Appendix for:

Designing Incentives for Impatient People: An RCT Promoting Exercise to Manage Diabetes

Shilpa Aggarwal
Indian School of Business

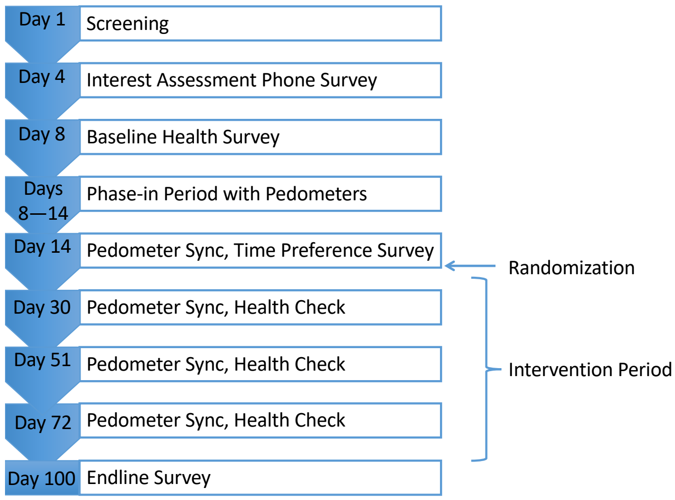
Rebecca Dizon-Ross
University of Chicago

Ariel Zucker
UC Santa Cruz

Table of Contents

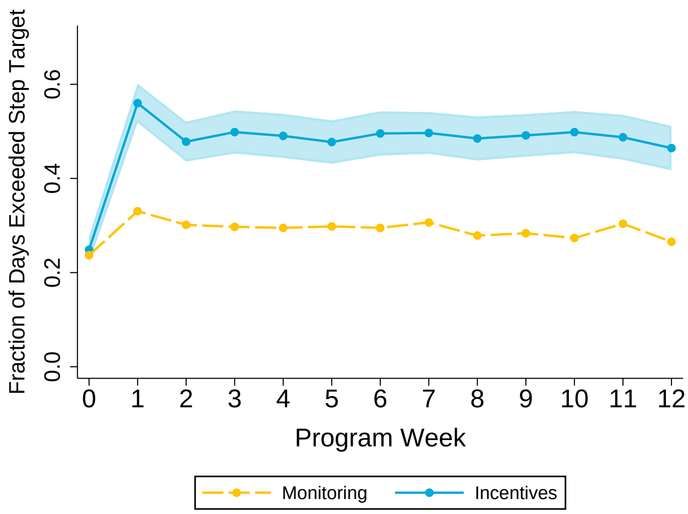
A Appendix Tables and Figures
B Theoretical Predictions Appendix
C Measuring the Effort and Payment Discount Factors
D Distributional Impacts of Thresholds
E Predicting Impatience with Policy Variables
F Additional Tables and Figures
G Misreporting Steps, Confusion, and Suspensions
H Theoretical Predictions: Additional Proofs
I CTB Time Preference Measurement
J Monitoring Treatment Impacts on Walking

A Additional Tables and Figures

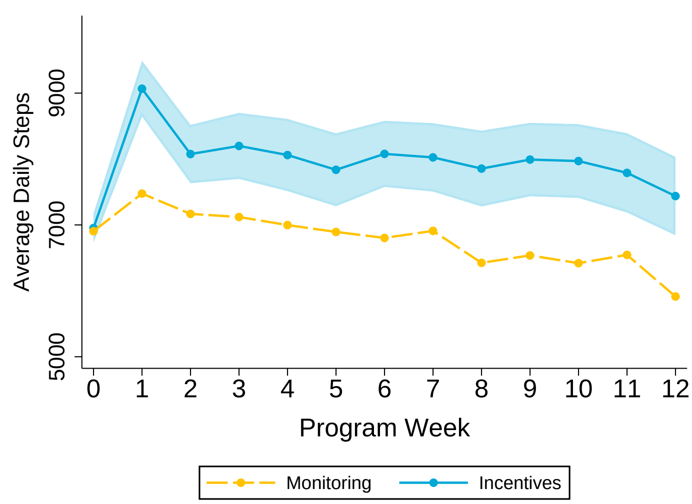


Appendix Figure A.1: Experimental Timeline for Sample Participant

Notes: This figure shows an experimental timeline for a participant. Visits were scheduled according to the participants' availability. We introduced variation in the timing of incentive delivery by delaying the start of the intervention period by one day for randomly selected participants. The intervention period was exactly 12 weeks for all participants.



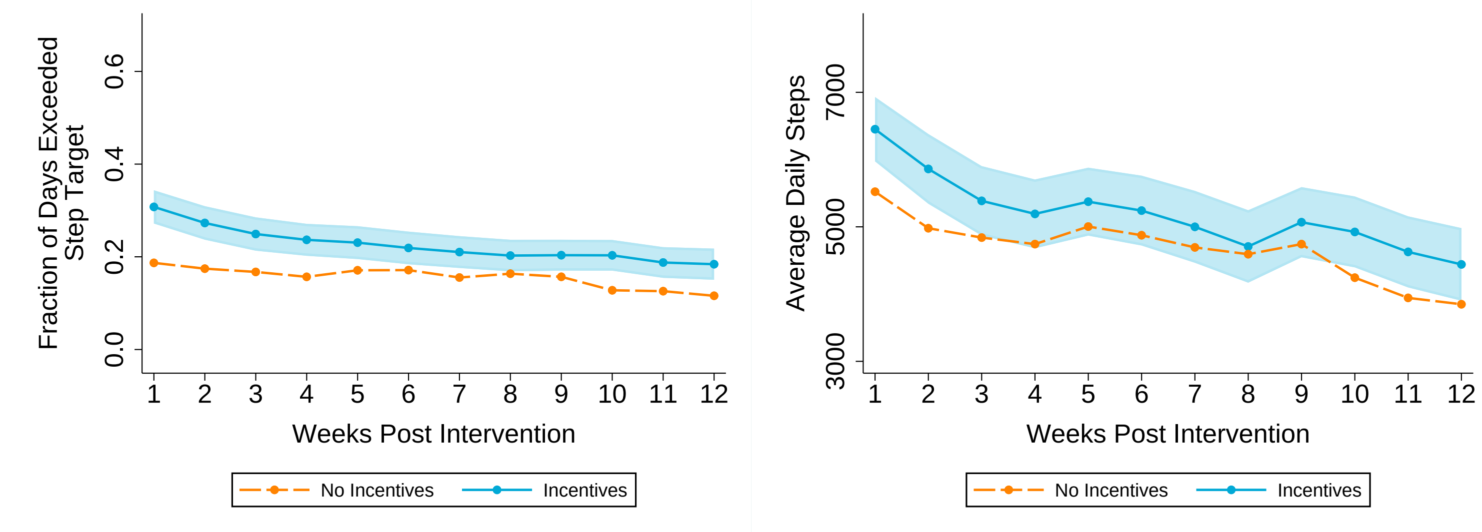
(a) Step-Target Compliance



(b) Daily Steps Walked

Appendix Figure A.2: Incentive Effects are Steady through the 12-Week Program

Notes: Panel (a) shows the average probability of exceeding the step target and Panel (b) shows the average daily steps walked, both during the intervention period. Week 0 is the phase-in period (before randomization). The shaded areas represent a collection of confidence intervals from tests of equality within each weekly period between the incentive and monitoring groups from regressions with the same controls as in Table 2. Data are at the individual-week level. Both graphs are unconditional on wearing the pedometer. Graphs look similar when condition on wearing the pedometer except that, in both groups, there is less downward trend over time.

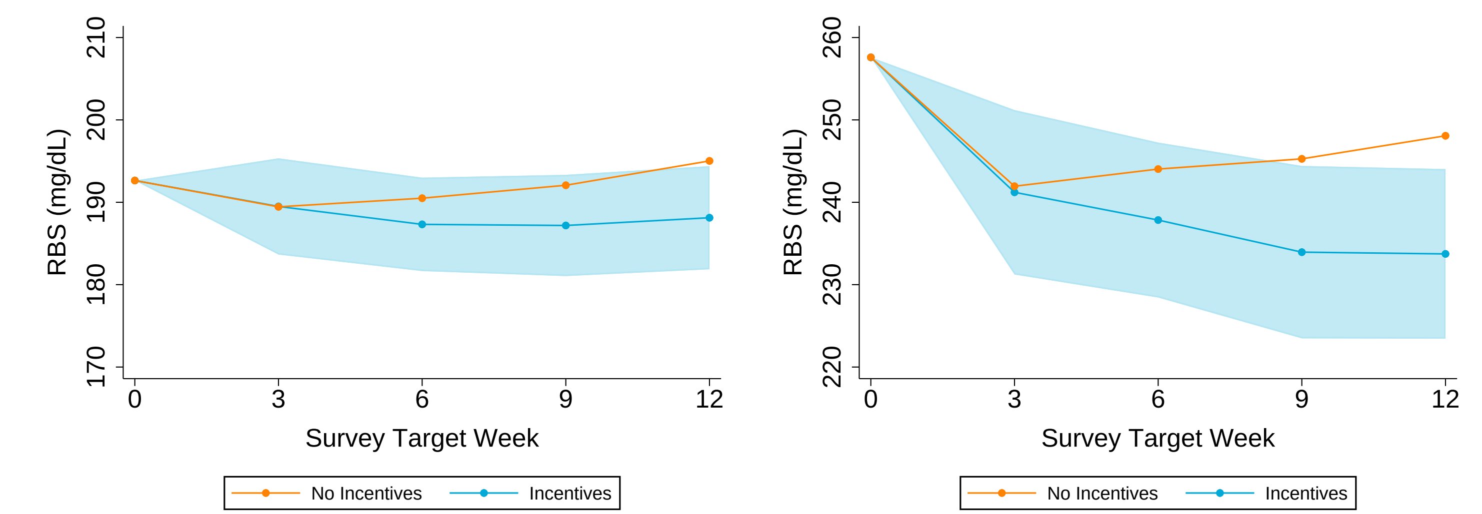


(a) Step-Target Compliance

(b) Daily Steps Walked

Appendix Figure A.3: Incentive Effects Persist After the 12-Week Program

Notes: Panel (a) shows the average probability of exceeding the step target and Panel (b) shows the average daily steps walked, both in the 12 weeks following the intervention. "No incentives" represents the pooled monitoring and control groups; the Panels look very similar when we compare with the control group only. The shaded areas represent a collection of confidence intervals from tests of equality within each weekly period between the incentive and no incentive groups from regressions with the same controls as in Table 2. All graphs are unconditional on wearing the pedometer. Data are at the individual-week level. Graphs look similar when condition on wearing the pedometer except that, in both groups, there is less downward trend over time.



(a) Full Sample

(b) Above-Median Blood Sugar Sample

Appendix Figure A.4: Blood Sugar Treatment Effects Grow Over Time

Notes: Figures show how the impact of incentives on random blood sugar (RBS) evolves over time by presenting the treatment effect of incentives on RBS separately for each time RBS was measured. Panel A shows the full sample and Panel B restricts to those with above-median baseline values of the blood sugar index. Survey week 0 was the baseline survey measurement; survey week 12 was the endline survey measurement; and survey weeks 3, 6, and 9 were the measurements at the pedometer sync visits held every three weeks during the intervention period. Observations are at the individual level. The "No incentives" group represents the pooled monitoring and control groups. As in our other graphs of trends over time, we pool the two comparison groups (control and monitoring) for power. Results are similar but slightly less precise if we compare incentives with control alone. For each survey, we regress random blood sugar on the incentives dummy and control for the same controls as in the random blood sugar specification in Table 4. The shaded areas represent a collection of 95% confidence intervals from those regressions. The p-values for the significance of the increase over time are .05 and .02 for the Panels A and B, respectively.

Appendix Table A.1: Measures of Effort Impatience Correlate with Baseline Exercise, Health, and Behavior

		Correlation with					
		Baseline exercise		Baseline indices			
	Mean	Daily steps	Daily exercise (min)	Negative health risk index	Negative vices index	Healthy diet index	# Individuals
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
A. Impatience index measures							
Impatience index	0.092	-0.080***	-0.070***	-0.016	-0.052	-0.181***	1,740
1. I'm always saying: I'll do it tomorrow	2.217	-0.059	-0.101***	-0.010	-0.031	-0.147***	1,740
2. I usually accomplish all the things I plan to do in a day	0.643	-0.054	-0.052	-0.012	-0.043*	-0.149***	1,740
3. I postpone starting on things I dislike to do	3.967	-0.042*	0.004	0.004	-0.052	0.050	1,740
4. I'm on time for appointments	0.468	-0.054	0.006	-0.021	0.008	-0.097***	1,740
5. I often start things at the last minute and find it difficult to complete them on time	2.506	-0.039	-0.069***	-0.009	-0.043*	-0.207***	1,740
B. Predicted index measures							
Predicted index	-0.052	0.000	-0.036	-0.064***	0.021	0.004	3,192
1. In the past week, how many times have you found yourself exercising less than you had planned?	0.526	0.015	-0.006	-0.064***	0.007	0.026	3,192
2. In the past 24 hours, how many times have you found yourself eating foods you had planned to avoid?	0.208	-0.001	0.050***	-0.058***	0.015	0.034*	3,192
3. Do you worry that if you kept a higher balance on your phone, you would spend more on talk time?	0.131	-0.027	-0.062***	-0.018	0.031*	-0.038	3,192
C. Simple CTB							
Simple CTB index	0.532	-0.120***	-0.028	-0.003	-0.018	-0.020	3,190
1. Chose 30 minutes today and 60 minutes in one week	0.508	-0.115***	-0.018	-0.006	-0.020	-0.021	3,190
2. Chose 20 minutes today and 60 minutes in one week	0.555	-0.120***	-0.037	0.000	-0.015	-0.019	3,190
D. Demand for commitment							
Chose commitment index	0.485	0.045	-0.005	-0.027	0.011	0.015	2,871
1. Chose 4-day threshold	0.511	0.027	-0.010	-0.021	0.017	0.017	2,881
2. Chose 5-day threshold	0.461	0.057***	-0.003	-0.030	0.004	0.015	2,889

Notes: This table displays the correlations between impatience measures and baseline behavior and health. Each coefficient represents results from a separate regression. We normalize variables such that a higher impatience measure value corresponds to greater impatience, and a higher health or behavior measure value corresponds to healthier behavior. Panel A shows the impatience index and its five components. Panel B shows the predicted index and its three components. Panel C shows the Simple CTB index and its two components. The Simple CTB index is the average of preferences for option (A) in the following two scenarios: 1. In exchange for 500 INR in 8 days, walk (A) 30 minutes today and 60 minutes in one week, or (B) 60 minutes today and 20 minutes in one week; 2. In exchange for 500 INR in 8 days, walk (A) 20 minutes today and 60 minutes in one week, or (B) 60 minutes today and 20 minutes in one week. Panel D shows the chose commitment index and its two components. The chose commitment index is defined as the average of preferring the Time Bundled contract in each of the following questions: "Which program would you prefer: The Weekly Recharge Program with a condition of 5 days, or the Basic Weekly Recharge Program with no condition?" and "Which program would you prefer: the Program with a minimum weekly condition of 4 days, or the Basic Weekly Program with no condition?"
Daily steps are from the phase-in period pedometer data. Daily exercise is self-reported. The health index is as in Table 4. The vices index includes an individual's daily cigarette, alcohol, and areca nut usage. The healthy diet index includes an individual's daily number of wheat, vegetable, and rice; spoonfuls of sugar; fruit, junk food, and sweets intake; and whether one avoids unhealthy foods. Data are at the individual level and include the full sample. Significance levels: * 10%, ** 5%, *** 1%.

Appendix Table A.2: Missing Pedometer Data During the Intervention Period

Dep. variable:	No Steps data	**Reason** no steps data		**Reason** no data from Fitbit			
		Did not wear Fitbit	No data from Fitbit	Lost data entire period	Immediate withdrawal	Mid-intervention withdrawal	Other reasons
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Incentives	-0.0140	-0.0287**	0.0155	-0.00203	0.00571	0.0166**	-0.00471
	[0.0174]	[0.0142]	[0.0124]	[0.00511]	[0.00731]	[0.00694]	[0.00594]
Monitoring mean	.19	.15	.047	.0049	.0099	.012	.02
# Individuals	2,607	2,559	2,607	2,607	2,607	2,607	2,607
# Observations	218,988	205,732	218,988	218,988	218,988	218,988	218,988

Notes: Each observation is an individual × day. The sample includes Incentives and Monitoring. Missing data have two sources: pedometer non-wearing (i.e., steps = 0) (column 2) or failure to retrieve pedometer data (column 3). Columns 2 + 3 = column 1 except column 2 conditions on there not being missing data (for consistency with our main step analyses, results are similar without this restriction), while columns 1 and 3 do not. Columns 4–7 summarize the reasons pedometer data in column 3 were missing. Controls are the same as in Table 2. Significance levels: * 10%, ** 5%, *** 1%.

Appendix Table A.3: Threshold Treatments Increase Cost-Effectiveness Relative to Base Case, With Similar Increases Among Those Who Are More and Less Impatient

		Sample defined by impatience indices			
	Full sample	Below-median (actual)	Above-median (actual)	Below-median (predicted)	Above-median (predicted)
Treatment group	(1)	(2)	(3)	(4)	(5)
Base Case	0.050	0.050	0.050	0.050	0.050
Threshold	0.056	0.056	0.057	0.057	0.056
4-Day Threshold	0.055	0.055	0.056	0.056	0.055
5-Day Threshold	0.059	0.059	0.059	0.059	0.058

Notes: The table displays the cost-effectiveness of different treatment groups (in rows) and different samples (in columns). Cost-effectiveness equals average compliance divided by the average payment per day, in units of days complied per INR. The sample includes Base Case and Threshold (Threshold pools 4- and 5-day Thresholds). We test for differences in cost-effectiveness using a mathematically equivalent test for differences in the fraction of days complied on which participants earned payment, shown in column 4 of Table 2 and Figure 2b.

Appendix Table A.4: Threshold Heterogeneity Results are Similar Among Naive Individuals

Dependent variable:	Exceeded step target (×100)				
	(1)	(2)	(3)	(4)	(5)
Impatience × Threshold	2.99	3.94	2.76	6.8	8.07**
	[-2.56, 8.55]	[-7.09,14.97]	[-1.31, 6.76]	[-1.31, 14.88]	[0.14, 16.01]
Threshold	-5.29*	-6.79*	-3.97*	-6.5**	-8.50***
	[-10.89, 0.32]	[-14.72,1.14]	[-7.98, 0.07]	[-11.14, -1.37]	[-14.19, -2.80]
Impatience	-3.11*	-6.70	-1.56	-5.15*	-4.06
	[-6.69, 0.47]	[-14.84,1.44]	[-4.51, 1.52]	[-11.18, 0.17]	[-9.73, 1.62]
Impatience measure:	Impatience index	Above-median impatience index	Predicted impatience index	Above-median predicted index	Simple CTB
Sample:	Late	Late	Full	Full	Full
Base Case mean	51.7	51.7	50.6	50.6	50.6
# Individuals	496	496	977	977	977
# Observations	39,562	39,562	78,096	78,096	78,096

Notes: This table is the same as Table 3 but limited to the subsample of participants who did not demand commitment (that is they did not prefer both the 4-day and 5-day threshold contract relative to the base case contract). Controls are the same as in Table 2. 95% confidence intervals are shown in brackets. For columns 1, 2, and 5, confidence intervals are based on standard errors clustered at the individual level. For columns 3 and 4, confidence intervals are constructed using bootstrap, with bootstrap draws clustered at the individual level; see the notes to Table 3 for a detailed description of the bootstrap procedure. Data are at the individual × day level. The sample includes Base Case and Threshold. Significance levels: * 10%, ** 5%, *** 1%.

Appendix Table A.5: Thresholds Are Similarly Cost-Effective Among Those with Higher Impatience

Dependent variable:	Earned payment when exceeded target					
	(1)	(2)	(3)	(4)	(5)	(6)
Impatience × Threshold	-0.00625	-0.0109	0.00386	0.0115	0.0202*	-6.96e-05
	[-0.02, 0.01]	[-0.04,0.02]	[-0.01, 0.01]	[-0.01, 0.03]	[-0.00, 0.04]	[-0.02, 0.02]
Threshold	-0.114***	-0.109***	-0.115***	-0.119***	-0.127***	-0.115***
	[-0.13, -0.10]	[-0.13,-0.09]	[-0.13, -0.10]	[-0.13, -0.10]	[-0.14, -0.11]	[-0.13, -0.10]
Impatience	0.00208	0.00544	-0.000834	-0.00275	-0.00233	0.00483*
	[-0.00, 0.01]	[-0.00,0.01]	[-0.00, 0.00]	[-0.01, 0.00]	[-0.01, 0.00]	[-0.00, 0.01]
Impatience measure:	Impatience index	Above-median impatience index	Predicted impatience index	Above-median predicted index	Chose commitment	Simple CTB
Sample:	Late	Late	Full	Full	Full	Full
Base Case mean	1	1	1	1	1	1
# Individuals	1,007	1,007	1,846	1,846	1,681	1,844
# Observations	42,830	42,830	79,248	79,248	71,525	79,150

Notes: This table shows heterogeneity in the impact of Threshold on the fraction of days on which participants received payment, conditional on meeting the step target, by different measures of impatience. A higher level of this outcome indicates lower cost-effectiveness among treatment groups that received the same payment per day (all groups except Small Payment). The impatience measure changes across columns. Controls are the same as in Table 2. 95% confidence intervals are shown in brackets. For columns 1–2 and 5–6, confidence intervals are based on standard errors clustered at the individual level. For columns 3 and 4, confidence intervals are constructed using bootstrap, with bootstrap draws clustered at the individual level; see the notes to Table 3 for a detailed description of the bootstrap procedure. Data are at the individual × day level. The sample includes Base Case and Threshold. Significance levels: * 10%, ** 5%, *** 1%.

Appendix Table A.6: The Effects of Incentives Persist After the Intervention Ends

	Post-intervention		
Dependent variable:	Exceeded step target	Daily steps	Daily steps (if > 0)
	(1)	(2)	(3)
Incentives	0.071***	537.2**	648.3***
	[0.01]	[220.90]	[195.82]
No incentives mean	0.156	4,674	6,773
# Individuals	1,122	1,122	1,122
# Observations	91,756	91,756	62,858

Note: This table shows the average treatment effect of Incentives relative to Control and Monitoring (pooled) during the "post-intervention period" (i.e., the 12 weeks after the intervention ended). Each observation is a person-day. Columns 1 and 2 include all days, and column 3 only includes days where the participant wore the pedometer (i.e., had step count > 0). Controls are the same as in Table 2. The number of individuals differs from the total number recruited for the post-intervention period because roughly 11% of participants withdrew immediately. The likelihood of immediate withdrawal is not significantly different between the incentive and comparison groups. Standard errors, in brackets, are clustered at the individual level. Significance levels: * 10%, ** 5%, *** 1%.

B Theoretical Predictions Appendix

We begin by presenting the formal model setup and assumptions in Section B.1. In Sections B.2 and B.3, we describe behavior under time-separable linear contracts and time-bundled contracts, respectively. Sections B.4 and B.5 present the formal mathematical results (labeled propositions) underlying our two key testable predictions regarding behavior in time-bundled relative to linear contracts. Section B.6 briefly analyzes the effects of payment frequency. Finally, Section B.7 considers the implications of adding a discounted health benefit to the model.

B.1 Full Model SetUp

Each day, an individual chooses whether to complete a binary action. Define w_t as an indicator for whether the individual *complies* (i.e., completes the action) on day t.

Incentive Contract Structure and Compliance We consider a principal who designs contracts to incentivize individuals for compliance over a sequence of T days. We call this sequence of days the payment period and index its days $t = 1, ..., T$.

Let m_t be the payment made by the principal to the individual on day t. Within each payment period, payments are delivered on day T only and depend on the individual's compliance decisions from day 1 through T of the payment period.

Define *compliance*, the expected fraction of days on which the individual complies, as $C = \frac{1}{T}\mathbb{E}[\sum_{t=1}^{T} w_t]$ and the expected per-day *payment* as $P = \frac{1}{T}\mathbb{E}[m_T]$. Define *cost-effectiveness* as compliance divided by expected per-day payment, C/P.

The Principal's Objective: Effectiveness We assume that the principal aims to maximize *effectiveness*, defined as the expected per-day benefit to the principal from compliance less the expected payment to agents. Maximizing effectiveness is analogous to the standard contract theory approach of maximizing output net of wage payments.[46] For the definition to be operable, we need to take a stand on the expected benefit function. We assume the expected benefit is linear in compliance, equal to λC for some $\lambda > 0$. This simplifying assumption is reasonable in our empirical setting since the estimated marginal health benefit of days of exercise is approximately linear (Warburton et al., 2006; Banach et al., 2023). With linear benefits, effectiveness becomes $\lambda C - P$.

We want to compare the effectiveness of different contracts even when we do not know λ. Rewriting effectiveness as $C\left(\lambda - \frac{1}{(C/P)}\right)$ shows that (assuming effectiveness is positive) one contract is more *effective* than another if it has strictly larger compliance and weakly larger cost-effectiveness, or weakly larger compliance and strictly larger cost-effectiveness.

Agent Utility Agent utility depends on the payments they receive from the principal and the cost of the effort of complying (if they comply), as captured by the following reduced-form utility function:

$$U = \mathbb{E}\left[\sum_{t=0}^{\infty} d^{(t)} m_t - \delta^{(t)} w_t e_t\right], \tag{8}$$

[46] This objective is often used in practice. For example, health policymakers and insurance companies often want to maximize the total health benefits of a program relative to its costs.

where e_t is the effort cost of complying on day t, $\delta^{(t)}$ is the discount factor over effort t days in the future, and $d^{(t)}$ is the discount factor over payments received t days in the future (for notational simplicity, we denote $\delta^{(1)}$ as δ and $d^{(1)}$ as d). Both $\delta^{(t)} \leq 1$ and $d^{(t)} \leq 1$, with $\delta^{(0)} = d^{(0)} = 1$. Neither $\delta^{(t)}$ nor $d^{(t)}$ are necessarily exponential functions of t; we assume only that they are weakly decreasing in t. We assume utility is linear in payments, which is likely a good approximation in our setting, as payments are small relative to overall consumption.

Importantly, this reduced-form utility function differentiates the discount factor over payments, $d^{(t)}$, from the discount factor over effort, $\delta^{(t)}$. The specification is consistent with a standard model of utility with a single structural discount factor over consumption and effort (e.g., Augenblick et al., 2015). In that case, $\delta^{(t)}$ is the structural discount factor, while $d^{(t)}$ depends on the availability of borrowing and savings. For example, in perfect credit markets, individuals should discount future payments at the interest rate r, and so $d^{(t)} = \left(\frac{1}{1+r}\right)^t$.

Time-Inconsistency and Sophistication Individuals will have time-inconsistent preferences either if $\delta^{(t)}$ or $d^{(t)}$ are non-exponential functions of t, or if $d^{(t)} \neq \delta^{(t)}$. Among time-inconsistent agents, we follow O'Donoghue and Rabin (1999a) in distinguishing sophisticates, who are aware of their discount factors (over both effort and money), from naifs, who "believe [their] future selves' preferences will be identical to [their] current self's." Thus, letting $w_{t,j}$ be the agent's prediction on day j about her compliance on day $t > j$, sophisticates accurately predict how their future selves will behave ($w_{t,j} = w_t$), while naifs may not.[47]

Effort Costs Let e_t be identically (but not necessarily independently) distributed across days, with the marginal distribution of e_t given by continuous cumulative distribution function (CDF) $F(\cdot)$. Individuals know the joint distribution of effort costs in advance but do not observe the realization of e_t until day t. e_t can be negative, as agents may comply without payment.

Agent Problem Given the notation and assumptions above, we can express the agent's problem as follows. On day t, the agent chooses compliance, w_t, to maximize expected discounted payments net of effort costs:

$$\max_{w_t \in \{0,1\}} \mathbb{E}\left[d^{(T-t)} m_T - \sum_{j=t+1}^{T} \delta^{(j-t)} w_{j,t} e_j \middle| e_1, .., e_t, w_1, .., w_t \right] - w_t e_t, \tag{9}$$

where the expectation over future discounted payment and future discounted effort depends on the history of effort costs $(e_1, .., e_t)$ and compliance decisions $(w_1, .., w_t)$ through time t, and where $w_{j,t}$ represents the agent's prediction on day t about her compliance on day j.

Denoting $\mathbb{E}\left[d^{(T-t)} m_T - \sum_{j=t+1}^{T} \delta^{(j-t)} w_{j,t} e_j \middle| e_1, .., e_t, w_1, .., w_t \right]$ as $V_t(w_t)$, the agent will thus choose to set $w_t = 1$ (i.e., comply on day t) if the following holds:

$$V_t(0) < V_t(1) - e_t \tag{10}$$

That is, on day t, the agent complies if the continuation value of complying net of the effort cost

[47] With domain-specific discounting, naivete can stem from misunderstanding how the future self will either (a) value current effort relative to money, or (b) discount effort or money further in the future.

is greater than the continuation value of not complying.

B.2 Time-Separable Linear Contracts (the Base Case)

We now solve for compliance and effectiveness under the base case contract. The contract is linear, paying m per day of compliance:

$$m_T^{\text{Base Case}} = m \sum_{t=1}^{T} w_t. \tag{11}$$

Agents comply on day t if the discounted payment outweighs the effort cost:

$$e_t < d^{(T-t)} m. \tag{12}$$

Expected payment per period P is then mC. As a result, effectiveness is $(\lambda - m)C$. Cost-effectiveness, C/P, is simply $\frac{1}{m}$ for any linear contract with positive compliance.

Observation 1. In a time-separable contract, holding all else constant, neither compliance, cost-effectiveness, nor effectiveness depend on $\delta^{(t)}$.[48]

B.3 Time-Bundled Contracts

Time-bundled contracts contain at least one period in which the payment for future compliance is increasing in current compliance. We focus on a *threshold* time-bundled contract, where there is a minimum threshold level of compliance K.[49] In a threshold contract, if the participant complies on fewer than K days in the payment period, no incentive is received. If they comply on at least K days, payment is a linear function of the number of days of compliance, with a rate of m' per day. Total payment in the threshold contract is thus:

$$m_T^{\text{Threshold}} = \begin{cases} m' \sum_{t=1}^{T} w_t & \text{if } (\sum_{t=1}^{T} w_t \geq K) \\ 0 & \text{otherwise.} \end{cases} \tag{13}$$

In the following two subsections, we theoretically examine the effect, relative to the Base Case, of adding a threshold while maintaining the same payment period length. Our results rest on the fact that, unlike in the Base Case, compliance, cost-effectiveness, and effectiveness in threshold contracts depends critically upon the discount factor over effort.

B.4 Thresholds versus Linear: Comparative Statics in the Effort Discount Rate

In this section, we present a series of propositions that provide the theoretical underpinning for Prediction 1 from Section 2.3. The prediction is that the lower is $\delta^{(t)}$, the higher are compliance and effectiveness in a threshold relative to time-separable contract. We have already seen that in time-separable contracts, compliance and effectiveness are flat in $\delta^{(t)}$ (Observation 1). The propositions demonstrate that in contrast, both compliance and effectiveness in time-bundled threshold contracts tend to decrease in $\delta^{(t)}$.

Specifically, Proposition 1 examines threshold contracts with $K = T$ (i.e., where one must comply on all days to receive payment). It shows that, for all T, *regardless of the effort cost distribution,* compliance is weakly decreasing in $\delta^{(t)}$.

[48]In linear contracts, compliance is $\frac{1}{T}\mathbb{E}\left[\sum_{t=1}^{T} w_t\right] = \frac{1}{T}\sum_{t=1}^{T} F(d^{(T-t)}m)$, which is not directly related to $\delta^{(t)}$.

[49]Our predictions hold for other types of time-bundled contracts in many circumstances.

To gain tractability to examine threshold effectiveness and threshold contracts with $K < T$, we then make assumptions about the effort cost distribution. Proposition 2 examines effectiveness when $K = T = 2$ and shows that, under relatively general conditions, effectiveness in the threshold contract is weakly decreasing in δ. Proposition 3 shows that, if costs are perfectly positively correlated over time, both compliance and effectiveness under the threshold are decreasing in $\delta^{(t)}$ for any $K \leq T$ and any T. Finally, Proposition 4 examines a simplified model where costs are binary and known from day 1, $K = 2$ and $T = 3$. We show that compliance and effectiveness are higher when $\delta^{(t)}$ is lower.

The propositions together suggest that Prediction 1 holds in many empirically-relevant conditions, including when either (a) K is high relative to T,[50] or (b) costs are positively correlated across periods. Both (a) and (b) hold in our empirical setting: our experiment uses relatively high levels of K relative to T, and costs are positively correlated across days.

Proposition 1 ($T = K$, Threshold Compliance and Impatience Over Effort)**.** *Let $T > 1$. Fix all parameters other than $\delta^{(t)}$. Take any threshold contract with threshold level $K = T$; denote the threshold payment M. Compliance in the threshold contract is weakly decreasing in $\delta^{(t)}$ for all $t \leq T - 1$.*

Proof. We provide the proof here for $T = 2$. The proof for $T > 2$ is in Appendix H.1.

