NBER WORKING PAPER SERIES

BIASED MEMORY AND PERCEPTIONS OF SELF-CONTROL

Afras Y. Sial
Justin R. Sydnor
Dmitry Taubinsky

Working Paper 30825
http://www.nber.org/papers/w30825

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
January 2023, Revised December 2025

We thank seminar and conference participants, as well as Mariana Carrera, Tristan Gagnon-Bartsch, Daniel Gottlieb, Davide Pace, Collin Raymond, Heather Royer, Mark Stehr, Andrej Woerner, and Sili Zhang for helpful comments. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Biased Memory and Perceptions of Self-Control
Afras Y. Sial, Justin R. Sydnor, and Dmitry Taubinsky
NBER Working Paper No. 30825
January 2023, Revised December 2025
JEL No. D9, I12

ABSTRACT

Using field-experimental data on gym attendance, we analyze the relationship between imperfect memory and people's awareness of their limited self-control. We develop a model that generates predictions about how imperfect recall is linked to forecasts and demand for behavior change, and we develop techniques for testing whether recall is selective. Empirically, we find that people overestimate both past and future attendance, and are more likely to recall attendances than absences. Larger overestimation of past attendance is associated with (i) more overestimation of future attendance, (ii) a lower willingness to pay to motivate higher future attendance, and (iii) a smaller gap between goal and forecasted attendance. Additionally, we find suggestive evidence that memory is motivated. We link biased memory to structural estimates of quasi-hyperbolic discounting with naivete.

Afras Y. Sial
University of California, Berkeley
afras@berkeley.edu

Justin R. Sydnor
University of Wisconsin - Madison
Wisconsin School of Business
and NBER
jsydnor@bus.wisc.edu

Dmitry Taubinsky
University of California, Berkeley
Department of Economics
and NBER
dmitry.taubinsky@berkeley.edu

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; Beshears et al., 2020; 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., DellaVigna and Malmendier, 2006; Kőszegi, 2006; Oster et al., 2013; 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; Gagnon-Bartsch et al., 2023). A number of different theories have been developed in recent years, suggesting that biases persist in part because of imperfect memory. In some models, people are motivated to hold positive beliefs and engage in biased information processing, which leads to memory distortions (e.g., Bénabou and Tirole, 2002, 2004; Gottlieb, 2014; Bénabou, 2015; Zimmermann, 2020; Gottlieb, 2021; Kőszegi et al., 2022). In other models, natural imperfections in memory formation and cognitive limits lead directly or indirectly to biased beliefs via failures in learning from experience (e.g., Mullainathan, 2002; Schwartzstein, 2014; Bordalo et al., 2018, 2023; Ba et al., 2023; Bordalo et al., 2024; Enke et al., 2024; Fudenberg et al., 2024; Graeber et al., 2024; Heidhues et al., 2024; Conlon, 2025). Taken together, these theories suggest that imperfect memory may be linked to the persistence of overconfidence and naivete about self-control problems.

Guided by a model of imperfect memory and biased beliefs, this paper provides new empirical evidence, from a field experiment on gym attendance, that recall biases predict (lack of) awareness of self-control problems. The gym attendance setting is a natural one for studying the link between memory and naivete about self-control problems because most gym members have significant past experience to draw on when forming beliefs about their future attendance.[1] We first show that memory is imperfect, systematically biased, and that people are more likely to recall attendances than absences. We then show that greater overestimation of past attendance is strongly linked to overestimation of future attendance and lack of awareness of self-control problems.

[1]Carrera et al. (2022) previously used these data to study gym members' self-control and naivete, but did not investigate biases in recall or their relationship with naivete about self-control.

To organize our empirical results, we start with a simple theoretical framework that encompasses multiple models (and mechanisms) where there is imperfect recall and systematically biased forecasts. In this framework, recall may be selective, where days on which the person attended the gym are more likely to be recalled than days on which the person didn't attend the gym. But the framework is agnostic about whether this asymmetry arises from motivated memory distortions or from the greater salience of attendances versus absences. People's perceptions about their past behavior are shaped both by what they remember, and by their perceptions of how they likely behaved on days that they do not remember. Thus, both imperfect recall and biased beliefs about oneself contribute to biased perceptions of past behavior. The framework encompasses models in which biased beliefs about oneself are caused by imperfect and selective recall, as well as models in which people start with biased priors that are not fully mitigated by learning.

In this framework, we show that people who overestimate their future attendance will also overestimate their past attendance, and that misperceptions of past attendance will be positively correlated with misperceptions of future attendance. Moreover, when there is heterogeneity in awareness of self-control problems, people who overestimate their past attendance more will have lower awareness of self-control problems. Additionally, we derive results about how the degree of selectivity in recalling attendances versus absences can be obtained from estimates of linear regressions of the perceived past attendance on actual past attendance. Finally, consistent with recent theoretical work (e.g., Fudenberg et al., 2024), this framework illustrates how selective recall, together with imperfect understanding of this selectivity, can lead to biased beliefs in the long run.

Guided by this framework, we begin our empirical investigation by studying people's recall of their past gym attendance. We document that recall of past gym attendance is strongly associated with actual past attendance, but that people on average overestimate their daily likelihood of visiting the gym in the past by about nine percentage points, off of a baseline likelihood of 23%. Moreover, we find evidence of selective recall: we estimate that on average, people are more likely to remember days on which they attended the gym than days on which they did not attend the gym by at least 30 percentage points.

We then investigate people's estimates of their future attendance and how they relate to biases in perceptions of past attendance. On average, gym members overestimate their future daily attendance likelihood by fifteen percentage points, and this bias in perceived future attendance is linked to bias in perceived past attendance. We find that overestimating the likelihood of past daily gym attendance by an additional 10 percentage points translates to overestimating future daily attendance likelihood by an additional 3 percentage points, on average. This replicates, in a very different field setting, the core finding in Huffman et al. (2022), who found a link between biased memories about past performance and biased predictions of future performance among managers participating in a tournament-incentive system.

What specifically generates the link between biased perceptions of the past and future? A key novel feature of our data is that it contains multiple measures of naivete about self-control problems, which allows us to study the link between imperfect memory and naivete about limited self-control. First, we examine the link between memory bias and a simple proxy for awareness of limited self-control using participants' self-reported goals for attendance. Average goal attendance is significantly higher than actual attendance rates, consistent with imperfect self-control. Goals are also 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, implying lower awareness of their imperfect self-control.

Second, we show that those with more upwardly-biased perceptions of past attendance 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 overestimation of past attendance are on average willing to pay \$0.97 to increase their future gym attendance by one visit, while those with below-median overestimation are willing to pay \$2.53. This suggests that those with inflated perceptions 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 perceptions of past attendances 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.

Building on these empirical results, we estimate a structural model of quasi-hyperbolic preferences and naivete, allowing the parameters to vary by misperception of the past. We find that individuals with above- and below-median overestimation of past attendance have essentially identical levels of time inconsistency.[2] However, individuals with above-median overestimation of past attendance are much less aware of their time inconsistency. Specifically, those with above-median overestimation of past attendance are aware of only about twenty-five percent of their degree of time inconsistency, while those with below-median overestimation of the past 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.

Finally, we investigate potential mechanisms that underlie our empirical results about selective memory and the link between recall and naivete about self-control. To investigate the possibility that recall is motivated, we examine whether people with greater self-control problems are more likely to remember attendances than absences. To investigate the possibility that attendances are simply more salient than absences, we investigate whether people with more regular schedules that concentrate attendance on particular days of the week exhibit less asymmetry in recalling attendances versus absences. We find support for the first mechanism, but not for the second.

Our work contributes to the small but growing literature we highlighted above that investigates the links between memory and behavioral biases. We provide new evidence of a link between memory and naivete about self-control specifically. These results, as well as our finding of selective recall, inform the theoretical literature on the persistence of behavioral biases.

Additionally, our paper contributes to empirical work in economics on the nature

[2]Our finding that present focus does not vary with memory bias is not a general prediction of our theoretical framework and is unlikely to hold generally across contexts. For example, 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.

of recall (see Amelio and Zimmermann, 2023, for a review). Much of the prior evidence relating to motivated memory has shown that lower past performance is more likely to be overestimated and incorrectly reported (Li, 2013; Saucet and Villeval, 2019; Chew et al., 2020; Zimmermann, 2020; Huffman et al., 2022; Caballero and López-Pérez, 2024; Roy-Chowdhury, 2024; Gödker et al., 2025), and this has been interpreted as evidence for selective recall. We show that such evidence could instead be consistent with biased priors, rather than selective recall.[3] To our knowledge, only a few laboratory experiments have directly shown that people are indeed more likely to forget negative events than positive events (Zimmermann, 2020; Chew et al., 2020; Caballero and López-Pérez, 2024; Li and Rong, 2023; Conlon, 2025), by showing, e.g., that people are more likely to remember auxiliary details in events with positive rather than negative outcomes. Our paper extends this small group of papers by providing estimates of selective recall in the field, and by providing a tractable methodology that could be fruitfully applied to other field settings.

In the rest of the paper, Section 1 presents the experimental design, Section 2 presents the conceptual framework, Section 3 presents the empirical results, Section 4 presents the structural estimates, Section 5 explores mechanisms, and Section 6 concludes. All proofs are gathered in the Appendix.

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.[4] 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 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

[3]For example, people may forget low and high past performance at the same rate, but because they use overly positive priors to construct their estimates of past performance, their reported perceptions of past performance are more accurate when their past performance was high rather than low.

[4]Our description of the experimental design mirrors Carrera et al. (2022).

component of the study.[5] 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.[6]

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.[7] 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.[8] Participants did not receive monetary incentives for accurate recall, which was a deliberate methodological decision. 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,

[5]While the 8-week minimum membership criterion was strictly enforced in the first two waves of the experiment, ten individuals in wave 3 with membership slightly shorter than 8 weeks participated due to laxer enforcement of the screening criterion by administrators.

[6]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.

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

[8]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).

we did not incentivize forecasts. We did not incentivize the memory question to keep it as comparable as possible to the forecast questions. 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 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.

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, and 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 and are excluded from our analysis.[9] 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.[10]

The online component also included questions to check for numeracy, compre-

[9] 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 ($n = 32$).

[10] 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.

hension, and attention. Fewer than 5% of participants failed to pass each of these checks, indicating high levels of engagement and understanding.[11] The final set of questions in the online component 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 1,242 participants,[12] 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 1,001 days.

Finally, although not a main focus of this paper, we also randomized some participants into an information treatment prior to eliciting forecasts and preferences over piece-rate incentives and commitment contracts (but after eliciting perceptions of past attendance). In wave 1, the 50% of 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, 50% of 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, but we summarize the impacts of the information treatments in Section 3.2.

2 Conceptual Framework

In this section we present a simple theoretical framework that we then use to organize our empirical analysis, and to obtain additional insights about recall. The framework

[11]4.9% 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.

[12]Our sample consists of six fewer participants than that of Carrera et al. (2022) since we could not reliably match these participants to pre-study attendance records.

encompasses multiple models (and mechanisms) where there is imperfect recall and systematically biased forecasts. In Section 2.2 we lay out the basic predictions about perceptions of past and future attendance that are common to these models. In Section 2.3 we show how regression estimates can be used to test for selective recall. In Appendix A we show how selective recall can pave the way for long-run biased beliefs.

2.1 Setup

We index time such that our experimental elicitations are made in period $t = 0$, with periods $t > 0$ representing the future and periods $t < 0$ representing the past. In particular, we think of periods $t = 1, \ldots, 28$ as corresponding to the 28 days of the experiment during which we randomize incentives and for which we elicit forecasts. We think of periods $t = -1, \ldots, -100$ as days for which we elicit perceptions of past attendance. None of our formal results below depend on the number of past periods before $t < 0$, or the number of future periods $t > 0$.

We assume that individuals face immediate stochastic costs $c_t \sim F$ of going to the gym on any day t and receive a fixed, delayed health benefit b from each visit. We assume that b accrues after the conclusion of the experiment (i.e., in the longer-run future). Similar to the health benefits, any financial attendance rewards p offered in our experiment accrue in periods $t > 28$.

We allow individuals to be uncertain about and systematically wrong about the cost distribution. In period 0 they believe that $c_t \sim \tilde{F}$, which may differ from the actual distribution F either because of memory and learning biases or because people start off with a biased prior (or both). That is, our framework does not take a stance on whether memory and learning biases *cause* biased beliefs, or whether individuals start with biased beliefs and the biases are not fully eliminated due to insufficient opportunities to learn or due to biases in memory and learning.

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 about their present focus. For simplicity and ease of exposition, we assume people have point beliefs about their present focus, so that their perceived present focus is $\tilde{\beta} \in [\beta, 1]$. The theoretical predictions hold if instead, as in Heidhues and Kőszegi (2009; 2010), people have dispersed beliefs $\tilde{G}$, supported on $[\beta, 1]$, about their present focus. 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 $U^{t'}$.[13] Thus, an individual visits the gym in period t if $\beta(b+p) > c_t$, and believes they will visit the gym in some future period t' if $\tilde{\beta}(b+p) > c_{t'}$.

Beliefs about the distribution of c_t and about present focus will generally be endogenous to the memory and learning process in the types of models our framework is meant to capture. But our core results are robust to any mechanism that generates biased beliefs. This is because we are simply formulating predictions given whatever beliefs result in period 0.

We let $\tilde{\mu}$ denote individuals' beliefs about the likelihood of attending the gym on any given day, in the absence of any additional experimental incentives. That is, $\tilde{\mu}$ corresponds to the forecasted attendance likelihood in the absence of additional financial incentives.

Suppose now that individuals remember days on which they attended or didn't attend the gym with probabilities ρ_1 and ρ_0, respectively, and that the likelihoods of remembering any two days are conditionally independent of each other. Selective recall, where $\rho_1 \neq \rho_0$, could either be the result of motivated memory (e.g., Bénabou and Tirole, 2002, 2004) or the result of salience bias, where "events" (e.g., going to the gym) are more likely to be remembered than "non-events" (e.g., Enke, 2020; Caballero and López-Pérez, 2024). Consistent with prior models and psychological evidence, to simplify exposition we assume that $\rho_1 \geq \rho_0$ (and do not consider the case $\rho_1 < \rho_0$) so that days with gym visits are weakly more likely to be remembered than those without, though this is not a critical assumption for most of our results.

A common assumption in models of selective memory is that people do not take the selectivity of their memory into account when forming beliefs (e.g., Mullainathan, 2002; Bordalo et al., 2018, 2023; Ba et al., 2023; Enke et al., 2024; Bordalo et al., 2024; Fudenberg et al., 2024; Graeber et al., 2024). Formally, let $\tilde{\rho}_1$ and $\tilde{\rho}_0$ be the perceived likelihoods of recalling attendances and absences. Then ignorance of selectivity (but not necessarily of imperfect recall itself) corresponds to $\tilde{\rho}_0 = \tilde{\rho}_1 \leq 1$. In this case, an individual's estimate $\tilde{\nu}$ of their likelihood of attendance on forgotten days is simply $\tilde{\mu}$, since the individual does not update prior beliefs based on the signal that they forgot

[13]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 of study are short.

what happened in a given period. More generally, a person may partially account for their selective recall, in the sense that $\rho_0 < \tilde{\rho}_0 < \tilde{\rho}_1 < \rho_1$. In this more general case,

$$\frac{\tilde{\nu}}{1-\tilde{\nu}} = \frac{1-\tilde{\rho}_1}{1-\tilde{\rho}_0} \cdot \frac{\tilde{\mu}}{1-\tilde{\mu}} \tag{1}$$

2.2 Memory Bias and Forecast Bias

Consider individuals i whose true likelihood of attending the gym is μ on a given day (in the absence of incentives), who attended the gym a fraction ϕ_i of the past 100 days, and who attend the gym a fraction γ_i of days in the next 28 days of the experimental period. Let $\tilde{\phi}_i$ denote i's estimate of the fraction of the past 100 days on which they attended the gym,[14] and note that $\tilde{F}\left(\tilde{\beta}b + \tilde{\beta}p\right)$ is the forecasted future daily attendance likelihood if given per-attendance incentive p. We refer to $\tilde{\phi}_i - \phi_i$ as *memory bias*, and to $\tilde{F}\left(\tilde{\beta}b + \tilde{\beta}p_i\right) - \gamma_i$ as *forecast bias*, where p_i is the per-attendance incentive assigned to individual i.

