hat{\xi}$ denote the parameter estimates, the GMM estimator chooses $\hat{\xi}$ to minimize

$$\left(m(\xi)-m(\hat{\xi})\right)' W \left(m(\xi)-m(\hat{\xi})\right), \tag{16}$$

where $m(\xi)$ are the theoretical moments, $m(\hat{\xi})$ are the empirical moments, and W is the optimal weighting matrix given by the inverse of the variance-covariance matrix of the moment

conditions. We approximate W using the two-step estimator outlined in Hall (2005). In the first step, we set W equal to the identity matrix,[18] and use this to solve the moment conditions for $\hat{\xi}$, which we denote $\hat{\xi}_1$. Since $\hat{\xi}_1$ is consistent, by Slutsky's theorem the sample residuals $\hat{u}$ will also be consistent. We then use these residuals to estimate the variance-covariance matrix of the moment conditions, S, given by $\text{cov}(\mathbf{z}u)$, where $\mathbf{z}$ is a vector of the instruments for the moment conditions. We then minimize

$$\left(m(\xi) - m(\hat{\xi})\right)' \hat{W} \left(m(\xi) - m(\hat{\xi})\right) \tag{17}$$

using $\hat{W} = \hat{S}^{-1}$, which gives the optimal $\hat{\xi}$ (Hansen, 1982).

C.1.1 Estimation Procedure with Misperceived Costs Parameter

To account for potential misperceptions of the distribution from which costs are randomly drawn, we can estimate a perceived cost parameter $\tilde{\lambda}$ by (i) assuming β is known or (ii) estimating the product $\lambda\beta$ of the actual cost and actual present focus parameters rather than each parameter separately. Following strategy (i), let $\xi = (\tilde{\beta}, b, \lambda, \tilde{\lambda})$ denote the vector of parameters that we are seeking to estimate, and let $\bar{\beta}$ denote the known value of the present focus parameter. We modify the first and second moment conditions in equations (13) and (14), respectively, to account for the fact that forecasted visits now depend on individual i's perception of the distribution of costs—characterized by rate parameter $\tilde{\lambda}$—and the present focus parameter is known.

The first set of moment conditions corresponding to forecasted attendance is:

$$\mathbb{E}\left[\left(28\left(1 - e^{-\tilde{\lambda}(\tilde{\beta}(b+p))}\right) - \tilde{\alpha}_i(p)\right) p^n\right] = 0 \tag{18}$$

for all $p \in \mathscr{P} = \{0, 1, 2, 3, 5, 7, 12\}$, and all $n \in \{0, 1, 2\}$.

The second set of moment conditions corresponding to actual attendance is:

$$\mathbb{E}\left[\left(28\left(1 - e^{-\lambda(\tilde{\beta}(b+p_i))}\right) - a_i\right) p_i^n\right] = 0 \tag{19}$$

for all $n \in \{0, 1, 2\}$. The third set of moment conditions is the same as in equation (15), and the remainder of the estimation procedure is unchanged.

Alternatively, following strategy (ii), let $\xi = (\lambda\beta, \tilde{\beta}, b, \tilde{\lambda})$ denote the vector of parameters that we are seeking to estimate. We modify the first and second moment conditions in equations (13) and (14), respectively, to account for the fact that forecasted attendance depends on individual

[18] One other common approach is to use $(\mathbf{z}\mathbf{z}')^{-1}$as the weighting matrix in the first stage, where $\mathbf{z}$ is a vector of the instruments in the moment equations. We confirmed our standard errors and point estimates are very similar under both choices.

i's perception of the distribution of costs and that we are no longer separately identifying the actual cost and actual present focus parameters.

The first set of moment conditions corresponding to forecasted attendance is the same as in equation (18). The second set of moment conditions corresponding to actual attendance is:

$$\mathbb{E}\left[\left(28\left(1-e^{-\lambda\beta(b+p_i)}\right)-a_i\right)p_i^n\right]=0 \tag{20}$$

for all $n \in \{0,1,2\}$. The third set of moment conditions is the same as in equation (15), and the remainder of the estimation procedure is unchanged.