Recall that the condition for complying on day 1 is to comply if $e_1 < V_1(1) - V_1(0)$ (equation (10)). Let $w_{t,j}$ be the agent's prediction on day j about her compliance on day $t > j$. With the threshold contract, we have that:

$$V_1(1) - V_1(0) = \mathbb{E}\left[(dM - \delta e_2)w_{2,1}|e_1, w_1 = 1\right] - \mathbb{E}\left[-\delta e_2 w_{2,1}|e_1, w_1 = 0\right] \quad (14)$$

We examine this expression separately for sophisticates and naifs.

For sophisticates, who accurately predict their own future behavior, $w_{2,1}|^{w_1=1} = \mathbb{1}\{e_2 < M\}$ and $w_{2,1}|^{w_1=0} = \mathbb{1}\{e_2 < 0\}$. Thus:

$$\begin{aligned} V_1(1) - V_1(0) &= \mathbb{E}\left[(dM - \delta e_2)w_{2,1}|e_1, w_1 = 1\right] - \mathbb{E}\left[-\delta e_2 w_{2,1}|e_1, w_1 = 0\right] \\ &= \mathbb{E}\left[(dM - \delta e_2)\mathbb{1}\{e_2 < M\} + \delta e_2 \mathbb{1}\{e_2 < 0\}|e_1\right] \end{aligned} \quad (15)$$

We show that this is weakly decreasing in δ by showing that the argument $(dM - \delta e_2)\mathbb{1}\{e_2 < M\} + \delta e_2 \mathbb{1}\{e_2 < 0\}$ is weakly decreasing in δ for all values of e_2. There are two cases:

1. $e_2 > 0$: In this case, $(dM - \delta e_2)\mathbb{1}\{e_2 < M\} + \delta e_2 \mathbb{1}\{e_2 < 0\} = (dM - \delta e_2)\mathbb{1}\{e_2 < M\}$, which is weakly decreasing in δ.
2. $e_2 \leq 0$: In this case, $(dM - \delta e_2)\mathbb{1}\{e_2 < M\} + \delta e_2 \mathbb{1}\{e_2 < 0\} = (dM - \delta e_2) + \delta e_2 = dM$, which is invariant to δ.

Since equation (15) is weakly decreasing in δ, day 1 compliance is decreasing in δ. The same is true for day 2 compliance, since $w_2 = 1$ if both $w_1 = 1$ and $e_2 < M$ (or if $e_2 < 0$), and

[50] Thresholds where K/T is very low may not always be better for impatient naifs than patient people because they include more days where current and future effort are substitutes, which can cause naifs to procrastinate.

w_1 is weakly decreasing in δ. Thus, compliance in the threshold contract is decreasing in δ for sophisticates.

We now turn to naifs. For naifs, who think their day 2 selves will share their day 1 preferences, $w_{2,1}|^{w_1=1} = \mathbb{1}\{\delta e_2 < dM\}$ and $w_{2,1}|^{w_1=0} = \mathbb{1}\{\delta e_2 < 0\}$. Thus:

$$\begin{aligned} V_1(1) - V_1(0) &= \mathbb{E}\left[(dM - \delta e_2)w_{2,1}|e_1, w_1 = 1\right] - \mathbb{E}\left[-\delta e_2 w_{2,1}|e_1, w_1 = 0\right] \\ &= \mathbb{E}\left[(dM - \delta e_2)\mathbb{1}\{\delta e_2 < dM\} + \delta e_2 \mathbb{1}\{\delta e_2 < 0\}|e_1\right] \\ &= \mathbb{E}\left[\max\{dM - \delta e_2, 0\} + \delta e_2 \mathbb{1}\{e_2 < 0\}|e_1\right] \end{aligned} \tag{16}$$

Again, we show that this is decreasing in δ by showing that the argument, $\max\{dM - \delta e_2, 0\} + \delta e_2 \mathbb{1}\{e_2 < 0\}$, is weakly decreasing in δ for all values of e_2. There are two cases:

1. $e_2 > 0$: In this case, $\max\{dM - \delta e_2, 0\} + \delta e_2 \mathbb{1}\{e_2 < 0\} = \max\{dM - \delta e_2, 0\}$, which is weakly decreasing in δ.
2. $e_2 \leq 0$: In this case, for $u = -e_2 \geq 0$, we have $\max\{dM - \delta e_2, 0\} + \delta e_2 \mathbb{1}\{e_2 < 0\} = \max\{dM + \delta u, 0\} - \delta u = (dM + \delta u) - \delta u = dM$ which is invariant to δ.

Since equation (16) is weakly decreasing in δ, day 1 compliance (and hence day 2 and total compliance) are also decreasing in δ for naifs. □

We now examine effectiveness when $T = K$. We examine the case where $T = 2$ and, to gain tractability, make a reasonable assumption on the cost function, assuming that e_2 is weakly increasing in e_1, in a first order stochastic dominance sense.[51] This assumption flexibly accommodates the range from IID to perfect positive correlation, just ruling out negative correlation. Under this assumption, we show that effectiveness is weakly decreasing in δ as long as there is not "too much" inframarginal behavior. When there is too much inframarginal behavior, not only will the effectiveness prediction not hold but incentives cease to be a cost-effective approach.

Proposition 2 ($T = 2, K = 2$, Threshold Effectiveness and Impatience Over Effort)**.** *Let $T = 2$. Let e_2 be weakly increasing in e_1, in a first order stochastic dominance sense. Fix all parameters other than $\delta^{(t)}$. Take any threshold contract with threshold level $K = 2$; denote the threshold payment M. As long as there is not "too much" inframarginal behavior,[52] the effectiveness of the threshold contract is weakly decreasing in δ.*

Proof. We first show that, if costs are positive, cost-effectiveness in the threshold is not increasing in δ. Because Proposition 1 showed that compliance is decreasing in δ, this establishes that effectiveness is decreasing in δ when costs are positive. We then show sufficient conditions for threshold effectiveness to decrease in δ when costs can be negative.

[51] $F_{e_2|e_1}(x)$ is weakly decreasing in e_1 for all x, with $F_{e_t|e_{t'}}(x)$ the conditional CDF of e_t given $e_{t'}$.

[52] See equation (20) for the exact condition. The intuition for why high levels of inframarginal behavior (combined with low $\frac{\lambda}{M}$) can flip the effectiveness prediction is as follows. If there is inframarginal behavior, then the principal effectively gets "free" compliance if people comply on day 2 only and not day 1. As we will show, lower δ increases compliance by making people more likely to comply on day 1. The benefit is extra compliance and the cost is extra payment. The cost will be particularly large if there is a lot of inframarginal behavior on day 2, because now the principal has to pay out for all of the day 2's on which day 1 compliance was induced, which the principal used to get for free.

To simplify notation, let e^* be the agent's cutoff value for complying in period 1, such that agents comply in period 1 if $e_1 < e^*$. From equations (15) and (16), we know that the value of e^* will depend on the agent's sophistication and, importantly, decrease in δ.

With our new notation, we can write the compliance decisions as:

$$\begin{aligned} w_1 &= \mathbb{1}\{e_1 < e^*\} \\ w_2 &= w_1\mathbb{1}\{e_2 < M\} + (1 - w_1)\mathbb{1}\{e_2 < 0\} \\ &= w_1\mathbb{1}\{0 < e_2 < M\} + \mathbb{1}\{e_2 < 0\} \end{aligned}$$

A Special Case: Positive Costs We first examine the restricted case where $e_1 > 0$ and $e_2 > 0$ and show that, in that case, C/P is not increasing in δ. In that case, $w_2 = w_1 w_2$. Therefore we have:

$$\begin{aligned} C/P &= \frac{1}{M}\frac{\mathbb{E}[w_1 + w_2]}{\mathbb{E}[w_1 w_2]} = \frac{1}{M}\frac{\mathbb{E}[w_1 + w_1 w_2]}{\mathbb{E}[w_1 w_2]} = \frac{1}{M}\left(\frac{\mathbb{E}[w_1]}{\mathbb{E}[w_1 w_2]} + 1\right) = \frac{1}{M}\left(\frac{\mathbb{E}[w_1]}{\mathbb{E}[w_1]\mathbb{E}[w_2|w_1 = 1]} + 1\right) \\ &= \frac{1}{M}\left(\frac{1}{\mathbb{E}[w_2|w_1 = 1]} + 1\right) \qquad (17) \end{aligned}$$

Consider the first term, $\frac{1}{\mathbb{E}[w_2|w_1=1]}$. To show this is not increasing in δ, we show that $\mathbb{E}[w_2|w_1 = 1] = \mathbb{E}[\mathbb{1}\{e_2 < M\}|w_1 = 1]$ is weakly increasing in δ. Call this expression p_2^*. If costs were IID, then $p_2^* = F(M)$, which is independent of δ. To see that p_2^* is also weakly increasing in δ under our more general assumption that e_2 is weakly increasing in e_1, note that higher δ means that $w_1 = 1$ will be associated with lower values of e_1 (since e^* is decreasing in δ). This implies lower values of e_2 conditional on $w_1 = 1$, since we assume that e_2 is weakly increasing in e_1. Lower values of e_2 then mean that $p_2^* = E[w_2|w_1 = 1]$ will be weakly higher. Hence, p_2^* is weakly increasing in δ and the first term is weakly decreasing in δ. Thus, we have shown that, with positive costs, C/P is weakly decreasing in δ.

General Case Instead of using cost-effectiveness as a means to prove the result for effectiveness, we turn to the expression for effectiveness directly: $\lambda C - P$. We show the conditions under which it is weakly increasing in e^*, and hence weakly decreasing in δ.

First, we rewrite the expression for effectiveness under the threshold given what we know about C and P. (For notational simplicity, we examine $2(\lambda C - P)$ instead of $\lambda C - P$.)

$$\begin{aligned} 2\left(\lambda C - P\right) &= \lambda\mathbb{E}[w_1 + w_2] - M\mathbb{E}[w_1 w_2] \\ &= \lambda\left(F(e^*) + \mathbb{E}[w_1\mathbb{1}\{0 < e_2 < M\} + \mathbb{1}\{e_2 < 0\}]\right) - M\mathbb{E}[w_1\mathbb{1}\{e_2 < M\}] \\ &= \lambda\left(F(e^*) + \mathbb{E}[\mathbb{1}\{e_1 < e^*\}\mathbb{1}\{0 < e_2 < M\} + \mathbb{1}\{e_2 < 0\}]\right) - M\mathbb{E}[\mathbb{1}\{e_1 < e^*\}\mathbb{1}\{e_2 < M\}] \\ &= \lambda\left(F(e^*) + Prob(e_1 < e^*, 0 < e_2 < M) + Prob(e_2 < 0)\right) - MProb(e_1 < e^*, e_2 < M). \qquad (18) \end{aligned}$$

We now take a derivative with respect to e^*. Let $g(e^*) = Prob(e_1 \leq e^*, e_2 \in S)$, where S is some set. It is straightforward to show that $g'(e^*) = f(e^*)\,Prob(e_2 \in S|e_1 = e^*)$.[53] Thus, we have

$$\frac{d}{de^*}[2\,(\lambda C - P)] = \lambda[f(e^*) + f(e^*)Prob(0 < e_2 < M|e_1 = e^*)] - Mf(e^*)Prob(e_2 < M|e_1 = e^*)$$

Hence, a sufficient condition for effectiveness to increase in e^* (and decrease in δ) is:

$$\lambda(1 + Prob(0 < e_2 < M|e_1 = e^*)) \geq MProb(e_2 < M|e_1 = e^*) \tag{19}$$

or

$$\frac{\lambda}{M}(1 + Prob(0 < e_2 < M|e_1 = e^*)) \geq Prob(e_2 < 0|e_1 = e^*) + Prob(0 < e_2 < M|e_1 = e^*)$$

or

$$Prob(e_2 < 0|e_1 = e^*) \leq \frac{\lambda}{M} + \left(\frac{\lambda}{M} - 1\right) Prob(0 < e_2 < M|e_1 = e^*)). \tag{20}$$

If $\lambda > M$, condition (20) will always hold. More broadly, the condition will be more likely to hold the greater λ relative to M. The condition essentially guarantees that there not be "too much" inframarginal behavior, which generally decreases the efficacy of incentives. For example, when $\lambda > M/2$, which is a reasonable condition as it guarantees that the payment to the agent for two days of compliance is less than the benefits to the principal, a sufficient condition is

$$Prob(e_2 < 0|e_1 = e^*) < Prob(e_2 > M|e_1 = e^*).$$

We have thus showed that, as long as there is not "too much" inframarginal behavior (i.e, as long as equation (20) holds), the effectiveness of a threshold contract is decreasing in δ. □

We now turn to examine threshold contracts with $K < T$. To gain tractability, we begin with the case where costs are perfectly correlated across periods, showing that both compliance and effectiveness under the threshold are increasing in impatience for any threshold level $K \leq T$.

Proposition 3 (Perfect Correlation, Threshold Effectiveness and Impatience over Effort)**.** *Let there be perfect correlation in costs across periods ($e_t = e_{t'} \equiv e$ for all t, t'). For simplicity, if $\delta^{(t)} < 1$ for any t, let $\delta^{(t)} < 1$ for all $t > 0$. Fix all parameters other than $\delta^{(t)}$ for some $t \leq T - 1$. Take any threshold contract with threshold level $K \leq T$. Compliance and effectiveness in the threshold contract will be weakly decreasing in $\delta^{(t)}$.*
Proof. See Appendix H.1. □

To make the problem more tractable when costs are not perfectly correlated, we now consider a simplified model where $T = 3$, $K = 2$, costs take on only two values (high or low), discount factors are exponential, and agents observe all future cost realizations on day 1. Again, threshold compliance and effectiveness are higher among those who are more impatient over effort.

[53] To show this, note that

$$\begin{aligned} g(e^* + \epsilon) - g(e^*) &= Prob(e^* < e_1 \leq e^* + \epsilon, e_2 \in S) = Prob(e^* < e_1 < e^* + \epsilon)Prob(e_2 \in S|e^* < e_1 \leq e^* + \epsilon) \\ &= (F(e^* + \epsilon) - F(e^*))\,Prob(e_2 \in S|e^* < e_1 \leq e^* + \epsilon). \end{aligned}$$

Dividing by ϵ gives us: $\frac{g(e^*+\epsilon)-g(e^*)}{\epsilon} = \frac{(F(e^*+\epsilon)-F(e^*))}{\epsilon}Prob(e_2 \in S|e^* < e_1 \leq e^* + \epsilon)$. Letting ϵ go to 0 and using the definition of the derivative gives that $g'(e^*) = f(e^*)\,Prob(e_2 \in S|e_1 = e^*)$.

Proposition 4. *Let $T = 3$. Let the cost of effort on each day be binary, taking on either a "high value" (e_H) or a "low value" (e_L), with $e_H \geq e_L$. Let agents observe the full sequence of costs e_1, e_2, e_3 on day 1. Let $\delta^{(t)} = \delta^t$ (i.e., let the discount factor over effort be exponential) and let $d^{(t)} = 1$. Fix all parameters other than δ. Consider a threshold contract with $K = 2$, where the agent must thus comply on at least 2 days in order to receive payment. Compliance and effectiveness in the threshold contract are weakly higher for someone with a discount factor $\delta < 1$ than for someone with discount factor $\delta = 1$.*

Proof. See Appendix H.1. □

For sophisticates, we can also show a stronger result. In simulations with most realistic cost distributions, this stronger result goes through for naifs as well.

Proposition 5. *Let $T = 3$. Let costs be weakly positive and let agents observe the full sequence of costs e_1, e_2, e_3 on day 1. Let $\delta^{(t)} = \delta^t$ (i.e., let the discount factor over effort be exponential) and let $d^{(t)} = 1$. Fix all parameters other than δ. Consider a threshold contract with $K = 2$, where the agent must thus comply on at least 2 days in order to receive payment. For sophisticates, compliance and effectiveness in the threshold contract are weakly decreasing in the discount factor δ.*

Proof. See Appendix H.1. □

B.5 Overall Effectiveness of Thresholds versus the Base Case

While Prediction 1 speaks to the heterogeneity in the performance of threshold relative to separable contracts by $\delta^{(t)}$, it is also important from a policy perspective to understand which type of contract performs better for any given level of $\delta^{(t)}$. The propositions in this section provide the theoretical underpinning for Prediction 2, which, while less general than Prediction 1, addresses this question. Specifically, Prediction 2 says that, under certain conditions, the most effective time-bundled threshold contract is more effective than the most effective linear contract if the discount factor over effort is sufficiently low, and less effective if the discount factor over effort is high.

Making some additional assumptions for tractability, we compare both optimized threshold and separable linear contracts, and threshold and linear contracts offering the same payment per day (as in our experiment),[54] paying particular attention to how the relative effectiveness of thresholds depends on δ. For simplicity, we assume that $T = 2$ and that $K = 2$ and denote the threshold payment as M (i.e., $M = 2m'$) throughout the section.

Our first proposition (Proposition 6) examines the relative performance of the contracts in the limit as δ goes to 0 under very general assumptions. It shows that, for sufficiently low δ, for any linear contract, there exists a threshold contract that achieves substantially higher cost-effectiveness with relatively little—and potentially even no—loss in compliance. In contrast, for any linear contract, one can always construct another *linear* contract with substantially

[54]In many empirical applications, constructing the optimal contract is not feasible as it requires knowledge of both the discount rate and the distribution of costs.

higher cost-effectiveness by decreasing the payment amount, but the loss in compliance may be arbitrarily large.

The next four propositions (Propositions 7a–8b) examine the full range of δ, not just the case where δ is sufficiently low. While we make additional assumptions on the effort cost distributions for tractability, the propositions demonstrate that thresholds can be effective for those who are impatient over effort in the two limiting cases of perfectly correlated and IID effort costs. IID effort costs is a common assumption in the literature (e.g., Garon et al., 2015). In each case, we begin with a testable comparison between threshold and linear contracts that offer the same payment per day before moving to more abstract comparisons that teach us about whether the optimal threshold contract or the optimal linear contract is more effective (and how that relationship depends on δ).[55]

Proposition 6. *Let $d = 1$ and $T = 2$. Fix all parameters other than δ, and take a linear contract that induces compliance $C > 0$.*
(a) If agents are naive and e_2 is weakly increasing in e_1, in a first order stochastic dominance sense,then for sufficiently small δ, there exists a threshold contract with $K = 2$ that has at least two times higher cost-effectiveness (and $1 + \frac{1}{C}$ times higher cost-effectiveness if costs are IID) and that generates compliance $\frac{1+C}{2}$ of the linear contract.
(b) If agents are sophisticated and costs are IID, then for sufficiently small δ, there exists a threshold contract with $K = 2$ that has at least $1 + C$ times higher cost-effectiveness and that generates compliance at least $\frac{1+C}{2}$ of the linear contract.

Proof. See Appendix H.2. □

The potential improvements from threshold contracts demonstrated by Proposition 6 are quantitatively large. For example, when costs are IID and agents are naive with sufficiently low δ, for a linear contract that generates $C = .9$, there exists a threshold contract that generates 95% as much compliance but for less than half the cost.

Proposition 7a (Perfect Correlation, $M = 2m$)**.** *Let $T = 2$. Fix all parameters other than δ. Consider a linear contract with payment m and a threshold contract with payment $2m$. Then, regardless of agent type, the threshold contract is more effective than the linear contract if $\delta < 2d - 1$. If $\delta \geq 2d - 1$, then the linear contract may be more effective.*

Proof. See Appendix H.2. □

Proposition 7b (Perfect Correlation, Optimal Contracts)**.** *Let $T = 2$. Fix all parameters other than δ, and take any linear contract that induces compliance $C > 0$. Let there be perfect correlation in costs across days ($e_1 = e_2$). Then, regardless of agent type, there exists a threshold contract that induces compliance of at least C and that has approximately $2\frac{d}{1+\delta}$ times greater cost-effectiveness than the linear contract. Hence, if $\delta < 2d - 1$, the most effective contract will always be a threshold contract.*

[55] Predictions about optimal contracts are hard to test since most policymakers do not have sufficient information about the cost function and δ to solve for the optimal contracts.

Proof. See Appendix H.2. □

Proposition 8a (IID Uniform, $M = 2m$)**.** *Let $d = 1$. Fix all parameters other than δ. Let costs be independently drawn each day from a uniform[0,1] distribution. Take any threshold contract paying $M < 2$ and compare it with the linear contract paying $m = \frac{M}{2}$.*
(a) If $M < 1$, the threshold contract is always more cost-effective, but whether it has higher compliance (and hence whether it is more effective) depends on δ. Define $\frac{2M^2}{1+M}$ as the cutoff value *for naifs and $2 - \frac{2}{M+M^2}$ as the* cutoff value *for sophisticates. If δ is less than the cutoff value for a given type, then the threshold contract is more effective, as it generates greater compliance.*
(b) If $1 \leq M < 2$,[56] then the threshold contract is more effective.

Proof. See Appendix H.2. □

Proposition 8b (IID Uniform, Optimal Contracts)**.** *Let $d = 1$. Fix all parameters other than δ. Let costs be independently drawn each day from a uniform[0,1] distribution. Whether the most effective threshold contract is more effective than the most effective linear contract depends on δ as well as λ, the principal's marginal return to compliance. For a wide and plausible range of values of λ,[57] there exists a* cutoff *value of δ such that the threshold contract is more effective when δ is below the cutoff, and the linear contract is more effective when δ is above the cutoff. For the remaining values of λ, either the threshold contract is always more effective, or the linear contract is always more effective, but in either case the effectiveness of the threshold relative to linear is decreasing in δ.*

Proof. See Appendix H.2. □

B.6 Payment Frequency

In this subsection, we first prove Prediction 3 from Section 2.4. Next, we present and prove a related prediction (Prediction 4) that follows Kaur et al. (2015) in showing an additional way to use empirical data to make inferences about the discount factor over payments, which we use in Section 5.4.

Before showing its proof, recall that Prediction 3 is the following: If agents are impatient over financial payments ($d^{(t)} < 1$), then the compliance and effectiveness of the base case linear contract are weakly increasing in the payment frequency. If agents are patient over financial payments ($d^{(t)} = 1$), then payment frequency does not affect compliance or effectiveness.

Proof. Equation (12) implies that, in a linear contract, $C = \frac{1}{T}\sum_{t=1}^{T} F(d^{(T-t)}m)$. Compliance is thus increasing in the discount factor over payment $d^{(T-t)}$. If agents are "impatient," then $d^{(T-t)}$ is weakly decreasing in the delay to payment $T-t$. Increasing payment frequency then decreases the average delay to payment, which weakly increases compliance. If agents are patient, then the discount factor is 1 irrespective of the delay to payment and increasing payment frequency has no effect on compliance. Effectiveness follows the same pattern as compliance since cost-effectiveness is invariant to payment frequency (it is always $\frac{1}{m}$). □

[56] Note that the principal would never pay $M > 2$ since $M = 2$ achieves 100% compliance regardless of δ.
[57] See proof in Appendix H.2 for specific ranges for both naifs and sophisticates.

Prediction 4 (Payday Effects)**.** *If the discount factor over payments $d^{(t)}$ is decreasing in t, then the probability of complying in the base case linear contract increases as the payday approaches. If the discount factor over payments $d^{(t)}$ is constant in t, then the probability of complying is constant as the payday approaches.*

Proof. Recall that, on day t, agents comply if $e_t < d^{(T-t)}m$. As the payment date approaches, the time to payment $T-t$ decreases. If $d^{(T-t)}$ is decreasing, this increases $d^{(T-t)}$ and hence increases the likelihood that $e_t < d^{(T-t)}m$. If $d^{(T-t)}$ is flat, then the likelihood that $e_t < d^{(T-t)}m$ remains constant. □

B.7 Modeling a Health Benefit to Compliance

Our model excludes a long-term health benefit of walking. In this section, we show that our theoretical predictions are robust to this change under a range of reasonable conditions: (1) with domain-specific discount factors over health and effort, or—even with a single discount factor for health and effort—(2) a small discounted health benefit relative to the discounted daily incentive payment, or (3) more periods of effort required to meet the threshold under the contract.

These conditions appear reasonable in our setting. (1) Empirical work suggests that domain specific discount factors for health and effort are likely. People tend to discount health more than money (Chapman and Elstein, 1995; Chapman, 1996; Hardisty and Weber, 2009), and people discount health gains more sharply than both losses in the health domain and gains in other domains (Hardisty and Weber, 2009; Chapman, 1996). (2) Even if people do discount health and effort with the same discount factor, the health benefit is much further in the future than the effort and payment we model, and so its discounted value is likely small. (3) The 4-day and 5-day thresholds used in our contracts are much higher than the 2-day thresholds that we examine in depth theoretically.

We first show how adding a future health benefit of compliance, b, impacts our predictions about behavior in linear vs. threshold contracts in a simple case where people discount health and effort with a single constant discount factor. We then briefly discuss how alternative, and arguably more realistic, specifications of the discount factor over health will dampen (or even eliminate) the impact of the health benefit to compliance on our predictions.

B.7.1 Predictions with One Discount Factor for Health and Effort

For simplicity, we restrict attention to the case where b is discounted with a simple quasi-hyperbolic discount factor that is identical to the discount factor over effort, $\delta^{(t)} = \delta$ if $t > 0$. We further restrict the discount factor over money, $d^{(t)}$, to be 1.

Compliance in Linear Contract Participants now comply on day j if the discounted payment *and* discounted benefit outweigh the effort cost:

$$e_j < m + \delta b. \tag{21}$$

Compliance is $F(m + \delta b)$, with $F(\cdot)$ the effort cost CDF. Compliance in the linear contract, and compliance without a contract, are thus no longer independent of δ: they increase in δ.

Relative Compliance in Threshold Contract The relationship between $\delta^{(t)}$ and behavior under threshold contracts is now more complex. We discuss the implications of adding b for Prediction 1 and then Prediction 2 in turn. Specifically, Prediction 1—that compliance and effectiveness in threshold relative to linear contracts are decreasing in the discount factor over effort—should still hold provided that: (a) the threshold payment m' is large relative to the future benefit of compliance b, or (b) a large number of periods of effort are required to meet the threshold (i.e., large K, which requires large T). Similarly, Prediction 2 can still hold: the discount factor over effort can still be pivotal to the relative effectiveness of threshold relative to linear contracts. Appendix H.3 presents the formal mathematical propositions underlying this discussion.

Prediction 1 with Discounted Benefit Prediction 1 is that compliance and effectiveness in threshold relative to linear contracts tend to decrease in $\delta(t)$. Adding b complicates this prediction and its underlying propositions (for simplicity, we focus here on the compliance implications). For example, without b, compliance in a threshold contract with $K = T$ is weakly decreasing in $\delta^{(t)}$ regardless of the cost distribution or other parameters (Proposition 1). With b, whether threshold compliance is weakly decreasing in $\delta^{(t)}$ now depends on parameters such as the cost distribution, the threshold payment m', and the threshold level K.

Simulation results show that two factors increase the likelihood that Prediction 1 holds: (a) a high threshold payment m' relative to the benefit b, and (b) a large number of periods until the threshold is reached K. We demonstrate these ideas more rigorously in the propositions presented in Appendix H.3, which for tractability assume perfect correlation in costs across periods.

First, Proposition 9 shows that, for a threshold contract with threshold level $T = K$, the sign of the derivative of compliance with respect to δ depends on the value of the daily threshold payment m' relative to $\frac{b}{K-1}$. When $m' \geq \frac{b}{K-1}$, compliance in the threshold contract decreases in δ, as it does in the model without b. This implies that compliance in the threshold relative to linear contract also decreases in δ (i.e., that Prediction 1 holds) since compliance in the linear contract is increasing in δ. The expression $m' \geq \frac{b}{K-1}$ is more likely to hold (a) the larger is m' relative to b, and (b) the larger is K, demonstrating the importance of these two factors.

In contrast, when $m' < \frac{b}{K-1}$, the derivative of threshold compliance with respect to δ is positive—making the derivative of *relative* compliance (compliance in the threshold relative to linear contract) ambiguous, as the derivative of linear compliance is also positive. Which derivative is more positive will depend on parameter values. To provide some results in this case, Proposition 10 makes further assumptions about the cost distribution (e.g., uniform costs across people), and shows that for high enough δ, relative compliance again tends to decrease in δ.[58] Simulation results support the findings from this simplified model.

Prediction 2 with Discounted Benefit Prediction 2 concerns the *level* of threshold relative to linear compliance, not just their comparative static in δ. Namely it states that δ can be pivotal to the relative effectiveness of threshold and linear contracts: when δ is sufficiently low,

[58] Relative compliance also tends to decrease in δ for a wider range of δ among naifs than sophisticates—a finding backed up by simulation results as well.

threshold contracts can be more effective than linear, whereas when δ is sufficiently high, linear can be more effective. Simulation results suggest that, even after adding a b term, under parameter assumptions that support Prediction 1, δ can also be pivotal to the relative effectiveness of threshold and linear contracts in models. To demonstrate this idea more rigorously, Proposition 11 shows that, in a simplified model with perfect correlation in costs and $T = 2$, a threshold contract offering the same per-day payment as a linear has the same compliance and effectiveness as the linear when $\delta = 1$, but weakly *higher* compliance and effectiveness when $\delta < 1$.

B.7.2 Alternative Health Discounting Models

While we cannot speak to all potential models, several other reasonable models for discounted benefits reduce the impact of b on our predictions. In particular, the assumption that $\delta^{(t)} = \delta$ produces a particularly large impact for three reasons: it assumes perfect correlation between discounting over the short- and long-run, it applies the same level of discounting in both the short- and long-run, and it applies the same discount rate to both effort costs and health benefits. Relaxing any of these assumptions mitigates the impact of b on our predictions.

Domain-Specific Discount Factors If discount factors are domain-specific across effort and health, then adding b does not change the results in Section 2. Specifically, if b is discounted by a health-specific discount factor other than $\delta^{(t)}$, the addition of b would leave our predictions unchanged.

More Flexible Quasi-Hyperbolic Discounting In practice, while our contracts incentivize effort in the near future, health benefits of compliance are realized far in the future (e.g., years rather than days). This is a critical distinction under a quasi-hyperbolic or "beta-delta" discount factor, where $\delta^{(t)} = \beta\delta^t$ for some $\delta < 1$. These conditions mitigate the impact of b for two reasons.

First, the magnitude of the discounted benefit of compliance will fade if it is further in the future: $\delta^{(t)}b = \beta\delta^t b$ approaches 0 for large enough t. As demonstrated above, the discounted benefit has a smaller impact on our predictions if its value is smaller. Second, while discounting over near-term effort would be primarily driven by the quasi-hyperbolic β term, discounting over future health benefits would depend more on the exponential δ term. This separation brings the comparative statics with respect to the short-run effort discount factor (holding all else constant) closer to the model without b.

More generally, the more people discount events far in the future (conditional on their short-run discount rates) and/or the lower the correlation between short and long-run discount rates, the smaller the impact of b on the comparative statics with respect to the short-run effort discount factor (holding all else constant). At the extremes, if short and long-run discount factors are uncorrelated or if discounted benefits of compliance approach zero, the situation resembles the domain-specific case above, leaving our predictions unchanged.

C Measuring the Effort and Payment Discount Factors

This section provides additional detail on measurements of impatience in our sample. We first describe how we validate the impatience index—our primary measure of effort discounting—using an incentivized effort task. We then present multiple estimates of the discount factors over effort and payment from our experimental context, showing substantial discounting of effort but not of payment. Finally, we show that there is limited correlation between the discount factors over effort and payment.

C.1 Validating the Impatience Index

We begin by describing the incentivized effort task data used for the validation exercise, along with other data collected. Next, we describe two effort discount rate measures obtained from these data. Third, we use these measures to validate our impatience index.

C.1.1 Data Collection in the Validation Sample

We validate our impatience index using a separate sample of 71 people who are very similar to our experimental sample (hereafter: the "validation sample").[59] The validation sample was randomly selected from a later evaluation of a similar incentive program for exercise (Dizon-Ross and Zucker, 2025) with nearly identical recruitment criteria,[60] and observable characteristics are balanced across the validation sample and experimental sample: walking levels, demographic characteristics, BMI, etc., are statistically indistinguishable (Appendix Table F.15).

In the validation sample, we collected the same *impatience index* described in this study and incentivized two tasks to measure impatience over effort and recharges, respectively.

Effort Task Respondents were incentivized to perform an effort task, which we call the "Effort Choice by Date" task, following the methodology of Augenblick (2018) and Augenblick and Rabin (2019), which John and Orkin (2022) previously adapted to a field setting. The task was to call into a toll-free automated phone line, listen to a useless 30-second recording, and answer a simple question to confirm that they listened. On the survey date (day 0), individuals chose how many calls to complete at time t for a piece rate w, where t is 0 (i.e., the same day), 1, 7, or 8 days from the time of the decision, and the piece rate is INR 10, 6, 2, or 0.[61] One choice was then randomly selected for implementation, and respondents received both the piece rate for the implemented choice as well as an additional 100 INR if they completed all the tasks they chose (in addition to one "mandatory task"). We refer to the measures we construct from these data as *effort impatience* measures.