Proposition 1. *Under the assumptions of Section 2.1,*

$$\mathbb{E}[\tilde{\phi}_i|\phi_i] = \left[\rho_1 - (\rho_1 - \rho_0)\,\tilde{\nu}\right]\phi_i + (1-\rho_0)\,\tilde{\nu} \tag{2}$$

Thus

1. *Perceptions of past attendance are a linear function of actual past attendance.*

2. *Suppose $\rho_0 < 1$. If $\tilde{\beta} > \beta$ or if F first-order stochastically dominates $\tilde{F}$, then the forecast bias and the memory bias will both be positive, on average.*

3. *Both memory bias and forecast bias are increasing in $\tilde{\beta}$, and decreasing in $\tilde{F}$ in the first-order stochastic dominance (FOSD) order.*

Proposition 1 formalizes three main results that organize our empirical analysis. First, it shows that the relationship between perceived and actual past attendance will be linear, which motivates the linear regression models in Section 3. Second, it shows that when memory is imperfect, misperceptions of one's behavior will generate not just forecast bias, but also memory bias. Note that this statement is about population

[14]For an individual who attended the gym a fraction ϕ_i of times and happens to recall $100 \cdot r_i^1$ attendances and $100 \cdot r_i^0$ non-attendances, $\tilde{\phi}_i = r_i^1 + \left(1 - r_i^1 - r_i^0\right)\tilde{\nu}$. By definition, $\mathbb{E}[r_i^1|\phi_i] = \rho_1\phi_i$ and $\mathbb{E}[r_i^0|\phi_i] = \rho_0(1-\phi_i)$.

averages: generally, both $\tilde{\phi}_i - \phi_i$ and $\tilde{F}\left(\tilde{\beta}b + \tilde{\beta}p\right) - \gamma_i$ will take on both positive and negative values at the individual level, because there is unpredictable variation in how many times a person will actually attend the gym.[15] Finally, part 3 of Proposition 1 shows that there will be a positive association between memory bias and forecast bias when there is variation in people's misperceptions of their behavior. Intuitively, this is because misperceptions of own behavior influence both forecasts and estimates of how likely one is to have visited the gym on forgotten days, as formalized in equation (1). In particular, when there is variation in $\tilde{\beta}$, part 3 of Proposition 1 makes the additional prediction that memory bias will be correlated with proxies for perceived time inconsistency and awareness of present focus.

An immediate corollary of Proposition 1 is that when the coefficient of ϕ_i in equation (2) is below 1, people's overestimation of their past performance will be greater for lower performance:

Corollary 1. *Suppose* $\rho_0 < 1$*.*

1. $\mathbb{E}\left[\tilde{\phi}_i - \phi_i | \phi_i\right]$ *is decreasing in* ϕ_i*.*
2. *Suppose that there is only one past period* $t = -1$*, so that* $\phi_i \in \{0, 1\}$*. Let* $\varepsilon_i = |\tilde{\phi}_i - \phi_i|$ *be the absolute recall error. Then for* $\tilde{\mu}$ *sufficiently high, the average absolute recall error,* $\mathbb{E}[\varepsilon_i|\phi_i]$*, will be lower for* $\phi_i = 1$ *than for* $\phi_i = 0$*.*

Corollary 1 shows that our model rationalizes past empirical results that (i) people overestimate low past performance more than high past performance and (ii) when asked about single past events (or multiple past events one by one), people are more likely to correctly recall good performance than bad performance (Li, 2013; Saucet and Villeval, 2019; Chew et al., 2020; Zimmermann, 2020; Huffman et al., 2022; Roy-Chowdhury, 2024; Gödker et al., 2025). What's important, however, is that these predictions do not require selective recall, $\rho_0 < \rho_1$. This motivates the need for the more rigorous tests of selective recall that we provide in Section 2.3.

The reason part 1 of the corollary holds when $\rho_0 = \rho_1 < 1$ is because people's perceptions of the past are not just shaped by what they recall, but also by what they think they did when they don't recall their actions. If people have upwardly biased beliefs about their performance, their guesses about how they behaved in

[15]That is, because the cost draws each day are random, by chance people might attend the gym a fraction larger than $\tilde{\nu}_i$ on forgotten days, which would lead to a negative value of $\tilde{\phi}_i - \phi_i$. Similarly, future gym attendance can by chance be more frequent than expected.

cases where they forget are likely to be overestimates when their past performance was particularly low. Part 2 additionally shows that perceptions of individual events will be more accurate for cases of good performance when people's beliefs $\tilde{\mu}$ about their overall performance are sufficiently biased upwards. Intuitively, this is because when people think that they tend to have high performance $\phi_i = 1$, then when they forget the event, they will be more likely to guess their performance correctly when, indeed, $\phi_i = 1$. Again, this prediction does not require that $\rho_0 < \rho_1$.

Note that to simplify exposition, we have formalized predictions for a group of individuals with a homogeneous attendance likelihood μ. With heterogeneity in μ, equation (2) can be generalized to

$$\mathbb{E}[\tilde{\phi}_i|\phi_i, \mu] = [\rho_1 - (\rho_1 - \rho_0)\,\mathbb{E}\,[\tilde{\nu}|\mu]]\,\phi_i + (1-\rho_0)\,\mathbb{E}\,[\tilde{\nu}|\mu]\,. \tag{3}$$

This generates two additional implications in the plausible case where actual attendance likelihood μ is positively associated with perceived attendance likelihood $\tilde{\mu}$, and thus $\tilde{\nu}$ (via equation (1)). The first is that controlling for μ is potentially important for obtaining an unbiased estimate of the (average) coefficient of ϕ_i in equation (3), because ϕ_i is positively associated with μ.[16] Empirically, we include a proxy for the overall attendance likelihood by controlling for past attendance outside of the 100-day window for which we elicit recalled attendance. The second is that even when controlling for past attendance in the 100-day window, actual attendance likelihood μ (or proxies for it) will be positively associated with *perceived* past attendance in the 100-day window, because it will be positively associated with the constant term $(1-\rho_0)\,\tilde{\nu}$ in equation (2) above.

2.3 Selective Recall

An estimate of the linear regression model in equation (2) from Proposition 1 can be used to test for selective recall. Intuitively, for there to be a bias in perceptions of past attendance, individuals cannot recall every day on which they did not attend the gym. Thus, the degree of bias provides an upper bound on ρ_0. Specifically, the estimated intercept $(1-\rho_0)\,\tilde{\nu}$ in equation (2) provides an upper bound on $(1-\rho_0)$ once perceived future attendance $\tilde{\mu}$ is elicited, as $\tilde{\nu} \leq \tilde{\mu}$ by equation (1). At the same time, the relationship between $\tilde{\phi}_i$, perceived past attendance, and ϕ_i, actual

[16] In particular, $\mathbb{E}[\phi_i|\mu] = \mu$, by definition.

past attendance, provides a lower bound on ρ_1: the more sensitive $\tilde{\phi}_i$ is to ϕ_i, the more likely a person is to remember an attendance, and thus the higher is ρ_1. We formalize this intuition below.

Proposition 2. *Let $\mathbb{E}[\tilde{\phi}_i|\phi_i] = b_0 + b_1\phi_i$, where b_0 and b_1 are obtained from a linear regression of $\tilde{\phi}_i$ on ϕ_i. With ignorance of selective recall, $\tilde{\rho}_1 = \tilde{\rho}_0$, the degree of selectivity is*

$$\rho_1 - \rho_0 = \frac{b_0 - \tilde{\mu}(1 - b_1)}{\tilde{\mu}(1 - \tilde{\mu})}. \tag{4}$$

With partial accounting of selective recall, $\tilde{\rho}_1 > \tilde{\rho}_0$, the expression in (4) is a lower bound on the actual degree of selectivity $\rho_1 - \rho_0$.

The identification result in Proposition 2 shows how simple regression results can be used to provide evidence of selective recall. This, in turn, can help address the puzzle of how biased beliefs can be sustained in the long run. To see this, consider an individual who believes that they forget at random. In the long run, this individual's beliefs $\tilde{\mu}$ will be consistent with the data and their mental model of recall if $\tilde{\mu}$ is equal to the fraction of attendances on the days that the individual remembers; that is, if $\tilde{\mu} = \rho_1\mu_1/\left(\rho_1\mu + \rho_0(1 - \mu)\right)$. This logic follows the model of Fudenberg et al. (2024), and is formalized in Appendix A. Proposition A1 in the appendix shows that as long as the person does not fully take the selectivity of their memory process into account, biased beliefs can be sustained in the long run, and the person will have no way of discovering that their mental model is wrong. However, long-run biased beliefs cannot be sustained when individuals are fully aware of and account for their selective recall.

3 Empirical Results

3.1 Memory bias

People's perception of their past gym attendance closely tracks their actual past attendance, except that past attendance is systematically overestimated. On average, individuals' perceived likelihood of attending the gym on a given day is 0.32, while in reality their gym attendance likelihood is 0.23, with the difference of 0.09 statistically significant at $p < 0.01$.

Figure 1 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. Consistent with Proposition 1, the relationship is clearly linear. If participants were unbiased, then the relationship between their perceived past attendance and actual past attendance would be on the dashed 45-degree line. Instead, while the best-fit line is nearly parallel to the 45-degree line, participants on average overestimate their past attendance.[17]

Appendix Figure A1 presents a histogram of perceived minus actual past attendance, $\tilde{\phi}_i - \phi_i$, 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. The presence of a small share of negative values of $\tilde{\phi}_i - \phi_i$ is consistent with our discussion in Section 2.2, and footnote 15.

Table 1 presents regression estimates of how the recalled likelihood of visiting the gym in the past 100 days relates to the actual visit likelihood. This provides an estimate of the linear model in Proposition 1, with additional demographic and time controls. The table reports an estimate of the constant term $(1 - \rho_0)\,\tilde{\nu}$ as well. Because our regressions include various controls, we estimate this constant term as the prediction when $\phi_i = 0$ and the controls are set at their average value. The even-numbered columns additionally control for a proxy of overall attendance likelihood μ using the longer-run past attendance rate prior to the 100-day look-back period, as motivated by the discussion around equation (3) in Section 2.2. While Columns 1 and 2 restrict to individuals with membership for at least 100 days, Columns 3 and 4 restrict to those who have been members for at least 200 days, such that our estimate of μ for these individuals is particularly precise. Consistent with Figure 1, Columns 1-4 shows that perceived past attendance closely tracks actual past attendance, but there is a level bias where individuals overestimate their past attendance likelihood by approximately 9 or 10 percentage points. Consistent with the discussion around equation (3), we also find that past attendance outside of the 100-day window is positively associated with perceived past attendance in the 100-day window.

[17]Notably, there is minimal attenuation, of the kind that would be implied by models of cognitive uncertainty. In our context, such models would predict attenuation that results in overestimation of attendance among those with low past attendance, and underestimation of attendance among those with high past attendance.

Columns 5 and 6 of Table 1 are motivated by Part 1 of Corollary 1, as well as the discussion around equation (3). Consistent with our model, past attendance outside the 100-day window informs individuals' priors about attendance in the 100-day window, and thus is positively associated with memory bias. By contrast, actual attendance in the 100-day window is negatively associated with memory bias, as in Corollary 1.

The results in Table 1 are robust to the inclusion of different sets of controls, with very similar point estimates when we omit demographic controls, or omit both demographic controls and wave fixed effects.[18]

Selective recall Proposition 2 shows how the estimated regression models can be used to estimate the degree of selectivity in recall, $\rho_1 - \rho_0$, under the assumption that individuals do not take this selectivity into account when forming beliefs about their past behavior. This is a psychologically plausible assumption that is commonly made in the literature (e.g., Mullainathan, 2002; Bordalo et al., 2018, 2023; Ba et al., 2023; Bordalo et al., 2024; Enke et al., 2024; Fudenberg et al., 2024; Graeber et al., 2024), and Proposition 2 also shows that this constitutes a lower bound for the general case in which people partially or completely account for their selective recall. We use *seemingly unrelated regression* (Zellner, 1962) to simultaneously estimate the parameters needed to compute these bounds, and the resulting standard errors.[19] Columns 1-4 of Table 1 imply the following values of $\rho_1 - \rho_0$, respectively, with standard errors in parentheses: 0.34 (0.02), 0.30 (0.02), 0.35 (0.02), and 0.32 (0.03). Our estimates suggest that on average gym members are at least 30 percentage points more likely to recall a day when they visited the gym than a day when they did not visit.

[18]These results are omitted from the Appendix for conciseness, but are included in the replication code.

[19]To account for potential covariance between our parameter estimates, we use a seemingly unrelated regression (SUR) framework, which employs the generalized least-squares algorithm described by Greene (2012). We first estimate a SUR system with regressions of the perceived past daily visit likelihood and forecasted future daily visit likelihood in the absence of incentives on the regressors in the relevant column of Table 1. Our estimate of b_1 is the coefficient estimate on actual daily visit likelihood 1-100 days prior, while our estimate of b_0 is the model-predicted value of the perceived past visit likelihood at zero actual past attendance 1-100 days prior and the means of all other regressors. Similarly, our estimate of $\tilde{\mu}$ is the model-predicted value of the forecasted future visit likelihood at the means of all regressors. We use the Delta method with the covariance matrix from the SUR system to estimate standard errors of $\frac{b_0 - \tilde{\mu}(1-b_1)}{\tilde{\mu}(1-\tilde{\mu})}$.

3.2 Link Between Memory Bias and Perceptions of Self-Control

We now move on to the second and third parts of Proposition 1, to investigate how memory bias relates to forecast bias and to proxies for naivete about self-control problems.

Forecast bias On average, individuals overestimate their future attendance likelihood. They expect to attend on a fraction 0.51 of days, but in reality attend on a fraction 0.36 of days, with the difference significant at $p < 0.01$. This implies an average forecast bias of 0.15, which is larger than the average memory bias of 0.09 ($p < 0.01$). Note that the higher attendance likelihood during the experimental period is due to many participants receiving incentives to attend the gym. Among participants who receive no incentives to attend the gym during the experiment, the perceived and actual attendance likelihoods are 0.41 and 0.26, respectively, with a difference that is also 0.15 ($p < 0.01$).

Figure 2a presents a binned scatter plot comparing participants' memory bias with their forecast bias: 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 association between memory bias and forecast bias. This relationship is also quantified in regression analysis in Column 1 of Table 2. 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.

Figure 2b 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. Thus, memory bias is associated with forecast bias rather than with preferences for gym attendance or with self-control.

Perceptions of falling short of one's goals Panel (a) of Figure 3 and Column 2 of Table 2 study 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. That is, this measure is negatively related to $\tilde{\beta}$. The mean gap between goal and forecasted daily attendance likelihood is 12 percentage points, which suggests some awareness of time inconsistency. The binned scatterplot in Figure 3a compares the gap between goal and forecasted attendance and memory bias, and shows a negative relationship between the two. In Column 2 of Table 2, 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 other words, those with more upwardly-biased recall of their past visits perceive that they will have less of a gap between their future attendance and their goal.