C.2 Additional Structural Estimates of Baseline Present Focus Model

We consider a version of our baseline model in Table 2 with four rather than two memory bias types in Appendix Table A8. The estimates in panel (a) are broadly consistent with those in our baseline model, with actual present focus, the perceived health benefits of going to the gym, and the mean costs of a gym visit at similar levels across all memory groups, while perceived present focus is decreasing in memory bias. The predicted moments from the model presented in rows 4-6 of panel (b) of Appendix Table A8 show that this version of the model performs similarly well compared to the baseline model in terms of model fit, with a slight improvement in the prediction of the average between-group difference in the behavior change premium.

Table A8: Present focus model, heterogeneity by quartile of memory bias

(a) Parameter estimates

	Memory bias	(1) $\hat{\beta}$	(2) $\hat{\tilde{\beta}}$	(3) $\hat{b}$	(4) $1/\hat{\lambda}$	(5) $\frac{(1-\hat{\tilde{\beta}})}{(1-\hat{\beta})}$
1	Quartile 1 (N=295)	0.55 (0.54, 0.56)	0.77 (0.76, 0.78)	9.27 (9.26, 9.28)	15.56 (12.97, 18.15)	0.51 (0.38, 0.63)
2	Quartile 2 (N=268)	0.49 (0.44, 0.55)	0.74 (0.69, 0.80)	8.76 (8.68, 8.83)	14.17 (11.62, 16.72)	0.50 (0.36, 0.65)
3	Quartile 3 (N=281)	0.57 (−0.58, 1.72)	0.86 (−0.52, 2.25)	9.14 (7.97, 10.32)	14.52 (12.51, 16.52)	0.32 (0.20, 0.43)
4	Quartile 4 (N=281)	0.54 (0.48, 0.60)	0.92 (0.87, 0.98)	11.07 (11.00, 11.14)	13.84 (11.91, 15.77)	0.17 (0.06, 0.28)
5	Test of equality, p-value	0.52	0.00	0.07	0.77	0.00

(b) Empirical and model-predicted moments

		Memory bias	(1) Behavior change premium ($)	(2) Actual attendance (likelihood)	(3) Forecasted – actual attend. (likelihood)
1	Empirical	Below med. (N=563)	1.81 (1.11, 2.50)	0.34 (0.32, 0.36)	0.12 (0.10, 0.14)
2		Above med. (N=562)	0.53 (0.06, 1.01)	0.39 (0.37, 0.41)	0.17 (0.16, 0.19)
3		Difference	1.27 (0.43, 2.11)	−0.05 (−0.08, −0.02)	−0.05 (−0.08, −0.02)
4	Predicted	Below med. (N=563)	2.29 (1.70, 2.89)	0.34 (0.32, 0.36)	0.11 (0.09, 0.13)
5		Above med. (N=562)	1.07 (0.67, 1.47)	0.40 (0.38, 0.42)	0.16 (0.14, 0.17)
6		Difference	1.22 (0.50, 1.94)	−0.05 (−0.08, −0.02)	−0.05 (−0.07, −0.02)

Notes: Panel (a) of this table modifies panel (a) of Table 2 by splitting the sample by quartile of memory bias and reporting in row 5 the p-values from tests of the equality of the parameter estimates across all four quartiles. Panel (b) of this table is analogous to panel (b) of Table 2.

C.3 Additional Structural Estimates with Misperceptions of Costs

In the model in Table 3, we fix the present focus parameter at the values estimated in panel (a) in order to achieve identification in the presence of misperception of the future costs of gym visits. We additionally assume that perceived present focus is homogenous across the population, an assumption which we remove in Appendix Table A9. Reassuringly, in panel (a) of Appendix Table A9, the estimates of $\tilde{\beta}$, b, and $1/\lambda$ in columns 2-4 are identical to those in panel (a) of Table 2, and the point estimates for $1/\lambda$ and $1/\tilde{\lambda}$ in columns 4 and 5, respectively, are almost exactly the same. Thus, even when we allow for misperception of both costs and present focus, the model only predicts misperception of present focus. Panel (b) of Appendix Table A9 reports the predicted moments from this model, which exhibit no improvement in model fit relative to the baseline model.