Patterns in the data indicate that respondents understood the exercise. For example, the average number of tasks chosen increases with the piece rate, with respondents choosing an average of 5.6, 7.1, 7.6, and 8.0 tasks when the piece rates were 0, 2, 6, and 10 INR, respectively. Our field team also reported limited respondent misunderstanding.

[59] The sample size is comparable to the number of people who completed choices in the two seminal papers measuring impatience with effort tasks: 99 in Augenblick (2018) and 100 in Augenblick and Rabin (2019).

[60] Both studies targeted participants from Coimbatore, Tamil Nadu, using public screening camps as the primary recruitment tool, and both focused on individuals aged 30-65 who were literate, comfortable using mobile phones, capable of receiving mobile recharge payments, and had or were at high risk of lifestyle disease. However, the later study enrolled participants with high blood pressure in addition to high blood sugar.

[61] We include a 0 INR piece rate following guidance from John and Orkin (2022) that it helped their model converge. However, our structural model does not converge with the 0 INR piece rate choice, so we exclude it when estimating the structural parameters.

Recharge Choices A secondary goal for the validation sample was to assess the relationship of the impatience index with recharge impatience. To do so, we measure impatience over recharges with a multiple price list (MPL) (Andreoni and Sprenger, 2012a; John and Orkin, 2022). Participants made 10 choices between receiving a recharge today and a later date (either 7 and 14 days from today). For simplicity, the recharge today was always 50 INR, and the later recharges were larger whole numbers: 60, 70, 90, 100, and 150 INR. One choice from the MPL was also randomly selected for implementation.

The MPL choices are not ideal for estimating a structural recharge discount factor: the later payment amounts are all meaningfully larger than the earlier payment (we cannot distinguish between one-week discount factors in the range from $\frac{50}{60} = 0.83$ to 1), and, as with all MPLs, any mistrust in receiving the payment will push participants toward earlier payment and bias implied discount factors downwards (Halevy, 2008; Andreoni and Sprenger, 2012b). Instead, we construct a reduced-form *recharge impatience measure* as the proportion of choices where the individual chose the smaller recharge on the sooner date.

C.1.2 Structural and Reduced-Form Effort Impatience Measures

The data from the effort task are consistent with positive discounting of future effort with some present bias. Consistent with positive discounting, the number of tasks chosen on days with $t > 0$ are all significantly greater than on $t = 0$. (Specifically, participants chose 7.4, 7.0 and 7.5 tasks on days 1, 7, and 8, respectively, and only 6.4 tasks on day 0.) Consistent with present bias, the biggest jump in task allocations appears between "today" and "tomorrow".

We thus parameterize a constant discount factor for all future days: $\delta^{(t)} = \delta_{QH}$ if $t > 0$. This is equivalent to a $\beta - \delta$ model in which $\delta = 1$. We use the effort task data to construct two measures, one structural and one reduced-form, for this parameter.

Structural Measure and Evidence Our structural estimation follows John and Orkin (2022).[62] (The estimating equation is in the notes to Table C.1.) We structurally estimate δ_{QH} at the group level. As in John and Orkin (2022), individual-level structural estimates converge for less than half of our sample.

Column 1 of Table C.1 shows that, in the full validation sample, we estimate a δ_{QH} of 0.572, which is significantly different from 1 and suggests a high degree of effort impatience. In column 2, we follow Augenblick and Rabin (2019) and remove "problematic" individuals with limited effort choice variation or effort choices that are not primarily monotonic in wage offers.[63] The discount factor estimate is similar and still significantly different from 1.

Reduced-Form Measure and Evidence Our reduced-form measure is based on the excess number of tasks chosen on future dates relative to day 0 at a given piece rate, following Augenblick (2018) and Augenblick and Rabin (2019). Specifically, for all task allocations made on future days ($t > 0$) at piece rate w, we construct a measure at the individual $\times$ choice level equal to the tasks allocated on day t minus the tasks allocated on day 0 at the same piece rate w. People who are more impatient (lower δ_{QH}) will choose more tasks on future days than today, and thus have higher average values of this measure.

[62] John and Orkin (2022) assumes quasilinear utility and a power effort cost function following Augenblick (2018), and includes a non-monetary per-task reward s in addition to the piece rate following DellaVigna and Pope (2018).

[63] We remove 28 of 71 respondents in a field setting; Augenblick and Rabin (2019) remove 28 of 100 in a lab setting for the same reasons. Our removal rates are not significantly different for those with below- vs. above-median impatience index.

Appendix Table C.1: Structural Estimates of the Effort Discount Factor

	Full validation sample		Below-median impatience sample		Above-median impatience sample	
	(1)	(2)	(3)	(4)	(5)	(6)
δ_{QH}	0.572	0.556	0.996	0.996	0.176	0.367
	[0.132]	[0.153]	[0.009]	[0.007]	[0.156]	[0.208]
P-value: $\delta_{QH} = 1$	0.001	0.004	0.674	0.597	<0.001	0.002
P-value: $\delta_{QH} = \delta_{QH}^{\text{Below}}$	0.001	0.004			<0.001	0.002
P-value: $\delta_{QH} = \delta_{QH}^{\text{Above}}$	<0.001	0.063	<0.001	0.002		
Sample	All	Changers + Monotone	All	Changers + Monotone	All	Changers + Monotone
# Individuals	71	43	32	24	39	19
# Observations	852	516	384	228	468	228

Notes: This table displays structural estimates of the effort discount factor, δ_{QH}, in the validation sample, estimated using data from the Effort Choice by Date task of Augenblick (2018) using an estimation approach similar to John and Orkin (2022). The optimal allocation of effort is given by: $e^* = argmax(s + d^{(11)} \cdot \phi \cdot w) \cdot e - \delta^{(t)}(\frac{1}{\gamma}e^{\gamma})$, where t is the time of effort provision, γ captures the convex cost of effort, s is a parameter that captures the non-monetary reward for each task, w is the monetary piece rate, $d^{(11)}$ captures the monetary discounting of the payment in 11 days, and ϕ is a slope parameter. We parametrize $\delta^{(t)} = \delta_{QH}$ (equivalent to a quasihyperbolic model with $\delta = 1$) and $d^{(11)} = 1$ and estimate s, ϕ, δ_{QH}, and γ. We present results using the full validation sample and the subsamples with below- and above-median impatience index, with or without inclusion restrictions from choice patterns. Columns 1, 3, and 5 have no inclusion restriction; columns 2, 4, and 6 restrict to individuals who changed their effort choice at least once and had at most 1 choice non-monotonicity in payment levels.

Overall, participants chose to complete 13% fewer tasks in the present than the future, suggesting meaningful effort discounting. The result is similar if we again remove problematic individuals: the restricted sample allocates 15% fewer tasks in the present than the future. Our results mimic Augenblick (2018) and Augenblick and Rabin (2019) which find that participants choose to complete 16% and 10-12% fewer tasks in the present than the future, respectively.

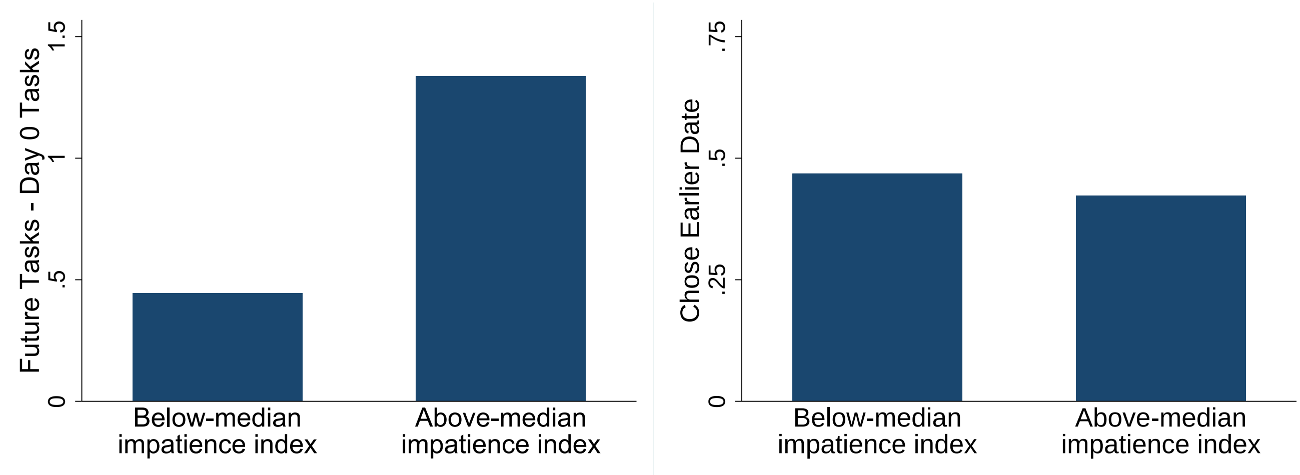
C.1.3 The Impatience Index Correlates with Effort Impatience Measures

In this section, we show that our impatience index correlates with the incentivized effort impatience measures in the validation sample. In contrast, it does not correlate with recharge impatience. Overall, this provides evidence that the impatience index proxies for impatience in the effort, but not payment, domain.

Correlation with Effort Impatience Measures Columns 3–6 of Table C.1 show that structural estimates of δ_{QH} are substantially higher among those with lower impatience index. Specifically, among individuals with below-median impatience index, our estimate of δ_{QH} is 0.996 and statistically indistinguishable from 1 (column 3). In contrast, we estimate that δ_{QH} is 0.176 for those with above-median impatience (column 5). We can reject equality of this estimate with 1 and with the corresponding estimate of δ_{QH} for those with below-median impatience. Columns 2, 4, and 6 show similar results after removing problematic respondents: our estimates of δ_{QH} are again significantly different for those with above- and below-median impatience index.

We summarize the reduced-form effort impatience measure separately for those with above- and below-median impatience index in Figure C.1(a). The above-median impatience sample has substantially higher average values of the reduced-form effort impatience measure: they allocate an average of 1.3 more tasks to future dates than today across piece rates, while those with below-median impatience index allocate only 0.4 more tasks to future days.

Appendix Figure C.1: Higher Impatience Index Predicts Higher Effort Impatience but Not Higher Recharge Impatience



(a) Effort Impatience (b) Recharge Impatience

Notes: Data come from the validation sample and are at the individual level. Panel (a) displays the average difference between the number of tasks chosen on all future dates minus the number of tasks chosen on the survey day (for the same payment amount) separately for the below- and above-median impatience index samples. In Panel (b), we display the average proportion of recharge MPL choices where the individual chose to get a smaller recharge today rather than a larger recharge in the future separately for the below- and above-median impatience index samples.

To test the significance of this difference, we estimate the following regression:

$$EffortImpatience_{itw} = \beta_0 + \beta_1 ImpatienceIndex_i + \beta_2 y_{i0w} + \tau_w + \tau_t + \varepsilon_{itw} \qquad (22)$$

where $EffortImpatience_{itw}$ is the reduced-form effort impatience measure for individual i allocating tasks on day t at piece rate w, $ImpatienceIndex_i$ is either the impatience index or an indicator for having an above-median impatience index, and y_{i0w} is the number of tasks chosen by individual i at piece rate w on day 0; controlling for this allows the effort impatience measure to vary with the overall number of chosen tasks and improves precision.[64] τ_w and τ_t are fixed effects for the piece rate and task day, respectively. The coefficient of interest is β_1.

Consistent with Figure C.1, Column 1 of Table C.2 shows that the difference in reduced-form effort impatience between those with above- and below-median impatience index is roughly 1.0 task, significant at the 10% level. Column 2 shows that the relationship is even stronger excluding

[64] Define y_{itw} as the number of tasks chosen by individual i on day t for piece rate w. Since $EffortImpatience_{itw} = y_{itw} - y_{i0w}$, the coefficients from this regression are exactly equivalent to a regression with y_{itw} as the dependent variable that includes the same controls. The specification in equation (22) allows the mean value of the dependent variable to be comparable to Figure C.1.

problematic individuals: the gap is 1.7 tasks, significant at the 5% level. Columns 3 and 4 show qualitatively similar but less precise patterns with the impatience index as the regressor.

Appendix Table C.2: Impatience Index Correlates With Effort (But Not Recharge) Impatience

	Effort impatience				Recharge impatience			
	Future tasks - day 0 tasks				Chose earlier date			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Above-median impatience index	1.000*	1.708**			-0.0457	-0.0397		
	[0.513]	[0.798]			[0.102]	[0.109]		
Impatience index			0.763	2.666*			-0.0425	-0.0559
			[0.629]	[1.490]			[0.0911]	[0.114]
P-value: Impatience	0.055	0.038	0.229	0.081	0.655	0.716	0.642	0.625
Sample	All	Changers + Monotone	All	Changers + Monotone	All	No violations	All	No violations
Dep. var. mean (below-median impatience)	0.445	0.596	0.445	0.596	0.469	0.455	0.469	0.455
Correlation (dep var, Impatience index)	0.15	0.19	0.13	0.25	-0.05	-0.05	-0.05	-0.06
# Individuals	71	43	71	43	71	64	71	64
# Observations	852	516	852	516	710	640	710	640

Notes: This table shows the relationship between the effort and recharge impatience measures and the impatience index in the validation sample. Each observation is an individual × effort or recharge choice. The dependent variable in columns 1–4 is the difference between the tasks allocated in the choice and the tasks allocated on day 0 (the survey date) for the same piece rate; controls include fixed effects for the piece rate and task day, as well as the number of tasks chosen for that same piece rate on day 0. The dependent variable in columns 5–8 is an indicator for choosing recharges today rather than in the future; controls include fixed effects for how many weeks in the future the individual will be paid for the later recharge option (either 1 or 2 weeks) and for the relevant payment amount. The "Changers + Monotone" sample restricts to individuals who changed their effort choice at least once and had fewer than two choice non-monotonicities in payment levels. The "No violations" sample represents people who do not switch multiple times on either price list. The regressor in columns 1, 2, 5 and 6 is an above-median impatience index dummy, while in columns 3, 4, 7 and 8 the regressor is the continuous index. Correlations shown at the bottom of each column are between the individual-level average of the dependent variable and the version of the impatience index used in that column. Significance levels: * 10%, ** 5%, *** 1%.

Excluding problematic individuals, the magnitudes of the correlations between effort impatience and the impatience index are relatively high for the (noisy) domain of effort impatience—0.3 and 0.2 for the continuous and binary indices, respectively. In comparison, Augenblick et al. (2015) and Augenblick (2018) find correlations of 0.2 and 0-0.2 between effort impatience estimates and demand for commitment or qualitative discounting questions, respectively.

Lack of Correlation with Recharge Impatience Measure Figure C.1(b) summarizes the recharge impatience measure separately for those with above- and below-median impatience index. Recharge impatience (i.e., choosing a smaller, sooner recharge over a larger, later recharge) is very similar across the subsamples; in fact, those with above-median impatience index have slightly lower recharge impatience. Columns 5 and 7 of Table C.2 confirm that there is no meaningful or significant relationship between recharge impatience and the impatience index using regression analysis. Columns 6 and 8 replicate the results without problematic respondents.

C.2 Additional Estimates of the Discount Factors Over Effort and Payment

In this section, we present two estimates of the discount factor over payment (recharges), and one additional estimate of the discount factor over effort, all from our main experimental sample. We then summarize these estimates alongside the effort discount factor estimated in the validation sample (the Section C.1.2 estimate based on the Effort Choice by Date data). While both estimates of the discount factor over effort are meaningfully below 1, both payment discount factor estimates are close to 1 and significantly higher than either effort discount factor estimate. We begin by describing the additional estimation procedures.

"Simple CTB" Estimates of the Discount Factors Over Effort and Payment Following Augenblick et al. (2015), we estimate the discount factors for effort and money using the "Simple CTB" choices in each domain described in Section 4.2. Our primary specifications parametrize each discount factor as a single quasihyperbolic discount factor on future events (e.g., $\delta^{(t)} = \delta_{QH}$) but we estimate exponential parameterizations for robustness (e.g., $\delta^{(t)} = \delta^{t}_{Exp}$).[65]

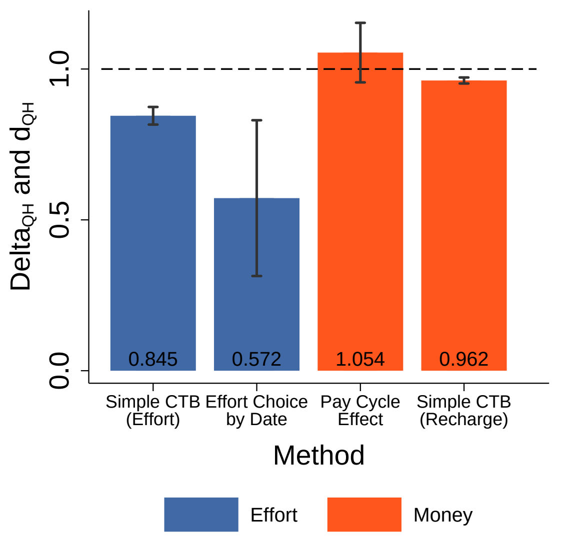
Paycycle Estimates of the Discount Factor Over Payment Since impatience over payment will lead effort to increase as the payday approaches, one can use the pattern of effort over the pay cycle to estimate the payment discount factor. We follow Kaur et al. (2015), which calculates the discount factor using the elasticity of walking to payment and the pattern of effort as the payday approaches. We calculate the payment discount factor with the equation $\frac{1}{d_{QH}} - 1 = \frac{1}{\varepsilon}\frac{w_T - w_{t<T}}{w_{t<T}}$. where ε is the elasticity of walking to payment, w_t is compliance in period t, day T is payday, and days $t < T$ all occur before payday. We calculate the percentage increase in compliance on payday, $\frac{w_T - w_{t<T}}{w_{t<T}}$ from the estimated "payday spike" in the base case group (column 1 of Appendix Table F.10), and we estimate ε from the compliance response to the payment variation between the small payment and base case groups.

Comparing the Discount Factors over Effort and Payment Figure C.2 shows the payment discount factor estimates from both the Simple Recharge CTB and the paycycle effects, as well as the effort discount factors estimated from the Simple Effort CTB and the Effort Choice by Date. In all cases, the figure presents the estimates with the discount factors parametrized as a single discount factor (δ_{QH}) applied to all future periods.

[65]The quasihyperbolic CTB discount factor over recharges is estimated with the equation $\ln\left(\frac{c_t+\omega_1}{c_{t+k}+\omega_2}\right) = \frac{\ln(d_{QH})}{\alpha-1}\mathbb{1}_{t=0} + \frac{1}{\alpha-1}(1+r)$ where c_t is money in the earlier period, c_{t+k} is money in the later period, ω_1 and ω_2 captures background consumption, and r is the interest rate for each choice. The estimating equation for the discount factor over effort is similar: c_t and c_{t+k} are replaced by e_t and e_{t+k} (minutes of walking on days t and $t+k$), ω_1 and ω_2 are background walking effort (10 minutes), and $1+r$ captures the marginal rate of substitution between sooner and later effort. Following Augenblick et al. (2015), we choose $\omega_1 = \omega_2 = \omega$, as a function of the base recharge consumption or base walking effort (we set the ω's at 50% of the base level for recharges and walking, so ω =50 INR and ω =10 minutes of walking, respectively), but the results are robust to a range of values from 25% to 200% of the base level.

The estimates of the payment discount factor are both near 1, with the payday effect estimate greater than (but not statistically significantly different from) 1, and the CTB estimate close to 1 (0.962) but significantly different from it. In contrast, both estimates of the effort discount factor are substantially smaller, at 0.572 from the validation sample and 0.845 from the Simple CTB in our main sample. Both are significantly less than either estimate of the payment discount factor (p-values for tests of equality are in the notes for Figure C.2.)

Appendix Figure C.2: The Discount Factors Over Effort Are Significantly Lower Than the Discount Factors Over Money



Notes: This figure presents four structural estimates of the discount factors over effort (blue bars) and payment (orange bars). From left to right, the estimates come from the Simple Effort CTB data from the experimental sample, the Effort Choice by Date data from the validation sample, the pay cycle method in the experimental sample, and the Simple Recharge CTB data from the experimental sample. The discount factor is parameterized as a single quasihyperbolic discount factor on the future ($\delta^{(t)} = \delta_{QH}$ or $d^{(t)} = d_{QH}$). The p-values for tests of equality between the effort discount factor (δ_{QH}) from the Effort Choice by Date methodology and the two monetary discount factors (d_{QH}) estimated via the Simple Recharge CTB and payday effects are 0.041 and 0.051, respectively. The p-values for tests of equality between the effort discount factor (δ_{QH}) from the Simple Effort CTB and the two monetary discount factors (d_{QH}) estimated via Simple Recharge CTB and payday effects are both <0.001. The respective samples for bars 1, 2, 3, and 4 include 852 choices of 71 individuals, 6,380 choices of 3,190 individuals, 71,672 days of 890 individuals, and 16,146 choices of 2,307 individuals.

Results are similar if we estimate exponential discount factors. We estimate daily exponential effort discount factors of 0.976 and 0.950 using Simple Effort CTB and Effort Choice by Date, respectively. Both are significantly less than 1 and significantly less than either estimate of the exponential payment discount factor (1.009 and 0.992 for Pay Cycle and Simple Recharge CTB estimates, respectively).

C.3 Measures of Effort and Recharge Impatience Are Uncorrelated

This section summarizes two types of evidence from our setting suggesting that discount factors over effort and recharge are relatively uncorrelated. First, survey measures of effort and recharge impatience are uncorrelated. Second, measures of effort impatience do not correlate with pay cycle effects.

Appendix Table C.3: No Correlation Between Measures of Impatience over Effort and Recharges

	Direct measure	Proxies for recharge impatience				
	Simple CTB (Recharge)	Negative mobile balance	Negative yesterday's talk time	Prefers daily (=1)	Prefers monthly (=-1)	# Individuals
	(1)	(2)	(3)	(4)	(5)	(6)
Impatience index	0.004	0.032	-0.068	-0.038	0.034	1740
Predicted impatience Index	0.000	0.021	-0.014	-0.005	-0.003	3192
Chose commitment	-0.006	0.009	-0.001	0.005	0.010	2871
Simple CTB	0.006	-0.011	-0.037	0.001	0.041	3190

Notes: This table displays the correlations in our experimental sample between our various measures of impatience in the effort domain (in the rows) and measures and proxies for impatience in the recharge domain (in columns). The "Simple CTB (Recharge)" measure is the average of the share of money allocated to today from the questions used in the Simple Recharge CTB. Proxies for recharge impatience in columns 2–5 were all measured at baseline. For columns 4 and 5: we asked participants whether they preferred daily, weekly, or monthly payments, and "Prefers Daily" ("Prefers Monthly") is an indicator that their most preferred frequency was daily (monthly). We normalize all impatience variables so that a higher value corresponds to greater impatience. Data are at the individual level. The sample in each row is the subset of participants we have each impatience measure for. Significance levels: * 10%, ** 5%, *** 1%.

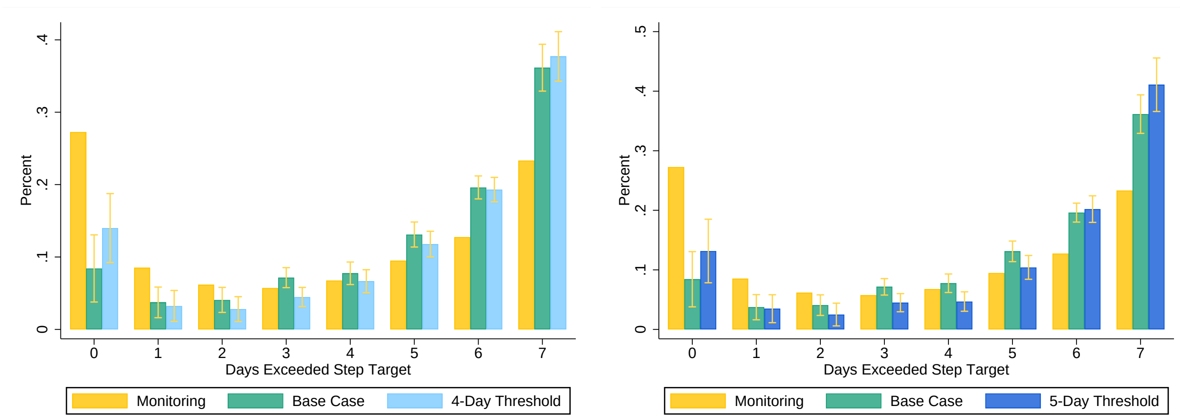
Table C.3 shows that there is no significant or meaningful correlation between any of the measures of impatience over effort and impatience over payment collected in the experimental sample. Similarly, we find that the correlation of the individual-level averages of the recharge impatience and effort impatience measures in the validation sample is only -0.05, which is statistically indistinguishable from 0 (p-value = 0.66).

As discussed in Section C.2, pay cycle effects also measure impatience over payments. Thus, we can test whether participants' impatience over payment relates to our measures of impatience over effort by testing whether effort impatience measures predict pay cycle effects.

Panel A of Appendix Figure F.2 shows that there are no meaningful payday spikes even among those with above-median impatience index. Moreover, the patterns across the pay cycle are very similar for those with below-median impatience, depicted in Panel B. Results are similar for the other measures of effort impatience (i.e., the predicted impatience index, demand for commitment, and simple CTB). Regression analysis confirms that there are no large or significant differences in pay cycle effects across any measure of effort impatience.

D Distributional Impacts of Thresholds

This section assesses the effect of thresholds on the distributions of weekly and intervention-average compliance. We first assess whether thresholds decrease intermediate effort just below the threshold. Panels (a) and (b) of Figure D.1 show histograms at the individual × week level of the number of days the individual met their step target in that week, for the 4-day or 5-day threshold group, respectively, relative to Base Case and Monitoring (confidence intervals are relative to Monitoring). Indeed, the threshold contracts do modestly decrease effort just below the threshold: the prevalence of walking 3 or 4 days is lower in 5-Day Threshold than either Base Case (p-value <0.001) or Monitoring (p-value = 0.008), and the prevalence of walking 2 or 3 days is lower in 4-Day Threshold than either reference group (p-values <0.001 for both Base Case and Monitoring).[66] Figure D.2 shows similar patterns for the subsets of people with above- and below-median impatience, showing that the overall distributional patterns we see are not predominantly explained by impatience.



(a) 4-Day Threshold

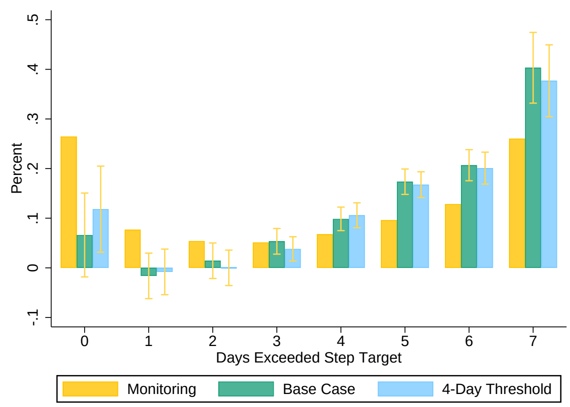
(b) 5-Day Threshold

Appendix Figure D.1: Thresholds Modestly Decrease Compliance Right Below the Threshold

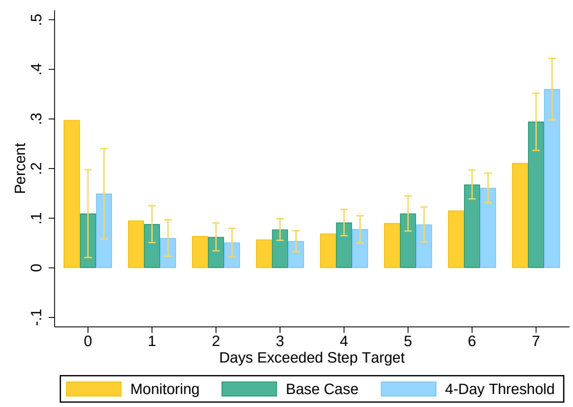
Notes: Figures show histograms of the number of days a participant exceeded the step target each week during the intervention period in the Base Case, 4-Day or 5-Day Threshold, and Monitoring. Data are at the respondent-week level. Confidence intervals represent a test of equality between Monitoring and each other group from regressions with the same controls as Table 2 except for day-of-week fixed effects (because data are weekly).

However, the magnitude of these differences are relatively small (especially compared to the differences from Monitoring), leading to only slight differences between Base Case and Threshold in the overall distribution of weekly compliance. Specifically, Panels (a) and (b) of Figure D.3 show the cumulative distribution functions (CDFs) of weekly compliance in 4-Day and 5-Day Threshold, respectively, relative to Base Case and Monitoring. While the distributions of weekly compliance in Base Case and both threshold groups all differ markedly from the distribution in Monitoring, the differences between Base Case and the threshold groups are small. Panels (c) and (d) of Figure D.3 shows similar results for the distribution of individual-level (instead of individual

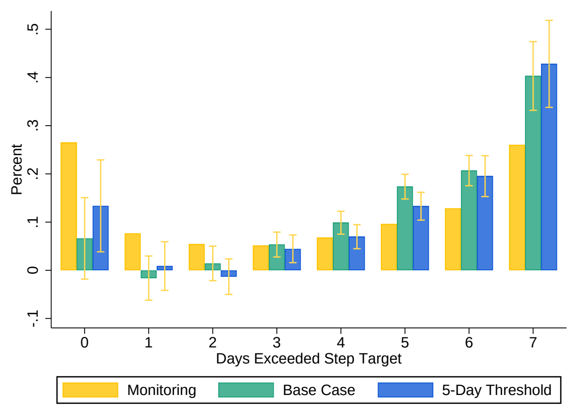
[66]Notably, neither threshold increases the likelihood of walking exactly the threshold number of days. Our model suggests this may reflect that the contracts pay for above-threshold compliance (e.g., the 4-day threshold pays for the 5th day of compliance). Additional explanations outside of the model include habit formation or that it is easier to schedule walking every day in a given week than on a subset of days.



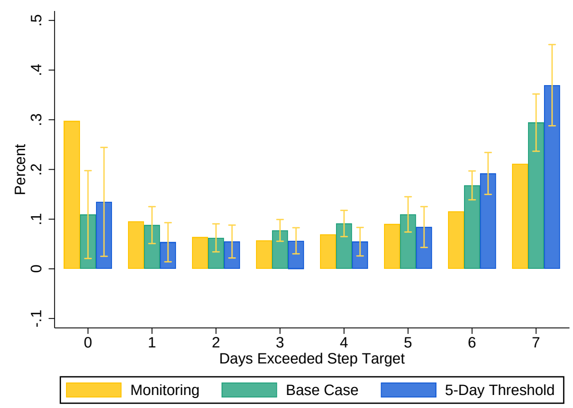
(a) 4-Day Threshold, Below-median Impatience



(b) 4-Day Threshold, Above-median Impatience



(c) 5-Day Threshold, Below-median Impatience

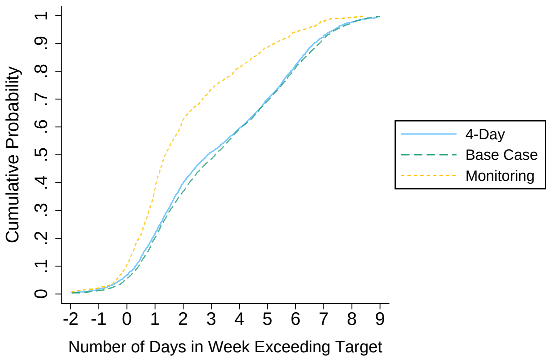


(d) 5-Day Threshold, Above-median Impatience

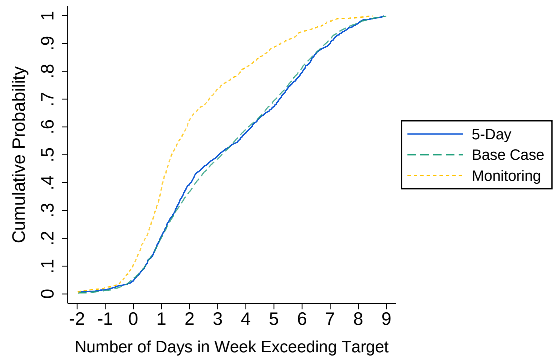
Appendix Figure D.2: Thresholds Modestly Decrease Compliance Right Below the Threshold

Notes: As in Figure D.1, Figures show histograms of the number of days a participant exceeded the step target each week during the intervention period in the Base Case, 4-Day or 5-Day Threshold, and Monitoring, but here we split the sample into below-median impatience index in Panels (a) and (c), and above-median impatience index in Panels (b) and (d). Data are at the respondent-week level. Confidence intervals represent a test of equality between Monitoring and each other group from regressions with the same controls as Table 2 except for day-of-week fixed effects (because data are weekly).