Behavior change premium Next, we consider 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 (i.e., their health benefits from a gym visit are augmented by a monetary incentive p), and is a measure of a person's *perceived* time inconsistency. Following Carrera et al. (2022) and Allcott et al. (2022b), the BCP at incentive p and an increment in the per-visit incentive Δ is defined as

$$BCP(p, \Delta) := \underbrace{\frac{w(p+\Delta) - w(p)}{\Delta}}_{\substack{\text{WTP per dollar} \\ \text{of incentive}}} - \underbrace{\frac{\tilde{\alpha}(p+\Delta) + \tilde{\alpha}(p)}{2}}_{\substack{\text{Forecasted earnings} \\ \text{per dollar of incentive}}} \tag{5}$$

where $w(\cdot)$ is the WTP for a given incentive and $\tilde{\alpha}(p) := 28 \cdot \tilde{F}(\tilde{\beta}b + \tilde{\beta}p)$ is the forecasted number of attendances during the 28-day experimental period, given incentive p. 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 C for further details.

To gain some intuition for the BCP measure and how it relates to memory bias, we begin in Figure 4 by graphing the measures that go into calculating the BCP at different levels of per-visit incentives. We split this figure into two panels, separating subjects with below and above-median memory bias in panels (a) and (b), respectively. The solid red line in the figure graphs the average willingness to pay (WTP) at each per-visit incentive level. The dashed black line shows the average subjective expected earnings, which is simply the average number of forecasted visits multiplied by the per-visit incentive level. Finally, the dotted blue line shows the average implied willingness to pay that a time-consistent agent would have for each incentive given the forecasted rate of visits.[20] The gap between the WTP and this time-consistent counterfactual benchmark is the measure of the behavior change premium. In both panels, the average WTP is higher than subjective expected earnings for small incentives, and above the time-consistent benchmark at all incentive levels, implying a positive BCP and some average perceived degree of time inconsistency. Importantly, however, we observe a larger gap in panel (a), indicating a lower BCP and hence a lower perceived level of time inconsistency for those with more positively biased memories of their past behavior.

To quantify these effects, we calculate each participant's BCP for different levels of incentives p and average these BCP values across different levels of incentives to get a single BCP estimate for each individual. Figure 3b presents a binned scatterplot comparing participants' average behavior change premium and their memory bias. It shows that there is a strong (and approximately linear) negative relationship between memory bias and the BCP. We quantify this relationship in the regression in Column

[20] This is constructed from the second-order approximation that a time-consistent person should be willing to pay $\Delta(\tilde{\alpha}(p + \Delta) + \tilde{\alpha}(p))/2$ for a Δ increase in incentives. See Appendix C for further details.

3 of Table 2 and estimate that a 10 percentage-point increase in a participant's bias in recollection of their past daily attendance likelihood is associated with a 44-cent decrease in their BCP (a 37% reduction at the mean). Thus, those with larger positive biases in memory tend to express less desire for behavior change, which is consistent with a lack of awareness of self-control problems.[21]

Take-up of commitment contracts 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. In particular, the higher is a person's expected attendance, the lower is the perceived likelihood of incurring a penalty. And because more naive people expect higher attendance, higher levels of naivete can lead to *higher* take-up of commitment contracts. Carrera et al. (2022) show that under a wide range of economic conditions that are plausible in our setting, more naive individuals will, on average, have higher take-up of commitment contracts because they are more optimistic about avoiding the penalty. Carrera et al. (2022) then confirm this theoretical prediction empirically, by showing that take-up of commitment contracts is strongly negatively related to the BCP and other reduced-form measures of awareness of present focus.

Motivated by the theoretical and empirical results of Carrera et al. (2022), we thus study how take-up of such commitment contracts relates to memory bias. The results of Carrera et al. (2022) suggest that memory bias should be positively related to take-up of commitment contracts. Consistent with this, panel (c) of Figure 3 and Column 4 of Table 2 show that memory bias is significantly positively associated with commitment contract take-up: a 10 percentage-point increase in overestimation of past attendance likelihood is associated with a 2.4 percentage-point increase in commitment contract take-up.

Robustness In the Appendix we provide additional evidence of the robustness of the link between memory bias and awareness of self-control problems. Appendix B.4

[21]Appendix Table A1 additionally controls for the forecasted change in attendance per dollar increase in incentive and fully replicates Column 3, and Appendix Table A2 shows that there is no statistically significant association between memory bias and noise in the BCP estimate.

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 2.

We study four variations of Table 2 to further test the robustness of this link. First, we study two variations of Table 2 with no controls and with fixed effects only. Next, we instead add a control for the actual daily visit likelihood over all days beyond 100 days prior, while restricting the sample to those with at least 100 days of membership. Lastly, we include the maximum number of observations available in each column, instead of using a constant sample. The results are largely the same as in Table 2 throughout all of these specifications.[22]

Information treatments Appendix B.3 summarizes the results of the information treatments. Because the information treatments were administered after the recall of past attendance was elicited, we cannot study their direct effect on recall. However, we can examine whether the treatment effects of our information provision covary with the degree of bias in people's recall. Ex-ante there is no theoretical prediction about this relationship. On the one hand, the more biased individuals have more scope to be debiased. On the other hand, those who have the most biased estimates of their past and future attendance might be biased precisely because they ignore information such as the information about past attendance in our treatments. This is consistent with theories of motivated beliefs, such as Bénabou and Tirole (2002) and the work that followed, as well as with theories of mis-specified learning (e.g., Gagnon-Bartsch et al., 2023), where people with the most strongly-held biases ignore information because they don't think they have anything else to learn.

Overall, the results are consistent with Allcott et al. (2025) and other recent work on information treatments, which finds statistically significant average effects in the "right" direction but no interaction with proxies for bias. The basic information treatment had no effect on forecasts and perceptions of self-control, and no interaction with the degree of memory bias. The enhanced information treatment lowered forecasts and increased awareness of present focus (see Carrera et al., 2022, for a detailed analysis), but like the basic information treatment it had no interaction with memory bias.

[22]These results are omitted from the Appendix for conciseness, but are included in the replication code.

4 Structural Estimates

In this section, we build on the empirical results of Section 3.2 to quantify the link between memory bias and perceived (and actual) self-control. Our strategy is to group people into high or low memory-bias subgroups, and to structurally estimate the quasi-hyperbolic model for each of those groups. We don't structurally estimate the recall parameters ρ_0, ρ_1 for groups with high versus low memory bias because measurement error in our classification could create systematic bias in those estimates. By contrast, because this measurement error is orthogonal to the moments we use to estimate the quasi-hyperbolic model, it does not bias our estimates of the time preference parameters, and can only attenuate estimates of the relationship between memory bias and the time preference parameters.[23]

In line with our theoretical framework, these structural estimates do not assume a particular direction of causality. As Proposition 1 clarifies, memory bias and naivete are linked through people's biased beliefs at the time of elicitation. These beliefs could be biased either because people have biased priors, which are not mitigated due to imperfect and selective memory, or because selective memory itself causes the biases.

4.1 Structural Model and Identification

Building on Carrera et al. (2022), we structurally estimate a model of quasi-hyperbolic discounting with imperfect perception. To do so, we parametrize the framework presented in Section 2.1. We assume that costs are on net always non-negative and distributed independently and identically according to the exponential distribution with mean $1/\lambda$.[24] We begin with the assumption that individuals correctly perceive

[23]Specifically, our measures of memory bias are noisy, which leads us to overestimate memory bias for those we classify in the above-median memory bias group, and to underestimate memory bias for those we classify in the below-median memory bias group. But because this measurement error is orthogonal to behavior during the experimental period, which we use to estimate the quasi-hyperbolic model, this *attenuates* any estimates of how the structural parameters correlate with mean memory bias. On the other hand, this measurement error is not orthogonal to the data on perceived and actual past attendance, which we use to estimate the degree of selectivity $\rho_1 - \rho_0$.

[24]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.

the cost distribution, and later provide evidence for the validity of this assumption. Given per-visit incentive p, these assumptions imply that the forecasted and actual number of attendances over the 28-day period 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 functions identify the parameters β, $\tilde{\beta}$, b, and λ. Proposition 1 of Carrera et al. (2022) (duplicated in Appendix C) shows that up to negligible higher-order terms, the BCP can be approximately expressed as a function of the structural parameters as follows:

$$BCP(p, \Delta) \approx (1 - \tilde{\beta})(b + p + \Delta/2)\frac{\tilde{\alpha}(p + \Delta) - \tilde{\alpha}(p)}{\Delta}. \tag{6}$$

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 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 D.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 $\tilde{\lambda}$.[25] 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

[25]Misprediction of the delayed benefits b of a gym visit could be another source of misprediction. However, misprediction of these benefits cannot generate a wedge between forecasted and actual behavior because an individual receives them and learns their true value in the future relative to both the time of the forecast and the decision of whether to attend the gym. Misprediction of immediate benefits, net of immediate costs, is accounted for in the difference between λ and $\tilde{\lambda}$.

identified. When analyzing this form of mis-prediction, we therefore either fix β at the estimates from the baseline model, or simply estimate $\lambda\beta$. See Appendix D.1.1 for further details.

4.2 Results

Table 3a 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 3, the degree of naivete—which affects forecasted attendance and the BCP—is significantly higher in the above-median memory bias group. The higher degree of naivete is reflected both in lower perceived resent focus $\tilde{\beta}$, as well as in the fraction of present focus that people are aware of, $(1-\tilde{\beta})/(1-\beta)$. Appendix Table A5 reports parameter estimates by quartile of memory bias, with largely the same conclusions.

Table 3b shows that the estimated model in Table 3a matches the empirical moments well. The estimated model almost perfectly matches average actual attendance and average misprediction of actual attendance. Appendix Figure A3 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 overestimated for the above-median memory bias group and more significantly for the below-median group. This mismatch is due to our baseline model understating the degree of heterogeneity. The more heterogeneous model in Appendix Table A5 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 4 shows that our data cannot be explained if we don't allow a relationship between memory bias and perceived present focus $\tilde{\beta}$, even if we allow misperceptions of costs to vary with memory bias. We begin with panel (a), which 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 4b and Appendix Figure A4 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 (6) 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 A1 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 4, Appendix Table A7 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 3.

Additional results and robustness To achieve identification in a model with potential misperceptions of both costs and present focus, without assumptions restricting parameter values, we can estimate the product $\lambda\beta$ in place of each parameter separately. Appendix Table A8 reports parameter estimates and predicted moments from this model, and again shows that allowing for heterogeneity in perceptions of $\tilde{\lambda}$ does not improve model fit relative to the baseline model in Table 3, and does not influence our estimates of $\tilde{\beta}$.

Appendix Table A6 presents a seven-parameter version of the model in Table 3, under the assumption that β is homogenous across memory bias groups. The model fit in Appendix Table A6 remains superior to that of the seven-parameter model in Table 4, further supporting the baseline modeling assumptions.

5 Mechanisms for Selective Recall

What might generate imperfect and selective recall? The first mechanism is motivated memory: people may prefer to remember attendances more than absences, perhaps because they prefer to have optimistic beliefs about their self-control (Bénabou and Tirole, 2002, 2004). The second mechanism is salience bias: attendances may simply be more salient and memorable than absences. To investigate these mechanisms, we study differences in the degree of selective recall, $\rho_1 - \rho_0$, across subsamples that differ in their motives for distorting their memories or in the likely salience of their past visits due to their attendance patterns. While our data do not provide a conclusive test of these different mechanisms, we find support for motivated memory, and little evidence consistent with salience/memorability bias.

We focus on mechanisms for selective recall because as we discuss in Section 2.3 and Appendix A, selective recall helps sustain both memory bias and forecast bias—including naivete about self-control. Understanding these mechanisms can thus help predict not only the degree of selective recall in other settings, but also memory and forecast bias. While in our setting attendances are plausibly both more salient and desirable than absences, in other settings the relationship between salience and desirability need not be positive.

5.1 Motivated memory

We begin by analyzing whether those with larger self-control problems are more motivated to distort their memory, and in particular have a larger degree of selective memory, $\rho_1 - \rho_0$. To do this, we first develop a proxy for the degree of time inconsistency. Our proxy is the *goal-behavior gap*, which we define as the difference between the self-stated goal attendance likelihood and the past attendance likelihood on days 1-100 prior to the survey. Appendix Table A9 validates this proxy by estimating our structural model for those with higher or lower values of this proxy. The estimates show that those with larger goal-behavior gaps are significantly more present-focused, but do not perceive themselves to be more present-focused. The lack of difference in perceived present focus could be due to selective memory.

Does the more present-focused subsample exhibit a greater degree of selective recall? Panel (a) of Figure 5 presents binned scatter plots that compare actual and perceived past visit likelihood for those with above- versus below-median goal-behavior gaps. As in Figure 1, the relationship is linear for both groups, and participants on average overestimate their past attendance. However, overestimation is higher for the above-median goal-behavior gap group ($p < 0.01$). In addition, the slope of the best-fit line is steeper for the above-median group, suggestive of greater asymmetry in recall. Panel (a) of Appendix Figure A5 shows that a similar pattern holds when we instead compare participants in the top and bottom quartiles of the goal-behavior gap.

To estimate the degree of selectivity in recall, $\rho_1 - \rho_0$, we again utilize Proposition 2 and the procedure in Section 3.1, assuming that individuals ignore their recall selectivity when predicting the past (recall that without this assumption, Proposition 2 implies that the estimates are lower bounds). For specifications corresponding to Columns 1 through 4 of Table 1, respectively, we estimate that $\rho_1 - \rho_0$ is 0.63 (se : 0.05), 0.54 (se : 0.06), 0.65 (se : 0.05), and 0.58 (se : 0.06) among participants with above-median goal behavior gaps, and is 0.23 (se : 0.02), 0.23 (se : 0.02), 0.25 (se : 0.02), and 0.24 (se : 0.03) among participants with below-median goal-behavior gaps. Across these four specifications, the degree of selectivity is significantly higher among participants with a higher goal-behavior gap, generating estimated differences of 0.40 (se : 0.05) , 0.31 (se : 0.06), 0.40 (se : 0.06), and 0.33 (se : 0.07) respectively, which are all significant at $p < 0.01$.[26]

These large and statistically significant differences suggest that selective recall is at least in part motivated. That is, the data suggest that those with more severe self-control problems are motivated to engage in more selective recall, so as to maintain a positive self-image of their self-control problems (Bénabou and Tirole, 2002, 2004). This helps explain the Appendix Table A9 finding that those with greater present focus do not perceive themselves to have greater present focus, and are thus significantly more naive.

5.2 Salience bias

The other mechanism we consider is that selective recall may arise because attendances are simply more salient and memorable than absences, because "events" (e.g., going to the gym) are more salient than "non-events" (e.g., Enke, 2020; Caballero and López-Pérez, 2024). Our approach to testing this is to consider individuals with more versus less regular schedules; i.e., individuals whose visits are concentrated on particular days of the week. Among those with more versus less regular schedules, absences are likely to be relatively more salient because they exhibit deviations from the norm, while attendances are likely to be relatively less salient because they are the

[26]When instead comparing participants in the top versus bottom quartile, the differences in the degree of selectivity are 0.63 (se : 0.10), 0.53 (se : 0.11), 0.63 (se : 0.10), and 0.57 (se : 0.12) respectively, which are significant at $p < 0.01$ in all specifications.

norm. The idea that salience is tied to deviations from the norm has a long history in psychology (e.g., Kahneman and Miller, 1986), and has played a prominent role in recent economic modeling (see Bordalo et al., 2022, for a review).

To quantify the "regularity" of past attendances, we define the *day-of-week Herfindahl-Hirschman Index (HHI)* using fourteen full weeks, Monday-Sunday, of past attendance behavior immediately prior to the survey. To construct the day-of-week HHI, we first compute the raw day-of-week HHI: the sum of squared shares of visits that occur on each day of the week, resulting in a measure of the concentration of visits on certain days of the week. Since the minimum and maximum attainable values of the raw day-of-week HHI vary with the total number of past visits,[27] we rescale by subtracting the minimum attainable value from the raw day-of-week HHI and divide by the difference between the maximum and minimum attainable HHI as a function of past visits. This procedure results in a unit-interval measure of how concentrated a participant's visits are on particular days of the week. See Appendix E.1 for additional details on the construction of this measure.