We also study the results under alternative model assumptions that allow for identification of a perceived cost parameter $\tilde{\lambda}$, which varies with memory bias. We implement an alternative adaptation of our GMM procedure, described as strategy (ii) in Appendix C.1.1. We estimate the product $\lambda\beta$ of the actual cost and present focus parameters rather than each parameter separately, eliminating our ability to estimate the degree of sophistication but avoiding imposing any additional homogeneity assumptions or fixing any parameter values. Appendix Table A10 reports parameter estimates and predicted moments from this model. Reassuringly, the estimates of $\tilde{\beta}$ and b in Appendix Table A10 are the same as those in our baseline model in Table 2, and the estimate $\widehat{\lambda\beta}$ in Appendix Table A10 is close to the product of the estimates $\hat{\lambda}$ and $\hat{\beta}$ from the model in Table 2.

Table A9: Present focus model with misperception of costs, $\hat{\beta}$ from Table 2

(a) Parameter estimates

	Memory bias	(1) $\hat{\beta}$	(2) $\hat{\tilde{\beta}}$	(3) $\hat{b}$	(4) $1/\hat{\lambda}$	(5) $1/\hat{\tilde{\lambda}}$
1	Below med. (N=563)	0.54 By assump.	0.78 (0.72, 0.85)	9.05 (8.24, 9.86)	15.38 (13.94, 16.81)	15.38 (13.48, 17.27)
2	Above med. (N=562)	0.55 By assump.	0.89 (0.85, 0.93)	10.04 (9.14, 10.94)	14.07 (12.78, 15.35)	14.07 (12.65, 15.48)
3	Difference	−0.01 By assump.	−0.11 (−0.18, −0.03)	−0.99 (−2.20, 0.22)	1.31 (−0.61, 3.24)	1.31 (−1.06, 3.68)

(b) Empirical and model-predicted moments

		Memory bias	(1) Behavior change premium ($)	(2) Actual attendance (likelihood)	(3) Forecasted – actual attend. (likelihood)
1	Empirical	Below med. (N=563)	1.81 (1.11, 2.50)	0.34 (0.32, 0.36)	0.12 (0.10, 0.14)
2		Above med. (N=562)	0.53 (0.06, 1.01)	0.39 (0.37, 0.41)	0.17 (0.16, 0.19)
3		Difference	1.27 (0.43, 2.11)	−0.05 (−0.08, −0.02)	−0.05 (−0.08, −0.02)
4	Predicted	Below med. (N=563)	2.07 (1.45, 2.69)	0.34 (0.32, 0.36)	0.11 (0.09, 0.13)
5		Above med. (N=562)	1.12 (0.70, 1.54)	0.39 (0.37, 0.41)	0.16 (0.14, 0.18)
6		Difference	0.95 (0.20, 1.70)	−0.05 (−0.08, −0.03)	−0.04 (−0.07, −0.02)

Notes: Panel (a) of this table modifies panel (a) of Table 3 by removing the restriction that the perceived present focus parameter $\tilde{\beta}$ is constant across the two memory bias groups. Panel (b) of this table is analogous to panel (b) of Table 3.