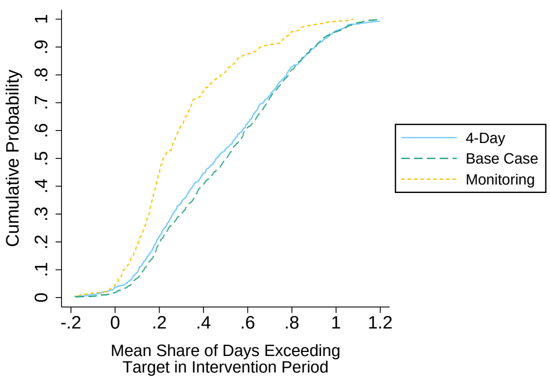
× week-level) compliance. Quantile regressions reveal no significant differences between the threshold groups and Base Case in the 25th, 50th, or 75th percentiles of the distributions of either individual × week-level or individual-level compliance (see Appendix Table F.7). Kolmogorov-Smirnov (KS) tests for the equivalence of the individual-level distributions also fail to reject the null of equal distributions (p-values 0.238 and 0.852 for the 4- and 5-Day Threshold, respectively, relative to Base Case).



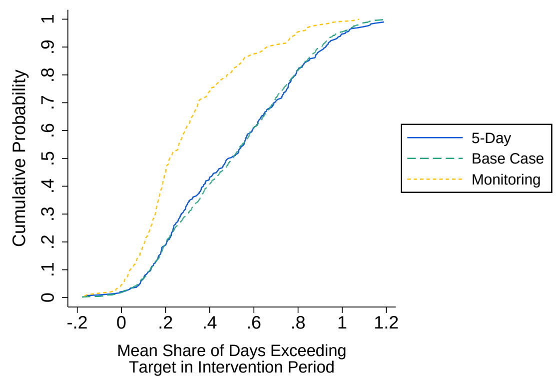
(a) 4-Day Threshold, Weekly Compliance



(b) 5-Day Threshold, Weekly Compliance



(c) 4-Day Threshold, Overall Compliance



(d) 5-Day Threshold, Overall Compliance

Appendix Figure D.3: Threshold and Base Case Have Similar Compliance Distributions

Notes: This figure shows the cumulative distribution functions (CDFs) of the distributions of weekly compliance (i.e., the number of days the individual exceeded the step target in a week) in Panels (a) and (b), and intervention-average compliance (i.e., the percentage of days the individual exceeded the step target during the intervention period) in Panels (c) and (d). All CDFs are plotted separately by treatment group for the monitoring, base case, 4-day (Panels (a) and (c)), and 5-day (Panels (b) and (d)) threshold groups. For Panels (a) and (b), data are at the individual × week level, limited to weeks where the individual has at least 4 days of data. For Panels (c) and (d), data are at the individual level, limited to individuals who had at least 21 days of data over the 12-week intervention period. Both weekly and intervention-average compliance are residualized using the same controls as in Table 2 except that we do not include day-of-week fixed effects because data are not at the day level.

E Predicting Impatience with Policy Variables

This appendix provides proof of concept that a policymaker could use hard-to-manipulate observable characteristics to predict impatience and effectively target the threshold contract.

Our Section 5.3 results suggest that a policymaker could improve our program's effectiveness by targeting threshold contracts only to more impatient individuals. However, impatience is challenging to observe; even were policymakers to field surveys on impatience, participants might game their responses to avoid a specific contract—especially a financially dominated one.

To address this concern, we construct a "policy prediction" of impatience: a prediction of the impatience index using demographics (e.g., age, labor force participation) and medical information (e.g., HbA1c, fatigue) that health policymakers would likely have access to. We show that there is significant heterogeneity in the effect of the threshold by the policy prediction. Hence, the policy prediction could be used to personalize contract assignment.

To prevent overfitting, we use a split sample approach. First, in a randomly-selected training sample, we fit a LASSO model to predict the impatience index with the variables listed in the Table E.1 notes. We then use the model to predict impatience out of sample for all other participants (the "regression sample"). Finally, in the regression sample, we estimate the heterogeneity in Threshold performance by the policy prediction using equation (6). To sufficiently power this regression, we allocate 2/3 of participants to the regression sample.

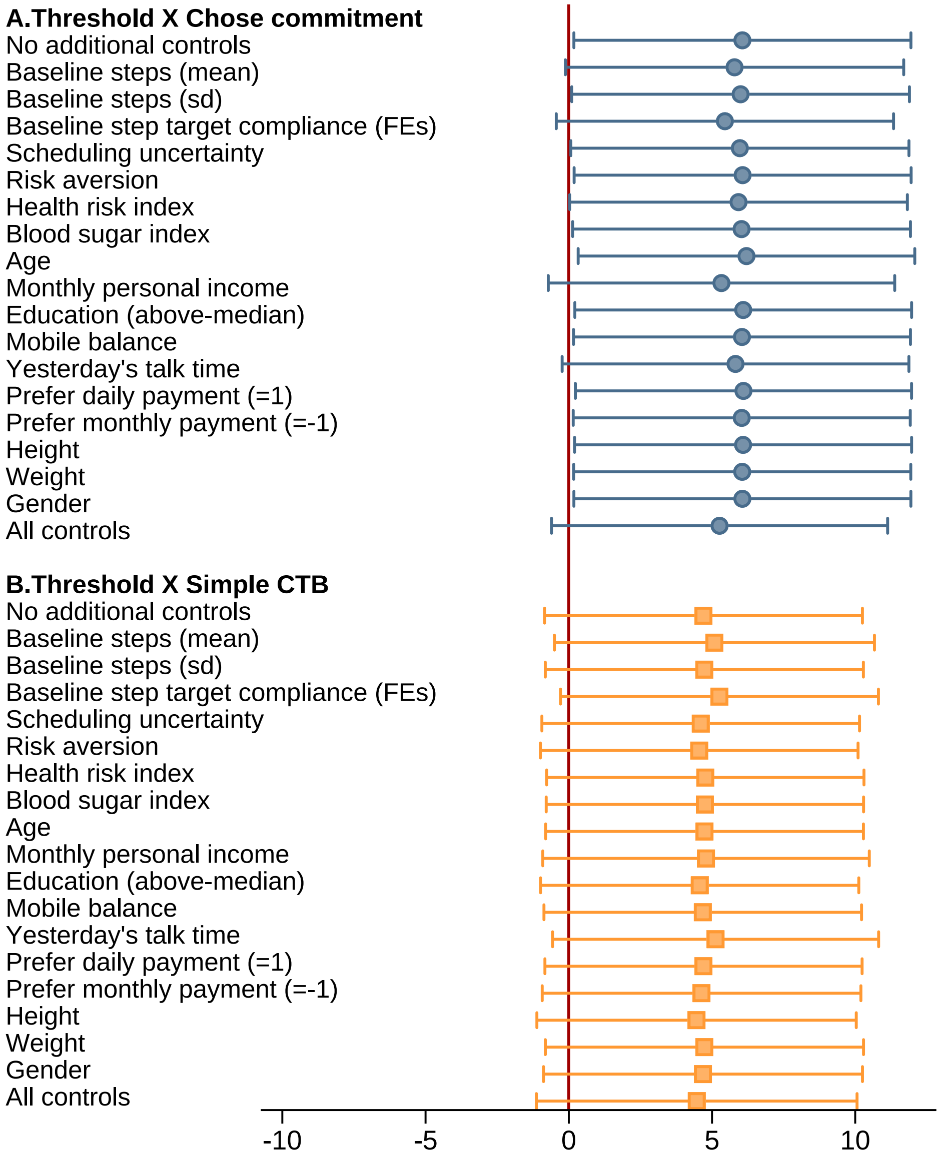
The results, in Table E.1, are similar to Table 3: Threshold has a higher treatment effect among people with higher predicted impatience. This suggests that personalizing thresholds using a policy prediction could significantly improve the effectiveness of incentives at scale.

Appendix Table E.1: Threshold Effect Varies with Policy Prediction of Impatience

Dependent variable:	Exceeded step target	
	(1)	(2)
Impatience × Threshold	0.03**	0.06**
	[0.00, 0.06]	[0.00, 0.12]
Threshold	-0.01	-0.03**
	[-0.04, 0.02]	[-0.07, -0.00]
Impatience	-0.02**	-0.05**
	[-0.04, -0.00]	[-0.09, -0.01]
Impatience measure:	Policy prediction	Above-median policy prediction
Base Case mean	.502	.502
# Individuals	1,969	1,969
# Observations	157,946	157,946

Notes: This table replicates Table 3 with an impatience index predicted out-of-sample with the following variables (and their interactions with above-median age, gender, and individual and household income): age; gender; labor participation; personal and household monthly income; household size; HbA1c; RBS; systolic and diastolic BP; BMI; waist circumference; walking speed; diagnosed diabetic or hypertensive; overweight; owns home; number of rooms and running water in home; has a bank account; hired help; number of scooters, cars, computers, smartphones, and mobile phones; mobile balance; hours of work on a weekday; consumes alcohol and cigarettes/bidis; has foot ulcer, rapid deterioration in eyesight, and pain or numbness in legs or feet. Significance levels: * 10%, ** 5%, *** 1%.

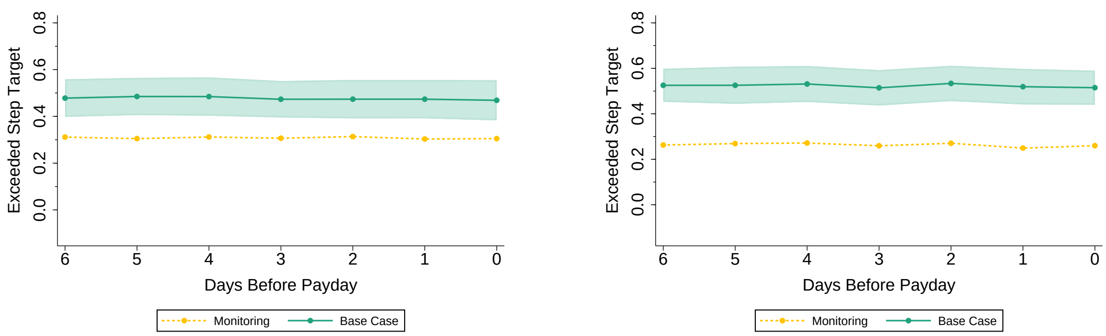
F Additional Tables and Figures



Appendix Figure F.1: Threshold Heterogeneity in Choosing Threshold and Choosing to Walk Later Is Robust to a Variety of Controls

Notes: This figure replicates Figure 3 using different impatience measures. Panel A uses demand for commitment and Panel B uses simple CTB. See the notes to Table 3 for more detail on these impatience measures. 95% confidence intervals are based on standard errors clustered at the individual level. All other details are the same as in Figure 3; see Figure 3 notes for more details.

Appendix Figure F.2: No Heterogeneity by Impatience in Compliance Pattern Across the Paycycle



(a) Above Median Impatience Index
(b) Below Median Impatience Index

Notes: The figures show the probability of exceeding the daily 10,000-step target for the base case relative to the monitoring group, according to days remaining until payday. Each Panel is limited to above/below-median values of the impatience index. Effects control for payday day-of-week fixed effects, day-of-week fixed effects, day-of-week relative to survey day-of-week fixed effects, and the same controls as in Table 2. The shaded area represents a collection of confidence intervals from tests of equality within each daily period between the incentive and monitoring groups from regressions with the same controls as in Table 2. p-values for the test that the payday spikes are equal across above/below-median samples for each impatience measure are: Impatience index: 0.462; Predicted impatience index: 0.803; Chose commitment: 0.647; Simple CTB: 0.100.

Appendix Table F.1: Participants Understood Their Assigned Contracts

Contract Type	Question	% Correct		
		At Contract Launch	First Call	Any Call
Base Case n=902	How many recharges would you receive on (*payment day of week*) if you walked 10,000 steps on exactly 1 day over the period (*payment day of week*) to (*payment day of week*–1)?	0.99	0.99	1.00
	How many times over the course of this week would you receive recharges if you walked 10,000 steps on exactly 5 days over the period (*payment day of week*) to (*payment day of week*–1)?	1.00	1.00	1.00
4-Day Threshold n=794	What is the minimum number of days that you need to walk to get a recharge?	-	0.90	0.95
	How many recharges would you receive at the end of this week if you walked 10, 000 steps on exactly 1 day this week?	0.93	0.98	1.00
	How many recharges would you receive at the end of this week if you walked 10, 000 steps on exactly 4 days this week?	0.99	0.99	1.00
	How many recharges would you receive at the end of this week if you walked 10, 000 steps on exactly 6 days this week?	0.98	1.00	1.00
5-Day Threshold n=312	What is the minimum number of days that you need to walk to get a recharge?	-	0.88	0.93
	How many recharges would you receive at the end of this week if you walked 10, 000 steps on exactly 1 day this week?	0.91	0.96	0.99
	How many recharges would you receive at the end of this week if you walked 10, 000 steps on exactly 5 days this week?	0.98	0.99	0.99
	How many recharges would you receive at the end of this week if you walked 10, 000 steps on exactly 6 days this week?	0.99	0.99	0.99
Daily n=166	How many times over the course of this week would you receive recharges if you walked 10, 000 steps on exactly 1 day ?	1.00	0.98	1.00
	How many times over the course of this week would you receive recharges if you walked 10, 000 steps on exactly 5 days ?	-	0.99	0.99
Monthly n=164	How many recharges would you receive on (*payment day of week*) if you walked 10, 000 steps on exactly 1 day over this week ?	0.99	0.99	1.00
	How many recharges would you receive on (*payment day of week*) if you walked 10, 000 steps on exactly 5 days in this week ?	1.00	1.00	1.00
Monitoring n=203	How do you report your steps to us?	1.00	0.99	1.00
	How large is the Fitbit wearing bonus?	-	0.78	0.99

Notes: This table shows the share of participants who correctly answered questions about their contract. Participants were initially asked these questions when contracts were first explained ("At Contract Launch"). Questions were asked again over the phone at a later date ("First Call"). Those who answered questions incorrectly were asked again in two subsequent follow-up calls. The "Any Call" column represents the proportion of participants who got the questions right at any of these phone calls. Some questions were not asked at the initial contract launch phase. Each participant in the monthly, base case, and threshold groups was always paid on the same day of the week, which is labeled *"payment day of week"*.

Appendix Table F.2: Threshold Heterogeneity Results are Robust to Ways of Constructing the "Chose Commitment" and "Simple CTB" Measures

Dependent variable:	Exceeded step target (×100)							
Impatience measure:	Chose commitment					Simple CTB		
	Average	Either	Both	4-Day	5-Day	Average	Either	Both
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Impatience × Threshold	6.06**	5.31*	5.74**	6.32**	4.75	4.70*	5.16*	3.82
	[0.18, 11.94]	[-0.33,10.95]	[0.07,11.41]	[0.16, 12.48]	[-3.55, 13.05]	[-0.84, 10.25]	[-0.27, 10.60]	[-1.60, 9.25]
Base Case mean	49.9	49.8	49.9	50	49.8	50.2	50.2	50.2
# Individuals	1,798	1,809	1,798	1,523	1,097	1,967	1,967	1,967
# Observations	144,099	145,005	144,099	122,277	87,990	157,799	157,799	157,799

Notes: This table shows robustness of results in columns 5 and 6 of Table 3 to different ways of constructing the Chose Commitment and Simple CTB variables. For Chose Commitment, "average" is the main specification in Table 3 and is the average of preference for 4-day and 5-day threshold contracts versus the linear contract. "Either" means preferring either 4-day or 5-day threshold, and "Both" means preferring both threshold contracts. "4-day" and "5-day" only look at the preference for 4-day and 5-day threshold respectively. For "Simple CTB", "Average" is the main specification and is the average between choosing to walk more earlier in two CTB-style walking choices, "Either" means choosing to walk earlier in either choice and "Both" means choosing to walk earlier in both choices. Controls are the same as in Table 2. The sample includes the base case and threshold groups. Data are at the individual × day level. 95% confidence intervals, in brackets, are constructed using standard errors clustered at the individual level. Significance levels: * 10%, ** 5%, *** 1%.

Appendix Table F.3: Lee Bounds on the Impacts of Incentives on Exercise

Definition of missing:	No steps data	Did not wear Fitbit	No data from Fitbit	Lost data entire period	Withdrew immediately	Mid-period withdrawal	Other reasons
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
A. Daily steps							
Regression estimate	1269	1269	1338	1338	1338	1338	1338
(conditional on nonmissing data)	[245]	[245]	[261]	[261]	[261]	[261]	[261]
Lee lower bound	1053	882	1230	1315	1297	1226	1303
	[290]	[234]	[268]	[242]	[280]	[218]	[288]
Lee upper bound	1426	1571	1572	1351	1430	1581	1358
	[349]	[306]	[312]	[243]	[288]	[213]	[285]
B. Met 10k step target							
Regression estimate	0.223	0.223	0.205	0.205	0.205	0.205	0.205
(conditional on nonmissing data)	[0.024]	[0.024]	[0.022]	[0.022]	[0.022]	[0.022]	[0.022]
Lee lower bound	0.215	0.208	0.200	0.204	0.203	0.200	0.204
	[0.024]	[0.025]	[0.022]	[0.025]	[0.021]	[0.026]	[0.021]
Lee upper bound	0.232	0.242	0.216	0.206	0.209	0.217	0.206
	[0.033]	[0.025]	[0.023]	[0.024]	[0.019]	[0.026]	[0.022]
# Individuals	2,607	2,559	2,607	2,568	2,598	2,561	2,566
# Observations	218,988	205,732	218,988	206,488	209,008	211,551	206,320

Notes: This table reports regression estimates and Lee bounds estimates (accounting for different types of missing pedometer data) of the effect of Incentives relative to Monitoring on exercise during the intervention period. Standard errors in parentheses. The regression estimates and Lee bounds condition on data not being missing, using different definitions of missing data in each column. Regression estimates are not comparable to those reported in Table 2 because each column conditions on the "type of missing" indicator in the first row being equal to 0 and does not include controls. Data are at the individual × day level.

Appendix Table F.4: Summaries From Minute-Level Pedometer Data

	Incentives	Monitoring	I - M	*p*-value: I=M
	(1)	(2)	(3)	(4)
A. Activity (by minute)				
Average daily activity	213	197	17	0.001
Average steps per minute	41	38	3	0.001
B. Time of day				
Average start time	07:11	07:16	5	0.441
Average end time	20:49	20:50	1	0.742
C. High step counts per minute (share)				
Steps > 242	0	0	0	.
Steps > 150	0	0	0	0.322
# Individuals	2,368	200		

Notes: This table presents various statistics at the respondent × minute level in the incentive and monitoring groups for the days on which minute-by-minute data were available (typically 10 days of minute-wise data prior to each sync). "Average daily activity" is the average number of minutes in which a step was recorded each day. "Average steps per minute" is the average steps per minute in which at least one step was recorded. Average start/end time is the average time the first/last step was recorded by the Fitbit on that day. The "High step counts per minute (share)" variables are the share of days on which we recorded steps-per-minute over the stated thresholds. High step count thresholds (242 and 150) were determined based on the average number of steps an individual takes when running at 5 mph and 8 mph, respectively. Only one individual's minute-by-minute data coincide with jogging at a pace greater than 5 miles per hour, and only for a total of 15 minutes over one day in the intervention period.

Appendix Table F.5: HbA1c and RBS Independently Predict One Another

Dependent variable:	Endline HbA1c	Endline RBS
	(1)	(2)
Baseline HbA1c (SDs)	0.60***	0.33***
	[0.045]	[0.057]
Baseline RBS (SDs)	0.25***	0.37***
	[0.044]	[0.055]
# Individuals	560	561

Notes: This table reports estimates from regressing standardized HbA1c (column 1) and standardized RBS (column 2) at endline on standardized HbA1c and standardized RBS at baseline. Standard errors in parentheses. The sample is the control group only. Data are at the individual level. No additional controls are included. Significance levels: * 10%, ** 5%, *** 1%

Appendix Table F.6: Table 2 Results Robust to Different Controls

	No controls				Stratum fixed effects				Lasso-selected controls			
Dependent variable:	Exceeded step target	Daily steps	Daily steps (if > 0)	Earned payment when target met	Exceeded step target	Daily steps	Daily steps (if > 0)	Earned payment when target met	Exceeded step target	Daily steps	Daily steps (if > 0)	Earned payment when target met
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
A. Pooled incentives												
Incentives	0.205***	1337.6***	1271.4***	0.950***	0.200***	1263.7***	1158.0***	0.952***	0.196***	1287.1***	1144.2***	0.952***
	[0.0224]	[261.1]	[246.1]	[0.00231]	[0.0185]	[208.7]	[188.1]	[0.00309]	[0.0180]	[211.4]	[190.3]	[0.00282]
B. Unpooled incentives												
Base Case	0.208***	1356.6***	1208.8***	1.000***	0.210***	1386.2***	1199.1***	1.006***	0.207***	1411.4***	1197.0***	1.005***
	[0.0241]	[277.0]	[258.6]	[1.62e-13]	[0.0201]	[222.0]	[199.4]	[0.00267]	[0.0196]	[225.0]	[201.8]	[0.00223]
Threshold	0.207***	1337.9***	1315.2***	0.890***	0.198***	1214.7***	1139.8***	0.892***	0.194***	1238.0***	1125.3***	0.892***
	[0.0240]	[277.1]	[259.3]	[0.00505]	[0.0199]	[220.8]	[198.0]	[0.00547]	[0.0194]	[223.2]	[200.3]	[0.00533]
Daily	0.207***	1202.7***	1363.9***	1.000***	0.200***	1120.7***	1279.2***	1.003***	0.199***	1126.7***	1245.0***	1.003***
	[0.0345]	[389.5]	[346.0]	[2.02e-13]	[0.0303]	[331.0]	[277.3]	[0.00365]	[0.0302]	[332.2]	[279.2]	[0.00296]
Monthly	0.198***	1568.6***	1482.3***	1.000***	0.177***	1265.7***	1174.2***	1.002***	0.179***	1302.6***	1152.4***	1.000***
	[0.0348]	[393.8]	[365.4]	[3.52e-13]	[0.0288]	[307.4]	[270.2]	[0.00335]	[0.0281]	[311.0]	[272.3]	[0.00271]
Small Payment	0.147***	820.5	658.5	1.000***	0.137***	728.1*	549.8	1.000***	0.128***	740.7*	510.6	0.999***
	[0.0485]	[524.0]	[477.9]	[5.58e-14]	[0.0383]	[386.1]	[334.9]	[0.00499]	[0.0382]	[381.0]	[331.3]	[0.00417]
p-value for Base Case vs												
Daily	0.981	0.634	0.571	.	0.706	0.347	0.725	0.335	0.786	0.314	0.834	0.563
Monthly	0.758	0.519	0.359	.	0.180	0.634	0.909	0.161	0.266	0.671	0.839	0.022
Threshold	0.982	0.914	0.479	<0.001	0.357	0.212	0.617	<0.001	0.354	0.210	0.548	<0.001
Small Payment	0.175	0.261	0.199	.	0.038	0.057	0.028	0.200	0.027	0.047	0.018	0.154
Monitoring mean	0.294	6,774	7,986	0	0.294	6,774	7,986	0	0.294	6,774	7,986	0
# Individuals	2,559	2,559	2,557	2,394	2,559	2,559	2,557	2,394	2,559	2,559	2,557	2,394
# Observations	205,732	205,732	180,018	99,406	205,732	205,732	180,018	99,406	205,732	205,732	180,018	99,406

Notes: This table replicates the Table 2 estimates with different sets of controls. Columns 1–4 do not use controls, columns 5–8 use the same controls as in 2 along with stratum fixed effects, and columns 9–12 use controls selected by double-LASSO. We allow LASSO to select from the following list of controls: female, age, age squared, weight, weight squared, indicator for missing weight, height, height squared, indicator for missing height, yearmonth and day of week fixed effects. In addition, column 9 controls for the number of days in phase-in the target was met, its square, and an SMS treatment indicator. Columns 10–12 control for average baseline steps, average baseline steps squared, an indicator for missing baseline steps, and an SMS treatment indicator. See the notes for Table 2 for more information. Significance levels: * 10%, ** 5%, *** 1%.

Appendix Table F.7: Quantile Regression Estimates Show That the Linear and Threshold Contracts Similarly Impact the Distribution of Individual-Level and Weekly Compliance

Dependent variable:	Share of days met step target in intervention period			Share of days met step target in week		
Percentile:	25	50	75	25	50	75
	(1)	(2)	(3)	(4)	(5)	(6)
5-Day Threshold	0.108***	0.238***	0.353***	0.105***	0.228***	0.351***
	[0.024]	[0.031]	[0.047]	[0.020]	[0.035]	[0.053]
4-Day Threshold	0.093***	0.206***	0.323***	0.092***	0.214***	0.327***
	[0.021]	[0.026]	[0.044]	[0.017]	[0.025]	[0.051]
Base Case	0.116***	0.245***	0.336***	0.111***	0.246***	0.328***
	[0.020]	[0.025]	[0.042]	[0.016]	[0.024]	[0.051]
p-value: 5-Day vs 4-Day	.48	.28	.29	.46	.69	.25
p-value: 5-Day vs Base Case	0.705	0.791	0.523	0.712	0.575	0.266
p-value: 4-Day vs Base Case	0.201	0.105	0.560	0.152	0.157	0.948
Monitoring mean	0.292	0.469	0.723	0.294	0.523	0.801
# Individuals	2,133	2,133	2,133	2,168	2,168	2,168
# Observations	2,133	2,133	2,133	24,864	24,864	24,864

Notes: This table shows quantile regressions where the dependent variable is the share of days a participant met their step target in a given week (columns 1–3) or during the intervention period (columns 4–6). Data in columns 1–3 are at the individual × week level; in columns 4–6 they are at the individual level. The sample includes the base case, threshold, and monitoring groups. Controls are the same as in Table 2, except that, because the data are not at the individual × day level, we do not include day-of-week fixed effects. Also, in columns 1–3 we include year-month fixed effects for the first year-month of the intervention period, and in columns 4–6, we include year-month fixed effects for the first year-month of the week. Significance levels: * 10%, ** 5%, *** 1%.

Appendix Table F.8: Threshold Heterogeneity in Chose Commitment Is Robust to Ways of Handling "No Preference" Responses

Dependent variable:	Exceeded step target (×100)			
	Excluding no preference	No preference as Threshold	No preference as Base Case	No preference as separate
	(1)	(2)	(3)	(4)
Impatience × Threshold	6.06**	5.67**	5.52*	6.13**
	[0.18, 11.94]	[0.06, 11.29]	[-0.12, 11.16]	[0.25, 12.02]
Base Case mean	49.9	50.2	50.2	50.2
# Individuals	1,798	1,969	1,969	1,969
# Observations	144,099	157,946	157,946	157,946

Notes: This table shows robustness of results in column 5 of Table 3 to different ways of handling participants with no preference between the 4- or 5-day threshold and base case contract. Column 1 uses the same specification as in column 5 of Table 3 by counting no preference as missing. Column 2 counts no preference as choosing Threshold and column 3 counts no preference as choosing Base Case. Column 4 counts no preference as a separate group by adding a dummy and its interaction with the indicator for threshold treatment. Controls are as in Table 2. 95% confidence intervals, in brackets, are constructed using standard errors clustered at the individual level. Data are at the individual × day level. The sample includes the threshold and base case groups. Significance levels: * 10%, ** 5%, *** 1%.

Appendix Table F.9: Threshold Heterogeneity Results Similar with Steps as Outcome or When Analyze Threshold Groups Separately

Impatience measure:	Impatience index	Above-median impatience index	Predicted impatience index	Above-median predicted index	Chose commitment	Simple CTB
	(1)	(2)	(3)	(4)	(5)	(6)
A. Dependent variable = steps						
Impatience × Threshold	289	576	238*	521**	580*	157
	[-89, 667]	[-167,1319]	[-36, 538]	[22, 1166]	[-9, 1168]	[-400, 715]
Threshold	-143	-401	-166	-360**	-457**	-253
	[-513, 228]	[-925, 124]	[-471, 121]	[-704, -28]	[-876, -37]	[-658, 151]
Impatience	-209	-444	-229**	-549***	-248	-41
	[-473, 54]	[-989, 101]	[-454, -25]	[-1079, -187]	[-671, 175]	[-444, 362]
Base Case mean	8,098	8,098	8,131	8,131	8,091	8,131
B. Dependent variable = exceeded step target (×100)						
Impatience × 5-Day Threshold	3.52*	6.00	3.66**	7.29***	6.10*	3.76
	[-0.54, 7.57]	[-2.04,14.04]	[0.70, 6.73]	[2.14, 15.46]	[-0.33, 12.52]	[-2.28, 9.80]
5-Day Threshold	-1.72	-4.31	-1.71	-4.42**	-4.92**	-3.79*
	[-5.71, 2.27]	[-9.94,1.32]	[-4.64, 1.46]	[-8.37, -0.97]	[-9.46, -0.39]	[-8.05, 0.46]
Impatience × 4-Day Threshold	5.00	7.95	1.76	2.51	6.20	7.10*
	[-1.05, 11.06]	[-3.14,19.03]	[-2.13, 5.77]	[-5.46, 11.13]	[-2.50, 14.90]	[-1.05, 15.25]
4-Day Threshold	-0.14	-3.71	0.17	-0.84	-2.81	-3.76
	[-5.66, 5.39]	[-11.19,3.78]	[-3.92, 4.66]	[-5.78, 4.37]	[-8.88, 3.27]	[-9.93, 2.41]
Impatience	-2.97**	-5.03*	-2.39**	-5.32***	-2.37	-2.68
	[-5.46, -0.48]	[-10.45,0.39]	[-4.52, -0.54]	[-10.84, -1.99]	[-6.66, 1.91]	[-6.73, 1.37]
Base Case mean	50.4	50.4	50.2	50.2	49.9	50.2
# Individuals	1,075	1,075	1,969	1,969	1,798	1,967
# Observations	86,215	86,215	157,946	157,946	144,099	157,799

Notes: Panel A shows that the Threshold heterogeneity reported in Table 3 is robust to using daily steps as the outcome. Panel B shows heterogeneity in the 4- and 5-day threshold treatments by impatience. The impatience measure changes across columns; its units in columns 1 and 3 are standard deviations. The sample includes the base case and threshold groups only. Specifications in columns 1 and 2 include only participants who were enrolled after we started measuring the impatience index; columns 3–6 include everyone. Threshold pools 4- and 5-day Thresholds. 95% confidence intervals are shown in brackets. For columns 1–2 and 5–6, confidence intervals are based on standard errors clustered at the individual level. For columns 3 and 4, which use the predicted impatience index, confidence intervals are constructed using bootstrap, with bootstrap draws clustered at the individual level. See the notes to Table 3 for a detailed description of the bootstrap procedure. Controls are the same as in Table 2. Data are at the individual × day level. Significance levels: * 10%, ** 5%, *** 1%.

Appendix Table F.10: Walking Does Not Vary Significantly Across the Pay Cycle

Dependent variable:	Exceeded step target (×100)				
Payment frequency:	Weekly		Monthly		
	(1)	(2)	(3)	(4)	(5)
Days before payday	0.11 [0.09]		0.11 [0.09]		
Payday		-0.63 [0.55]		-0.63 [0.55]	
Payweek					-0.12 [1.02]
Sample mean	50.19	50.19	50.19	50.19	49.28
# Individuals	890	890	890	890	163
# Observations	71,672	71,672	71,672	71,672	13,333

Notes: The columns show the effect of days until payday on the probability of meeting the step target in the base case and monthly groups; the sample in columns 1 and 2 is restricted to the base case group, and the sample in columns 3–5 is restricted to the monthly group. We control for payday day-of-week fixed effects, day-of-week fixed effects, day-of-week relative to launch survey day-of-week fixed effects, a day-of-contract-period time trend, and the same controls as in Table 2. Data are at the individual × day level. Standard errors, in brackets, are clustered at the individual level. Significance levels: * 10%, ** 5%, *** 1%.

Appendix Table F.11: Effect of Incentives on BMI, Blood Pressure, and Waist Circumference

Sample:	Full sample effects			Above-median baseline blood sugar effects		
	Body mass index	Mean arterial BP	Waist circumference	Body mass index	Mean arterial BP	Waist circumference
	(1)	(2)	(3)	(4)	(5)	(6)
Incentives	-0.0525 [0.0409]	0.0884 [0.426]	-0.211 [0.284]	0.0195 [0.0570]	-0.0811 [0.605]	-0.275 [0.396]
Monitoring	0.0657 [0.0838]	1.121 [0.739]	-0.0352 [0.438]	0.00127 [0.0830]	-0.478 [1.083]	0.345 [0.590]
Sample	Full	Full	Full	Above-median blood sugar	Above-median blood sugar	Above-median blood sugar
Control mean	26.45	103.02	94.44	26.09	103.96	94.57
# Individuals	3,058	3,056	3,059	1,527	1,529	1,525

Notes: This table shows the effect of incentives on the endline components of the health risk index not included in Table 4. Columns 4–6 restricts to the above-median blood sugar index sample. The blood sugar index is constructed as in Table 4. Controls are as described in Table 4 notes. The sample includes the incentive, monitoring, and control groups. Data are at the individual level. Standard errors are in brackets. Significance levels: * 10%, ** 5%, *** 1%.