Panel (b) of Figure 5 presents binned scatter plots that compare actual and perceived past visit likelihoods for those with above- versus below-median values of the HHI.[28] As in panel (a), both groups exhibit a linear relationship between the actual and perceived likelihood of attendance, and both overestimate past attendance. In contrast to panel (a), however, the relationship between the actual and perceived visit likelihood is nearly identical across the two groups. This similarity in perceptions of past attendance suggests a more limited role for salience bias in explaining selective recall. Panel (b) of Appendix Figure A5 shows the same pattern across the top and bottom quartile of the HHI.

To quantify the implications for selective recall, we again estimate $\rho_1 - \rho_0$, assuming that individuals ignore their recall selectivity when predicting the past. For specifications corresponding to Columns 1 through 4 of Table 1, respectively, we estimate that $\rho_1 - \rho_0$ is 0.36 (se : 0.03), 0.35 (se : 0.03), 0.36 (se : 0.03), and 0.36 (se : 0.04) among participants with an above-median HHI, and is 0.31 (se : 0.04),

[27]For example, an individual with only one past visit has all of their past visits concentrated on one day of the week, mechanically resulting in a high HHI. Meanwhile, an individual with eight past visits cannot have all of their past visits concentrated on a single day of the week, mechanically decreasing their HHI.

[28]Since the day-of-week HHI is constructed as a function of past visits, the sample is split at the median and into quartiles conditional on ventiles of attendance in the fourteen weeks used to construct the day-of-week HHI.

0.25 (se : 0.04), 0.34 (se : 0.04), and 0.28 (se : 0.04) among participants with a below-median HHI. The degree of selectivity does not meaningfully vary between the HHI groups, and is not statistically significantly different between the HHI groups at conventional levels. The results do not materially change when we instead compare the top and bottom quartile of HHI; the differences in $\rho_1 - \rho_0$ between the top and bottom quartile are again not statistically significant.[29]

Consistent with these findings, we also find that differences in the HHI do not predict differences in naivete about present focus—in contrast to our results about the goal-behavior gap. Appendix Table A10 estimates our structural model for participants with above- versus below-median HHI (panel (a)), and those with top- versus bottom-quartile HHI (panel (b)). In both panels, actual and perceived present focus are similar across the two groups, and the differences are not statistically significant at conventional levels.

Taken together, the results of this analysis, along with the estimates in Section 5.1, suggest that motivated memory plays a larger role in explaining selective recall than does salience bias.

6 Conclusion

This paper contributes new evidence of a link between memory biases and awareness of self-control limitations. We find that people with more upwardly-biased perceptions of their past gym attendance are also more naive about their self-control issues, that recall is selective, which can help sustain biased beliefs, and that this selectivity of recall is particularly pronounced for people with larger self-control problems. This provides empirical support for recent theoretical models in which biases in learning and memory formation support persistent overconfidence and naivete. An implication of our results and these models 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.

[29]The differences in the degree of selectivity, $\rho_1 - \rho_0$, between the top and bottom quartile are -0.03 (se : 0.06), 0.04 (se : 0.07), -0.03 (se : 0.06), and 0.01 (se : 0.07), respectively, none of which are statistically significant.

While our findings lend support to theories that biased beliefs (such as naivete about self-control problems) persist in part due to imperfect memory, they should not be misinterpreted as showing that biased recall *causes* naivete. Our findings are also consistent with theories where people start with biased beliefs, and these biases are not completely eliminated due to imperfections in recall. Biased perceptions of the past then result from biased priors. At the same time, we show that in order to hold biased beliefs that are internally consistent, people must exhibit both selective recall and under-appreciate the extent of that selectivity. In this way, our framework suggests that biased beliefs help to generate biased memories and that biased memory may help to support the persistence of biased beliefs. More research is needed to explore whether and when treatments that affect recall lead to changes in 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, Daniel Cohen, William Morrison, and Dmitry Taubinsky. 2025. "When Do "Nudges" Increase Welfare?" *American Economic Review* 115 (5): 1555–1596. 10.1257/aer.20231304.

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.

Amelio, Andrea, and Florian Zimmermann. 2023. "Motivated Memory in Economics–A Review." *Games* 14 (1): 15.

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.

Ba, Cuimin, J. Aislinn Bohren, and Alex Imas. 2023. "Over- and Underreaction to Information." Working Paper.

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.

Bénabou, Roland. 2015. "The Economics of Motivated Beliefs." *Revue d'économie politique* 125 (5): 665–685.

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

Bénabou, Roland, and Jean Tirole. 2004. "Willpower and Personal Rules." *Journal of Political Economy* 112 (4): 848–886.

Beshears, John, James J. Choi, Christopher Harris, David Laibson, Brigitte C. Madrian, and Jung Sakong. 2020. "Which Early Withdrawal Penalty Attracts the Most Deposits to a Commitment Savings Account?" *Journal of Public Economics* 183 104144.

Bordalo, P, G Burro, K Coffman, N Gennaioli, and A Shleifer. 2024. "Imagining the future: memory, simulation and beliefs." *The Review of Economic Studies.*

Bordalo, Pedro, John J Conlon, Nicola Gennaioli, Spencer Y Kwon, and Andrei Shleifer. 2023. "Memory and Probability." *The Quarterly Journal of Economics* 138 (1): 265–311.

Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer. 2018. "Diagnostic Expectations and Credit Cycles." *The Journal of Finance* 73 (1): 199–227.

Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer. 2022. "Salience." *Annual Review of Economics* 14 (1): 521–544. 10.1146/annurev-economics-051520-011616.

Caballero, Adrián, and Raúl López-Pérez. 2024. "Memory Bias Beyond Ego: Selective Recall of Positive Financial Outcomes." *Journal of Economic Psychology* 105 102771.

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.

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

Conlon, John. 2025. "Memory Rehearsal and Belief Biases." Working Paper.

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

Enke, Benjamin. 2020. "What You See Is All There Is." *The Quarterly Journal of Economics* 135 (3): 1363–1398.

Enke, Benjamin, Frederik Schwerter, and Florian Zimmermann. 2024. "Associative Memory and Belief Formation." *Journal of Financial Economics* 157 103853.

Fudenberg, Drew, Giacomo Lanzani, and Philipp Strack. 2024. "Selective Memory Equilibrium." *Journal of Political Economy* 132 (12): 3978–4020.

Gagnon-Bartsch, Tristan, Matthew Rabin, and Joshua Schwartzstein. 2023. "Channeled Attention and Stable Errors." Working Paper.

Gödker, Katrin, Peiran Jiao, and Paul Smeets. 2025. "Investor Memory." *The Review of Financial Studies* 38 (6): 1595–1640.

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.

Graeber, Thomas, Christopher Roth, and Florian Zimmermann. 2024. "Stories, Statistics, and Memory." *The Quarterly Journal of Economics* 139 (4): 2181–2225.

Greene, William H. 2012. *Econometric Analysis*. Prentice Hall.

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, and Botond Kőszegi. 2009. "Futile Attempts at Self-Control." *Journal of the European Economic Association* 7 (2-3): 423–434.

Heidhues, Paul, and Botond Kőszegi. 2010. "Exploiting Naïvete about Self-Control in the Credit Market." *American Economic Review* 100 (5): 2279–2303.

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

Heidhues, Paul, Botond Kőszegi, and Philipp Strack. 2024. "Misinterpreting Yourself." Working Paper.

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.

Kahneman, Daniel, and Dale T. Miller. 1986. "Norm theory: Comparing reality to its alternatives." *Psychological Review* 93 (2): 136–153. 10.1037/0033-295X.93.2.136.

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

Kőszegi, Botond, George Loewenstein, and Takeshi Murooka. 2022. "Fragile Self-Esteem." *The Review of Economic Studies* 89 (4): 2026–2060.

Li, King King. 2013. "Asymmetric memory recall of positive and negative events in social interactions." *Experimental Economics* 16 248–262.

Li, King King, and Kang Rong. 2023. "Real-Life Investors' Memory Recall Bias: A Lab-in-the-Field Experiment." *Journal of Behavioral and Experimental Finance* 37 100760.

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.

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

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.

Roy-Chowdhury, Vivek. 2024. "Self-Confidence and Motivated Memory Loss: Evidence From Schools." Working Paper.

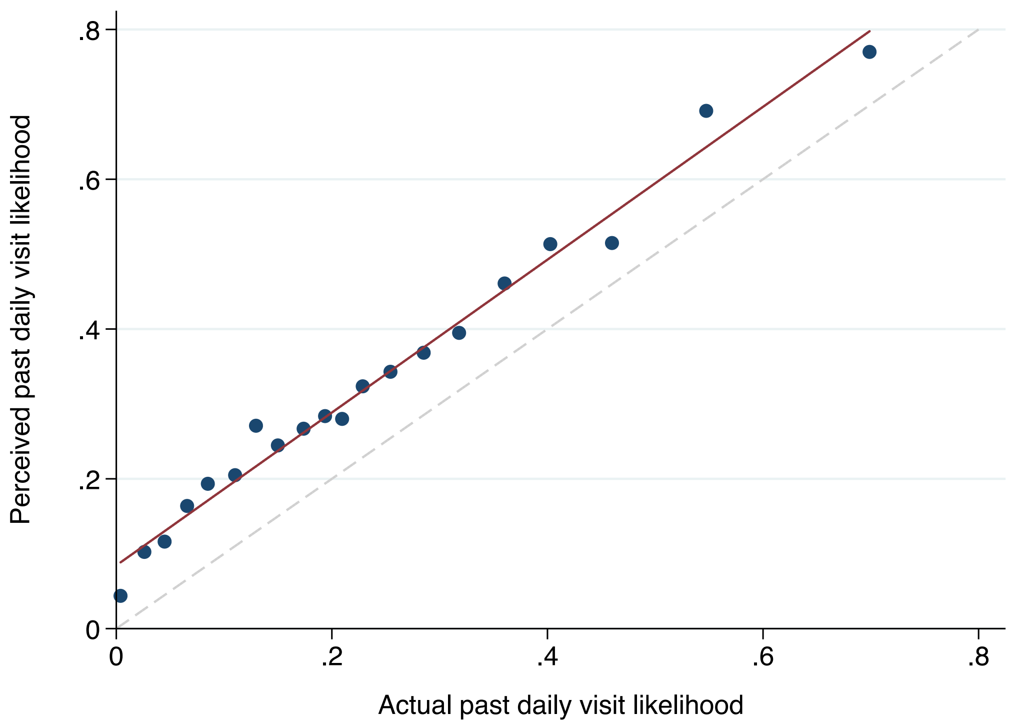
Saucet, Charlotte, and Marie Claire Villeval. 2019. "Motivated Memory in Dictator Games." *Games and Economic Behavior* 117 250–275.

Schwartzstein, Joshua R. 2014. "Selective Attention and Learning." *Journal of the European Economic Association* 12 (6): 1423–1452.

Zellner, Arnold. 1962. "An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias." *Journal of the American Statistical Association* 57 (298): 348–368.

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

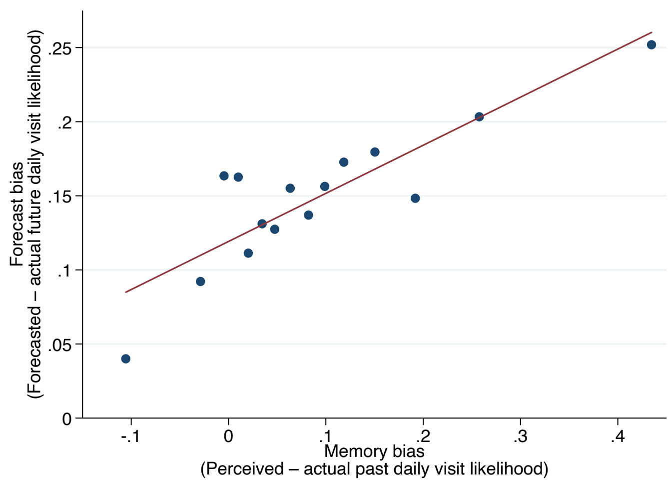
Figure 1: Perception of past daily likelihood of visiting the gym



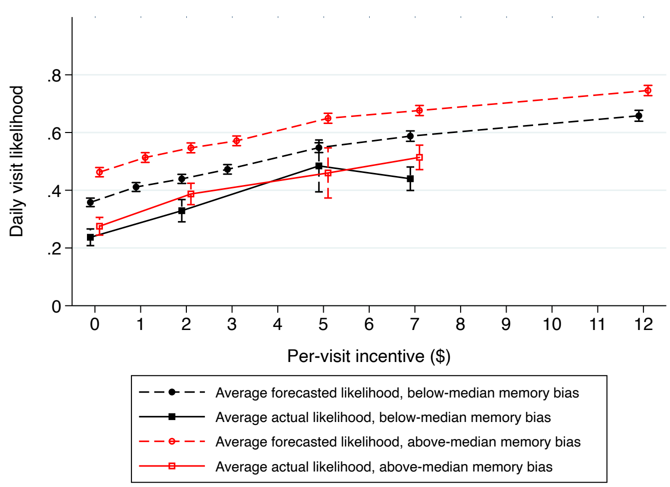
Notes: This figure presents a binned scatterplot comparing participants' actual past daily likelihood of visiting the gym and their perception of their past daily likelihood of visiting the gym. Actual past daily visit likelihood is the fraction of days—either out of the past 100 or out of the total membership duration when the duration was lower than 100—on which the participant attended the gym. Perceived past daily visit likelihood is defined analogously, but using the participant's recollection in the numerator. No additional sample restrictions are made. A dashed 45-degree line is included for reference.

Figure 2: Biases in forecasts vs. biases in memory of gym attendance

(a) Forecast bias by memory bias



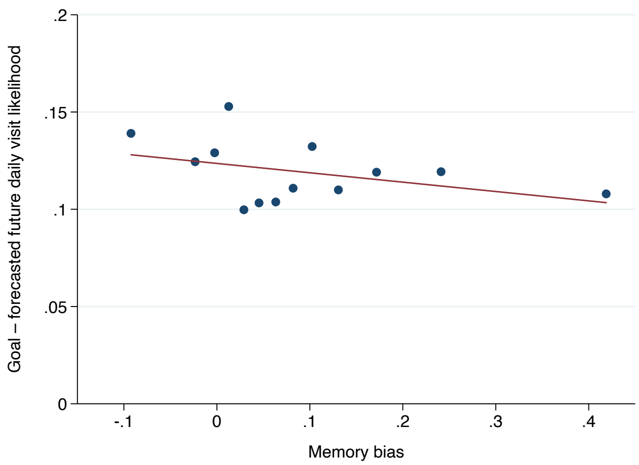
(b) Forecasted and actual attendance by memory bias and incentive



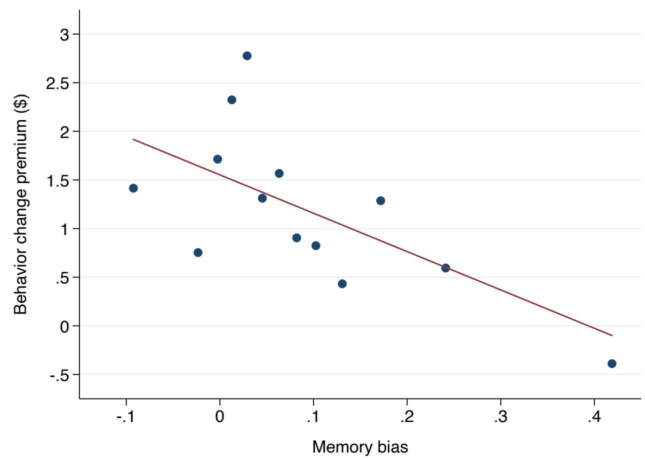
Notes: Panel (a) of this figure presents a binned scatterplot comparing participants' memory bias and forecast bias. Memory bias is defined on a scale from 0 to 1 as the difference between participants' perceived and actual daily likelihood of visiting the gym in the past 100 days, both as defined in Figure 1. Forecast biases are similarly defined as the difference between participants' forecasted and actual daily likelihood of visiting the gym during the experiment under their randomly-assigned incentive. The sample excludes 121 participants assigned a commitment contract since forecasted attendance under commitment contracts was not elicited. Panel (b) of this figure compares the means and 95% confidence intervals of participants' forecasted and actual daily gym visit likelihood during the experiment under their randomly-assigned incentive for the subsamples with below- and above-median memory bias. The sample median of memory bias is 0.06. Forecasted visit likelihoods are averaged over all participants in each subsample, while actual visit likelihoods are averaged over the participants within each subsample randomly assigned each 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= 412; $2: N= 292; $5: N= 75; $7: N= 339).