Table A10: Present focus model with naivete about present focus and misperception of costs

(a) Parameter estimates

	Memory bias	(1) $\widehat{\lambda\beta}$	(2) $\hat{\tilde{\beta}}$	(3) $\hat{b}$	(4) $1/\hat{\tilde{\lambda}}$
1	Below med. (N=563)	0.03 (0.03, 0.04)	0.78 (0.72, 0.85)	9.05 (8.24, 9.86)	15.38 (13.48, 17.27)
2	Above med. (N=562)	0.04 (0.04, 0.04)	0.89 (0.85, 0.93)	10.04 (9.14, 10.94)	14.07 (12.65, 15.48)
3	Difference	−0.00 (−0.01, 0.00)	−0.11 (−0.18, −0.03)	−0.99 (−2.20, 0.22)	1.31 (−1.06, 3.68)

(b) Empirical and model-predicted moments

		Memory bias	(1) Behavior change premium ($)	(2) Actual attendance (likelihood)	(3) Forecasted – actual attend. (likelihood)
1	Empirical	Below med. (N=563)	1.81 (1.11, 2.50)	0.34 (0.32, 0.36)	0.12 (0.10, 0.14)
2		Above med. (N=562)	0.53 (0.06, 1.01)	0.39 (0.37, 0.41)	0.17 (0.16, 0.19)
3		Difference	1.27 (0.43, 2.11)	−0.05 (−0.08, −0.02)	−0.05 (−0.08, −0.02)
4	Predicted	Below med. (N=563)	2.07 (1.45, 2.69)	0.34 (0.32, 0.36)	0.11 (0.09, 0.13)
5		Above med. (N=562)	1.12 (0.70, 1.54)	0.39 (0.37, 0.41)	0.16 (0.14, 0.18)
6		Difference	0.95 (0.20, 1.70)	−0.05 (−0.08, −0.03)	−0.04 (−0.07, −0.02)

Notes: Panel (a) of this table modifies panel (a) of Table 2 by allowing the actual mean costs of a gym visit to differ from the perceived mean costs of a gym visit. The product of the actual cost and present focus parameters $\lambda\beta$ is estimated in place of the present focus parameter β and actual mean costs of a gym visit $1/\lambda$. Panel (b) of this table is analogous to panel (b) of Table 2.

C.4 Additional Structural Estimates Under a Homogeneity Assumption

Table 2 presents parameter estimates from a model with eight estimated parameters. We compare these results to those in Table 3, which presents parameter estimates from a model with

only seven estimated parameters due to its assumption that the perceived present focus parameter is homogeneous across the population. In panel (a) of Appendix Table A11 below, we present a seven-parameter version of the model in panel (a) of Table 2, assuming that the actual present focus parameter β is homogeneous across the population. These estimates are close to those in Table 2.

Panel (b) of Appendix Table A11 reveals that actual present focus parameter heterogeneity is not necessary to produce a superior fit of predicted moments to empirical moments relative to the seven-parameter model in Table 3.

Table A11: Present focus model with misperception of costs, homogeneous present focus parameter

(a) Parameter estimates

	Memory bias	(1) $\hat{\beta}$	(2) $\hat{\tilde{\beta}}$	(3) $\hat{b}$	(4) $1/\hat{\lambda}$	(5) $\frac{(1-\hat{\tilde{\beta}})}{(1-\hat{\beta})}$
1	Below med. (N=563)	0.55 (0.51, 0.58)	0.80 (0.75, 0.85)	9.13 (8.32, 9.93)	15.68 (14.06, 17.31)	0.45 (0.35, 0.55)
2	Above med. (N=562)	0.55 (0.51, 0.58)	0.88 (0.84, 0.92)	10.03 (9.14, 10.91)	13.96 (12.61, 15.30)	0.26 (0.18, 0.34)
3	Difference	0 By assump.	−0.09 (−0.14, −0.03)	−0.90 (−2.08, 0.28)	1.73 (−0.17, 3.62)	0.19 (0.08, 0.31)