Appendix Table F.12: Impacts of Incentives on Health, Robustness to Different Controls

	Full sample effects				Above-median baseline blood sugar sample effects			
	Blood sugar index	HbA1c	Random blood sugar	Health risk index	Blood sugar index	HbA1c	Random blood sugar	Health risk index
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A. No controls								
Incentives	-0.044	-0.068	-5.53	-0.055	-0.092*	-0.15	-11.4*	-0.13**
	[0.043]	[0.11]	[4.37]	[0.047]	[0.053]	[0.14]	[6.12]	[0.060]
Monitoring	0.0073	-0.078	4.62	0.058	-0.070	-0.30	-1.35	-0.11
	[0.074]	[0.19]	[7.94]	[0.078]	[0.088]	[0.22]	[10.8]	[0.10]
p-value: I = M	0.435	0.952	0.153	0.102	0.770	0.463	0.294	0.803
Control mean	0.00	8.44	193.83	0.00	0.64	10.09	248.26	0.45
# Individuals	3,067	3,066	3,067	3,068	1,530	1,529	1,530	1,531
Panel B. Stratum fixed effects								
Incentives	-0.05*	-0.07	-6.70*	-0.05*	-0.10**	-0.13	-14.01**	-0.09**
	[0.03]	[0.07]	[3.44]	[0.03]	[0.05]	[0.12]	[5.85]	[0.04]
Monitoring	-0.02	-0.14	2.10	0.02	-0.06	-0.31	-0.24	-0.05
	[0.05]	[0.12]	[6.36]	[0.04]	[0.08]	[0.19]	[10.37]	[0.07]
p-value: I = M	0.492	0.515	0.124	0.120	0.576	0.278	0.138	0.546
Control mean	0.00	8.44	193.83	0.00	0.64	10.09	248.26	0.45
# Individuals	3,067	3,066	3,067	3,068	1,530	1,529	1,530	1,531
Panel C. Lasso-selected controls								
Incentives	-0.05**	-0.08	-6.04*	-0.05*	-0.10**	-0.15	-11.95**	-0.08**
	[0.03]	[0.07]	[3.52]	[0.02]	[0.05]	[0.12]	[5.90]	[0.04]
Monitoring	-0.03	-0.14	1.29	0.01	-0.07	-0.33*	0.85	-0.05
	[0.05]	[0.12]	[6.61]	[0.04]	[0.08]	[0.20]	[10.48]	[0.07]
p-value: I = M	0.517	0.573	0.220	0.129	0.631	0.306	0.170	0.553
Control mean	0.00	8.44	193.83	0.00	0.64	10.09	248.26	0.45
# Individuals	3,067	3,066	3,067	3,068	1,530	1,529	1,530	1,531

Notes: This table reports the results of the specifications displayed in Table 4 with different controls. Panel A include no controls, Panel B include the same controls as 4 along with stratum fixed effects, Panel C include controls selected by double-LASSO. We allow LASSO to select from the following list of controls: female, age, age squared, weight, weight squared, weight missing indicator, height, height squared, height missing indicator, completed endline survey indicator, and date and hour of endline completion fixed effects. Panel C also control for the baseline value of the outcome (or index components for indices), along with an SMS treatment indicator. Standard errors are in brackets. Data are at the individual level. The sample includes the incentive, monitoring, and control groups. p-value: I = M is the p-value for incentives vs monitoring. See Table 4 for more information on outcome variables and controls. Significance levels: * 10%, ** 5%, *** 1%.

Appendix Table F.13: Impact of Incentives on Fitness and Mental Health

A. Mental Health	Mental health index	Felt happy	Less nervous	Peaceful	Energy	Less blue	Less worn	Less harm to social life
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Incentives	0.095**	0.088*	0.026	0.054	0.062	0.016	0.090**	0.053
	[0.045]	[0.045]	[0.044]	[0.047]	[0.048]	[0.044]	[0.042]	[0.032]
Monitoring	0.16**	0.074	0.13	0.095	0.032	0.13*	0.17***	0.049
	[0.073]	[0.075]	[0.077]	[0.083]	[0.082]	[0.075]	[0.066]	[0.053]
p-value: M = I	0.34	0.82	0.14	0.59	0.68	0.09	0.14	0.93
Control mean	0.00	3.06	3.48	3.35	3.30	3.86	4.40	4.71
# Individuals	3,068	3,068	3,068	3,068	3,068	3,068	3,068	3,068

B. Fitness	Fitness time trial index	Seconds to walk 4m	Seconds for 5 sit-stands
	(1)	(2)	(3)
Incentives	0.024	0.042	-0.10
	[0.045]	[0.043]	[0.12]
Monitoring	0.069	0.080	-0.088
	[0.077]	[0.076]	[0.19]
p-value: M = I	0.50	0.57	0.94
Control mean	0.00	3.88	13.18
# Individuals	2,890	2,825	2,793

Notes: The Mental health index averages the values of seven questions adapted from RAND's 36-Item Short Form Survey. A large value of the Fitness time trial index indicates low fitness. The sample includes the incentive, monitoring, and control groups. Controls are the same as described in the Table 4 notes, along with the same set of additional controls described in the Table F.14 notes. Robust standard errors are in brackets. Data are at the individual level. Significance levels: * 10%, ** 5%, *** 1%.

Appendix Table F.14: Impacts of Incentives on Diet and Addictive Consumption

A. Healthy diet

	Healthy diet index	Wheat meals	Meals with vegeta-bles	Servings of fruit	Negative of rice meals	Negative of junkfood pieces	Negative of spoons sugar in coffee	Negative of sweets yester-day)	Avoid un-healthy food
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Incentives	0.052	0.028	0.060**	0.038	0.029	-0.020	-0.019	-0.028	0.0037
	[0.044]	[0.029]	[0.030]	[0.035]	[0.033]	[0.066]	[0.047]	[0.038]	[0.018]
Monitoring	0.023	0.019	0.082	0.062	-0.0068	0.13	-0.026	-0.048	-0.040
	[0.085]	[0.053]	[0.054]	[0.062]	[0.060]	[0.10]	[0.081]	[0.082]	[0.033]
p-value: M = I	0.71	0.85	0.66	0.68	0.51	0.08	0.92	0.80	0.14
Control mean	0.00	0.49	0.58	0.53	-2.34	-0.91	-1.12	-0.35	0.83
# Individuals	3,068	3,068	3,068	3,068	3,068	3,068	3,068	3,068	3,068

B. Addictive consumption

	Addictive good consumption index	Average daily areca	Average daily alcohol	Average daily cigarettes
	(1)	(2)	(3)	(4)
Incentives	-0.014	0.034	-0.036	-0.056
	[0.037]	[0.037]	[0.028]	[0.095]
Monitoring	-0.0036	0.015	-0.016	-0.018
	[0.060]	[0.068]	[0.038]	[0.14]
p-value: M = I	0.85	0.76	0.46	0.77
Control mean	0.00	0.13	0.11	1.02
# Individuals	3,068	3,068	3,068	3,068

Notes: The Healthy Diet Index is composed of the average values of eight diet questions, standardized by their average and standard deviation in the control group; a larger value indicates a healthier diet. The Addictive Good Consumption Index is an index created by the average self-reported daily consumption of areca, alcoholic drinks, and cigarettes, standardized by their average and standard deviation in the control group; a larger value indicates higher consumption. The omitted category is Control. All specifications control for the baseline value of the dependent variable (or index components for indices), the baseline value of the dependent variable squared (or index components squared for indices), an SMS treatment indicator, and the following controls: age, weight, height, gender, and their second-order polynomials, as well as endline completion date, hour of endline completion, and dummy for late completion. Standard errors, in brackets, are clustered at the individual level. The sample includes the incentive, monitoring, and control groups. Data are at the individual level. Significance levels: * 10%, ** 5%, *** 1%.

Appendix Table F.15: Main Sample and Validation Sample Have Similar Characteristics

	Main sample	Validation sample	p-value	Norm. Diff.
	(1)	(2)	(3)	(4)
A. Demographics				
Age	49.56	50.55	0.331	-0.120
	(8.51)	(7.98)		
Female (=1)	0.42	0.41	0.815	0.028
	(0.49)	(0.50)		
Labor force participation (=1)	0.74	0.75	0.963	-0.006
	(0.44)	(0.44)		
Household size	3.91	3.73	0.370	0.126
	(1.62)	(1.08)		
B. Health				
Overweight (=1)	0.61	0.65	0.513	-0.079
	(0.49)	(0.48)		
BMI	26.42	27.06	0.226	-0.131
	(4.34)	(5.31)		
Systolic BP (mmHg)	133.38	135.83	0.290	-0.134
	(19.16)	(17.21)		
Diastolic BP (mmHg)	88.48	91.17	0.045	-0.246
	(11.10)	(10.76)		
C. Walking - phase-in				
Exceeded step target (=1)	0.25	0.21	0.321	0.116
	(0.32)	(0.34)		
Average daily steps	7004.04	6539.98	0.331	0.119
	(3981.43)	(3837.98)		
F-test for joint orthogonality				
p-value			0.48	
Sample size				
Number of individuals	3232	71		

Notes: Means are reported for each variable and standard deviations are in parentheses. Main sample is our primary experimental sample. Validation sample is the sample used to validate our impatience index as described in Appendix C. Norm. Diff. is normalized differences. All variables are as in Table 1. The number of individuals with pedometer data differs from the total number of individuals because a few participants withdrew immediately. The F-statistic is obtained by running regressions with all characteristics. Data are at the individual level.

G Misreporting Steps, Confusion, and Suspensions

Procedures to Curb Misreporting Because incentive payments were determined by self-reported data and not pedometer data, we implemented a number of checks to ensure integrity of step reporting. Within each 28-day sync period, respondents who incorrectly over-reported meeting a 10k step target on more than 25% of days were flagged for cheating and suspended from receiving recharges for 7 days, and those who over-reported on 10–25% of days were flagged for cheating but only given a warning. Those who were flagged for cheating more than once were terminated from the program. Fewer than 5% of Incentive participants were suspended for cheating and only 1 was terminated (Table G.1)

During the intervention, we also attempted to flag participants who appeared to be confused about how to read their pedometers or report properly. We flagged those whose reported steps were either more than 10% higher than their pedometer steps or more than 15% lower than their pedometer steps on 40% of days as "confused" (unless their misreporting was indicative of cheating). Those who were flagged received a refresher from the surveyors on how to use the step-reporting system. We did not require pedometer and reported steps to match exactly because our pedometers record daily steps until midnight, but respondents typically reported their daily steps before midnight. As a result, we expected pedometer and reported steps to diverge slightly, either because respondents continued to walk after reporting their steps or because respondents (incorrectly) estimated the number of additional steps they would take post-reporting, and reported that amount instead.

We also took measures to encourage regular reporting for all groups. We offered a 50 INR "pedometer wearing and reporting bonus" to participants during the pre-intervention period if they wore the pedometer and reported steps on 80% of days to ensure that all participants were familiar with the step reporting system. At contract launch, we also briefly encouraged all but Control participants to report steps regularly during the intervention period, and offered a larger 200 INR pedometer wearing and reporting bonus for wearing and reporting during the intervention period. Finally, if participants did not report for a number of consecutive days, we would send them a text message reminder to report.

Rates of Misreporting and Confusion Our analysis only uses pedometer data (not reported data), so misreporting would not bias our conclusions. However, it is still interesting to examine the prevalence of misreporting. The prevalence of misreporting, defined as reporting steps above 10,000 when the pedometer itself records fewer than 10,000 steps, is less than 5% and, interestingly, balanced across incentive and monitoring groups (column 1 of Table G.2). The balance with the monitoring group, who had no incentives to over-report, suggests that over-reporting was mainly unintentional participant mistakes. The incentive group also appeared to put more effort into making correct step reports, with fewer divergences in either the positive or the negative directions (columns 2-4 of Table G.2).

Appendix Table G.1: Summary Statistics on Audits and Suspensions

	Count		Share	
	Incentives	Monitoring	Incentives	Monitoring
	(1)	(2)	(3)	(4)
Shared Fitbit ever	3	0	0.004	0.000
Suspended for cheating	100	N/A	0.042	N/A
Terminated for cheating	1	N/A	0.000	N/A
Total:	2,404	203	0.92	0.08

Notes: We randomly audited around 1,000 individuals from both the incentive and monitoring groups to look for evidence of pedometer sharing. The first row in columns 3 and 4 is conditional on being audited.

Appendix Table G.2: Misreporting, Confusion and Cheating by Treatment Group

Variable type:	Reporting	Confusion		
Dependent variable:	Incorrectly reported over 10k steps	Over-reported or under-reported	Over-reported by at least 10%	Under-reported by at least 15%
	(1)	(2)	(3)	(4)
Incentives	0.0079	-0.081***	-0.059***	-0.022**
	[0.01]	[0.02]	[0.02]	[0.01]
Monitoring mean	0.049	0.272	0.167	0.104
# Individuals	2,542	2,542	2,542	2,542
# Observations	173,131	173,131	173,131	173,131

Notes: Each observation is a respondent $\times$ day. Column 2 shows whether a respondent over-reported by at least 10% or under-reported by at least 15%. The omitted group is the monitoring group. Controls are the same as Table 2. Standard errors, in brackets, are clustered at the individual level. The sample includes the incentive and monitoring groups. Significance levels: * 10%, ** 5%, *** 1%.

H Theoretical Predictions: Additional Proofs

H.1 Proofs of Section B.4 Propositions

We begin by proving Proposition 1 for $T \geq 2$. We then prove Propositions 3, 4, and 5.

Proposition 1 ($T = K$, Threshold Compliance and Impatience Over Effort)**.** *Let $T > 1$. Fix all parameters other than $\delta^{(t)}$. Take any threshold contract with threshold level $K = T$; denote the threshold payment M. Compliance in the threshold contract will be weakly decreasing in $\delta^{(t)}$ for all $t \leq T - 1$.*

Proof. Let $V_{t,j}^{(1)}$ be the value of being on day t having complied on all previous days 1 through $t-1$, where the value is evaluated from the perspective of the agent on day $j \leq t$. Let $V_{t,j}^{(0)}$ be the value of being on day t having *not* complied on at least one of the previous days 1 through $t-1$, again evaluated from the day j perspective. And let $V_{t,j}^{(1-0)} = V_{t,j}^{(1)} - V_{t,j}^{(0)}$. Correspondingly, let $w_t(e_t, 1)$ be the compliance decision on day t if the person has effort cost e_t and has complied on all prior days, and let $w_t(e_t, 0)$ be the compliance decision on day t if the person has effort cost e_t and has not complied on all prior days. If the person has complied on all previous days, we thus have that day t compliance is as follows:

$$w_t(e_t, 1) = \begin{cases} 1 & \text{if } e_t < V_{t+1,t}^{(1-0)} \\ 0 & \text{otherwise} \end{cases} \tag{23}$$

and as follows if the person has not complied on all previous days

$$w_t(e_t, 0) = \begin{cases} 1 & \text{if } e_t < 0 \\ 0 & \text{otherwise} \end{cases} \tag{24}$$

We look at naifs first and then sophisticates. For both types, we begin by examining day T and then use the day T result to show results for days $t < T$. On day j, naifs think that, on day T, conditional on complying on days 1 through $T-1$, their day-T self will comply if $\delta^{(T-j)} e_T < d^{(T-j)} M$, or equivalently if $d^{(T-j)} M - \delta^{(T-j)} e_T > 0$. Their value if they comply is the discounted payment net of discounted effort costs, $d^{(T-j)} M - \delta^{(T-j)} e_T$. Hence, we have

$$\begin{aligned} V_{T,j}^{(1)} &= \mathbb{E}\left[\left(d^{(T-j)} M - \delta^{(T-j)} e_T\right) \mathbb{1}\{d^{(T-j)} M - \delta^{(T-j)} e_T > 0\} \middle| e_1, \ldots, e_j\right], \ j = 1, \ldots, T \\ &= \mathbb{E}\left[\max\{d^{(T-j)} M - \delta^{(T-j)} e_T, 0\} | e_1, \ldots, e_j\right], \ j = 1, \ldots, T \end{aligned} \tag{25}$$

They also think that, on any day t including T, if they haven't complied on all days through $t-1$, they will comply if $\delta^{(t-j)} e_t < 0$, which is equivalent to $e_t < 0$, which yields

$$V_{t,j}^{(0)} = \mathbb{E}\left[-\delta^{(t-j)} e_t \mathbb{1}\{e_t < 0\} | e_1, \ldots, e_j\right], \ j = 1, \ldots, t \tag{26}$$

As a result, we have that:

$$V_{T,j}^{(1-0)} = \mathbb{E}\left[\max\{d^{(T-j)} M - \delta^{(T-j)} e_T, 0\} + \delta^{(T-j)} e_T \mathbb{1}\{e_T < 0\} | e_1, \ldots, e_j\right] \tag{27}$$

To show that this expectation is decreasing in $\delta^{(T-j)}$, we show that the argument, $\max\{d^{(T-j)} M - \delta^{(T-j)} e_T, 0\} + \delta^{(T-j)} e_T \mathbb{1}\{e_T < 0\}$, is decreasing in δ for all values of e_T. Consider two cases:

1. Case 1: $e_T > 0$. In this case,

$$\max\{d^{(T-j)}M - \delta^{(T-j)}e_T, 0\} + \delta^{(T-j)}e_T\mathbb{1}\{e_T < 0\} = \max\{d^{(T-j)}M - \delta^{(T-j)}e_T, 0\},$$

which is decreasing in $\delta^{(T-j)}$.

2. Case 2: $e_T \leq 0$ In this case, letting $u = -e_T \geq 0$, we have

$$\begin{aligned} &\max\{d^{(T-j)}M - \delta^{(T-j)}e_T, 0\} + \delta^{(T-j)}e_T\mathbb{1}\{e_T < 0\} \\ &= \begin{cases} \max\{d^{(T-j)}M + \delta^{(T-j)}u, 0\} - \delta^{(T-j)}u & \text{if } e_T \neq 0 \\ d^{(T-j)}M & \text{if } e_T = 0 \end{cases} \\ &= d^{(T-j)}M, \end{aligned}$$

which is invariant to $\delta^{(T-j)}$.

Thus, $\max\{d^{(T-j)}M - \delta^{(T-j)}e_T, 0\} + \delta^{(T-j)}e_T\mathbb{1}\{e_T < 0\}$ is weakly decreasing in $\delta^{(T-j)}$ for all e_t, and so, by taking expectations, equation (27) must also be decreasing in $\delta^{(T-j)}$.

In addition, on day j, naifs think that, conditional on having complied on days 1 through $t-1$, they will comply on day $t \geq j$, if $\delta^{(t-j)}e_t < V^{(1-0)}_{t+1,j}$. So, for $t \leq T-1$ we have:

$$\begin{aligned} V^{(1)}_{t,j} &= \mathbb{E}\left[\left(V^{(1-0)}_{t+1,j} - \delta^{(t-j)}e_t\right)\mathbb{1}\left\{\delta^{(t-j)}e_t < V^{(1-0)}_{t+1,j}\right\}|e_1,\ldots,e_j\right],\ j = 1,\ldots,t \\ &= \mathbb{E}\left[\max\{V^{(1-0)}_{t+1,j} - \delta^{(t-j)}e_t, 0\}|e_1,\ldots,e_j\right],\ j = 1,\ldots,t \end{aligned}$$

Combined with equation (26) this yields:

$$V^{(1-0)}_{t,j} = \mathbb{E}\left[\max\{V^{(1-0)}_{t+1,j} - \delta^{(t-j)}e_t, 0\} + \delta^{(t-j)}e_t\mathbb{1}\{\delta^{(t-j)}e_t < 0\}\big|e_1,\ldots,e_j\right],\ j = 1,\ldots,t \quad (28)$$

Equations (27) and (28) thus recursively define all of the $V^{(1-0)}_{t,j}$ for any $t \leq T$ and $j \leq t$. Since we already showed that $V^{(1-0)}_{T,j}$ is weakly decreasing in $\delta^{(T-j)}$ for all $j \leq T$ (equation (27)), we can then use reverse induction from $t = T,\ldots,j$ using equations (27) and (28) to see that $V^{(1-0)}_{t,j}$ is decreasing in all $\delta^{(T-j)},\ldots,\delta^{(t-j)}$ for any $t \leq T$ and $j \leq t$.[67]

[67]We make the induction hypothesis that $V^{(1-0)}_{t+1,j}$ is weakly decreasing in $\delta^{(1)},\ldots,\delta^{(t)}$ and show that, under this hypothesis, $V^{(1-0)}_{t,j}$ is also weakly decreasing in $\delta^{(1)},\ldots,\delta^{(t)}$. Since we have already shown that $V^{(1-0)}_{T,j}$ is decreasing in all $\delta^{(1)},\ldots,\delta^{(T-1)}$, the result then follows. To show that $V^{(1-0)}_{t,j}$ is weakly decreasing in all $\delta^{(1)},\ldots,\delta^{(t+1)}$, we show that the argument of equation (28), $\max\{V^{(1-0)}_{t+1,j} - \delta^{(t-j)}e_t, 0\} + \delta^{(t-j)}e_t\mathbb{1}\{\delta^{(t-j)}e_t < 0\}$, is decreasing in $\delta^{(t-j)}$ for all e_t. Again there are two cases:

1. Case 1: $e_t > 0$. In this case,

$$\max\{V^{(1-0)}_{t+1,j} - \delta^{(t-j)}e_T, 0\} + \delta^{(t-j)}e_t\mathbb{1}\{e_t < 0\} = \max\{V^{(1-0)}_{t+1,j} - \delta^{(t-j)}e_T, 0\},$$

which is weakly decreasing in $\delta^{(t-j)}$ under the induction hypothesis.

2. Case 2: $e_t \leq 0$ In this case, letting $u = -e_t \geq 0$, we have

$$\max\{V^{(1-0)}_{t+1,j} - \delta^{(t-j)}e_T, 0\} + \delta^{(t-j)}e_t\mathbb{1}\{e_t < 0\} = \max\{V^{(1-0)}_{t+1,j} + \delta^{(t-j)}u, 0\} - \delta^{(t-j)}u = V^{(1-0)}_{t+1,j},$$

which is again weakly decreasing in $\delta^{(t-j)}$ under the induction hypothesis.

The fact that $V_{t,j}^{(1-0)}$ is decreasing in all $\delta^{(T-j)},\dots,\delta^{(t-j)}$ for any $t \leq T$ and $j \leq t$ shows that day t compliance is also weakly decreasing in all $\delta^{(T-t)},\dots,\delta^{(t-t)}$, since one complies on day t if $e_t < V_{t+1,t}^{(1-0)}$ (equation (23)). Hence, overall compliance C from days $1,\dots,T$, is weakly decreasing in $\delta^{(1)},\dots,\delta^{(T-1)}$ for naifs.

Sophisticates know that, conditional on complying on all prior days, on day T they will comply if $e_T < M$. Thus, equation (27) becomes:

$$V_{T,j}^{(1-0)} = \mathbb{E}\left[\left(d^{(T-j)}M - \delta^{(T-j)}e_T\right)\mathbb{1}\left\{e_T < M\right\} + \delta^{(T-j)}e_T\mathbb{1}\{e_T<0\}\middle|e_1,\dots,e_j\right] \quad j=1,\dots,T \tag{29}$$

This is weakly decreasing in $\delta^{(T-j)}$ since the argument is weakly decreasing in $\delta^{(T-j)}$ for all e_T:

1. $e_T > 0$: In this case, $\left(d^{(T-j)}M - \delta^{(T-j)}e_T\right)\mathbb{1}\left\{e_T < M\right\} + \delta^{(T-j)}e_T\mathbb{1}\{e_T<0\} = \left(d^{(T-j)}M - \delta^{(T-j)}e_T\right)\mathbb{1}\left\{e_T < M\right\}$, which is weakly decreasing in $\delta^{(T-j)}$.
2. $e_T \leq 0$: In this case, $\left(d^{(T-j)}M - \delta^{(T-j)}e_T\right)\mathbb{1}\left\{e_T < M\right\} + \delta^{(T-j)}e_T\mathbb{1}\{e_T<0\} = \left(d^{(T-j)}M - \delta^{(T-j)}e_T\right) + \delta^{(T-j)}e_T = d^{(T-j)}M$, which is invariant to $\delta^{(T-j)}$.

Thus, $V_{T,j}^{(1-0)}$ is weakly decreasing in $\delta^{(T-j)}$.

Sophisticates also know that, on day $t \leq T-1$, if they have complied on all previous days, they will comply if $e_t < V_{t+1,t}^{(1-0)}$ and so equation (28) becomes:

$$V_{t,j}^{(1-0)} = \mathbb{E}\left[\left(V_{t+1,j}^{(1-0)} - \delta^{(t-j)}e_t\right)\mathbb{1}\left\{e_t < V_{t+1,t}^{(1-0)}\right\} + \delta^{(t-j)}e_t\mathbb{1}\{e_t<0\}\middle|e_1,\dots,e_j\right] \quad j=1,\dots,t \tag{30}$$

Since we showed above that $V_{T,j}^{(1-0)}$ is weakly decreasing in $\delta^{(T-j)}$ for all $j \leq T$, one can thus use equation (30) and the same reverse induction argument as for naifs to show this implies that $V_{t,j}^{(1-0)}$ is decreasing in all $\delta^{(T-j)},\dots,\delta^{(t-j)}$ for all $j \leq t \leq T$.[68] By the same argument used for naifs, this then implies overall compliance C is weakly decreasing in $\delta^{(1)},\dots,\delta^{(T-1)}$ for sophisticates. □

Proposition 3 (Perfect Correlation, Threshold Effectiveness and Impatience Over Effort)**.** *Let there be perfect correlation in costs across periods ($e_t = e_{t'} \equiv e$ for all t,t'). For simplicity, let $\delta^{(t)} < 1$ for all $t > 0$ if $\delta^{(t)} < 1$ for any t. Fix all parameters other than $\delta^{(t)}$ for some $t \leq T-1$.*

---

[68]Again the induction hypothesis is that $V_{t+1,j}^{(1-0)}$ is weakly decreasing in $\delta^{(1)},\dots,\delta^{(t)}$. One can then use equation (30) to show that this implies that $V_{t,j}^{(1-0)}$ is weakly decreasing in $\delta^{(1)},\dots,\delta^{(t)}$ because the argument, $\left(V_{t+1,j}^{(1-0)} - \delta^{(t-j)}e_t\right)\mathbb{1}\left\{e_t < V_{t+1,j}^{(1-0)}\right\} + \delta^{(t-j)}e_t\mathbb{1}\{e_t<0\}$, is weakly decreasing in $\delta^{(1)},\dots,\delta^{(t)}$ for all e_t. There are two cases::

1. $e_t > 0$: In this case, $\left(V_{t+1,j}^{(1-0)} - \delta^{(t-j)}e_t\right)\mathbb{1}\left\{e_t < V_{t+1,j}^{(1-0)}\right\} + \delta^{(t-j)}e_t\mathbb{1}\{e_t<0\} = \left(V_{t+1,j}^{(1-0)} - \delta^{(t-j)}e_t\right)\mathbb{1}\left\{e_t < V_{t+1,j}^{(1-0)}\right\}$, which is weakly decreasing in $\delta^{(t-j)}$ under the induction hypothesis.
2. $e_t \leq 0$: In this case, $\left(V_{t+1,j}^{(1-0)} - \delta^{(t-j)}e_t\right)\mathbb{1}\left\{e_t < V_{t+1,j}^{(1-0)}\right\} + \delta^{(t-j)}e_t\mathbb{1}\{e_T<0\} = V_{t+1,j}^{(1-0)}$, which is weakly decreasing in $\delta^{(t-j)}$ under the induction hypothesis.

Since we have already shown that $V_{T,j}^{(1-0)}$ is weakly decreasing in $\delta^{(1)},\dots,\delta^{(T-1)}$ the result is thus shown.

Take any threshold contract with threshold level $K \leq T$. Compliance and effectiveness in the threshold contract will be weakly decreasing in $\delta^{(t)}$.

Proof. We first examine compliance and then examine effectiveness.

To gain intuition for the compliance result, first think about a person who is fully patient over both effort and payment: $\delta^{(t)} = 1$ and $d^{(t)} = 1$ for all t. That person will comply on all days if $e < m'$ (with m' the per-day reward in the threshold contract) and on no days if $e \geq m'$. In contrast, we now show that when people are impatient over effort, they often will comply even when $e > m'$.

When people are impatient, there are two cases. The first (less interesting) case is where it would be worthwhile for the agent to comply on at least K days in a separable contract paying m': $e < d^{(T-K+1)}m'$. In that case, the threshold does not "bind" and the person just complies on all days t for which $e < d^{(T-t)}m'$. Compliance is just like in the separable contract paying m' and is invariant to $\delta^{(t)}$.

The second (interesting) case is where the agent would not comply on at least K days in a separable contract paying m' ($e \geq d^{(T-K+1)}m'$) and so the threshold "binds." In this case, note that agents will never comply more than K days total.[69]

A naif who is impatient over effort (i.e., for whom $\delta^{(t)} < 1$ for all $t > 0$) will never comply before day $T - K + 1$ (i.e., before the last K days). In period $T - K + 1$, the naif will comply if on day $T - K + 1$:

$$\sum_{t=T-K+1}^{T} \delta^{(t-(T-K+1))} e \leq d^{(K-1)} K m' \tag{31}$$

Compliance on day $T-K+1$ is thus decreasing in $\delta^{(t)}$ for all t from 1 to K. If the naif complies on day $T-K+1$, the naif will then comply on all future days. Hence, compliance is decreasing in $\delta^{(t)}$ for all t from 1 to K.

A sophisticate who is impatient over effort will always comply when a naif with the same discount rates would. In addition, the sophisticate may comply before the last K days as well.[70]

To formalize the sophisticate's conditions for compliance, consider all combinations of size K taken from the days 1 through T. There will be $\binom{T}{K}$ such combinations.[71] Order each combination chronologically and index the ordered days as days $j = 1, ..., K$ with values t_1 through t_k (e.g., if the combination is day 1 and day 3, then $t_1 = 1$ and $t_2 = 3$). A sophisticate will comply exactly K times if, for *any* of the $\binom{T}{K}$ combinations, *all* of the following K constraints

[69] Once people have reached the threshold, they will only comply on the other days if they would have complied on those days for a piece rate of m' and, since the agent would not have complied K days in a separable contract pay m', there will be no additional days that satisfy that criterion after they have reached the threshold.

[70] For example, take the case where $T = 3$ and $K = 2$. There may be cases where the individual would not find it worthwhile to comply on day 2, since $(1 + \delta^{(1)})e > 2dm'$, but would find it worthwhile to comply on day 1, since $(1 + \delta^{(2)})e < 2dm'$. In that case, the sophisticate would comply on days 1 and 3.

[71] In our example with $T = 3$ and $K = 2$, the combinations would be $1, 3$ and $2, 3$.

hold:

$$\sum_{j=1}^{K} \delta^{(t_j - t_1)} e \leq d^{(T-t_1)} K m'$$

$$\sum_{j=2}^{K} \delta^{(t_j - t_2)} e \leq d^{(T-t_2)} K m'$$

$$.....$$

$$\sum_{j=K}^{K} \delta^{(t_j - t_K)} e \leq d^{(T-t_K)} K m' \tag{32}$$

Since any of these constraints is weakly more likely to hold the lower any $\delta^{(t_j - t_1)}$, the result is thus shown for sophisticates as well.

Having shown that compliance in the threshold contract is weakly decreasing in $\delta^{(t)}$, we now just need to show that cost-effectiveness is not increasing in $\delta^{(t)}$ and the effectiveness result follows. To show this, we note that, in the perfect correlation case, regardless of $\delta^{(t)}$, any agent who complies on at least one day will always follow through to reach the threshold and achieve payment. Payments will thus be $m'C$ and cost-effectiveness will thus be $\frac{1}{m'}$ regardless of the discount factors. This is invariant to $\delta^{(t)}$. □

Proposition 4. *Let $T = 3$. Let the cost of effort on each day be binary, taking on either a "high value" (e_H) or a "low value" (e_L), with $e_H \geq e_L$. Let agents observe the full sequence of costs e_1, e_2, e_3 on day 1. Let $\delta^{(t)} = \delta^t$ (i.e., let the discount factor over effort be exponential) and let $d^{(t)} = 1$. Fix all parameters other than δ. Consider a threshold contract with $K = 2$, where the agent must thus comply on at least 2 days in order to receive payment. Compliance and effectiveness in the threshold contract are weakly higher for someone with a discount factor $\delta < 1$ than for someone with discount factor $\delta = 1$.*

Proof. We first consider different values of e_H and e_L. First, if $e_H < m'$, then $\sum_{t=1}^{3} w_t = 3$ for all δ and so the prediction trivially goes through. Second, if $e_L \geq m'$, then $\sum_{t=1}^{3} w_t = 0$ for $\delta = 1$. However, some people with $\delta < 1$ may walk in at least one period due to the standard cost-bundling effect (e.g., if they have costs of e_L every period and if $e_L + \delta e_L < 2m'$, then they would walk twice). Thus, the prediction goes through in that case as well. We thus have proved the prediction in the cases where $e_H < m'$ and $e_L \geq m'$ and so we next consider the cases where $e_H \geq m'$ and $e_L < m'$.

To prove the prediction, we examine all 8 potential sequences of costs and prove it separately for each case. Note that we only consider the cases where $e_H \geq m'$ and $e_L < m'$.