Figure 3: Proxies for awareness of present focus by memory bias

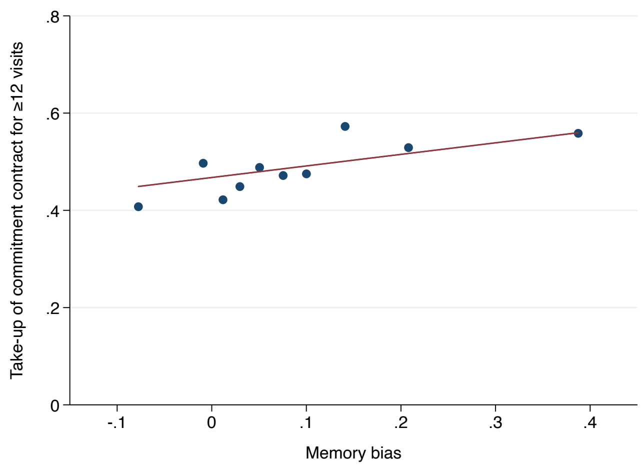
(a) Goal-forecasted attendance gap



(b) Behavior change premium



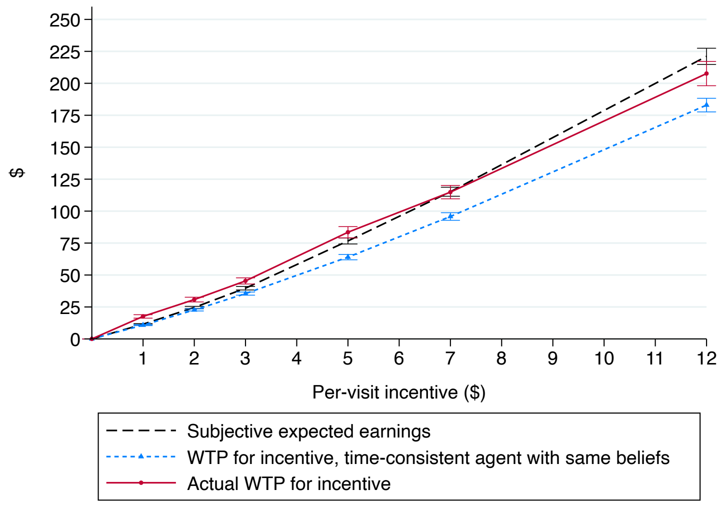
(c) Commitment contract take-up



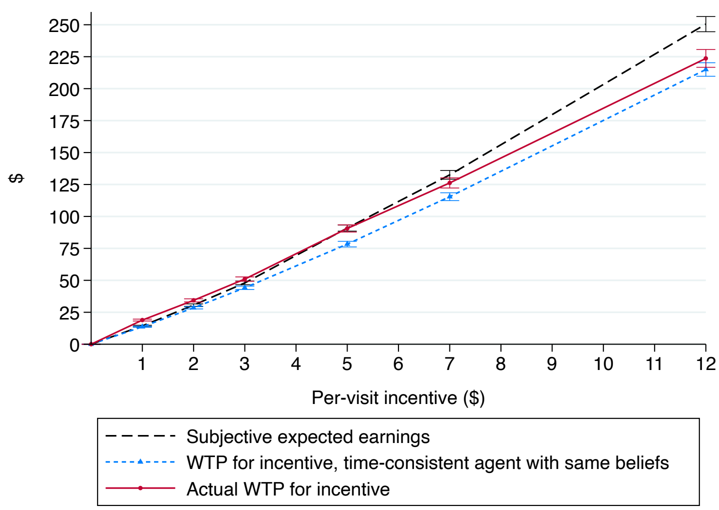
Notes: This figure presents three binned scatterplots comparing proxies for awareness of present focus to memory bias. Panel (a) compares the difference between participants' goals and forecasts of their likelihood of visiting the gym on a given day during the experiment to their memory bias. Goals and forecasts are for attendance in the absence of incentives. Panel (b) compares participants' estimated behavior change premium to their memory bias. Panel (c) compares participants' take-up of the commitment contract for at least 12 gym visits to their memory bias. As in Figure 2, memory bias is the difference between participants' perceived and actual past daily likelihood of visiting the gym. See Appendix C for additional information about the behavior change premium.

Figure 4: Willingness to pay for per-visit incentives by memory bias

(a) Below-median memory bias



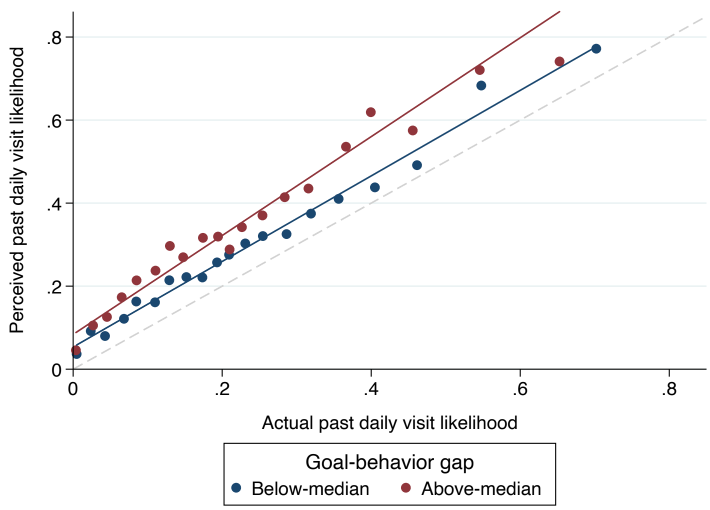
(b) Above-median memory bias



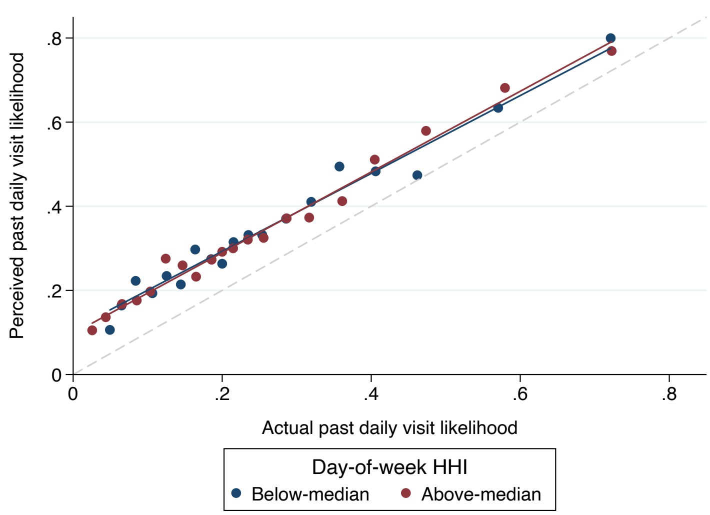
Notes: This figure compares means and 95% confidence intervals of participants' subjective expected earnings under each per-visit incentive to their actual willingness to pay for that incentive, as well as the willingness to pay for a time-consistent agent with the same beliefs. Subjective expected earnings are the product of the per-visit incentive and participants' subjective beliefs about the number of days they would visit under that incentive. The willingness to pay for a time-consistent agent is approximated by numerically integrating the area under the forecasted attendance by per-visit incentive curve. The samples in panels (a) and (b) are restricted to participants with below- and above-median memory bias, respectively. As in Figure 2, memory bias is the difference between participants' perceived and actual past daily likelihood of visiting the gym.

Figure 5: Heterogeneity in perceptions of past daily likelihood of visiting the gym

(a) Above- vs. below-median goal-behavior gap



(b) Above- vs. below-median day-of-week HHI



Notes: Panel (a) of this figure modifies Figure 1 by splitting the sample at the median value of the *goal-behavior gap*. The *goal-behavior gap* is the difference between participants' goal for their daily likelihood of visiting the gym during the experiment and their actual past daily likelihood of visiting the gym in the 100 days prior, or out of the total membership duration when the duration was lower than 100. No additional sample restrictions are made. Panel (b) of this figure modifies Figure 1 by splitting the sample at the median value of the *day-of-week Herfindahl-Hirschman Index (HHI)*, conditional on the ventile of attendance in the past fourteen weeks. The *day-of-week HHI* is a measure of the concentration of participants' gym visits on particular days of the week over the fourteen weeks prior, with higher values indicating more concentrated visits conditional on the total number of visits over this period. The sample is restricted to those with at least fourteen full weeks of membership prior to the survey. See Section E.1 for a formal definition of the *day-of-week HHI*.

Table 1: Perceived past attendance and memory bias by actual past attendance

	Perceived past daily visit likelihood				Memory bias	
	(1)	(2)	(3)	(4)	(5)	(6)
Daily visit likelihood,	0.99***	0.94***	1.02***	0.98***	–0.06**	–0.02
1-100 days prior	(0.02)	(0.03)	(0.02)	(0.03)	(0.03)	(0.03)
Daily visit likelihood,		0.11***		0.10**	0.11***	0.10**
>100 days prior		(0.04)		(0.05)	(0.04)	(0.05)
Dep. var. prediction,	0.09	0.10	0.08	0.09	0.10	0.09
at 0 past attendance	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Membership days >:	100	100	200	200	100	200
Demographic controls	Yes	Yes	Yes	Yes	Yes	Yes
Wave fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
N	1,025	1,025	821	821	1,025	821

Notes: This table reports the association between participants' perceived likelihood of visiting the gym in the 100 days prior and their actual daily likelihood of visiting the gym during the same period and prior to that period. Each column presents coefficient estimates from OLS regressions, with heteroskedasticity-robust standard errors reported in parentheses. In Columns 1-4, the dependent variable is participants' perceived likelihood of visiting the gym in the 100 days prior. In Columns 5-6, the dependent variable is memory bias, defined as in Figure 2 as the difference between participants' recalled and actual past daily likelihood of visiting the gym in the 100 days prior. *Daily visit likelihood, 1-100 days prior* is participants' daily likelihood of visiting the gym in the 100 days prior. *Daily visit likelihood, >100 days prior* is participants' daily likelihood of visiting the gym on any given day more than 100 days prior. In each column, predicted values of the dependent variable at zero past actual attendance and the means of all other regressors are reported, with standard errors in parentheses. In all columns, "demographic controls" include gender, student status, age, and the natural log of membership duration. The sample in each column excludes 6 participants who declined to state their gender or age. The sample in Columns 1, 2, and 5 is restricted to participants with a membership greater than 100 days, while the sample in Columns 3, 4, and 6 is restricted to participants with a membership greater than 200 days. *, **, ***: statistically significantly different from 0 at the 10%, 5%, and 1% level, respectively.

Table 2: 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**	–4.38***	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	1,115	1,115	1,115	2,795
Clusters	1,115	1,115	1,115	1,115

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 2, memory bias is the difference between participants' perceived and actual past 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. In Columns 1-3, heteroskedasticity-robust standard errors are reported. 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 prior to the experiment. The sample excludes 121 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 3: 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=561)	0.54 $(0.48, 0.59)$	0.78 $(0.72, 0.85)$	9.09 $(8.28, 9.91)$	15.44 $(13.53, 17.35)$	0.47 $(0.37, 0.57)$
2	Above med. (N=560)	0.55 $(0.51, 0.59)$	0.89 $(0.85, 0.93)$	10.01 $(9.11, 10.91)$	14.00 $(12.59, 15.40)$	0.25 $(0.17, 0.33)$
3	Difference	-0.01 $(-0.08, 0.06)$	-0.10 $(-0.18, -0.03)$	-0.92 $(-2.13, 0.30)$	1.45 $(-0.93, 3.82)$	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=561)	1.82 $(1.12, 2.52)$	0.34 $(0.32, 0.36)$	0.12 $(0.10, 0.14)$
2		Above med. (N=560)	0.53 $(0.05, 1.00)$	0.39 $(0.37, 0.41)$	0.17 $(0.16, 0.19)$
3		Difference	1.29 $(0.45, 2.14)$	-0.05 $(-0.08, -0.02)$	-0.05 $(-0.08, -0.03)$
4	Predicted	Below med. (N=561)	2.08 $(1.45, 2.70)$	0.34 $(0.32, 0.36)$	0.11 $(0.10, 0.13)$
5		Above med. (N=560)	1.14 $(0.72, 1.56)$	0.39 $(0.37, 0.41)$	0.16 $(0.14, 0.17)$
6		Difference	0.94 $(0.19, 1.69)$	-0.05 $(-0.08, -0.02)$	-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. The parameters β, $\tilde{\beta}$, b, and $1/\lambda$ denote, respectively, present focus, perceived present focus, perceived health benefits of a gym visit, and the mean costs of a gym visit. Standard errors are clustered at the participant level. See Appendix D.1 for details about the GMM estimation procedure. Panel (b) reports empirical and model-predicted means, differences in means, and 95% confidence intervals for the following moments: the average behavior change premium, actual daily attendance likelihood during the experiment, and the difference between forecasted and actual daily attendance likelihood under assigned incentives. 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). The Delta method is used for inference on the statistics in Columns 4-5 of panel (a) and rows 4-6 of panel (b). The sample excludes 121 participants assigned a commitment contract since forecasted attendance under commitment contracts was not elicited.

Table 4: 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=561)	0.54 By assump.	0.85 $(0.82, 0.89)$	9.50 $(8.64, 10.36)$	15.88 $(14.39, 17.37)$	17.20 $(15.58, 18.83)$
2	Above med. (N=560)	0.55 By assump.	0.85 $(0.82, 0.89)$	9.65 $(8.81, 10.50)$	13.73 $(12.48, 14.97)$	13.18 $(11.93, 14.44)$
3	Difference	-0.01 By assump.	0 By assump.	-0.15 $(-1.34, 1.04)$	2.15 $(0.22, 4.08)$	4.02 $(2.26, 5.78)$

(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=561)	1.82 $(1.12, 2.52)$	0.34 $(0.32, 0.36)$	0.12 $(0.10, 0.14)$
2	Empirical	Above med. (N=560)	0.53 $(0.05, 1.00)$	0.39 $(0.37, 0.41)$	0.17 $(0.16, 0.19)$
3	Empirical	Difference	1.29 $(0.45, 2.14)$	-0.05 $(-0.08, -0.02)$	-0.05 $(-0.08, -0.03)$
4	Predicted	Below med. (N=561)	1.42 $(1.08, 1.76)$	0.34 $(0.32, 0.36)$	0.11 $(0.10, 0.13)$
5	Predicted	Above med. (N=560)	1.50 $(1.14, 1.85)$	0.39 $(0.37, 0.41)$	0.16 $(0.14, 0.17)$
6	Predicted	Difference	-0.08 $(-0.10, -0.05)$	-0.05 $(-0.08, -0.02)$	-0.04 $(-0.07, -0.02)$

Notes: Panel (a) of this table modifies panel (a) of Table 3 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 3. 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 3.