(b) Empirical and model-predicted moments

		Memory bias	(1) Behavior change premium ($)	(2) Actual attendance (likelihood)	(3) Forecasted – actual attend. (likelihood)
1	Empirical	Below med. (N=563)	1.81 (1.11, 2.50)	0.34 (0.32, 0.36)	0.12 (0.10, 0.14)
2		Above med. (N=562)	0.53 (0.06, 1.01)	0.39 (0.37, 0.41)	0.17 (0.16, 0.19)
3		Difference	1.27 (0.43, 2.11)	−0.05 (−0.08, −0.02)	−0.05 (−0.08, −0.02)
4	Predicted	Below med. (N=563)	1.95 (1.47, 2.44)	0.34 (0.32, 0.36)	0.11 (0.10, 0.13)
5		Above med. (N=562)	1.19 (0.80, 1.58)	0.39 (0.38, 0.41)	0.16 (0.14, 0.17)
6		Difference	0.77 (0.25, 1.28)	−0.05 (−0.07, −0.03)	−0.05 (−0.06, −0.03)

Notes: Panel (a) of this table modifies panel (a) of Table 2 by restricting the present focus parameter β to be constant across the two memory bias groups. Panel (b) of this table is analogous to panel (b) of Table 2.

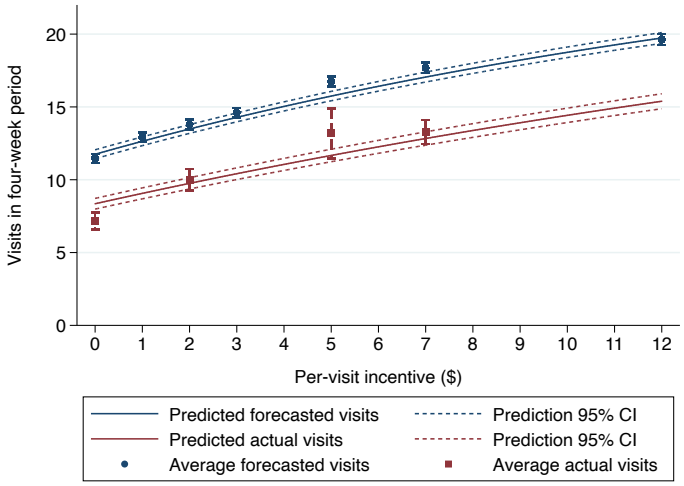
C.5 In-Sample Fit of Structural Models

In this section, we examine the in-sample fit of certain structural models to the forecasted and actual attendance curves. In Appendix Figure A4, we show the in-sample fit of the baseline structural model presented in Table 2, as well as a modification of that model presented in Appendix Table A8 with additional memory bias types. The model with only two types appears to fit the attendance data as well as the model with four types.

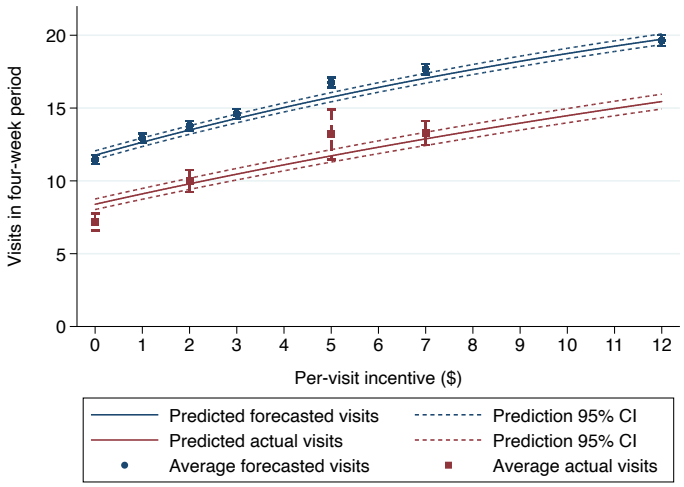
In Appendix Figure A5, we show the in-sample fit of the primary alternative model presented in Table 3, as well as a modification of that model presented in Appendix Table A9, which removes the restriction that perceived present focus may not vary with memory bias. Neither of these models improves on the fit of the attendance data relative to the results in Appendix Figure A4. One might expect that including separate parameters defining the perceived and actual cost distributions would add flexibility to the model that aids in separately fitting the forecasted and actual attendance curves. These figures show that allowing for misperceptions of the future costs of gym visits does not yield any noticeable improvement in terms of fitting either the forecasted or actual attendance data. They also highlight the importance of fitting the behavior change premium data, which confirms that the baseline model best explains participant beliefs and behavior.