1. Cases 1 and 2: e_L, e_L, e_L and e_H, e_H, e_H. Since in these cases, costs are constant across periods, the prediction goes through by using the same arguments as in the proof for the case when costs are perfectly correlated across periods (Proposition 7b).
2. Case 3: e_H, e_H, e_L: Again, neither sophisticates nor naifs walk in period 1 but both walk in period 2 and period 3 if $e_H + \delta e_L < 2m'$ (note that by the assumptions above, since

$e_L < m'$, they will always follow-through so there is no follow-through constraint). Thus total compliance is decreasing in δ.

3. Case 4: e_H, e_L, e_H. Again, nobody walks in period 1. Sophisticates walk in periods 2 and 3 if $e_L + \delta e_H < 2m'$ and $e_H < 2m'$. Naifs walk in period 2 if $e_L + \delta e_H < 2m'$ and in period 3 if they've walked in period 2 and $e_H < 2m'$. Again, total compliance is decreasing in δ.
4. Case 5: e_L, e_H, e_H. Sophisticates walk in period 1 if $e_L + \delta^2 e_H < 2m'$ and they know they will follow through ($e_H < 2m'$). Naifs walk in period 1 if $e_L + \delta^2 e_H < 2m'$. Neither type walks in period 2 since $e_H \geq m'$. Both types walk in period 3 if they walked in period 1 and $e_H < 2m'$. Again total compliance is thus decreasing in δ.
5. Cases 6, 7, and 8: e_L, e_H, e_L; e_L, e_L, e_H; and e_H, e_L, e_L. All people, regardless of δ, walk in the two periods where the cost is e_L, since $e_L + e_L < 2m'$. Nobody walks in the period where the cost is e_H since they know they will walk in the other periods and $e_H \geq m'$. Thus, the prediction (trivially) holds.

To prove the effectiveness part of the result, we examine sophisticates first and then naifs and show that cost-effectiveness is non-increasing in δ for both types. Sophisticates will always get paid for every day they comply. Thus, regardless of δ, if compliance is non-0, cost-effectiveness will be $\frac{1}{m'}$, and hence non-increasing in δ. In contrast with sophisticates, naifs can sometimes not receive payment for a day on which they comply. In case 4, naifs will walk on day 2 if $e_L + \delta e_H < 2m'$ but not walk on day 3—and hence not be paid—if $e_H > 2m'$. Those two conditions are more likely to hold in conjunction the lower is δ. Similarly in case 5, naifs will walk on day 1 if $e_L + \delta^2 < 2m'$ but not receive payment if $e_H > 2m'$, which is again more likely to occur the lower is δ. Since having days of compliance that the principal does not have to pay for increases cost-effectiveness, this means that the lower is δ, the weakly higher cost-effectiveness is for naifs.

Hence, since we have shown that compliance is decreasing in δ whereas cost-effectiveness is non-increasing (and in particular, flat for sophisticates and weakly decreasing for naifs), then we have shown that *effectiveness* is also weakly decreasing in δ.

□

For sophisticates, we can also show a stronger result. In simulations with most realistic cost distributions, this stronger result goes through for naifs as well.

Proposition 5. *Let $T = 3$. Let costs be weakly positive and let agents observe the full sequence of costs e_1, e_2, e_3 on day 1. Let $\delta^{(t)} = \delta^t$ (i.e., let the discount factor over effort be exponential) and let $d^{(t)} = 1$. Fix all parameters other than δ. Consider a threshold contract with $K = 2$, where the agent must thus comply on at least 2 days in order to receive payment. For sophisticates, compliance and effectiveness in the threshold contract are weakly decreasing in the discount factor δ.*

Proof. We begin by examining compliance and then turn to effectiveness. For the compliance result, we first define some useful notation. Let X_t be the "walking stock" coming into period

t (i.e., sum from period 1 to period $t-1$ of whether the person complied $X_t = \sum_{i=1}^{t-1} w_i$). Let $w_t(X_t)$ be a dummy for whether the person complies in period t as a function of the walking stock coming into period t.

To examine compliance, we work backward. In period 3, behavior will depend on the walking stock X_3:

$$\begin{aligned} w_3(2) &= \mathbb{1}\{e_3 < m'\} \\ w_3(1) &= \mathbb{1}\{e_3 < 2m'\} \\ w_3(0) &= \mathbb{1}\{e_3 < 0\}. \end{aligned}$$

We show that the prediction holds by showing that it holds under all potential cases for e_3.

Case 1: $m' \leq e_3 < 2m'$ In this case, walking in period 3 is

$$\begin{aligned} w_3(2) &= 0 \\ w_3(1) &= 1 \\ w_3(0) &= 0. \end{aligned}$$

Note that this implies the person will never walk three times. Walking in period 2 is

$$\begin{aligned} w_2(1) &= \mathbb{1}\{e_2 \leq \delta e_3\} \\ w_2(0) &= \mathbb{1}\{e_2 + \delta e_3 < 2m'\}. \end{aligned}$$

In period 1, consider two cases:

1. $e_2 + \delta e_3 < 2m'$: she knows she will walk at least twice, and the only question is whether to walk now or later. If $e_1 < min\{\delta e_2, \delta^2 e_3\}$, then she will walk in period 1; if not, then she will wait and walk in periods 2 and 3. Either way, she walks twice.
2. $e_2 + \delta e_3 \geq 2m'$: she knows she will not walk later, so she will walk if $e_1 + min\{\delta e_2, \delta^2 e_3\} < 2m'$.

Thus we can see that when $m' \leq e_3 < 2m'$, overall compliance is as follows:

$$\text{Days walked} = \begin{cases} 2 & \text{if } e_2 + \delta e_3 \leq 2m' \text{ OR } e_1 + \delta min\{e_2, \delta e_3\} \leq 2m' \\ 0 & \text{otherwise.} \end{cases}$$

Thus, compliance is obviously decreasing in δ.

Case 2: $e_3 \geq 2m'$ In this case, the person will never walk in period 3 regardless of the walking stock. Thus, overall compliance is as follows:

$$\text{Days walked} = \begin{cases} 2 & \text{if } e_1 + \delta e_2 < 2m' \text{ AND } e_2 < 2m' \\ 0 & \text{otherwise.} \end{cases}$$

This is again decreasing in δ.

Case 3: $e_3 < m'$ In this case, walking in period 3 is

$$w_3(2) = 1$$
$$w_3(1) = 1$$
$$w_3(0) = 0.$$

There are two cases to consider for e_2:

1. $e_2 < m'$: in this case (for $\delta \leq 1$), discount rates do not matter since the person will walk regardless in periods 2 and 3. Then they walk in period 1 if $e_1 < m'$.
2. $e_2 \geq m'$: in this case, the person will not walk in period 2 with walking stock 1. Thus, the maximum the person will ever walk is two periods (the first or the second and then the third).

$$\text{Days walked} = \begin{cases} 2 & \text{if } (e_1 + \delta^2 e_3 < 2m' \ \&\ e_3 < 2m') \text{ or } (e_2 + \delta e_3 < 2m' \ \&\ e_3 < 2m') \\ 0 & \text{otherwise.} \end{cases}$$

Thus days walked is again weakly decreasing in δ.

Thus, we have shown the compliance portion of the result, as we have shown that compliance is weakly decreasing in δ for all potential values of e_3.

To prove the effectiveness part of the result, note that sophisticates will always get paid for every day they comply. Thus, regardless of δ, if compliance is non-0, cost-effectiveness will be $\frac{1}{m'}$. Hence, since compliance is decreasing in δ whereas cost-effectiveness is non-increasing, then effectiveness is also decreasing in δ.

□

H.2 Proofs of Section B.5 Propositions

We now provide the proofs for Propositions 6–8b.

Proposition 6. *Let $d = 1$ and $T = 2$. Fix all parameters other than δ, and take a linear contract that induces compliance $C > 0$.*
(a) If agents are naive and e_2 is weakly increasing in e_1, in a first order stochastic dominance sense,[72] *then for sufficiently small δ, there exists a threshold contract with $K = 2$ that has at least two times higher cost-effectiveness (and $1 + \frac{1}{C}$ times higher cost-effectiveness if costs are IID) and that generates compliance $\frac{1+C}{2}$ of the linear contract.*
(b) If agents are sophisticated and costs are IID, then for sufficiently small δ, there exists a threshold contract with $K = 2$ that has at least $1+C$ times higher cost-effectiveness and that generates compliance at least $\frac{1+C}{2}$ of the linear contract.

Proof. Take a linear contract with payment m that induces compliance $C > 0$. Equation (12) implies that compliance in a linear contract is $C = \frac{1}{T} \sum_{t=1}^{T} F(d^{(T-t)} m)$, which simplifies to $C = F(m)$when $d^{(T-t)} = 1$. Recall that the cost-effectiveness of a linear contract is $\frac{1}{m}$ (see Section B.2).

[72]Note that this assumption flexibly accommodates the range from IID to perfect positive correlation, just ruling out negative correlation.

(a) Naifs: Consider a threshold contract that pays $M = m + \varepsilon$. On day 1, the naive agent thinks that, conditional on complying on day 1, she will comply on day 2 if $\delta e_2 < M$. The perceived probability of day 2 compliance conditional on day 1 compliance is $F_{e_2|e_1}(\frac{m+\varepsilon}{\delta})$. For $\delta \simeq 0$, $F_{e_2|e_1}(\frac{m+\varepsilon}{\delta}) \simeq 1$. Hence, for $\delta \simeq 0$, on day 1, the naive agent will comply if $e_1 + \delta E[e_2|e_1] < m + \varepsilon$; the probability of effort on day 1 thus approaches $F(m)$ as $\delta \to 0, \varepsilon \to 0$. Conditional on complying on day 1, the probability of compliance on day 2 then approaches $F_{e_2|e_1<m}(m)$. This is equal to $F(m)$ if costs are IID and is weakly greater than $F(m)$ under our more general assumption that e_2 is weakly increasing in e_1. Overall compliance is thus equal to $0.5(F(m) + F(m)F_{e_2|e_1<m}(m)) = 0.5(C + CF_{e_2|e_1<m}(m)) \geq 0.5C(1 + C)$. Expected payment per period then approaches $0.5mF(m)F_{e_2|e_1<m}(m) = 0.5mCF_{e_2|e_1<m}(m)$. Cost-effectiveness thus approaches $\frac{1}{m}\left(1 + \frac{1}{F_{e_2|e_1<m}(m)}\right) \geq 2/m$. This means the contract generates compliance of at least $(1 + C)/2$ times that of the linear contract and has at least 2 times higher cost-effectiveness. If costs are IID, $F_{e_2|e_1<m}(m) = F(m) = C$, and so cost-effectiveness approaches $\frac{1}{m}\left(1 + \frac{1}{C}\right)$, which is $1 + 1/C$ times larger than the cost-effectiveness of the linear contract.

(b) Sophisticates with IID costs: Now consider a threshold contract that pays $M = m/p' + \varepsilon$ for p' defined as a fixed point to $F(m/p') = p'$. The intermediate value theorem tells us that such a solution exists for $p' \in [C, 1]$ because F is continuous, $F(m/1) \leq 1$, and $F(m/C) \geq F(m) = C$.

Under this threshold contract, conditional on working in the first period, the probability of working in the second period is $F(M) = F(m/p' + \varepsilon) \geq F(m/p') = p'$, with $F(M) \simeq p'$ for $\varepsilon \simeq 0$. Hence, the expected payment conditional on working in the first period is $MF(M) \geq \frac{m}{p'}p' = m$, with this payment approximately m for $\varepsilon \simeq 0$. Therefore, for $\delta \simeq 0$, the probability of effort in the first period is at least $C = F(m)$, and approaches $F(m)$ for $\varepsilon \to 0, \delta \to 0$.

Taking $\varepsilon \to 0$ and then $\delta \to 0$: Total compliance in this contract is approximately $\frac{1}{2}(F(m) + F(m)F(M)) = \frac{1}{2}C(1 + p')$, with $\frac{1}{2}C(1 + p') \geq \frac{1}{2}C(1 + C)$ since $p' \geq C$. Payment per period is approximately $\frac{1}{2}MCp'$, with C the probability of working in the first period and p' the probability of working in the second period conditional on working in the first period; we have $\frac{1}{2}MCp' \simeq \frac{1}{2}\frac{m}{p'}Cp' = \frac{1}{2}mC$. Hence, cost-effectiveness is approximately $(\frac{1}{2}C(1 + p'))/(\frac{1}{2}mC) = (1 + p')/m \geq (1 + C)/m$. □

Proposition 7a (Perfect Correlation, $M = 2m$). *Let $T = 2$. Fix all parameters other than δ. Consider a linear contract with payment m and a threshold contract with payment $2m$. Then, regardless of agent type, the threshold contract is more effective than the linear contract if $\delta < 2d - 1$. If $\delta \geq 2d - 1$, then the linear contract may be more effective.*

Proof. As before, with perfect correlation, the agent takes effort either in both periods or in neither of a threshold contract. Thus the cost-effectiveness of the threshold contract will be $1/m$ and is thus the same as the cost-effectiveness of the linear contract. Therefore, whichever contract has higher compliance will be more effective. On day 1 of the linear, the agent complies if $e_1 < dm$, and on day 2 if $e_2 < m$, and so compliance in the linear contract is $\frac{1}{2}(F(dm) + F(m)) \leq F(m)$. In the threshold contract, on day 1 (and consequently day 2) the agent complies if $e_1(1 + \delta)d2m$, and so compliance is $F\left(\frac{2d}{1+\delta}m\right)$. Thus, if $\frac{2d}{1+\delta}m > m$ (i.e., if $\delta < 2d - 1$), the threshold contract has higher compliance (and hence effectiveness) than the linear. If that is not true, then the linear could have higher effectiveness. □

Proposition 7b (Perfect Correlation). *Let $T = 2$. Fix all parameters other than δ, and take any linear contract that induces compliance $C > 0$. Let there be perfect correlation in costs across days*

($e_1 = e_2$). Then, regardless of agent type, there exists a threshold contract that induces compliance of at least C and that has approximately $2\frac{d}{1+\delta}$ times greater cost-effectiveness than the linear contract. Hence, if $\delta < 2d - 1$, the most effective contract will always be a threshold contract.

Proof. With perfect correlation, the agent takes effort either in both periods or in neither of a threshold contract. Therefore, as long as the agent ever exerts effort, the cost-effectiveness is equal to 2 divided by the threshold payment.

Suppose a linear contract paying m induces $C > 0$ and has cost-effectiveness $\frac{1}{m}$. Note that, because $C = \frac{1}{2}(F(dm) + F(m))$, this implies that $F(m) \geq C$.

Consider a threshold contract with payment $M = m\frac{1+\delta}{d}$. Note that this contract will have cost effectiveness of $2\frac{d}{(1+\delta)m}$, which is $2\frac{d}{(1+\delta)}$ times the cost-effectiveness of the linear contract. On day 1 (and consequently day 2), the agent complies under the threshold contract if $e_1(1+\delta) < dM$ (where the left side comes from the fact that $e_1 = e_2$). With payment $M = m\frac{1+\delta}{d}$, the agent thus complies if $e_1 < m$. Thus, the threshold contract achieves compliance of $F(m) \geq C$. □

Proposition 8a (IID Uniform, $M = 2m$)**.** *Let $d = 1$. Fix all parameters other than δ. Let costs be independently drawn each day from a uniform[0,1] distribution. Take any threshold contract paying $M < 2$ and compare it with the linear contract paying $m = \frac{M}{2}$.*
(a) If $M < 1$, the threshold contract is always more cost-effective, but whether it has higher compliance (and hence whether it is more effective) depends on δ. There is a type-specific "cutoff value" such that if δ is less than the cutoff value for a given type, then the threshold contract is more effective, as it generates greater compliance.
(b) If $1 \leq M < 2$,[73] then the threshold contract is more effective.

Proof. Note that we take the general solution for compliance and payments for threshold contracts from the proof for Proposition 8b.

For a linear contract with payment level $\frac{M}{2}$, we have:

$$C = \frac{M}{2}$$
$$P = \frac{M^2}{4}$$
$$\frac{C}{P} = \frac{2}{M}$$
$$E = \lambda\frac{M}{2} - \frac{M^2}{4}$$

Now we consider multiple cases for what the threshold contract compliance and payments would be depending on the parameters.

(a) $0 < M < 1$ We begin with naifs and then move to sophisticates. For naifs, there are two cases:

Case 1: $M < \delta$ for Naifs In this case, $E[e_2|e_2 < M/\delta] = \frac{M}{2\delta}$, giving that

$$e_1^* = (M - \delta\frac{M}{2\delta})\frac{M}{\delta} = \frac{M^2}{2\delta}$$

[73]Note that the principal would never pay $M > 2$ since $M = 2$ achieves 100% compliance regardless of δ.

Thus,

$$
\begin{aligned}
C &= .5\left[\frac{M^2}{2\delta}+\frac{M^3}{2\delta}\right] \\
P &= .5\frac{M^4}{2\delta}
\end{aligned}
$$

Thus, cost-effectiveness is:

$$\frac{C}{P}=\frac{1+M}{M^2}$$

and effectiveness is:

$$E=.5\lambda\left[\frac{M^2}{2\delta}+\frac{M^3}{2\delta}\right]-.5\frac{M^4}{2\delta}$$

The threshold has higher cost-effectiveness if:

$$\frac{2}{M}<\frac{1+M}{M^2}.$$

This holds if $2M < 1 + M$ which is always true for $M < 1$. Thus, the threshold is always more cost-effective in this case.

The threshold has higher compliance if:

$$\frac{M}{2}<.5\left[\frac{M^2}{2\delta}+\frac{M^3}{2\delta}\right]$$

which simplifies to

$$\delta<\left[\frac{M}{2}+\frac{M^2}{2}\right].$$

This expression is not satisfied because $M < \delta$. Therefore, in this case, the threshold has lower compliance, and may have lower effectiveness. In fact, for $M < \delta$, whether the threshold has higher effectiveness depends on λ, the principal's marginal return to compliance: the higher λ, the more likely the threshold is to have higher effectiveness. Thus, in this range of relatively large δ we are above the cutoff value for naif types, and it is possible that the threshold will have either higher or lower effectiveness.

Case 2: $\delta < M$ for Naifs Because $M > \delta e_2$,

$$e_1^*=E[(M-\delta e_2)\mathbb{1}\{M-\delta e_2>0\}]=E[M-\delta e_2]=M-\delta/2$$

Thus,

$$
\begin{aligned}
C &= .5(M-\delta/2)(1+M) \\
P &= .5(M-\delta/2)M^2
\end{aligned}
$$

giving cost-effectiveness of

$$\frac{C}{P}=\frac{1+M}{M^2}$$

and effectiveness of

$$E=.5\lambda(M-\delta/2)(1+M)-.5(M-\delta/2)M^2.$$

The cost-effectiveness of the threshold contract is the same as in case 1, and so the threshold contract is again always more cost-effective.

Compliance of the threshold contract is higher than in the linear if:

$$.5M < .5(M - \frac{\delta}{2})(1 + M)$$

which simplifies to:

$$\begin{aligned}
M &< (M - \frac{\delta}{2})(1 + M) \\
M &< M(1 + M) - \frac{\delta}{2}(1 + M) \\
M &< M + M^2 - \frac{\delta}{2}(1 + M) \\
0 &< M^2 - \frac{\delta}{2}(1 + M) \\
\frac{\delta}{2}(1 + M) &< M^2 \\
\delta &< \frac{2M^2}{1 + M}
\end{aligned}$$

Note that, for $M < 1$, it is always true that $\frac{2M^2}{1+M} < M$.

Hence we can see that $\frac{2M^2}{1+M}$ is the cutoff value for naifs. For naifs, if $\delta < \frac{2M^2}{1+M}$ and $M < 1$, the threshold will always be more effective than the linear contract.

For sophisticates, there is just one case:

Case 3: $M < 1$ for Sophisticates In this case,

$$e_1^* = \left(M - \delta\frac{M}{2}\right)M = M^2(1 - \delta/2)$$

Thus,

$$\begin{aligned}
C &= .5(M^2 + M^3)(1 - \delta/2) \\
P &= .5(M^4)(1 - \delta/2)
\end{aligned}$$

Thus, cost-effectiveness is:

$$\frac{C}{P} = \frac{1 + M}{M^2}$$

The cost-effectiveness of the threshold contract is the same as in cases 1 and 2, and so, again, the threshold is always more cost-effective.

The compliance of the threshold contract is higher if:

$$.5M < .5(M^2 + M^3)(1 - \delta/2)$$

which holds if all of the following hold:

$$1 < (M + M^2)(1 - \delta/2)$$
$$\frac{1}{M + M^2} < 1 - \delta/2$$
$$\delta < 2 - \frac{2}{M + M^2}$$

Thus, the cutoff value for sophisticates is $2 - \frac{2}{M+M^2}$. If $\delta < 2 - \frac{2}{M+M^2}$, the threshold contract is more effective. For larger δ, the linear contract may be more effective.

(b) $1 \leq M < 2$ Here naifs and sophisticates behave the same and there are two cases.

Case 4: $1 < M < 1 + \delta/2$ In this case, because $M > \delta e_2$ and $M > e_2$

$$e_1^* = M - \delta/2$$

Because $M - \delta/2 < 1$,

$$\begin{aligned} C &= (M - \delta/2) \\ P &= .5M(M - \delta/2) \end{aligned}$$

giving

$$\frac{C}{P} = \frac{2}{M}.$$

This is the same cost-effectiveness as the linear contract. Hence, whichever contract has higher compliance will have higher effectiveness. Threshold compliance will be higher if:

$$\begin{aligned} M/2 &< (M - \delta/2) \\ \delta/2 &< M/2 \\ \delta &< M \end{aligned}$$

which is always true assuming that $\delta \leq 1$, since $M > 1$. Hence the threshold is always more effective.

Case 5: $1 + \delta/2 < M < 2$ Again, because $M > \delta e_2$ and $M > e_2$

$$e_1^* = M - \delta/2$$

Because $M - \delta/2 > 1$,

$$\begin{aligned} C &= 1 \\ P &= .5M \end{aligned}$$

giving

$$\frac{C}{P} = \frac{2}{M},$$

which is again the same as the cost-effectiveness of the linear contract. Hence, the threshold will have higher effectiveness if it has higher compliance, which is true if

$$M/2 < 1,$$

which will always be the case for $M < 2$. Hence, the threshold is always more effective. □

Proposition 8b (IID Uniform, Optimal Contracts)**.** *Let $d = 1$. Fix all parameters other than δ. Let costs be independently drawn each day from a uniform[0,1] distribution. Whether the most effective threshold contract is more effective than the most effective linear contract depends on δ as well as λ, the principal's marginal return to compliance. For a wide and plausible range of values of λ,[74] there exists a "cutoff" value of δ such that the threshold contract is more effective when δ is below the cutoff, and the linear contract is more effective when δ is above the cutoff. For the remaining values of λ, either the threshold contract is always more effective, or the linear contract is always more effective, but in either case the effectiveness of the threshold relative to linear is decreasing in δ.*

Proof. We begin with a more precise statement of the result, before proceeding to prove the result. Specifically, the following describes how the effectiveness of optimal threshold contract relative to the optimal linear one depends on the value of δ in different ranges of λ values:

(a) Naifs for $0 < \lambda < 0.225$, and naifs and sophisticates for $0.225 \leq \lambda < 1$ and $3 \leq \lambda \leq 2 + \sqrt{2}$. In these cases, there is a "cutoff" value of δ such that the threshold contract is more effective when δ is below the cutoff, and the linear contract is more effective when δ is above the cutoff.

(b) Naifs and Sophisticates for $1 \leq \lambda < 3$. In this case, the threshold contract is more effective than the linear contract for all δ, with the gap decreasing in δ.

(c) Sophisticates for $\lambda < 0.225$ and naifs and sophisticates for $\lambda > 2 + \sqrt{2}$. In this case, the linear contract is always more effective, with the gap increasing in δ.

To prove the result, we begin by calculating the optimal linear and threshold contracts. For both, we proceed in two steps: we first solve for the compliance, effectiveness, and cost-effectiveness of any given linear or threshold contract, and then we solve for the optimal contract. Finally, we compare the optimal linear and threshold contracts within different ranges of λ.

Linear Contract Compliance and Effectiveness: Consider a linear contract with payment level $\frac{M}{2}$. Substituting this into the formulas from Section 2, we have the following values for compliance, daily payment, cost-effectiveness, and effectiveness, respectively:

$$C = \frac{M}{2}$$
$$P = \frac{M^2}{4}$$
$$\frac{C}{P} = \frac{2}{M}$$
$$E = \lambda\frac{M}{2} - \frac{M^2}{4}$$

[74] See the beginning of the proof for specific ranges for both naifs and sophisticates.

Optimal Linear Contract: We want to choose the payment level to maximize contract effectiveness. The first-order condition for maximizing effectiveness is:

$$\frac{\partial E}{\partial M} = \frac{\lambda}{2} - \frac{M}{2} = 0$$

Denoting the arg max as M^{L*}, the payment level in the optimal linear contract is thus:

$$M^{L*} = \lambda$$

and the effectiveness of the optimal linear contract (which we will denote as E^{L*}) is:

$$\begin{aligned} E^{L*} &= \lambda \frac{M^{L*}}{2} - \frac{M^{L*2}}{4} \\ &= \lambda \frac{\lambda}{2} - \frac{\lambda^2}{4} \\ &= \frac{\lambda^2}{4} \end{aligned}$$

Threshold Contract Compliance and Effectiveness: We begin by solving for compliance, payments, and effectiveness in a two period threshold contract with payment level M. In the two-period IID threshold case, the agent complies in period 2 if they complied in period 1 and $e_2 < M$. Moreover, equation (10) implies that the agent will comply in period 1 if:

$$e_1 < E[(M - \delta e_2) w_{2,1} | w_1 = 1]. \tag{33}$$

Let $e_1^* = E[(M - \delta e_2) w_{2,1} | w_1 = 1]$ be the maximum effort cost that results in compliance. For naifs, for whom $w_{2,1}|^{(w_1=1)} = \mathbb{1}\{M - \delta e_2 > 0\}$,

$$\begin{aligned} e_1^* &= E[(M - \delta e_2) \mathbb{1}\{M - \delta e_2 > 0\}] \\ &= E[M - \delta e_2 | \delta e_2 < M] \times Prob(\delta e_2 < M) \\ &= (M - \delta E[e_2 | e_2 < M/\delta]) F(M/\delta) \end{aligned}$$

For sophisticates, for whom $w_{2,1}|^{(w_1=1)} = \mathbb{1}\{M - e_2 > 0\}$,

$$\begin{aligned} e_1^* &= E[(M - \delta e_2) \mathbb{1}\{M - e_2 > 0\}] \\ &= E[M - \delta e_2 | e_2 < M] \times Prob(e_2 < M) \\ &= (M - \delta E[e_2 | e_2 < M]) F(M) \end{aligned}$$

Compliance and payments are functions of e_1^*:

$$\begin{aligned} C &= .5[F(e_1^*) + F(e_1^*) F(M)] \\ P &= .5 M F(e_1^*) F(M) \end{aligned}$$

Effectiveness depends on the size of M and δ. When $\mathbf{0 < M < 1}$, we explore two cases for naifs and a single case for sophisticates based on the relative size of δ.:

Case 1: $0 < M < \delta < 1$ for Naifs In this case, $E[e_2|e_2 < M/\delta] = \frac{M}{2\delta}$, giving the following values for e_1^*, C, and P :

$$\begin{aligned} e_1^* &= (M - \delta\frac{M}{2\delta})\frac{M}{\delta} \\ &= \frac{M^2}{2\delta} \\ C &= .5\left[\frac{M^2}{2\delta} + \frac{M^3}{2\delta}\right] \\ P &= .5\frac{M^4}{2\delta} \end{aligned}$$

Thus, cost-effectiveness and effectiveness, respectively, are:

$$\frac{C}{P} = \frac{1+M}{M^2}$$

and

$$E = .5\lambda\left[\frac{M^2}{2\delta} + \frac{M^3}{2\delta}\right] - .5\frac{M^4}{2\delta}$$

Case 2: $0 < \delta < M < 1$ for Naifs In this case, because $M > \delta e_2$, the value e_1^* is:

$$e_1^* = E[(M - \delta e_2)\mathbb{1}\{M - \delta e_2 > 0\}] = E[M - \delta e_2] = M - \delta/2$$

This yields compliance and payments of:

$$\begin{aligned} C &= .5(M - \delta/2)(1 + M) \\ P &= .5(M - \delta/2)M^2 \end{aligned}$$

This gives cost-effectiveness and effectiveness, respectively, of:

$$\frac{C}{P} = \frac{1+M}{M^2}$$

and

$$E = .5\lambda(M - \delta/2)(1 + M) - .5(M - \delta/2)M^2.$$

Case 3: $0 < M < 1$ for Sophisticates In this case, the value e_1^* is:

$$e_1^* = \left(M - \delta\frac{M}{2}\right)M = M^2(1 - \delta/2)$$

So compliance and payments are:

$$\begin{aligned} C &= .5(M^2 + M^3)(1 - \delta/2) \\ P &= .5(M^4)(1 - \delta/2) \end{aligned}$$

and cost-effectiveness and effectiveness, respectively, are:

$$\frac{C}{P} = \frac{1+M}{M^2}$$

and

$$E = .5\lambda(M^2 + M^3)(1-\delta/2) - .5(M^4)(1-\delta/2).$$

For larger values of M, such that $\mathbf{1 \leq M < 2}$, naifs and sophisticates behave the same way. We consider two more cases.

Case 4: $1 < M < 1 + \delta/2$ for Naifs and Sophisticates In this case, because $M > \delta e_2$ and $M > e_2$, the value e_1^* is:

$$e_1^* = M - \delta/2$$

Furthermore, because $M - \delta/2 < 1$, compliance and payments are:

$$\begin{aligned} C &= (M - \delta/2) \\ P &= .5M(M - \delta/2) \end{aligned}$$

giving cost-effectiveness and effectiveness, respectively, of

$$\frac{C}{P} = \frac{2}{M}$$

and

$$E = \lambda(M - \delta/2) - .5M(M - \delta/2).$$

Case 5: $1 + \delta/2 < M < 2$ for Naifs and Sophisticates Again, because $M > \delta e_2$ and $M > e_2$, the value e_1^* is:

$$e_1^* = M - \delta/2$$

Because in this case $M - \delta/2 > 1$, compliance and payments are:

$$\begin{aligned} C &= 1 \\ P &= .5M \end{aligned}$$

giving cost-effectiveness and effectiveness, respectively, of

$$\frac{C}{P} = \frac{2}{M}$$

and

$$E = \lambda - .5M.$$

Having solved for compliance, payments, and effectiveness for naifs and sophisticates and for all M between 0 and 2, we now derive the payment level of the optimal threshold contract, which we denote as M^{T*}, and its effectiveness, which we denote as E^{T*}. We first consider sophisticates and then naifs.

Optimal threshold contract for sophisticates:
Aggregating cases 3–5 above, we have that effectiveness for sophisticates is as follows:

$$E = \begin{cases} .5(1-\delta/2)\left(\lambda(M^2+M^3)-M^4\right) & \text{if } M<1 \\ \lambda(M-\delta/2)-M^2/2+\delta M/4 & \text{if } 1\le M<1+\delta/2 \\ \lambda - M/2 & \text{if } 1+\delta/2\le M \end{cases}$$

The derivative of effectiveness with respect to the payment level M is:

$$\frac{\partial E}{\partial M} = \begin{cases} .5(1-\delta/2)\left(\lambda(2M+3M^2)-4M^3\right) & \text{if } M<1 \\ \lambda - M+\delta/4 & \text{if } 1\le M<1+\delta/2 \\ -1/2 & \text{if } 1+\delta/2\le M \end{cases}$$

The payment level of the optimal threshold contract, M^{T*}, will set this derivative equal to zero. Note that if $1+\delta/2 \le M$, it follows that $\frac{\partial E}{\partial M} < 0$ (since $M = 1+\delta/2$ achieves full compliance). Hence, M^{T*} is always smaller than $1+\delta/2$. However, the exact value of M^{T*} depends on the value of λ. We consider three cases, (A) - (C).

Case A: $\lambda \ge 1+\delta/4$
In this case, we have that $\frac{\partial E}{\partial M}|^{1\le M<1+\delta/2} = \lambda - M + \delta/4 > 0$ for $1 \le M < 1+\delta/2$. In addition, $\frac{\partial E}{\partial M}|^{M<1} = .5(1-\delta/2)\left(\lambda(2M+3M^2)-4M^3\right)$ is always positive.[75] Combined with the fact that $\frac{\partial E}{\partial M}|^{M>1+\delta/2} < 0$, the optimal payment is:

$$M^{T*}|^{\lambda>1+\delta/4} = 1+\delta/2.$$

and the effectiveness of the optimal threshold contract is

$$\begin{aligned} E^{T*}|^{\lambda>1+\delta/4} &= \lambda - M^*/2 \\ &= \lambda - .5 - \delta/4 \end{aligned}$$

Case B: $\lambda < 1-\delta/4$
In this case, $\frac{\partial E}{\partial M}|^{1\le M<1+\delta/2} = \lambda - M+\delta/4 < 0$ for all $1 \le M < 1+\delta/2$. Recall that $\frac{\partial E}{\partial M}|^{M>1+\delta/2} < 0$ in all cases. Hence $\frac{\partial E}{\partial M}|^{M>1} < 0$, which implies that the optimum must have $M \le 1$.