Appendix: Proofs of Propositions

Proof of Proposition 1

Proof. For an individual who attended the gym a fraction ϕ_i of times and happens to recall $100 \cdot r_i^1$ attendances and $100 \cdot r_i^0$ non-attendances,

$$\tilde{\phi}_i = \underbrace{r_i^1}_{\substack{\text{recalled}\\\text{attendance}}} + \underbrace{\left(1 - r_i^1 - r_i^0\right)}_{\substack{\text{fraction of days}\\\text{that are forgotten}}} \underbrace{\tilde{\nu}}_{\substack{\text{perceived}\\\text{attendance on}\\\text{forgotten days}}}$$

By definition, $\mathbb{E}[r_i^1|\phi_i] = \rho_1\phi_i$ and $\mathbb{E}[r_i^0|\phi_i] = \rho_0(1-\phi_i)$, and thus

$$\mathbb{E}\left[\tilde{\phi}_i|\phi_i\right] = \rho_1\phi_i + \left[(1-\rho_1)\phi_i + (1-\rho_0)(1-\phi_i)\right]\tilde{\nu} \tag{7}$$

$$= \left[\rho_1 - (\rho_1 - \rho_0)\tilde{\nu}\right]\phi_i + (1-\rho_0)\tilde{\nu}. \tag{8}$$

Equation (2) and part 1 immediately follow from the above.

To establish part 2, first note that $\tilde{F}(\tilde{\beta}(b+p)) > F(\beta(b+p))$ for all $p \geq 0$ under the conditions stated there. This shows that average forecast bias, $\tilde{F}(\tilde{\beta}(b+p)) - \mathbb{E}[\gamma_i] = \tilde{F}(\tilde{\beta}(b+p)) - F(\beta(b+p))$, will be positive. To establish that memory bias will be positive, first note that when $\tilde{\mu} = \mu$, $\tilde{\rho}_0 = \rho_0$, and $\tilde{\rho}_1 = \rho_1$,

$$\begin{aligned}
\mathbb{E}\left[\tilde{\phi}_i\right] &= \rho_1\mu + \left[(1-\rho_1)\mu + (1-\rho_0)(1-\mu)\right]\tilde{\nu} \\
&= \rho_1\mu + \frac{(1-\rho_1)\,\mu\left[(1-\rho_1)\mu + (1-\rho_0)(1-\mu)\right]}{(1-\rho_1)\,\mu + (1-\rho_0)\,(1-\mu)} \\
&= \rho_1\mu + (1-\rho_1)\,\mu = \mu
\end{aligned}$$

Next, note that (i) $\tilde{\mu} > \mu$ under the asusmptions in the Proposition; (ii) $\tilde{\nu}$ is strictly monotonic in $\tilde{\mu}$ by equation (1); (iii) $\tilde{\phi}_i$ is strictly monotonic in $\tilde{\nu}$ by equation (7); and (iv) $\tilde{\nu}$ is increasing in $\tilde{\rho}_0$ and decreasing in $\tilde{\rho}_1$. Thus, when $\tilde{\mu} > \mu$, $\mathbb{E}\left[\tilde{\phi}_i\right] > \mathbb{E}\left[\phi_i\right] = \mu$, which shows that memory bias is positive.

To establish part 3, simply note that $\tilde{F}(\tilde{\beta}(b+p))$ is increasing in $\tilde{\beta}$ and decreasing in $\tilde{F}$ (in the FOSD order) for all p, which shows that average forecast bias will be increasing in $\tilde{\beta}$ and decreasing in $\tilde{F}$. When $p = 0$, this shows that $\tilde{\mu}$ is increasing in $\tilde{\beta}$

and decreasing in $\tilde{F}$, which by the logic above shows that average memory bias will be increasing in $\tilde{\beta}$ and decreasing in $\tilde{F}$. $\square$

Proof of Proposition 2

Proof. By Proposition 1,

$$\rho_0 = 1 - \frac{b_0}{\tilde{\nu}} = \frac{\tilde{\nu} - b_0}{\tilde{\nu}}$$
$$\rho_1 = \frac{b_1 - \rho_0 \tilde{\nu}}{1 - \tilde{\nu}} = \frac{b_1 - \tilde{\nu} + b_0}{1 - \tilde{\nu}}$$

and thus

$$\begin{aligned} \rho_1 - \rho_0 &= \frac{b_1 + b_0 - \tilde{\nu}}{1 - \tilde{\nu}} + \frac{b_0}{\tilde{\nu}} - 1 \\ &= \frac{b_0 - \tilde{\nu}(1 - b_1)}{\tilde{\nu}(1 - \tilde{\nu})} \end{aligned}$$

Now

$$\begin{aligned} \frac{d\rho_0}{d\tilde{\nu}} &= \frac{b_0}{\tilde{\nu}^2} = \frac{1 - \rho_0}{\tilde{\nu}} > 0 \\ \frac{d\rho_1}{d\tilde{\nu}} &= \frac{b_0 + b_1 - 1}{(1 - \tilde{\nu})^2} \\ &= \frac{(\rho_1 - 1)(1 - \tilde{\nu})}{(1 - \tilde{\nu})^2} = \frac{\rho_1 - 1}{1 - \tilde{\nu}} < 0 \end{aligned}$$

Thus, $\rho_1 - \rho_0$ is decreasing in $\tilde{\nu}$. Finally, note that

$$\begin{aligned} \tilde{\nu} &= \frac{(1 - \tilde{\rho}_1)\,\tilde{\mu}}{(1 - \tilde{\rho}_1)\,\tilde{\mu} + (1 - \tilde{\rho}_0)\,(1 - \tilde{\mu})} \\ &\leq \frac{(1 - \tilde{\rho}_1)\,\tilde{\mu}}{(1 - \tilde{\rho}_1)\,\tilde{\mu} + (1 - \tilde{\rho}_1)\,(1 - \tilde{\mu})} \\ &= \frac{(1 - \tilde{\rho}_1)\,\tilde{\mu}}{(1 - \tilde{\rho}_1)} = \tilde{\mu} \end{aligned}$$

with equality obtained only if the individual is unaware of their selective recall ($\tilde{\rho}_1 = \tilde{\rho}_0$). $\square$

A Sustaining Biased Beliefs via Asymmetric Recall

To motivate our consistency conditions, note that an individual who attends the gym a fraction μ of days would end up recalling attendance on a fraction $\rho_1\mu$ of days and recalling non-attendance on a fraction $\rho_0(1-\mu)$ of days. For the individual's mental model to be internally consistent with their recalled data in the long run, it must satisfy the following consistency condition:

Definition A1. The agent has long-run-consistent beliefs if

$$\tilde{\rho}_1\tilde{\mu} = \rho_1\mu \tag{9}$$

$$\tilde{\rho}_0(1-\tilde{\mu}) = \rho_0(1-\mu) \tag{10}$$

That is, the recalled number of attendances and non-attendances, respectively, must correspond to what would be implied by the individual's mental model. These consistency conditions are motivated by Heidhues et al.'s (2024) model, where individuals' beliefs about themselves must be consistent with their (recalled) history. The difference is that Heidhues et al. assume perfect memory but allow for multiple sources of mis-specification. Our consistency conditions are also related to the selective memory equilibrium concept of Fudenberg et al. (2024), with the key difference being that Fudenberg et al. (2024) do not require that people's perceptions of their memory process are consistent with the recalled events.[30]

Note that when the consistency conditions are satisfied, individuals have no way of realizing that their perceived attendance likelihood $\tilde{\mu}$ is different from their actual attendance likelihood μ, and they have no way of even realizing that $\tilde{\rho}_1 \neq \rho_1$ or $\tilde{\rho}_0 \neq \rho_0$. For example, individuals believe that they will recall what happened on a fraction $r = \tilde{\rho}_1\tilde{\mu} + \tilde{\rho}_0(1-\tilde{\mu})$ of days, and they indeed remember what happened on a fraction $\rho_1\mu + \rho_0(1-\mu) = r$ of days.

We now characterize the set of all possible beliefs $\tilde{\mu}$ that are consistent with Definition A1.

[30] This differs from the selective memory equilibrium concept of Fudenberg et al. (2024), including the generalization in Appendix A.3, which does not require that people's perceived memory process is consistent with their actual memory process. For example, Fudenberg et al. (2024) allow individuals to believe that they never forget. In our setting, this corresponds to $\tilde{\rho}_0 = \tilde{\rho}_1 = 1$. This violates the conditions of Definition A1 because equation (9) would imply that $\tilde{\mu} = \rho_1\mu$ while equation (10) would imply that $\tilde{\mu} = 1-\rho_0(1-\mu)$; thus, $\rho_1\mu = 1-\rho_0(1-\mu) = \rho_0\mu + 1 - \rho_0$, or $\mu = \frac{1-\rho_0}{\rho_1-\rho_0}$. In other words, if $\tilde{\rho}_0 = \tilde{\rho}_1 = 1$, then Definition A1 is violated for all values μ that don't equal $\frac{1-\rho_0}{\rho_1-\rho_0}$.

Proposition A1. *When the person is fully sophisticated about selective recall, $\tilde{\rho}_1 = \rho_1$ and $\tilde{\rho}_0 = \rho_0$, long-run consistency requires $\tilde{\mu} = \mu$. When the person is naive about recall, $\tilde{\rho}_0 = \tilde{\rho}_1$, $\tilde{\mu} = \frac{\rho_1}{\mu(\rho_1-\rho_0)+\rho_0}\mu$. For any $\tilde{\mu} \in \left[\mu, \frac{\rho_1}{\mu(\rho_1-\rho_0)+\rho_0}\mu\right]$, there exist perceived recall parameters $\tilde{\rho}_0$ and $\tilde{\rho}_1$ satisfying $\tilde{\rho}_0 \leq \tilde{\rho}_1$ such that $\tilde{\mu}$, $\tilde{\rho}_0$, $\tilde{\rho}_1$ satisfy the long-run consistency requirements (9) and (10). The maximum is attained when $\tilde{\rho}_0 = \tilde{\rho}_1$; that is, when individuals believe that they forget at random. The minimum is attained when $\tilde{\rho}_0 = \rho_0$ and $\tilde{\rho}_1 = \rho_1$; that is, when individuals correctly understand their forgetting process.*

Proposition A1 generates several insights about the long-run persistence of biased beliefs. First, selective recall, $\rho_1 > \rho_0$, is a necessary condition. When $\rho_1 = \rho_0$, Proposition A1 shows that $\tilde{\mu} = \mu$. Second, misperceptions of selective recall are also necessary: otherwise Proposition A1 also implies that $\tilde{\mu} = \mu$. Third, the belief that recall is symmetric, $\tilde{\rho}_0 = \tilde{\rho}_1$, generates the most biased long-run beliefs.

To illustrate Proposition A1, observe that if the actual attendance likelihood is $\mu = 0.25$ and $\rho_1 = 1$ while $\rho_0 = 0.5$, then $\tilde{\mu}$ can be as high as $\mu/(0.125 + 0.5) = 1.6\mu = 0.4$. Alternatively, when $\rho_0 = 0$, so that the individual only remembers days on which they attended the gym, then $\tilde{\mu}$ can be as high as 1; that is, the belief that the individual will always attend the gym can be sustained.

A.1 Proof of Proposition A1

Proof. First, note that when $\tilde{\rho}_0 = \rho_0$ and $\tilde{\rho}_1 = \rho_1$, the consistency conditions (9) and (10) imply that $\tilde{\mu} = \mu$.

Next, note that condition (9) implies that $\mu = \frac{\tilde{\rho}_1}{\rho_1}\tilde{\mu}$. Plugging this into (10) implies

$$\begin{aligned}
\tilde{\rho}_0 (1 - \tilde{\mu}) &= \rho_0 \left(1 - \frac{\tilde{\rho}_1}{\rho_1}\tilde{\mu}\right) \\
\Leftrightarrow \tilde{\mu} \left(\tilde{\rho}_0 - \frac{\tilde{\rho}_1}{\rho_1}\rho_0\right) &= \tilde{\rho}_0 - \rho_0 \\
\Leftrightarrow \tilde{\mu} &= \frac{\tilde{\rho}_0 - \rho_0}{\tilde{\rho}_0\rho_1 - \tilde{\rho}_1\rho_0}\rho_1 \qquad (11)
\end{aligned}$$

Equation (11) shows that in the quadrant where $\tilde{\rho}_1 \geq \tilde{\rho}_0$, $\tilde{\mu}$ is continuous in $\tilde{\rho}_1$ and $\tilde{\rho}_0$ and is decreasing in $\tilde{\rho}_1$ and increasing in $\tilde{\rho}_0$. Thus, the maximum value of $\tilde{\mu}$, call it $\tilde{\mu}^*$, is obtained when $\tilde{\rho}_1 = \tilde{\rho}_0$, and any value in the set $[\mu, \tilde{\mu}^*]$ can be obtained by an appropriate choice of $(\tilde{\rho}_0, \tilde{\rho}_1)$.

We now compute $\tilde{\mu}$ when $\tilde{\rho}_0 = \tilde{\rho}_1 \equiv \tilde{\rho}$. In this case, conditions (9) and (10) imply that

$$\begin{aligned}
\frac{1-\tilde{\mu}}{1-\mu} &= \frac{\rho_0}{\tilde{\rho}} \\
\Leftrightarrow \frac{1-\rho_1\mu/\tilde{\rho}}{1-\mu} &= \frac{\rho_0}{\tilde{\rho}} \\
\Leftrightarrow \frac{\tilde{\rho}-\rho_1\mu}{1-\mu} &= \rho_0 \\
\Leftrightarrow \tilde{\rho}-\rho_1\mu &= \rho_0 - \rho_0\mu \\
\Leftrightarrow \tilde{\rho} &= \mu(\rho_1-\rho_0)+\rho_0
\end{aligned}$$

while (11) reduces to

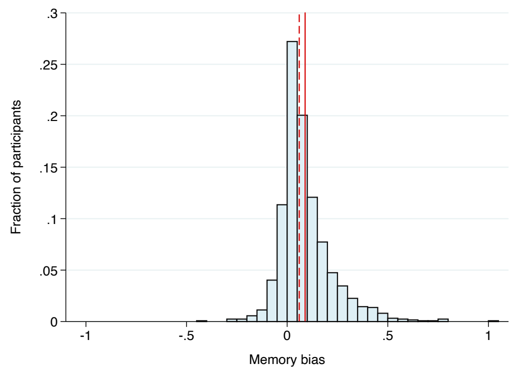
$$\begin{aligned}
\tilde{\mu} &= \frac{\tilde{\rho}-\rho_0}{\tilde{\rho}(\rho_1-\rho_0)}\rho_1 \\
&= \frac{\mu(\rho_1-\rho_0)}{\tilde{\rho}(\rho_1-\rho_0)}\rho_1 \\
&= \frac{\rho_1}{\tilde{\rho}}\mu \\
&= \frac{\rho_1}{\mu(\rho_1-\rho_0)+\rho_0}\mu
\end{aligned}$$

□

B Reduced-Form Results

B.1 Distributions of Memory Bias and Membership Duration

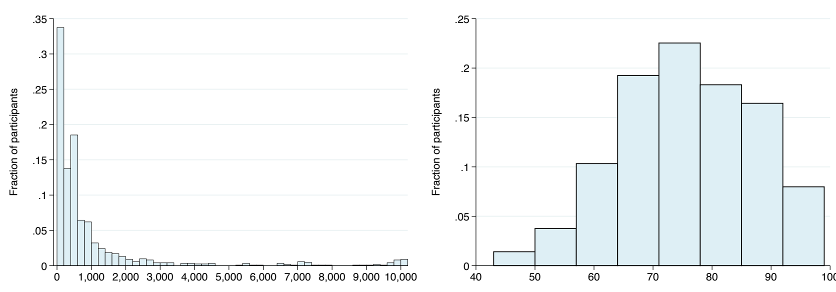
Figure A1: Histogram of biases in memory of attendance likelihood



Notes: This figure presents a histogram showing the distribution of memory biases in our sample. As in Figure 2, memory bias is the difference between participants' perceived and actual past 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 perceives their past number of days of gym attendance as greater 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.