Figure A4: In-sample fit of baseline model to forecasted and actual attendance

(a) Heterogeneity by above- vs. below-median memory bias



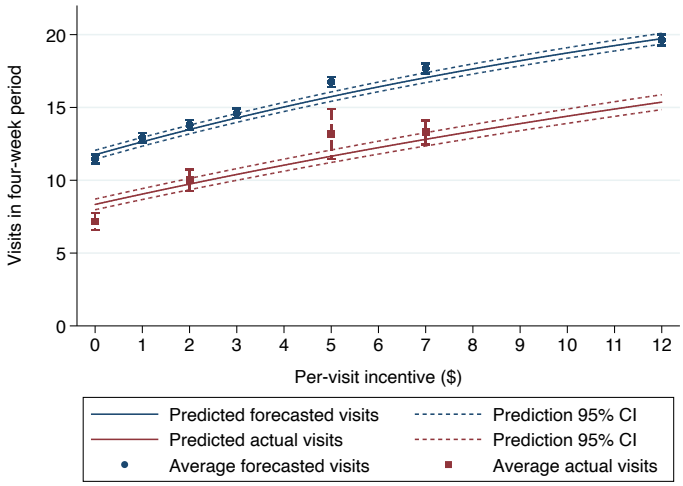
(b) Heterogeneity by quartile of memory bias



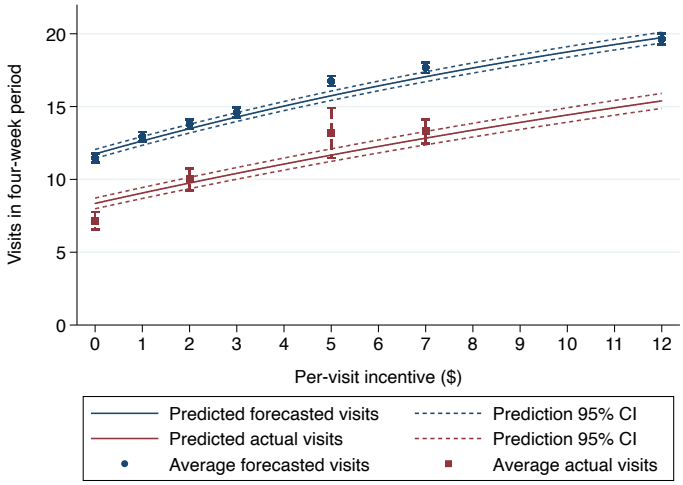
Notes: This figure compares the means and 95% confidence intervals of participants' subjective forecasts of gym visits during the four-week experimental period and their actual attendance under their assigned per-visit incentive, as empirically observed and predicted by two structural models. Panel (a) considers the structural model with two types, as in Table 2. Panel (b) considers the structural model with four types, as in Appendix Table A8. The empirical estimates of forecasted and actual attendance are computed across the full sample used to generate the respective structural estimates.

Figure A5: In-sample fit of alternative models with misperceptions of costs to forecasted and actual attendance

(a) Homogeneous perceptions of present focus



(b) Heterogeneous perceptions of present focus



Notes: This figure compares the means and 95% confidence intervals of participants' subjective forecasts of gym visits during the four-week experimental period and their actual attendance under their assigned per-visit incentive, as empirically observed and predicted by two structural models. Panel (a) considers the structural model that allows the actual mean costs of a gym visit to differ from the perceived mean costs of a gym visit under the restriction that perceived present focus does *not* vary with memory bias, as in Table 3. Panel (b) considers the analogous structural model that allows perceived present focus to vary with memory bias, as in Appendix Table A9. The empirical estimates of forecasted and actual attendance are computed across the full sample used to generate the respective structural estimates.