We hence set the $\frac{\partial E}{\partial M}|^{M<1} = 0$, which yields:

$$\frac{\partial E}{\partial M}|^{M<1} = .5(1-\delta/2)\left(\lambda(2M+3M^2)-4M^3\right) = 0$$

which implies

$$\lambda(2M+3M^2)-4M^3 = 0$$

[75]This is because, given $\lambda \ge 1$, the function $\lambda(2M+3M^2)-4M^3$ increases at $M=0$ and is never 0 in (0,1].

or that

$$\lambda(2+3M) - 4M^2 = 0$$

The solution to this quadratic is:

$$M = \lambda\left(\frac{3}{8} + \sqrt{\frac{9}{64} + \frac{1}{2\lambda}}\right)$$

This M falls in the region $M < 1$ whenever $\lambda < \frac{4}{5}$. When $\lambda \geq \frac{4}{5}$, $\frac{\partial E}{\partial M}|^{M<1} > 0$ for all $M < 1$, which (combined with the fact that $\frac{\partial E}{\partial M}|^{M>1} < 0$) implies that the optimal M must be at the "kink point" where M=1:

$$M^{T*}|^{\lambda<1-\delta/4 \ \& \ \lambda>4/5} = 1$$

Note that having $\lambda \geq \frac{4}{5}$ while $\lambda < 1 - \delta/4$ implies a relatively low δ.

Thus we have:

$$M^{T*}|^{\lambda<1-\delta/4} = \begin{cases} \lambda\left(\frac{3}{8} + \sqrt{\frac{9}{64} + \frac{1}{2\lambda}}\right) & \text{if } \lambda < 4/5 \& \lambda < 1-\delta/4 \\ 1 & \text{if } \lambda \geq 4/5 \& \lambda < 1-\delta/4 \end{cases}$$

This implies that maximized effectiveness when $\lambda < 4/5$ is:

$$\begin{aligned} E^{T*}|^{\lambda<1-\delta/4 \ \& \ \lambda<4/5} &= .5(1-\delta/2)\left(\lambda(M^2+M^3) - M^4\right)|^{M=\lambda\left(\frac{3}{8}+\sqrt{\frac{9}{64}+\frac{1}{2\lambda}}\right)} \\ &= \frac{1}{16}(2-\delta)\left(\frac{3}{8} + \frac{1}{8}\sqrt{9+\frac{32}{\lambda}}\right)^2 \lambda^3 \left(4 + \lambda\left(\frac{15}{16} + \frac{1}{16}\left(-9 - \frac{32}{\lambda}\right) + \frac{1}{8}\sqrt{9+\frac{32}{\lambda}}\right)\right) \end{aligned}$$

When $\lambda \geq 4/5$, maximized effectiveness is:

$$\begin{aligned} E^{T*}|^{\lambda<1-\delta/4 \ \& \ \lambda\geq4/5} &= \lambda(M-\delta/2) - M^2/2 + \delta M/4|^{M=1} \\ &= \lambda(1-\delta/2) - 1/2 + \delta/4 \\ &= \lambda - 1/2 - \delta(\lambda/2 - 1/4) \end{aligned}$$

Note that both of these are decreasing in δ (where the latter holds because $\lambda/2 - 1/4 > 0$ when $\lambda > 4/5$.

Case C: $1 - \delta/4 \leq \lambda < 1 + \delta/4$ In this case, we have that $\frac{\partial E}{\partial M}|^{1\leq M<1+\delta/2} = \lambda - M + \delta/4 = 0$ somewhere in the region of $1 \leq M < 1 + \delta/2$—that is, there is a local max in this region.

There are two subcases.

Subcase C(i): $1 - \delta/4 \leq \lambda < 1 + \delta/4$ **and** $\lambda \geq 4/5$

If $\lambda \geq 4/5$, then $\frac{\partial E}{\partial M}|^{M<1} > 0$, which means that the optimum must be the local max in the region of $1 \leq M < 1 + \delta/2$.

We thus solve for this local maximum by finding the M at which $\frac{\partial E}{\partial M}|^{1\leq M<1+\delta/2}$ is 0:

$$\frac{\partial E}{\partial M}|^{1\leq M<1+\delta/2} = \lambda - M^* + \delta/4 = 0$$

which implies that

$$M^* = \lambda + \delta/4$$

which means that

$$\begin{aligned} E^{T*}\big|^{1-\delta/4<\lambda<1+\delta/4 \text{ \& } \lambda>4/5} &= \lambda(M^* - \delta/2) - M^{*2}/2 + \delta M^*/4 \\ &= \lambda(\lambda + \delta/4 - \delta/2) - (\lambda + \delta/4)^2/2 + \delta(\lambda + \delta/4)/4 \\ &= \lambda^2 - \lambda\delta/4 - \lambda^2/2 - \lambda\delta/4 - \delta^2/32 + \lambda\delta/4 + \delta^2/16 \\ &= \lambda^2/2 - \lambda\delta/4 + \delta^2/32 \end{aligned}$$

Note again that this is decreasing in δ for all $\lambda > 4/5$ and $\delta \leq 1$.[76]

Subcase C(ii): $1 - \delta/4 \leq \lambda < 1 + \delta/4$ **and** $\lambda < 4/5$

In this case, there are two local maxima: one when $M < 1$ and one when $1 \leq M < 1+\delta/2$. The global maximum thus is the larger of those two values:

$$E^{T*} = \max\Bigg\{ \lambda^2/2 - \lambda\delta/4 + \delta^2/32, \\ \frac{1}{16}(2-\delta)\left(\frac{3}{8} + \frac{1}{8}\sqrt{9 + \frac{32}{\lambda}}\right)^2 \lambda^3 \left(4 + \lambda\left(\frac{15}{16} + \frac{1}{16}\left(-9 - \frac{32}{\lambda}\right) + \frac{1}{8}\sqrt{9 + \frac{32}{\lambda}}\right)\right)\Bigg\}$$

We next aggregate the cases into a single solution for the effectiveness of the most effective threshold for sophisticates as a function of δ. We then compare the most effective threshold and linear contracts as δ changes. However, the solution function depends on λ.

Threshold vs. Linear Effectiveness with $\lambda \geq 4/5$.

When $\lambda \geq 4/5$, we aggregate the effectiveness function of the optimal threshold contract from cases A-C as:

$$E^{T*}\big|^{\lambda\geq 4/5} = \begin{cases} \lambda - 1/2 - \delta(\lambda/2 - 1/4) & \text{if } \lambda < 1 - \delta/4 \\ \lambda^2/2 - \lambda\delta/4 + \delta^2/32 & \text{if } 1 - \delta/4 \leq \lambda < 1 + \delta/4 \\ \lambda - .5 - \delta/4 & \text{if } \lambda > 1 + \delta/4. \end{cases}$$

We can rewrite effectiveness more transparently as a function of δ. If $4/5 \leq \lambda < 1$, we have:

$$E^{T*} = \begin{cases} \lambda - 1/2 - \delta(\lambda/2 - 1/4) & \text{if } \delta < 4(1-\lambda) \\ \lambda^2/2 - \lambda\delta/4 + \delta^2/32 & \text{if } \delta \geq 4(1-\lambda) \end{cases}$$

and if $1 \leq \lambda$, we have

$$E^{T*} = \begin{cases} \lambda^2/2 - \lambda\delta/4 + \delta^2/32 & \text{if } \delta > 4(\lambda - 1) \\ \lambda - .5 - \delta/4 & \text{if } \delta \leq 4(\lambda - 1) \end{cases}$$

[76] This is because the function $-\lambda/4 + \delta/16$ is negative for all $\lambda \geq 4/5$ as long as $\delta < 16/5$.

Note that each of these functions is continuous in δ. Moreover, because each segment is decreasing in δ, we achieve the important result: $\frac{\partial E^{T*}}{\partial \delta} < 0$. That is, the effectiveness of the most effective threshold contract is decreasing in δ.

Now we compare the effectiveness of the optimal threshold and linear contracts in the region $\lambda \geq 4/5$. First consider the case where $4/5 \leq \lambda < 1$. For $\delta < 4(1-\lambda)$, $E^{T*} > E^{L*}$ would require $\delta > \frac{\lambda^2/4-\lambda+1/2}{1/4-\lambda/2}$, but this value is greater than $4(1-\lambda)$ for $4/5 \leq \lambda < 1$. So the linear contract is always more effective if $\delta < 4(1-\lambda)$. For $\delta \geq 4(1-\lambda)$, in order for $E^{T*} > E^{L*} = \lambda^2/4$, it would require that $\lambda^2/2 - \lambda\delta/4 + \delta^2/32 > \lambda^2/4$ or $\delta < (4-2\sqrt{2})\lambda$. Since $(4-2\sqrt{2})\lambda > 1$ if $\lambda > \frac{1}{4-2\sqrt{2}} \approx 0.85$, the threshold contract will always be more effective for $\lambda > 0.85$ and $\delta \geq 4(1-\lambda)$. And then for $\lambda \leq 0.85$, which contract is more effective depends on the exact value of δ.

In case where $\lambda \geq 1$, if $\delta > 4(\lambda-1)$, $E^{T*} > E^{L*}$ would require $\delta < (4-2\sqrt{2})\lambda$, which is always true for $\lambda \geq 1$. If $\delta \leq 4(\lambda-1)$, $E^{T*} > E^{L*}$ would require $\delta < -\lambda^2+4\lambda-2$. This holds for all $\delta \in [0,1]$ if $\lambda < 3$, for some δ if $3 \leq \lambda < 2+\sqrt{2}$, and no δ if $\lambda \geq 2+\sqrt{2}$.

Threshold vs. Linear Effectiveness with $\lambda < 4/5$ Now, we write the effectiveness of the optimal threshold contract as a function of λ and δ when $\lambda < 4/5$.

Let $\xi(\lambda) = \frac{1}{16}\left(\frac{3}{8}+\frac{1}{8}\sqrt{9+\frac{32}{\lambda}}\right)^2 \lambda^3 \left(4+\lambda\left(\frac{15}{16}+\frac{1}{16}\left(-9-\frac{32}{\lambda}\right)+\frac{1}{8}\sqrt{9+\frac{32}{\lambda}}\right)\right)$. Then we have

$$E^{T*}|^{\lambda<4/5.} = \begin{cases} (2-\delta)\,\xi(\lambda) & \text{if } \lambda < 1-\delta/4 \\ \max\left\{(2-\delta)\,\xi(\lambda), \lambda^2/2 - \lambda\delta/4 + \delta^2/32\right\} & \text{if } 1-\delta/4 \leq \lambda \end{cases}$$

or equivalently:

$$E^{T*}|^{\lambda<4/5.} = \begin{cases} (2-\delta)\,\xi(\lambda) & \text{if } \delta < 4(1-\lambda) \\ \max\left\{(2-\delta)\,\xi(\lambda), \lambda^2/2 - \lambda\delta/4 + \delta^2/32\right\} & \text{if } 4(1-\lambda) \leq \delta \end{cases}$$

If $\lambda < 0.75$, then $\delta < 4(1-\lambda)$ and so we have the $E^{T*} = (2-\delta)\,\xi(\lambda)$. This function is continuous and decreasing in δ.

Threshold effectiveness will in this case be higher than linear if

$$(2-\delta)\,\xi(\lambda) > \lambda^2/4.$$

This implies that threshold effectiveness is higher if

$$\delta < 2 - \frac{\lambda^2}{4\xi(\lambda)}.$$

Since the function $2 - \frac{\lambda^2}{4\xi(\lambda)}$ is negative for $\lambda \leq 0.225$, the linear contract is always more effective for this range of λ. For $\lambda > 0.225$, there is a cutoff value for δ where the optimal threshold contract is more effective for δ below the threshold.

If $0.75 \leq \lambda < 0.8$, we need some additional analysis on the function E^{T*}. Both $(2-\delta)\,\xi(\lambda)$ and $\lambda^2/2 - \lambda\delta/4 + \delta^2/32$ are continuous for $\delta \in [0,1]$, and E^{T*} is continuous at $\delta = 4(1-\lambda)$ since $(2-\delta)\,\xi(\lambda) > \lambda^2/2 - \lambda\delta/4 + \delta^2/32$ at $\delta = 4(1-\lambda)$. Also the maximum of two continuous functions is continuous, so E^{T*} is continuous in δ. Then if $E^{T*} > E^{L*}$ when $\delta = 0$ and $E^{T*} < E^{L*}$ when $\delta = 1$, there is some threshold value of δ for which the linear and threshold contracts will have the same effectiveness, and above that the threshold will have higher effectiveness and below that the linear will. This is true as long as $\lambda > 0.225$, which holds for all λ in this interval. So again there is a cutoff value for δ below which the threshold contract is more effective.

Optimal threshold contract for naifs: Again using the formulas from the proof of Proposition 8a, we have that effectiveness is as follows:

$$E = \begin{cases} .5\lambda(M-\delta/2)(1+M) - .5(M-\delta/2)M^2 & \text{if } \delta < M < 1 \\ \lambda(M-\delta/2) - M^2/2 + \delta M/4 & \text{if } 1 \leq M < 1+\delta/2 \\ \lambda - M/2 & \text{if } 1+\delta/2 \leq M \end{cases}$$

The derivative of effectiveness w.r.t. M is hence:

$$\frac{\partial E}{\partial M} = \begin{cases} 0.5\lambda\left[\frac{M}{\delta} + \frac{3M^2}{2\delta}\right] - \frac{M^3}{\delta} & \text{if } M \leq \delta \\ -0.5M^2 + 0.5\lambda(1+M) + 0.5\lambda(-(\delta/2)+M) - M(-(\delta/2)+M) & \text{if } \delta < M < 1 \\ \lambda - M + \delta/4 & \text{if } 1 \leq M < 1+\delta/2 \\ -1/2 & \text{if } 1+\delta/2 \leq M \end{cases}$$

Note that this is the same as sophisticates when $M \geq 1$.

Again we derive the payment and effectiveness of the optimal contract based on the value of λ. We consider two cases, (D) and (E).

Case D: $\lambda \geq 4/5$

When $\lambda > 4/5$, $\frac{\partial E}{\partial M} > 0$ for all $M \leq \delta$, and we have the following cases:

- If $\lambda \geq \frac{1.5-\delta/2}{1.5-\delta/4}$, $\frac{\partial E}{\partial M} > 0$ for all $M < 1$ and the sophisticate results go through. Note that $\lambda \geq \frac{1.5-\delta/2}{1.5-\delta/4}$ implies $\lambda \geq 1-\delta/4$.
- If $\lambda < 1-\frac{\delta}{4}$, $\lambda < \frac{1.5-\delta/2}{1.5-\delta/4}$. $\frac{\partial E}{\partial M} < 0$ for $M \geq 1$ and also for some $M \in (\delta, 1)$, so there is an optimum in $(\delta, 1)$ and it is global, so the sophisticate results go through as well.
- If $1-\frac{\delta}{4} \leq \lambda < \frac{1.5-\delta/2}{1.5-\delta/4}$, there are two local optima, one in $(\delta, 1)$ and another in $[1, 1+\frac{\delta}{2})$, so the global optimum is the maximum between the two. Also threshold efficiency is decreasing in δ for a given λ, and increasing in λ for a given δ. Also, at $\lambda = 4/5$, there is a cutoff value of δ when linear contract becomes more effective. So we can let $\delta = \frac{1.5-1.5\lambda}{1/2-\lambda/4}$, and solve for the λ value such that $E^{T*} = E^{L*}$, and the solution is $\lambda = 0.81$. So there is a cutoff value of δ for when linear contract becomes more effective if $\lambda < 0.81$; otherwise the threshold contract is always more effective.

Case E: $\lambda < 4/5$

From the discussion of sophisticates, we know in this case that if $\lambda < 1-\delta/4$, the optimum will have $M < 1$; if $1-\delta/4 \leq \lambda < 4/5$, there will be another local optimum in $[1, 1+\delta/2)$, and the global optimum will be the maximum between the two. Explicitly, in case $\lambda < 1-\delta/4$, we have

$$M^* = \begin{cases} \frac{\frac{3}{4}\lambda + \sqrt{\frac{9}{16}\lambda^2 + 2\lambda}}{2} & \text{if } M^* \leq \delta \\ \frac{\lambda + \delta/2 + \sqrt{\lambda^2 - \frac{1}{2}\delta\lambda + 3\lambda + \frac{\delta^2}{4}}}{3} & \text{if } M^* > \delta \end{cases}$$

Let $\delta^* = \frac{\frac{3}{4}\lambda + \sqrt{\frac{9}{16}\lambda^2 + 2\lambda}}{2}$. This turns out to be the solution for $\delta = \frac{\lambda + \delta/2 + \sqrt{\lambda^2 - \frac{1}{2}\delta\lambda + 3\lambda + \frac{\delta^2}{4}}}{3}$, so we have

$$M^* = \begin{cases} \frac{\frac{3}{4}\lambda + \sqrt{\frac{9}{16}\lambda^2 + 2\lambda}}{2} & \text{if } \delta \geq \delta^* \\ \frac{\lambda + \delta/2 + \sqrt{\lambda^2 - \frac{1}{2}\delta\lambda + 3\lambda + \frac{\delta^2}{4}}}{3} & \text{if } \delta < \delta^* \end{cases}$$

So

$$E^{T*} = \begin{cases} .5\lambda \left[\frac{M^{*2}}{2\delta} + \frac{M^{*3}}{2\delta}\right] - .5\frac{M^{*4}}{2\delta} & \text{if } \delta \geq \delta^* \\ .5\lambda(M^* - \delta/2)(1 + M^*) - .5(M^* - \delta/2)M^{*2} & \text{if } \delta < \delta^* \end{cases}$$

When $1-\delta/4 \leq \lambda < 4/5$, there is another optimum at $M = \lambda+\delta/4 \in [1, 1+\delta/2)$. For simplicity, let $E_1 = .5\lambda \left[\frac{M^{*2}}{2\delta} + \frac{M^{*3}}{2\delta}\right] - .5\frac{M^{*4}}{2\delta}$, $E_2 = .5\lambda(M^*-\delta/2)(1+M^*) - .5(M^*-\delta/2)M^{*2}$, and $E_3 = \frac{1}{2}\lambda - \frac{\delta\lambda}{4} - \frac{\delta^2}{16}$ which is $\lambda(M - \delta/2) - M^2/2 + \delta M/4$ evaluated at $\lambda + \delta/4$. We have

$$E^{T*} = \begin{cases} \max\{E_1, E_3\} & \text{if } \delta \geq \delta^* \\ \max\{E_2, E_3\} & \text{if } \delta < \delta^* \end{cases}$$

We know that E^{T*} is continuous in δ since the maximum of a function is continuous in the parameter if it is maximized on a compact domain, and in this case we are considering $M \in [0, 1 + \delta/2]$. So we can compare E^{T*} and E^{L*} by analyzing their values at $\delta = 0$ and $\delta = 1$. If $E^{T*} < E^{L*}$ at one endpoint and $E^{T*} > E^{L*}$ at another, we can conclude that there is a cutoff δ where threshold contract becomes more effective beyond.

If $\delta = 0$, then $E = 0.5\lambda(M - \delta/2)(1 + M) - 0.5(M - \delta/2)M^2$. This function is maximized on the region from $0 < M < 1$ (i.e., $\frac{\partial E}{\partial M} = 0$) when $M = \frac{1}{6}\left(\delta + 2\lambda + \sqrt{\delta^2 + 12\lambda - 2\delta\lambda + 4\lambda^2}\right)$. The corresponding maximized value of effectiveness is greater than the effectiveness of the optimal linear contract, $\lambda^2/4$, when $\delta = 0$, for all $\lambda > 0$.

If $\delta = 1$, then $E = \max\{E_1, E_3\}$, which is less than the effectiveness of the optimal linear contract, $\lambda^2/4$ for all λ.

Hence, we have that maximized effectiveness from the threshold is greater than maximized effectiveness from the linear, $E^{T*} > E^{L*}$, when $\delta = 0$, while the opposite is true, $E^{T*} < E^{L*}$, when $\delta = 1$. Since maximized effectiveness is continuous in δ,[77] this implies that there is a cutoff δ for which the effectiveness of the optimal threshold is the same as the effectiveness of the optimal linear, and that the effectiveness of the optimal threshold is above the linear for δ below the threshold (and vice verse for δ above the threshold).

□

H.3 Supporting Propositions for Section B.7 Discussion

We now present and prove the propositions described in Section B.7.

Proposition 9 (Adding discounted health benefit)**.** *Let $d^{(t)} = \delta$ for all $t > 1$ and let $d^{(t)} = 1$ for all t. Let there be perfect correlation in costs across days. In addition, assume that, in addition to a present cost, compliance has a future benefit b that the participant discounts with the discount factor*

[77]This follows because $.5\lambda\left[\frac{M^2}{2\delta} + \frac{M^3}{2\delta}\right] - .5\frac{M^4}{2\delta} = .5\lambda(M - \delta/2)(1 + M) - .5(M - \delta/2)M^2$ when $M = \delta$.

over effort, δ. Fix all parameters other than δ. Take any threshold contract with threshold level $K = T$ and threshold payment $m'T$. Take any linear contract with daily payment $m \leq m'T$.

(a) If $m' \geq \frac{b}{T-1}$,[78] *compliance in the threshold contract is decreasing in δ. Because compliance in the linear compliance is increasing in δ, compliance in the threshold relative to linear contract is decreasing in δ.*

(b) If $\frac{b}{T-1} > m' \geq \frac{b(1-\delta)}{T}$,[79] *then compliance in the threshold contract is increasing in δ. Whether compliance in the threshold relative to linear contract is decreasing or increasing in δ depends on the parameters of the cost distribution.*

Proof. For notational simplicity, denote the threshold payment $m'T$ as M. In period 1, the participant will comply if (a) the present discounted benefits outweigh costs, and (b) they expect to follow-through in all future periods. In period 1, condition (a) requires:

$$e - \delta b + (T-1)(\delta e - \delta b) < M \tag{34}$$

$$e + \delta(T-1)e < \delta T b + M \tag{35}$$

$$e < \frac{\delta T b + M}{1 + (T-1)\delta} \tag{36}$$

In period j, condition (a) requires:

$$e - \delta b + (T-j)(\delta e - \delta b) < M \tag{37}$$

$$e + \delta(T-j)e < \delta(T-j+1)b + M \tag{38}$$

$$e < \frac{\delta(T-j+1)b + M}{1 + (T-j)\delta} \tag{39}$$

Because people sink costs as they go but the marginal incentive M remains constant, under relatively broad conditions, people will be more likely to want to comply (i.e., condition (a) will be more likely to hold) if j increases. Specifically, condition (a) will hold in period $j+1$ if it holds in period j as long as $M > b(1-\delta)$.[80]

Thus, assuming $M \geq b(1-\delta)$, compliance will be 100% if $e < \frac{\delta T b + M}{1+(T-1)\delta}$, and 0% otherwise. For the remainder of the proof, for simplicity, we thus assume $M \geq b(1-\delta)$ (or, alternatively, that $m \geq \frac{b(1-\delta)}{T}$),

[78] Equivalently, if $m' \geq \frac{b}{K-1}$, since $T = K$.

[79] Equivalently, if $\frac{b}{K-1} > m' \geq \frac{b(1-\delta)}{K}$. The final case, where $m' < \frac{b(1-\delta)}{T}$, is more complex to analyze as participants no longer necessarily comply in all periods if they comply in the first.

[80] We show here that the threshold from equation 39 for period j will be less than the corresponding threshold for period $j+1$ if $M > b(1-\delta)$. That is, we want to show that the following inequality holds if $M > b(1-\delta)$:

$$\frac{\delta(T-j+1)b + M}{1 + (T-j)\delta} < \frac{\delta(T-j)b + M}{1 + (T-j-1)\delta}$$

Since $0 < \delta \leq 1$ and $j \leq T$, both denominators are positive, and so this becomes:

$$[\delta(T-j+1)b + M][1 + (T-j-1)\delta] < [\delta(T-j)b + M][1 + (T-j)\delta] \tag{40}$$

Expanding the left side of equation 40:

$$\begin{aligned}&[\delta(T-j+1)b + M][1 + (T-j-1)\delta] \\ &= \delta(T-j+1)b + \delta^2(T-j+1)(T-j-1)b + M + M(T-j-1)\delta\end{aligned}$$

and hence parts (a) and (b) of the proposition only address the case where $M \geq b(1-\delta)$.

Defining

$$f(\delta) = \frac{\delta T b + M}{1 + (T-1)\delta}$$

as the cutoff value as a function of δ, we first check whether the cutoff value is increasing or decreasing in δ.

Using the quotient rule:

$$f'(\delta) = \frac{Tb \cdot [1 + (T-1)\delta] - [\delta T b + M] \cdot (T-1)}{[1 + (T-1)\delta]^2} \tag{41}$$

Expanding the numerator:

$$Tb[1 + (T-1)\delta] - [\delta T b + M](T-1) \tag{42}$$
$$= Tb + Tb(T-1)\delta - \delta T b(T-1) - M(T-1) \tag{43}$$
$$= Tb + T^2 b\delta - Tb\delta - \delta T^2 b + \delta T b - M(T-1) \tag{44}$$
$$= Tb - M(T-1) \tag{45}$$

Therefore:

$$f'(\delta) = \frac{Tb - M(T-1)}{[1 + (T-1)\delta]^2} \tag{46}$$

For $f'(\delta) < 0$, we require the numerator to be negative (as the denominator is always positive):

$$Tb - M(T-1) < 0 \tag{47}$$

Expanding the right side of equation 40:

$$[\delta(T-j)b + M][1 + (T-j)\delta]$$
$$= \delta(T-j)b + \delta^2(T-j)^2 b + M + M(T-j)\delta$$

Substituting back into equation 40:

$$\delta(T-j+1)b + \delta^2(T-j+1)(T-j-1)b + M + M(T-j-1)\delta$$
$$< \delta(T-j)b + \delta^2(T-j)^2 b + M + M(T-j)\delta$$

Rearranging and cancelling like terms:

$$\delta b[(T-j+1) - (T-j)] + \delta^2 b[(T-j+1)(T-j-1) - (T-j)^2]$$
$$+ M\delta[(T-j-1) - (T-j)] < 0$$

Simplifying:

$$\delta b + \delta^2 b[(T-j+1)(T-j-1) - (T-j)^2] - M\delta < 0$$

To compute the quadratic term, $(T-j+1)(T-j-1) - (T-j)^2$, set $u = T - j$ for clarity. Then we have $(u+1)(u-1) - u^2 = u^2 - 1 - u^2 = -1$.

Substituting back in, our inequality becomes: $\delta b - \delta^2 b - M\delta < 0$, or, since $\delta > 0$, $b - \delta b - M < 0$, which is equivalent to $b(1-\delta) < M$.

This is equivalent to:

$$M > \frac{T}{T-1}b \tag{48}$$

or alternatively, representing M in terms of the per-period payment m' ($M = m'T$), we have

$$m' > \frac{b}{T-1}. \tag{49}$$

So, if $m' > \frac{b}{T-1}$, the derivative of the cutoff—and hence the derivative of threshold compliance—with respect to δ is negative. Since the derivative of linear compliance with respect to δ is positive, then the derivative of compliance in the threshold relative to linear contract is also negative. In contrast, if $m' < \frac{b}{T-1}$, the derivative of threshold compliance, like linear compliance, is negative, and so the derivative of relative compliance is ambiguous, dependent on parameters.

□

Proposition 10 (Adding discounted health benefit, T=2)**.** *Let $T = 2$ and let $d = 1$. Fix all parameters other than δ. Let there be perfect correlation in costs across days. In addition, assume that, in addition to a present cost, compliance has a future benefit b that the participant discounts with the discount factor over effort, δ. Take a threshold contract with threshold level $K = 2$ and payment level m'. Compare it with a linear contract with payment level m, with $m \leq Tm'$. We refer to compliance in the threshold contract relative to the linear contract as "relative compliance." Then:*

(a) If $m' > b$, relative compliance is strictly decreasing in δ for all cost distributions.

(b) Additionally, if costs are uniformly distributed across people U[0,$\bar{e}$], with $\bar{e} > Tm' + b$,[81] then:

(i) For sophisticates: If $m' > \frac{b(1-\delta)}{T}$, then relative compliance is strictly decreasing in δ. If $m' < \frac{b(1-\delta)}{T}$, then relative compliance is flat in δ.

(ii) For naifs: If $m' > \frac{b\left(2-(\delta+1)^2\right)}{T}$, then relative compliance is strictly decreasing in δ. Since $\frac{b\left(2-(\delta+1)^2\right)}{T} < \frac{b(1-\delta)}{T}$, relative compliance is decreasing over a larger range of δ for naifs than sophisticates. If $m' < \frac{b\left(2-(\delta+1)^2\right)}{T}$, then relative compliance is increasing in δ.

Proof. **Threshold Contract** For notational simplicity, we denote the threshold payment $2m'$ as M. We work backwards. On day 2:

- If $w_1 = 1$, the participant will comply if $e < \delta b + M$.
- If $w_1 = 0$, the participant will comply if $e < \delta b$.

On day 1, the participant will comply if both:

(a) The discounted costs are less than the discounted benefits, $e(1+\delta) < 2\delta b + M$, which is equivalent to:

$$e < \frac{2\delta b + M}{1+\delta} \tag{50}$$

(b) they think their future self will follow-through. They think their future self will follow-through if:

[81] This is for tractability, and guarantees that no participant will have an 100% likelihood of complying in any period. Further, note that for this and all expressions in the Proposition, T can be replaced with K since they are assumed to be equal.

– Sophisticates: $e < \delta b + M$.

– Naifs: $\delta e < \delta b + M$, which holds whenever equation 50 holds.[82]

Sophisticates thus comply in periods 1 and 2 if both (a) $e < \frac{2\delta b+M}{1+\delta}$ and (b) $e < \delta b + M$, and so their compliance is 100% if $e < \min\left\{\frac{2\delta b+M}{(1+\delta)\bar{e}}, \frac{\delta b+M}{\bar{e}}\right\}$ and 0% otherwise.

Naifs thus comply in period 1 if $e < \frac{2\delta b+M}{1+\delta}$ and in period 2 if both $e < \frac{2\delta b+M}{1+\delta}$ and $e < \delta b + M$.

Given these compliance conditions, it is useful to know the conditions under which $\frac{2\delta b+M}{1+\delta}$ is greater than or less than $\delta b + M$:

$$\frac{2\delta b + M}{(1+\delta)} \overset{?}{<} \delta b + M$$

$$(2\delta b + M) \overset{?}{<} (\delta b + M)(1+\delta)$$

$$(2\delta b + M) \overset{?}{<} \delta b + M + \delta^2 b + \delta M$$

$$2\delta b \overset{?}{<} \delta b + \delta^2 b + \delta M$$

$$\delta b \overset{?}{<} \delta^2 b + \delta M$$

$$b \overset{?}{<} \delta b + M \tag{51}$$

$$b(1-\delta) \overset{?}{<} M \tag{52}$$

Thus, $e < \frac{2\delta b+M}{1+\delta}$ implies that $e < \delta b + M$ as long as $M > b(1-\delta)$. Thus, when $M > b(1-\delta)$, for both sophisticates and naifs, $e < \frac{2\delta b+M}{1+\delta}$ will be the single binding constraint. That is, participants will comply in both periods if $e < \frac{2\delta b+M}{1+\delta}$ holds and not otherwise. We thus analyze results first for the case where $M > b(1-\delta)$ before returning to the case where $M \leq b(1-\delta)$.

Case 1: $M > b(1-\delta)$

In this case, compliance is 100% if $e < \frac{2\delta b+M}{1+\delta}$ and 0% otherwise. Defining $f(\delta) = \frac{2\delta b+M}{1+\delta}$ as the cutoff value as a function of δ, we first check when the cutoff value is increasing or decreasing in δ, i.e., whether $f'(\delta)$ is positive or negative.

Using the quotient rule:

$$f'(\delta) = \frac{(1+\delta)(2b) - (2\delta b + M)(1)}{(1+\delta)^2} \tag{53}$$

$$= \frac{2b + 2\delta b - 2\delta b - M}{(1+\delta)^2} \tag{54}$$

$$= \frac{2b - M}{(1+\delta)^2} \tag{55}$$

Since $(1+\delta)^2 > 0$ for all $\delta \geq 0$, the sign of the derivative is determined by $(2b - M)$:

[82]If $e < \delta b$, then the person complies in all periods without incentives and so both equations hold. If $e > \delta b$, then they sink costs as they go and so the period 2 equation will hold whenever the period 1 equation holds.

- If $M < 2b$, then $f'(\delta) > 0$.
- If $M > 2b$, then $f'(\delta) < 0$.

Case 1a: $M > 2b$

Thus, if $M > 2b$, $f'(\delta) < 0$ and the cutoff value is decreasing with δ. That means compliance in the threshold will be decreasing with δ. Since we showed above that compliance in the linear contract is increasing with δ, this shows that compliance in the threshold relative to linear is decreasing with δ, thus proving part (a) of the proposition. This finding also follows directly from Proposition 9.