B.2 The Behavior Change Premium and Memory Bias

Appendix Table A1 reproduces the result from a regression of the average behavior change premium on memory bias from Column 3 of Table 2, adds two variations of this regression with fewer controls, and compares these results to those from an alternative specification. In the alternative specification, we include an additional control for the forecasted change in attendance per dollar increase in attendance incentives. This variable corresponds to the term $\frac{\tilde{\alpha}(p+\Delta)-\tilde{\alpha}(p)}{\Delta}$ in equation (6), 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 2, the coefficients of memory bias are not statistically distinguishable across the columns with different controls.

Table A1: Control for forecasted attendance elasticity in BCP regressions

	Behavior change premium					
	(1)	(2)	(3)	(4)	(5)	(6)
Memory bias	–4.38***	–4.01**	–4.16**	–3.94**	–4.49***	–4.31**
	(1.63)	(1.65)	(1.65)	(1.66)	(1.68)	(1.69)
$\frac{\Delta \text{ forecasted attendance}}{\Delta \text{ incentive}}$		1.25***		1.53***		1.48***
		(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	1,115	1,115	1,115	1,115	1,115	1,115

Notes: This table reproduces Column 3 from Table 2 in Column 1, and modifies the controls in Column 1 by including only fixed effects—omitting the past attendance and demographic controls—and omitting all controls in Columns 3 and 5, respectively. It also modifies the controls in Columns 1, 3, and 5 by controlling for the average forecasted change in attendance per dollar increase in incentives in Columns 2, 4, and 6, respectively.

Table A2: Noise in behavior change premium measure and memory bias

	Standard deviation of behavior change premium		
	(1)	(2)	(3)
Memory bias	1.52	1.39	1.26
	(1.58)	(1.58)	(1.59)
Dependent var. mean	7.80	7.80	7.80
	(0.22)	(0.22)	(0.22)
Past attendance control	Yes	No	No
Demographic controls	Yes	No	No
Information fixed effects	Yes	Yes	No
Wave fixed effects	Yes	Yes	No
N	1,115	1,115	1,115

Notes: This table modifies Columns 1, 3, and 5 of Appendix Table A1 in Columns 1, 2, and 3, respectively, by replacing the average behavior change premium across incentive levels with the within-participant standard deviation of the behavior change premium.

B.3 Information Treatments

Carrera et al. (2022) report that the enhanced information treatment (but not the basic one) decreased overestimation of future visits, increased the average behavior change premium, and decreased take-up of commitment contracts. In a structural model, consistent with our discussion in the prior section, they estimate that the enhanced information treatment was associated with an increase in awareness of time inconsistency (i.e., a reduction in naivete about self-control problems).

Appendix Table A3 presents regression estimates for the same set of proxies of awareness of self-control problems used in our main reduced form estimates in Table 2 on indicators for above-median memory bias (versus below-median bias as the omitted category), indicators for the information treatment groups, and the interactions between above-median memory bias and the information treatment groups. In Columns 1 and 3, consistent with the results reported in Carrera et al. (2022), we find that the main estimated effect of the enhanced information treatment was to significantly decrease participants' forecast bias and increase the behavior change premium, respectively. Both of these results are consistent with the enhanced information treatment increasing awareness of present focus. The bottom row in the table shows the interaction term between the enhanced information treatment and a participant having above-median memory bias. We estimate a very small and statistically insignificant interaction for the gap between forecasted and actual attendance. There is a somewhat more sizable interaction for the behavior change premium in Column 3, indicating that the enhanced information treatment increased the BCP less for those with more positively biased memories, but this difference is imprecisely estimated and not statistically significant.

Table A3: 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	–0.95*	0.09**
	(0.02)	(0.01)	(0.48)	(0.04)
Basic info. treatment	0.00	–0.02	–0.28	–0.01
	(0.03)	(0.02)	(0.68)	(0.05)
Basic info. treatment	–0.03	–0.02	1.04	–0.01
× above-med. memory bias	(0.04)	(0.02)	(0.94)	(0.07)
Enhanced info. treatment	–0.06***	0.00	2.04**	–0.06
	(0.02)	(0.01)	(0.96)	(0.04)
Enhanced info. treatment	0.01	–0.02	–1.35	–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	1,115	1,115	1,115	2,795
Clusters	1,115	1,115	1,115	1,115

Notes: This table modifies Table 2 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. As in Figure 2, memory bias is the difference between participants' perceived and actual past daily likelihood of visiting the gym. See Section 1 for a description of the basic and enhanced information treatments.

B.4 Bias-Adjusted Estimates of the Effects of Memory Bias

We implement the method developed by Oster (2019) to obtain consistent estimates of the association between proxies for awareness of present focus and memory bias—

our "treatment" of interest—adjusted for bias from the omission of unobserved confounders.[31] We use the conservative estimate of $\delta = 1$ in the notation of Oster (2019); a value of 1 indicates that "the unobservable and observables are equally related to the treatment."

We use the suggested value $R^2_{max} = 1.3\tilde{R}^2$. In addition to this benchmark, 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 prior 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 prior 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.

Appendix Table A4 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 2. 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.

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

Table A4: 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=1,115)	0.31 [0.04]	0.30 [0.10]	0.30 [0.13]
2	Forecasted – actual attendance (N=1,115)	0.31 [0.04]	0.30 [0.10]	0.20 [0.50]
3	Goal – forecasted attendance (N=1,115)	−0.04 [0.00]	−0.07 [0.06]	−0.08 [0.08]
4	Goal – forecasted attendance (N=1,115)	−0.04 [0.00]	−0.07 [0.06]	−0.31 [0.50]
5	Behavior change premium (N=1,115)	−4.15 [0.02]	−4.38 [0.04]	−4.50 [0.05]
6	Take-up of "more" visits contract (N=2,795)	0.28 [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 treatment status, 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 2. 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 prior 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 prior 100 days. In Row 6, observations are pooled across the three types of visit-threshold contracts. The sample excludes 121 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.

C 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 quasi-linear 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 $1 \leq t \leq T$ 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 A2. *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*

$$\frac{w(p+\Delta) - w(p)}{\Delta} \approx \frac{\tilde{\alpha}(p+\Delta) + \tilde{\alpha}(p)}{2}. \tag{12}$$

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

$$\frac{w(p+\Delta)-w(p)}{\Delta} \approx \underbrace{\frac{\tilde{\alpha}(p+\Delta)+\tilde{\alpha}(p)}{2}}_{\textit{Surplus if time-consistent}} + \underbrace{(1-\tilde{\beta})(b+p+\Delta/2)\frac{\tilde{\alpha}(p+\Delta)-\tilde{\alpha}(p)}{\Delta}}_{\textit{Behavior change premium}}. \tag{13}$$

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

$$BCP(p,\Delta) := \frac{w(p+\Delta)-w(p)}{\Delta} - \frac{\tilde{\alpha}(p+\Delta)+\tilde{\alpha}(p)}{2}, \tag{14}$$

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.[32]

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.

[32]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 (14) 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.

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 (13) becomes

$$\mathbb{E}\left[\frac{w_i(p+\Delta)-w_i(p)}{\Delta}\right]=\mathbb{E}\left[\frac{\tilde{\alpha}_i(p+\Delta)+\tilde{\alpha}_i(p)}{2}+(1-\tilde{\beta}_i)(b_i+p+\Delta/2)\frac{\tilde{\alpha}_i(p+\Delta)-\tilde{\alpha}_i(p)}{\Delta}\right]. \tag{15}$$

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

D Structural Results

D.1 Details on GMM Estimation of Parameters

The following discussion on generalized method of moments estimation of parameters is adapted 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, which result from forecasted attendance, actual attendance, and the behavior change premium, respectively:

$$\mathbb{E}\left[\left(28\left(1-e^{-\lambda(\tilde{\beta}(b+p))}\right)-\tilde{\alpha}_i(p)\right)p^n\right]=0 \tag{16}$$

$$\mathbb{E}\left[\left(28\left(1-e^{-\lambda(\beta(b+p_i))}\right)-a_i\right)p_i^n\right]=0 \tag{17}$$

$$\mathbb{E}\left[(1-\tilde{\beta})(b+(p_k+p_{k+1})/2)\frac{\tilde{\alpha}_i(p+\Delta_k)-\tilde{\alpha}_i(p)}{\Delta_k}-\left(\frac{w_i(p+\Delta_k)-w_i(p)}{\Delta_k}-\frac{\tilde{\alpha}_i(p+\Delta_k)+\tilde{\alpha}_i(p)}{2}\right)\right]=0 \tag{18}$$

where $p\in\mathcal{P}=\{0,1,2,3,5,7,12\}$, $n\in\{0,1,2\}$ and in the third set p_k and p_{k+1} are one of five pairs of adjacent incentives from the set $\mathcal{P}\setminus\{0\}$, and $\Delta_k:=p_{k+1}-p_k$.

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

$$\left(m(\xi) - m(\hat{\xi})\right)' W \left(m(\xi) - m(\hat{\xi})\right), \tag{19}$$

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,[33] 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 $\mathrm{cov}(\mathbf{z}u)$, where $\mathbf{z}$ is a vector of the instruments for the moment conditions. We then minimize using the weighting matrix $\hat{W} = \hat{S}^{-1}$, which gives the optimal $\hat{\xi}$ (Hansen, 1982).

D.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 (16) and (17), 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. In the first and second set of moments, in equations (16) and (17), we thus replace λ with $\tilde{\lambda}$ and β with $\bar{\beta}$, respectively.

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 (16) and (17), respectively, to account for the fact that forecasted attendance depends on individual i's perception of the distribution of costs and that we are no longer separately identifying the actual cost and actual present

[33] 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.

focus parameters.

D.2 Additional Structural Estimates of Baseline Present Focus Model

The results in this section study two versions of our baseline model in Table 3. We consider a version of our baseline model with four rather than two memory bias types in Appendix Table A5. The predicted moments from the model presented in Rows 4-6 of panel (b) of Appendix Table A5 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.

In panel (a) of Appendix Table A6, we present a seven-parameter version of the model in panel (a) of Table 3, assuming that the actual present focus parameter β is homogeneous across the population. These estimates are close to those in Table 3. Panel (b) of Appendix Table A6 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 4.

Table A5: Model with naivete about present focus, 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=266)	0.50 $(0.44, 0.56)$	0.74 $(0.69, 0.80)$	8.85 $(8.78, 8.93)$	14.31 $(11.72, 16.91)$	0.51 $(0.37, 0.65)$
3	Quartile 3 (N=280)	0.57 $(-0.58, 1.72)$	0.86 $(-0.53, 2.26)$	9.14 $(7.97, 10.31)$	14.50 $(12.49, 16.50)$	0.31 $(0.20, 0.43)$
4	Quartile 4 (N=280)	0.54 $(0.48, 0.60)$	0.92 $(0.86, 0.97)$	11.10 $(11.03, 11.17)$	13.82 $(11.88, 15.76)$	0.18 $(0.07, 0.29)$
5	Test of equality, p-value	0.56	0.00	0.08	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=561)	1.82 $(1.12, 2.52)$	0.34 $(0.32, 0.36)$	0.12 $(0.10, 0.14)$
2	Empirical	Above med. (N=560)	0.53 $(0.05, 1.00)$	0.39 $(0.37, 0.41)$	0.17 $(0.16, 0.19)$
3	Empirical	Difference	1.29 $(0.45, 2.14)$	-0.05 $(-0.08, -0.02)$	-0.05 $(-0.08, -0.03)$
4	Predicted	Below med. (N=561)	2.29 $(1.69, 2.89)$	0.34 $(0.32, 0.36)$	0.11 $(0.09, 0.13)$
5	Predicted	Above med. (N=560)	1.09 $(0.69, 1.49)$	0.40 $(0.38, 0.42)$	0.16 $(0.14, 0.17)$
6	Predicted	Difference	1.20 $(0.48, 1.92)$	-0.05 $(-0.08, -0.02)$	-0.05 $(-0.07, -0.02)$

Notes: Panel (a) of this table modifies panel (a) of Table 3 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 3.

Table A6: Model with naivete about present focus, 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=561)	0.55 (0.51, 0.58)	0.79 (0.74, 0.84)	9.17 (8.35, 9.98)	15.71 (14.07, 17.34)	0.46 (0.36, 0.55)
2	Above med. (N=560)	0.55 (0.51, 0.58)	0.88 (0.84, 0.92)	10.00 (9.11, 10.89)	13.91 (12.57, 15.25)	0.26 (0.18, 0.34)
3	Difference	0 By assump.	−0.09 (−0.14, −0.04)	−0.84 (−2.02, 0.35)	1.80 (−0.10, 3.70)	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=561)	1.82 (1.12, 2.52)	0.34 (0.32, 0.36)	0.12 (0.10, 0.14)
2	Empirical	Above med. (N=560)	0.53 (0.05, 1.00)	0.39 (0.37, 0.41)	0.17 (0.16, 0.19)
3	Empirical	Difference	1.29 (0.45, 2.14)	−0.05 (−0.08, −0.02)	−0.05 (−0.08, −0.03)
4	Predicted	Below med. (N=561)	1.97 (1.49, 2.46)	0.34 (0.32, 0.36)	0.11 (0.10, 0.13)
5	Predicted	Above med. (N=560)	1.19 (0.80, 1.58)	0.39 (0.38, 0.41)	0.16 (0.14, 0.17)
6	Predicted	Difference	0.78 (0.27, 1.29)	−0.05 (−0.07, −0.03)	−0.05 (−0.07, −0.03)

Notes: Panel (a) of this table modifies panel (a) of Table 3 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 3.

D.3 Additional Structural Estimates with Misperceptions of Costs

In the model in Table 4, we fix the present focus parameter at the values estimated in Table 3 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 A7. Reassuringly, in panel (a) of Appendix Table A7, the estimates of $\tilde{\beta}$, b, and $1/\lambda$ in Columns 2-4 are identical to those in panel (a) of Table 3, 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 A7 reports the predicted moments from this model, which exhibit no improvement in model fit relative to the baseline model.

We also implement an alternative adaptation of our GMM procedure, described as strategy (ii) in Appendix D.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 A8 reports parameter estimates and predicted moments from this model. Reassuringly, the estimates of $\tilde{\beta}$ and b in Appendix Table A8 are the same as those in our baseline model in Table 3, the estimate $\widehat{\lambda\beta}$ in Appendix Table A8 is close to the product of the estimates $\hat{\lambda}$ and $\hat{\beta}$ from the model in Table 3, and the estimate $1/\hat{\tilde{\lambda}}$ is almost identical to the estimate $1/\hat{\lambda}$ in Table 3.

Table A7: Model with naivete about present focus and misperceptions of costs, $\hat{\beta}$ from Table 3

(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=561)	0.54 By assump.	0.78 $(0.72, 0.85)$	9.09 $(8.28, 9.91)$	15.44 $(14.00, 16.89)$	15.44 $(13.53, 17.35)$
2	Above med. (N=560)	0.55 By assump.	0.89 $(0.85, 0.93)$	10.01 $(9.11, 10.91)$	14.00 $(12.72, 15.28)$	14.00 $(12.59, 15.40)$
3	Difference	-0.01 By assump.	-0.10 $(-0.18, -0.03)$	-0.92 $(-2.13, 0.30)$	1.45 $(-0.48, 3.38)$	1.45 $(-0.93, 3.82)$

(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=561)	1.82 $(1.12, 2.52)$	0.34 $(0.32, 0.36)$	0.12 $(0.10, 0.14)$
2	Empirical	Above med. (N=560)	0.53 $(0.05, 1.00)$	0.39 $(0.37, 0.41)$	0.17 $(0.16, 0.19)$
3	Empirical	Difference	1.29 $(0.45, 2.14)$	-0.05 $(-0.08, -0.02)$	-0.05 $(-0.08, -0.03)$
4	Predicted	Below med. (N=561)	2.08 $(1.45, 2.70)$	0.34 $(0.32, 0.36)$	0.11 $(0.09, 0.13)$
5	Predicted	Above med. (N=560)	1.14 $(0.72, 1.56)$	0.39 $(0.37, 0.41)$	0.16 $(0.14, 0.18)$
6	Predicted	Difference	0.94 $(0.19, 1.69)$	-0.05 $(-0.08, -0.02)$	-0.05 $(-0.07, -0.02)$

Notes: Panel (a) of this table modifies panel (a) of Table 4 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 4.

Table A8: Alternative model with naivete about present focus and misperceptions 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=561)	0.03 $(0.03, 0.04)$	0.78 $(0.72, 0.85)$	9.09 $(8.28, 9.91)$	15.44 $(13.53, 17.35)$
2	Above med. (N=560)	0.04 $(0.04, 0.04)$	0.89 $(0.85, 0.93)$	10.01 $(9.11, 10.91)$	14.00 $(12.59, 15.40)$
3	Difference	-0.00 $(-0.01, 0.00)$	-0.10 $(-0.18, -0.03)$	-0.92 $(-2.13, 0.30)$	1.45 $(-0.93, 3.82)$

(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=561)	1.82 $(1.12, 2.52)$	0.34 $(0.32, 0.36)$	0.12 $(0.10, 0.14)$
2	Empirical	Above med. (N=560)	0.53 $(0.05, 1.00)$	0.39 $(0.37, 0.41)$	0.17 $(0.16, 0.19)$
3	Empirical	Difference	1.29 $(0.45, 2.14)$	-0.05 $(-0.08, -0.02)$	-0.05 $(-0.08, -0.03)$
4	Predicted	Below med. (N=561)	2.08 $(1.45, 2.70)$	0.34 $(0.32, 0.36)$	0.11 $(0.09, 0.13)$
5	Predicted	Above med. (N=560)	1.14 $(0.72, 1.56)$	0.39 $(0.37, 0.41)$	0.16 $(0.14, 0.18)$
6	Predicted	Difference	0.94 $(0.19, 1.69)$	-0.05 $(-0.08, -0.02)$	-0.05 $(-0.07, -0.02)$

Notes: Panel (a) of this table modifies panel (a) of Table 3 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 3.

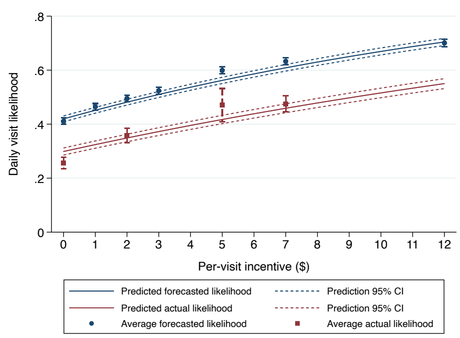
D.4 In-Sample Fit of Structural Models

In this section, we examine the in-sample fit of certain structural models to the forecasted and actual attendance curves. In Appendix Figure A3, we show the in-sample fit of the baseline structural model presented in Table 3, as well as a modification of that model presented in Appendix Table A5 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.

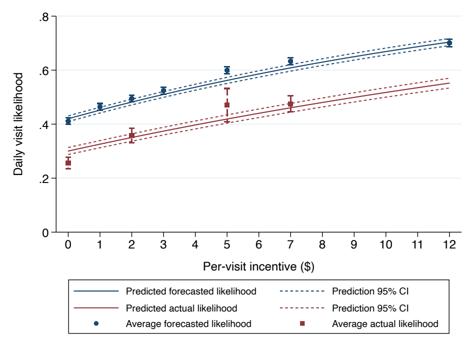
In Appendix Figure A4, we show the in-sample fit of the primary alternative model presented in Table 4, as well as a modification of that model presented in Appendix Table A7, which removes the restriction that perceived present focus may not vary with memory bias. Neither of these models improves on the fit of the attendance data relative to the results in Appendix Figure A3.

Figure A3: In-sample fit of baseline model to forecasted and actual attendance

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



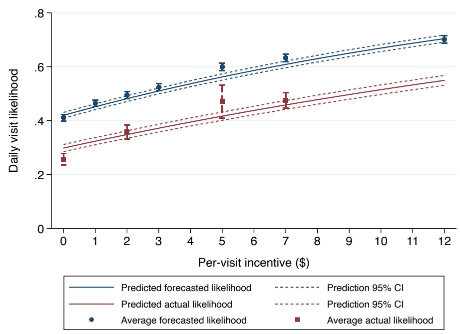
(b) Heterogeneity by quartile of memory bias



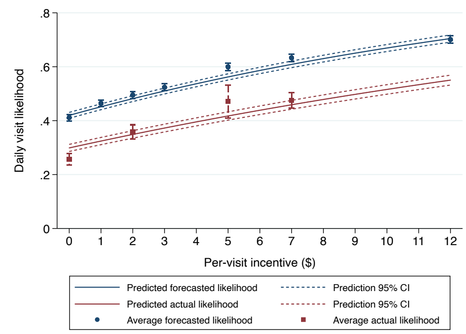
Notes: This figure studies the in-sample fit to participants' forecasted and actual attendance. Inference for the model-predicted attendance likelihoods is conducted using the Delta method. Panel (a) considers the structural model with two types, as in Table 3. Panel (b) considers the structural model with four types, as in Appendix Table A5.

Figure A4: 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 studies the in-sample fit to participants' forecasted and actual attendance. Inference for the model-predicted attendance likelihoods is conducted using the Delta method. 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 4. Panel (b) considers the analogous structural model that allows perceived present focus to vary with memory bias, as in Appendix Table A7.

E Mechanisms

E.1 Definition of Day-of-Week Herfindahl-Hirschman Index

To compute the day-of-week Herfindahl-Hirschman Index (HHI), we first compute the raw day-of-week HHI, which is the sum of squared day-of-week visit shares across fourteen weeks:

$$\mathrm{HHI}_{\mathrm{raw}} = \sum_{d=0}^{6} \left(\frac{c_d}{n}\right)^2, \tag{20}$$

where c_d is the number of visits on day-of-week d (i.e., for Sunday $d = 0$, Monday $d = 1$, ..., and Saturday $d = 6$), and $n = \sum_{d=0}^{6} c_d$ is the total number of visits.

The minimum possible value of the raw day-of-week HHI, as a function of the number of past visits, is achieved with the most even distribution of visits across the days of the week: $\mathrm{HHI}_{\mathrm{min}}(n) = \frac{(7-r_7)\times k_7^2 + r_7 \times (k_7+1)^2}{n^2}$,where $k_7 = \left\lfloor \frac{n}{7} \right\rfloor$ and $r_7 = n \bmod 7$.

The maximum possible value of the raw day-of-week HHI, as a function of the number of past visits, is achieved with the most concentrated distribution of visits across days of the week. A maximum of fourteen visits can be concentrated on a single day of the week since we use fourteen weeks of data: $\text{HHI}_{\max}(n) = \frac{k_{14} \times 14^2 + r_{14}^2}{n^2}$, where $k_{14} = \left\lfloor \frac{n}{14} \right\rfloor$ and $r_{14} = n \bmod 14$.

We rescale the raw day-of-week HHI on a scale from 0 to 1 to obtain our measure of the concentration of visits across days of the week as follows:

$$\text{Day-of-week HHI}(n) = \frac{\text{HHI}_{\text{raw}} - \text{HHI}_{\min}(n)}{\text{HHI}_{\max}(n) - \text{HHI}_{\min}(n)}. \tag{21}$$

E.2 Additional Reduced-Form Results

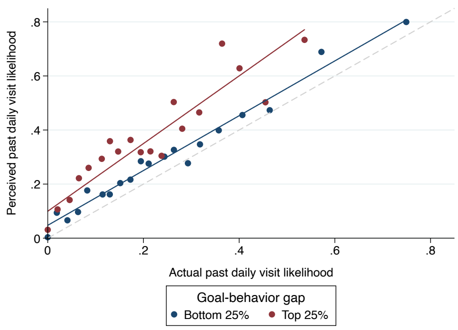
In Appendix Figure A5, we confirm that the same patterns in participants' perceptions of their past daily likelihood of visiting the gym as in Figure 5 hold when comparing participants in the top and bottom quartiles of the goal-behavior gap and day-of-week HHI rather than above and below the median. In both panels, the relationship between the perceived and actual past visit likelihood is linear for all subsamples, and all subsamples overestimate past attendance.

As in panel (a) of Figure 5, in panel (a) of Appendix Figure A5, the group with a more positive goal-behavior gap overestimates past attendance to a greater extent. However, in panel (a) of Appendix Figure A5, the difference in recall patterns between participants in the top and bottom quartiles of the goal-behavior gap is greater than between those above and below the median. This pattern suggests that the difference in selectivity of recall is increasing in the difference in the goal-behavior gap. The estimates reported in Section 5.1 confirm this pattern. This result supports the motivated memory mechanism, wherein those with a greater motive to distort their memories—having a larger gap between their desired and past behavior—engage in more memory distortion.

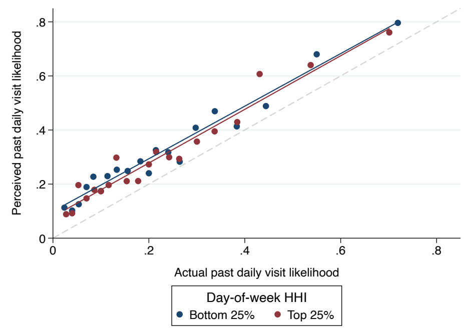
As in panel (b) of Figure 5, in panel (b) of Appendix Figure A5, recall patterns do not vary with the day-of-week HHI, and the estimates reported in Section 5.2 do not show statistically significant differences in the bounds on selectivity of recall for these subsamples at conventional levels. This result does not support the salience bias mechanism, wherein those with easier-to-remember or more salient past visit patterns—having past visits more concentrated on particular days of the week—recall visits and non-visits at more similar rates.

Figure A5: Heterogeneity in perceptions of past daily likelihood of visiting the gym, quartiles

(a) Top- vs. bottom-quartile goal-behavior gap



(b) Top- vs. bottom-quartile day-of-week HHI



Notes: Panel (a) of this figure modifies panel (a) of Figure 5 by splitting the sample into quartiles of the *goal-behavior gap*, and comparing the top and bottom quartiles. Panel (b) of this figure modifies panel (b) of Figure 5 by splitting the sample into quartiles of the *day-of-week Herfindahl-Hirschman Index (HHI)* conditional on the ventile of attendance over the past fourteen weeks, and comparing the top and bottom quartiles. See Section E.1 for a formal definition of the *day-of-week HHI*.

E.3 Additional Structural Estimates

Appendix Table A9 studies differences in parameter estimates for participants with an above- and below-median goal-behavior gap in panel (a) and top- and bottom-quartile goal-behavior gap in panel (b). In both panels, participants with a more positive goal-behavior gap have a lower estimated β.

Appendix Table A10 studies differences in parameter estimates for participants with an above- and below-median day-of-week HHI in panel (a) and top- and bottom-quartile day-of-week HHI in panel (b). In both panels, the estimates of β and $\tilde{\beta}$ are similar across the two subsamples, suggesting no differences in the motivation to distort memories.

Table A9: Model with naivete about present focus, heterogeneity by goal-behavior gap

(a) Above- vs. below-median

		(1)	(2)	(3)	(4)	(5)
	Goal-behavior gap	$\hat{\beta}$	$\hat{\tilde{\beta}}$	$\hat{b}$	$1/\hat{\lambda}$	$\frac{(1-\hat{\tilde{\beta}})}{(1-\hat{\beta})}$
1	Below med. (N=552)	0.67 $(0.62, 0.72)$	0.82 $(0.78, 0.86)$	11.50 $(10.36, 12.65)$	19.55 $(17.47, 21.64)$	0.55 $(0.46, 0.64)$
2	Above med. (N=569)	0.43 $(0.39, 0.47)$	0.82 $(0.76, 0.89)$	8.09 $(7.42, 8.76)$	10.95 $(9.66, 12.24)$	0.31 $(0.21, 0.41)$
3	Difference	0.24 $(0.18, 0.30)$	-0.01 $(-0.08, 0.07)$	3.41 $(2.08, 4.74)$	8.60 $(6.15, 11.05)$	0.24 $(0.11, 0.38)$

(b) Top- vs. bottom-quartile

		(1)	(2)	(3)	(4)	(5)
	Goal-behavior gap	$\hat{\beta}$	$\hat{\tilde{\beta}}$	$\hat{b}$	$1/\hat{\lambda}$	$\frac{(1-\hat{\tilde{\beta}})}{(1-\hat{\beta})}$
1	Bottom quartile (N=281)	0.75 $(0.68, 0.82)$	0.81 $(0.75, 0.87)$	12.68 $(10.76, 14.60)$	22.50 $(18.84, 26.15)$	0.76 $(0.57, 0.95)$
2	Top quartile (N=283)	0.36 $(0.31, 0.41)$	0.80 $(0.72, 0.87)$	7.49 $(6.62, 8.37)$	8.77 $(7.40, 10.13)$	0.32 $(0.21, 0.42)$
3	Difference	0.39 $(0.30, 0.47)$	0.01 $(-0.08, 0.11)$	5.19 $(3.08, 7.30)$	13.73 $(9.82, 17.63)$	0.44 $(0.23, 0.66)$

Notes: Panel (a) of this table modifies panel (a) of Table 3 by defining the two subsamples by splitting at the median value of the *goal-behavior gap*. Panel (b) of this table modifies panel (a) of Table 3 by defining the two subsamples by splitting by quartile of the *goal-behavior gap* and comparing the top and bottom quartiles.

Table A10: Model with naivete about present focus, heterogeneity by day-of-week HHI

(a) Above- vs. below-median

	Day-of-week HHI	(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=395)	0.59 $(0.54, 0.65)$	0.86 $(0.81, 0.91)$	9.37 $(8.39, 10.35)$	12.95 $(11.42, 14.47)$	0.34 $(0.24, 0.44)$
2	Above med. (N=483)	0.56 $(0.50, 0.62)$	0.81 $(0.74, 0.88)$	10.32 $(9.30, 11.34)$	16.34 $(14.09, 18.59)$	0.44 $(0.31, 0.56)$
3	Difference	0.03 $(-0.05, 0.11)$	0.05 $(-0.04, 0.14)$	-0.95 $(-2.37, 0.47)$	-3.39 $(-6.11, -0.67)$	-0.10 $(-0.26, 0.06)$

(b) Top- vs. bottom-quartile

	Day-of-week HHI	(1) $\hat{\beta}$	(2) $\hat{\tilde{\beta}}$	(3) $\hat{b}$	(4) $1/\hat{\lambda}$	(5) $\frac{(1-\hat{\tilde{\beta}})}{(1-\hat{\beta})}$
1	Bottom- quartile (N=279)	0.53 $(0.47, 0.59)$	0.82 $(0.76, 0.88)$	8.89 $(7.77, 10.00)$	12.24 $(10.49, 13.99)$	0.39 $(0.29, 0.49)$
2	Top quartile (N=236)	0.56 $(0.46, 0.67)$	0.76 $(0.63, 0.90)$	10.70 $(9.22, 12.19)$	17.05 $(13.09, 21.01)$	0.54 $(0.34, 0.73)$
3	Difference	-0.03 $(-0.15, 0.09)$	0.05 $(-0.09, 0.20)$	-1.82 $(-3.68, 0.04)$	-4.81 $(-9.14, -0.48)$	-0.15 $(-0.37, 0.07)$

Notes: Panel (a) of this table modifies panel (a) of Table 3 by defining the two subsamples by splitting at the median value of the *day-of-week Herfindahl-Hirschman Index (HHI)*, conditional on the ventile of attendance over the past fourteen weeks. Panel (b) of this table modifies panel (a) of Table 3 by defining the two subsamples by splitting by quartile of the *day-of-week HHI*, conditional on the ventile of attendance over the past fourteen weeks, and comparing the top and bottom quartiles. See Section E.1 for a formal definition of the *day-of-week HHI.*