Case 1b: $2b \geq M > b(1-\delta)$

In this case, compliance is increasing with δ, and so one must make distributional assumptions to determine whether compliance in the threshold relative to linear contracts is increasing or decreasing with δ.

We assume for tractability that costs are uniformly distributed across people U$[0,\bar{e}]$, with $\bar{e} > M+b$. Note that this implies that $\bar{e} > \frac{2\delta b+M}{1+\delta}$, since $M+b > \frac{2\delta b+M}{1+\delta}$. As such, since threshold compliance is 100% when $e < \frac{2\delta b+M}{1+\delta}$, threshold compliance is $\frac{2\delta b+M}{(1+\delta)\bar{e}}$. In contrast, linear compliance is $\frac{\delta b+m}{\bar{e}}$.

Thus, compliance in the threshold relative to linear contract is:

$$\frac{2\delta b+M}{(1+\delta)\bar{e}} - \frac{\delta b+m}{\bar{e}} \tag{56}$$

The derivative of this w.r.t. δ is:

$$\begin{aligned}
\frac{\partial}{\partial\delta}\left(\frac{2\delta b+M}{(1+\delta)\bar{e}} - \frac{\delta b+m}{\bar{e}}\right) &= \frac{2b(1+\delta)\bar{e} - (2\delta b+M)\bar{e}}{(1+\delta)^2\bar{e}^2} - \frac{b}{\bar{e}} &(57)\\
&= \frac{2b+2b\delta-2\delta b-M}{(1+\delta)^2\bar{e}} - \frac{b}{\bar{e}} &(58)\\
&= \frac{2b-M}{(1+\delta)^2\bar{e}} - \frac{b}{\bar{e}} &(59)\\
&= \frac{2b-M-b(1+\delta)^2}{(1+\delta)^2\bar{e}} &(60)\\
&= \frac{2b-M-b-2b\delta-b\delta^2}{(1+\delta)^2\bar{e}} &(61)\\
&= \frac{b-M-2b\delta-b\delta^2}{(1+\delta)^2\bar{e}} &(62)
\end{aligned}$$

We examine the conditions under which the derivative is negative. Since $\bar{e} > 0$ and $(1+\delta)^2 > 0$, the denominator is always positive. Therefore, the derivative is negative when the numerator is negative:

$$b - M - 2b\delta - b\delta^2 < 0$$

which means when:

$$b\delta^2 + 2b\delta + (M-b) > 0 \tag{63}$$

When $M > b$, all terms in the expression are positive (since $\delta \geq 0$, $b > 0$, and $M - b > 0$), making the expression always positive. Therefore, the derivative is always negative when $M > b$ (our maintained assumption in Case 1b).

When $M < b$, this expression may not be always positive. To solve for the transition point, we use the quadratic formula:

$$\delta = \frac{-2b \pm \sqrt{4b^2 - 4b(M-b)}}{2b} \tag{64}$$

$$= \frac{-2b \pm \sqrt{4b^2 + 4b^2 - 4bM}}{2b} \tag{65}$$

$$= \frac{-2b \pm \sqrt{8b^2 - 4bM}}{2b} \tag{66}$$

$$= \frac{-2b \pm 2\sqrt{2b^2 - bM}}{2b} \tag{67}$$

$$= -1 \pm \sqrt{\frac{2b - M}{b}} \tag{68}$$

Since $0 \leq \delta \leq 1$, we're interested in the solution: $\delta = -1 + \sqrt{\frac{2b-M}{b}}$. Moreover, for the derivative to be negative, we need:

$$\delta > -1 + \sqrt{\frac{2b - M}{b}} \tag{69}$$

This condition always holds when $M > b(1-\delta)$ (which is true for Case 1b).[83]

This means that throughout Case 1b, **the derivative of relative compliance in δ is negative.**

Case 2: $M \leq b(1-\delta)$, for Sophisticates and Naifs, Respectively

We now turn to the cases where $M \leq b(1-\delta)$. Since the threshold compliance conditions differ for naifs and sophisticates when $M \leq b(1-\delta)$, we analyze sophisticates and naifs in turn. In both cases, we maintain the assumption we made in Case 1b regarding uniform costs.

Case 2a: $M \leq b(1-\delta)$ for Sophisticates

For sophisticates, under our distributional assumptions, compliance in the threshold relative to linear contract is:

$$\min\left\{\frac{2\delta b + M}{(1+\delta)\bar{e}}, \frac{\delta b + M}{\bar{e}}\right\} - \frac{\delta b + m}{\bar{e}}$$

Moreover, equation 52 showed that, when $M \leq b(1-\delta)$ (as is the case in Case 2), then $\min\left\{\frac{2\delta b+M}{(1+\delta)\bar{e}}, \frac{\delta b+M}{\bar{e}}\right\} = \frac{\delta b+M}{\bar{e}}$.

Thus, relative compliance is:

$$\frac{\delta b + M}{\bar{e}} - \frac{\delta b + m}{\bar{e}} = \frac{M - m}{\bar{e}}$$

[83] $M > b(1-\delta)$ implies that $\delta > \frac{b-M}{b}$. Hence to show that equation 69 holds when $M > b(1-\delta)$, we can show that $\frac{b-M}{b} > -1 + \sqrt{\frac{2b-M}{b}}$. To do so, let $x = \frac{b-M}{b}$, which gives $\frac{2b-M}{b} = 1 + x$. We want to compare:

$$-1 + \sqrt{1+x} \overset{?}{<} x$$

This is true when $\sqrt{1+x} < 1 + x$, which is always true for $x > 0$. Since $x = \frac{b-M}{b} > 0$ for $M < b$, we have: $-1 + \sqrt{\frac{2b-M}{b}} < \frac{b-M}{b}$, as desired.

Taking the derivative:

$$\frac{d}{d\delta}\left[\frac{M-m}{\bar{e}}\right]=0$$

Therefore, in this range, **the function is constant with respect to δ.**

Case 2b: $M \leq b(1-\delta)$ for Naifs

Recall from our original discussion of thresholds that naifs comply in period 1 if $e < \frac{2\delta b+M}{1+\delta}$ and in period 2 if both $e < \frac{2\delta b+M}{1+\delta}$ and $e < \delta b + M$.

As a result, naif compliance under the threshold is:

$$F\left(\frac{2\delta b+M}{1+\delta}\right)\left(0.5+0.5F\left(\delta b+M \middle| e<\frac{2\delta b+M}{1+\delta}\right)\right)$$

Moreover, the value of the $F\left(\delta b+M \middle| e<\frac{2\delta b+M}{1+\delta}\right)$ term hinges on whether $\delta b+M$ is smaller or larger than $\frac{2\delta b+M}{1+\delta}$. When $M \leq b(1-\delta)$, we showed above that $\delta b+M < \frac{2\delta b+M}{1+\delta}$ (equation 52).

In this case, $F\left(\delta b+M \middle| e<\frac{2\delta b+M}{1+\delta}\right) = \frac{F(\delta b+M)}{F\left(\frac{2\delta b+M}{1+\delta}\right)}$[84] and so compliance is:

$$F\left(\frac{2\delta b+M}{1+\delta}\right)\left(0.5+0.5\frac{F(\delta b+M)}{F\left(\frac{2\delta b+M}{1+\delta}\right)}\right)=0.5F\left(\frac{2\delta b+M}{1+\delta}\right)+0.5F(\delta b+M) \tag{71}$$

Bringing in our uniform distributional assumption, this becomes:

$$0.5\left(\frac{2\delta b+M}{(1+\delta)\bar{e}}\right)+0.5\frac{\delta b+M}{\bar{e}}$$

And relative compliance under thresholds relative to linear becomes:

$$0.5\left(\frac{2\delta b+M}{(1+\delta)\bar{e}}\right)-0.5\frac{\delta b+M}{\bar{e}}$$

Note that this relative compliance function (when $\delta < \frac{b-M}{b}$) is just twice the relative compliance function when $\delta \geq \frac{b-M}{b}$ from equation 56; thus the derivatives in δ have the same sign. In the Case 2a section, equation 69 showed us that the derivative is negative when:

$$\delta > -1+\sqrt{\frac{2b-M}{b}}$$

And positive when:

$$\delta < -1+\sqrt{\frac{2b-M}{b}}$$

[84]This is because, by definition: $\mathrm{P}\left(e \leq \delta b+M \middle| e<\frac{2\delta b+M}{1+\delta}\right) = \frac{P\left(e\leq \delta b+M \text{ AND } e<\frac{2\delta b+M}{1+\delta}\right)}{P\left(e<\frac{2\delta b+M}{1+\delta}\right)}$

If $\delta b+M < \frac{2\delta b+M}{1+\delta}$, then $P\left(e \leq \delta b+M \text{ AND } e<\frac{2\delta b+M}{1+\delta}\right) = P(e \leq \delta b+M)$. As a result,

$$P\left(e \leq \delta b+M \middle| e<\frac{2\delta b+M}{1+\delta}\right) = \frac{P(e\leq \delta b+M)}{P\left(e<\frac{2\delta b+M}{1+\delta}\right)} = \frac{F(\delta b+M)}{F\left(\frac{2\delta b+M}{1+\delta}\right)} \tag{70}$$

Since we're analyzing a case where $M \leq b$, then $-1+\sqrt{\frac{2b-M}{b}}$ will always fall between 0 and 1. Moreover, when $M \leq b$, we also have that $-1+\sqrt{\frac{2b-M}{b}} < \frac{b-M}{b}$, so there will always be regions in case 2b when the derivative is positive and regions where it is negative. Specifically:

- When $\delta \geq -1+\sqrt{\frac{2b-M}{b}}$—or equivalently, when $M > b\left(2-(\delta+1)^2\right)$— **relative compliance is decreasing in** δ.
- When $\delta < -1+\sqrt{\frac{2b-M}{b}}$—or equivalently, when $M < b\left(2-(\delta+1)^2\right)$— **relative compliance is increasing in** δ.

□

Proposition 11 (Adding discounted health benefit, main effect, $T=2$, $m'=m$)**.** *Let $T=2$ and let $d=1$. Fix all parameters other than δ. Take any threshold contract with threshold level $K=2$. Consider a linear contract with payment m and a threshold contract with per-period payment m. Let there be perfect correlation in costs across days 1 and 2. Finally, assume that compliance has a future benefit b (in addition to a cost) that the participant discounts by δ.*
(a) If $\delta=1$, the threshold and linear contract will have the same compliance and effectiveness.
(b) If $\delta<1$, the threshold contract will have weakly higher compliance and effectiveness than the linear.

Proof. Under the linear contract, each participant complies in each period if $e<\delta b+m$.

Under the threshold contract, on day 2:

- If $w_1=1$, the participant will comply if $e<\delta b+2m$.
- If $w_1=0$, the participant will comply if $e<\delta b$.

On day 1, the participant will comply if

(a) $e(1+\delta)<2\delta b+2m$, which is equivalent to $e<\frac{2}{(1+\delta)}(\delta b+m)$, and

(b) They think their future self will follow-through, which requires the following for sophisticates and naifs:

  - Sophisticates: $e<\delta b+2m$.
  - Naifs: $\delta e<\delta b+2m$, which holds whenever condition (b) holds.[85]

Case 1: $\delta=1$. If $\delta=1$, in both the linear and threshold contracts, the compliance conditions simplify to complying in both periods if $e+b<m$ and not otherwise. Therefore compliance in the contracts is the same; moreover, since the payment will be $2m$ in both cases for full compliance and 0 otherwise, the effectiveness is also the same.

Case 2: $\delta<1$. However, if $\delta<1$, the participant may comply in the threshold contract even if they would not comply in the linear.

In particular, if the linear compliance condition holds ($e<\delta b+m$), then both conditions (a) and (b) will hold as well—a participant will never comply in the linear but not the threshold.

However, they may comply in the threshold even if the linear compliance condition does not hold ($e>\delta b+m$), as conditions (a) and (b) may still hold.

That is, if $e>\delta b+m$, it is still possible for (a) $e<\frac{2}{(1+\delta)}(\delta b+2m)$, and (b) $e<\delta b+2m$ or $e<b+\frac{2}{\delta}m$. This is because (a) $\frac{2}{(1+\delta)}(\delta b+2m)>\delta b+m$ when $\delta<1$, and (b) both $\delta b+2m>\delta b+m$

[85] If $e<\delta b$, then the person complies in all periods without incentives and so both equations hold. If $e>\delta b$, then they sink costs as they go and so the period 2 equation will hold whenever the period 1 equation holds.

and $b + \frac{2}{\delta}m > \delta b + m$ as well.

Thus, compliance for both sophisticates and naifs will be weakly larger under the threshold contract than linear (and strictly larger when there are some cost realizations e that satisfy $\delta b + m < e < \min\{\frac{2}{(1+\delta)}(\delta b + 2m), \delta b + 2m\}$ for sophisticates or $\delta b + m < e < \min\{\frac{2}{(1+\delta)}(\delta b + 2m), b + \frac{2}{\delta}m\}$ for naifs).

Moreover, cost-effectiveness is the same for sophisticates in the threshold and linear contracts (both pay 2m for 2 days of compliance), and weakly lower for naifs in the threshold contract (because occasionally they will comply in the first period and not receive payment).

Hence, compliance and effectiveness are weakly higher in the threshold than linear when $\delta < 1$.

$\square$

I CTB Time Preference Measurement

We adapted the convex time budget (CTB) methodology of Andreoni and Sprenger (2012a) to try to measure time preferences in two domains, walking and mobile recharges. Unfortunately, it did not work for either domain. As a result, we do not use the full CTB measures for analysis and instead use the simple versions of CTB described in Section 4.2. In Section I.1 we summarize why we believe our full CTB measurement was not a reliable measure of time preferences in this setting. In Section I.2 we briefly summarize evidence that the Simple CTB measures worked better. Section I.3 further expands upon section I.1 and provides additional evidence.

I.1 Performance of the Full CTB

We believe our implementation of the full CTB methodology of Andreoni and Sprenger (2012a) was unsuccessful because respondents did not understand it. The complex methodology was difficult to explain to our participants, who had limited familiarity with screens, sliders, or complicated exercises. Due to survey length constraints, we also included fewer questions (and gave less practice) than previous laboratory studies.

A number of patterns in the data suggest that participant understanding was limited. First, law of demand violations are far more common than in previous studies.[86] As shown in Table I.1, a markedly large 57% of the sample violated the law of demand at least once. In comparison, only 16% of participants in Augenblick et al. (2015) violated the law of demand at least once despite 16 opportunities to do so (relative to just two opportunities in our study).

Appendix Table I.1: Law of Demand Violations in Effort Allocations

	# of violators	% of sample
	(1)	(2)
Violates 0/7	1,318	41.3
Violates 7/14	1,493	46.8
Violates at least once	1,805	56.6
Violates both	1,006	31.5
Total:	3,192	100

Notes: This table summarizes law of demand violations in the full CTB in the recharge domain. Violators allocate more steps to the future date when we increase the interest rate from 1 to 1.25. We varied the exchange rate for two questions: today versus 7 days from now, and 7 versus 14 days; rows 1 and 2 show violations for these two questions separately and row 3 and 4 show percentages of people who violated at least once or both.

Second, the CTB estimates do not correlate with any of the behaviors one would expect

[86] We can only examine law of demand violations in the effort domain because we did not include exchange rate variation in the recharge domain, so cannot estimate the demand curve.

them to. The CTB estimates in the steps/effort domain do not correlate with exercise and health, and the estimates in the recharge domain do not correlate consistently with our proxies for impatience over recharges (e.g., balances).

Finally, there are a number of other problems with the full CTB data, such as low follow-through on the incentivized activity and low convergence of the parameters. We describe these issues in more depth in Section I.3.

For all of these reasons, we do not think our CTB estimates are a reliable measure of discount rates in this setting and do not use them for analysis.

I.2 Performance of the Simple CTB

The Simple CTB measures seem to have performed better than the full CTB exercise. For example, only 18% of the participants had any law-of-demand violations in these simpler questions, much lower than the 57% in the full CTB, even though participants had the same number of opportunities for violations in both question sets. The 18% estimate is much closer to the 16% found in Augenblick et al. (2015). The percent of future-biased choices (19%) is also closer to what is found in Augenblick et al. (2015) (which finds 17%) than to the higher estimates from the full CTB (26%).

Note that these estimates come from the performance of the simple CTB over recharges but not over effort; given the specific questions we asked in the effort domain, we cannot calculate law of demand violations nor future bias, so we cannot compare the measures on that front. However, as shown in Table A.1, the simple CTB over effort correlates in the expected direction with exercise (i.e., people who look more impatient under the simple CTB have lower steps). In contrast, the full CTB estimates do not correlate in the expected direction with any behaviors. Hence, the simple CTB still appears to be the better measure for our context.

I.3 Implementation of the Full CTB

We first discuss the methodology used for the full CTB. We then show that the full CTB measures do not correlate with the behaviors that we would expect. Finally, we describe additional problems with the full CTB implementation.

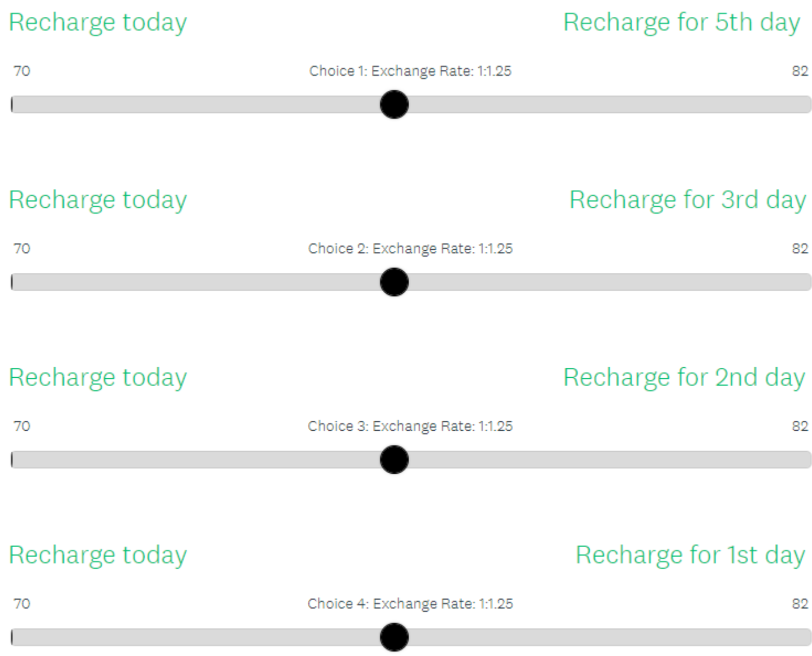
I.3.1 Estimation Methodology

Our full CTB uses the full CTB methodology of Augenblick et al. (2015). In each CTB choice in our full CTB module, the participant is asked to allocate a fixed budget of either steps or mobile recharges between a "sooner" and a "later" date using a slider bar. In particular, each choice allows the respondent to choose an allocation of consumption on the sooner and later dates, c_t, c_{t+k} that satisfies the budget constraint

$$c_t + \frac{1}{r}c_{t+k} = m \tag{72}$$

where the sooner date t, the later date $t+k$, the interest rate r, and the budget m change between each choice. A sample slider screen allowing for such choices is shown in Figure I.1.

We asked participants to make six allocations in the recharge domain, and eight allocations in the step domain, as summarized in Table I.2. We assume a time-separable and good-separable



Appendix Figure I.1: Sample Decision Screen for Mobile Recharges

Notes: In this example, the interest rate, r, is 1.25; the total budget, m, is 140; the "sooner" date is Today; and the "later" date decreases from 5 days from today in the first choice to 1 day from today in the final choice. The sliders are shown positioned at the choice $(c_t = 70, c_{t+k} = 82)$.

CRRA utility function with quasihyperbolic discounting[87]. In the domain of recharges, individuals will then seek to maximize utility,

$$U\left(c_t, c_{t+k}\right) = \frac{1}{\alpha}\left(c_t - \omega\right)^\alpha + \beta\delta^k \frac{1}{\alpha}\left(c_{t+k} - \omega\right)^\alpha \tag{73}$$

and in the step domain, individuals will seek to minimize costs of effort

$$C\left(c_t, c_{t+k}\right) = \frac{1}{\alpha}\left(c_t + \omega\right)^\alpha + \beta\delta^k \frac{1}{\alpha}\left(c_{t+k} + \omega\right)^\alpha \tag{74}$$

The variation in consumption choices as the budget constraint varies identify the time preference parameters—in particular, the daily discount factor δ and the present-bias parameter β—as well as the concavity or convexity of preferences α. Due to budget and time constraints, we had to keep the module short and so did not implement interest rate variation for the recharge tradeoffs, only for the step tradeoffs. Thus α is identified for the effort estimation only, not the recharge one; for the recharge estimation, we calibrate α using the estimate of α from Augenblick et al. (2015) in the financial payment domain.

We recover individual-level structural estimates of time preference and concavity parameters from the allocations (c_t, c_{t+k}) using a two-limit Tobit specification of the intertemporal Euler

[87]Unlike in Appendix C.2 where the quasihyperbolic discounting model we used only has one parameter δ_{QH} or d_{QH}, here we use both β and δ since we estimated them simultaneously.

condition following Augenblick et al. (2015).

$$\log\left(\frac{c_t+\omega}{c_{t+k}+\omega}\right)=\frac{\log(\beta)}{\alpha-1}1_{t=0}+\frac{\log(\delta)}{\alpha-1}k-\frac{1}{\alpha-1}\log(r) \tag{75}$$

Details on the estimation strategy can be found in the Online Appendix of Augenblick et al. (2015). Because our predictions concern overall impatience, not whether an individual is time-consistent, on the time preference side, we want one single summary measure capturing impatience. To do so, we estimate two different variants. In one, we set $\beta = 1$ for everyone at the estimation stage and simply estimate δ at the individual level. In the second, we estimate the equation as above, allowing both β and δ to vary at the individual level, and use $\beta \times \delta$ as our measure of individual-level impatience. In both estimation procedures, we allow α to vary at the individual-level in the steps domain, since we considered individual-level convexity of the step function to be an important potential confound.[88] However, the results we describe next are similar if we do not allow α to vary at the individual-level for steps.

Appendix Table I.2: CTB Allocation Parameters

Summary of convex time budget allocations					
Question no.	t	k	r	Recharge domain	Step domain
1	7	7	1	X	X
2	0	7	1	X	X
3	0	5	1	X	X
4	0	3	1	X	X
5	0	2	1	X	X
6	0	1	1	X	X
7	7	7	1.25		X
8	0	7	1.25		X

Notes: This table summarizes the parameters of the six CTB allocations made over recharges, and the eight CTB allocations made over steps.

Our CTB environment builds on a number of features from previous studies. First, the choices are made after the one-week phase-in period in which all participants have pedometers and report their daily steps, ensuring that participants are familiar with the costs of walking. This allows for meaningful allocations of steps between sooner and later dates. Second, the responses are designed to be incentive compatible; all respondents were informed that we would implement their choice from a randomly selected survey question. We set the probabilities such that for most respondents the randomly selected survey question was a multiple price list of

[88]Indeed, when we estimate impatience (e.g., δ) but do not allow α to vary, that estimated δ correlates as strongly with α as it does with the δ estimated allowing α to vary, suggesting that convexity is an important confound indeed.

lotteries over money (which measures risk preferences), but for a few a CTB allocation was selected. Because the allocations might have interfered with any walking program offered, we excluded the 40 respondents who were randomly selected to receive one of their allocations from the experimental sample.[89] To try to ensure that participants complete the allocated steps, we offer a large cash completion bonus of 500 INR in the step domain if the allocation is selected to be implemented, and the steps are completed as allocated, with the bonus to be delivered 15 days from the date of the survey (which is 1 day after the latest "later" day used).

We also take a number of precautions to avoid various potential confounds, including confounds reflecting fixed costs or benefits of taking an action, or confounds due to the time of day of measurement.[90] However, we were not able to address one potential confound to our estimates of time-preferences across individuals fully: variation across people in the cost of walking over time, or in the benefit of receiving a recharge over time. For example, an individual with a particularly busy week after the time-preference survey, and therefore relatively high costs to steps in the near-term relative to the distant future, will appear to be particularly impatient over steps in our data (he will wish to put off walking). An individual with a relatively free week just after the time-preference survey will instead appear particularly forward-looking (he will not wish to put off walking). The same concerns can also arise with recharges.

I.3.2 CTB Estimates: Problems with Convergence and Lack of Correlation

Table I.3 displays the summary statistics as well as the convergence statistics. The CTB parameter estimates themselves are not robust and are inconsistent with typical priors. First, we do not have estimates for a large, endogenous share of the sample. The estimates do not converge (i.e., we are unable to estimate discount rate parameters) for 38 to 44% of the sample

[89] This means we have CTB data from a total of 3,232 people: the 3,192 in the experimental sample plus the 40 selected to receive "real-stakes" allocations. In this section we exclude the "real-stakes" observations but the results are similar if we include them.

[90] To avoid confounds related to fixed costs or benefits, such as the effort of wearing a pedometer or the psychological benefit of receiving a free recharge, we include minimum allocations on both sooner and later days in each domain. The minimum allocations were chosen to be high enough that any fixed costs would be included (e.g. one could not easily achieve the minimums by simply shaking the pedometer) but low enough to avoid corner solutions. In the step domain, this required a modification of the CTB methodology: individual-specific minimum allocations. Our step allocations also featured individual-specific total step budgets m, which were chosen to be large enough that achieving them would require some effort beyond simply wearing the pedometer but small enough that participants would certainly achieve them in exchange for the completion bonus. Specifically, minimum steps on each day are calculated as $\frac{X}{10}$, and the total step budget m is $X + 2\frac{X}{10}$, respectively, where $X \in \{3000, 4000, 5000\}$ is the element closest to the participant's average daily walking during the phase-in period. That is, minimum steps are one of 300, 400, or 500 on each day, and the total step budget is one of 3,600, 4,800, or 6,000. To avoid confounding impatience with the time of day that the baseline time-preference survey was administered (which could influence the desirability of walking and/or recharges delivered in the next 24 hours), as well as to capture heterogeneity in time preferences including any present-bias for very short beta-windows, we required that all walking on any date be conducted within a 2 hour period, which was chosen to start at the time immediately after the time-preference survey would end (e.g., if the survey ended at 4pm, the time period for any day's walking would be 5-7pm). The short window could potentially bias our overall measures of impatience downwards, as uncertainty about future schedules in a short time window could lead participants to want to get their walking done early when they had more certainty over their schedule. However, our primary purpose was to capture heterogeneity in time-preferences, and we considered the potential loss in validity of aggregate time preference estimates to be worth the ability to capture heterogeneity in time preferences in the time frames near to the present.

in the recharge domain, and 23 to 44% of the sample in the steps domain. Moreover, many of the participants with estimates that converge in the effort domain have an estimated $\alpha < 1$, which violates the first order conditions for estimation and is often associated with non-sensible δ and β estimates. When we exclude these estimates, we are left with estimates for only 34 to 38% of the sample in the effort domain. Second, we have a high rate of negative estimated discount rates: 42% for steps and 61% for recharges. This is more than the usual rate of negative individual-level estimates.

Appendix Table I.3: Summary Statistics For CTB Parameters

	Full sample		$\alpha > 1$	
Parameters estimated:	$\beta\delta$	δ	$\beta\delta$	δ
	(1)	(2)	(3)	(4)
A. Effort				
Beta	2.079	–	1.590	–
Delta	0.843	0.998	1.015	0.999
Alpha	0.775	0.729	1.708	1.575
% of sample:	77.3	56.4	34.1	38.0
# Individuals:	2,466	1,799	1,088	1,213
B. Recharges				
Beta	0.972	–	–	–
Delta	0.990	0.996	–	–
% of sample:	56.1	62.4	–	–
# Individuals:	1,789	1,991	–	–

Notes: This table displays means and convergence rates of individual-level CTB parameters in both the effort and recharge domains. Columns 1–2 display average values for the parameters from the full sample of individuals with parameters that converged. In the effort domain, in columns 3 - 4, we ignore all individuals whose estimated *alpha* was below 1, as handled similarly in Andreoni and Sprenger (2012a), as that is inconsistent with the first order conditions. We winsorize all parameters at the top and bottom 1 percentiles. We allow α to vary at the individual level in the effort domain, and in the recharge domain, we calibrate α to be 0.975, which is the estimated value in Augenblick et al. (2015). Delta is estimated by allowing δ to vary at the individual level and setting β to 1. Beta-delta is estimating by allowing both δ and β to vary. We derive these two parameters from an estimation that allows δ and β to vary at the individual level. Significance levels: * 10%, ** 5%, *** 1%.

Tables I.4 and I.5 show that the estimated CTB parameters do not correlate in the expected direction with measured behaviors. In particular, Table I.4 shows that the CTB estimates in the steps/effort domain do not correlate with exercise and health,[91] and Table I.5 shows that

[91]Table I.4 shows the correlations when we exclude the effort estimates from participants with estimated

the estimates in the recharge domain do not correlate with recharge balances, usage, or credit constraint proxies. The CTB measures do correlate at the 1% level with our measure of marginal propensity to consume recharges, but the correlations go in opposite directions for the two CTB measures (δ from an estimation setting $\beta = 1$ vs. $\beta\delta$ estimated allowing both parameters to vary) so is likely noise.

Appendix Table I.4: CTB Estimates of Discount Factors Over Steps Do Not Correlate With Measured Behaviors

Covariate type:	Baseline exercise		Baseline indices			
Dependent variable:	Daily steps	Daily exercise (min)	Health index	Negative vices index	Healthy diet index	# Individuals
Delta	-0.019	0.009	-0.040	0.011	0.027	1,213
Beta-delta	0.016	0.018	0.014	0.010	0.027	1,086

Notes: This table displays the correlations between CTB parameters in the effort domain and a few baseline health covariates. We normalize impatience variables so that a higher value corresponds to greater impatience, and we normalize health outcomes so that higher values correspond to healthier outcomes. All CTB parameters have been winsorized at the top and bottom 1 percentile to remove outliers. Delta is measured from an estimation that allows δ and α to vary at the individual level, while excluding β. Beta-delta is a measure of beta times the average delta over one week. We estimate the two parameters by allowing β, δ, and α to vary at the individual level. Significance levels: * 10%, ** 5%, *** 1%.

Appendix Table I.5: CTB Estimates of Discount Factors Over Recharges Do Not Correlate With Other Proxies for Impatience Over Recharges

Covariate type:	Recharge variables				Credit constraint proxies		
Dependent variable:	Negative mobile balance	Negative yesterday's talk time	Prefers daily (=1)	Prefers monthly (=-1)	Negative wealth index	Negative monthly household income	# Individuals
Delta	0.026	0.013	-0.141***	0.045	-0.010	-0.008	1,836
Beta-delta	-0.002	-0.022	0.145***	-0.019	-0.015	0.033	1,652

Notes: This table displays the correlations between CTB parameters in the recharge domain and baseline measures that should be related to credit constraints and discount rates over recharges. We normalize impatience variables so that a higher value corresponds to greater impatience, and we normalize the proxies so that higher values correspond to higher expected discount rates; hence, the prediction is that coefficients should be positive. All CTB parameters have been winsorized at the top and bottom 1 percentile to remove outliers. We use two main estimation specifications, and to identify parameters, we calibrate α to be 0.975, the value of α estimated in Augenblick et al. (2015). Delta is estimated by allowing δ to vary at the individual level and excluding β. Beta-delta is a measure of the average delta over one week multiplied by beta. We derive these two parameters from an estimation that allows δ and β to vary at the individual level. Significance levels: * 10%, ** 5%, *** 1%.

$\alpha < 1$, but the results are similar when we include all estimates together.

I.3.3 Additional Problems With the Full CTB Data

Finally, we provide more detail on other problems with the Full CTB, in addition to the law of demand violations, the lack of convergence, and the lack of correlation described earlier.

First, in the effort task, there was low follow-through on the incentivized activity: fewer than 50% of participants selected to complete the step task did so despite large rewards (500 INR) for completion. While this partly reflects a logistical glitch (we failed to give respondents intended reminder calls the day before their activity), the lack of follow-through may also indicate a lack of respondent understanding. Regardless, the poor follow-through is problematic methodologically: identification requires that, when participants make their allocation decisions, they think they will follow-through with certainty, which seems unrealistic given how few followed through in practice.

Second, respondents on average allocated more steps to today than the future even when the interest rate was 1:1. Although they could be future-biased, the following other potential explanations are concerning for interpretation: respondents were confused; they saw steps as consumption instead of a cost (violating the first order conditions underlying estimation); or uncertainty over future walking costs and schedules led participants to want to finish steps sooner, which would confound discount rate estimates with risk aversion and uncertainty.

Third, day-specific shocks appear to be important in practice. 19% of respondents' allocations of steps to the sooner date are neither monotonically weakly increasing nor monotonically weakly decreasing across questions which feature the same sooner date (today) but a monotonically decreasing later date (questions 2-6). These allocations cannot be rationalized with a discount rate that is either weakly decreasing *or* increasing with lag length without day-specific utility shocks. The same holds for 24% of respondents in the recharge domain. These types of shocks would also confound estimation.

J Monitoring Treatment Impacts on Walking

The health results suggest that the monitoring treatment had limited impact, although the results are somewhat imprecise. Did the monitoring treatment not affect exercise, or were the exercise impacts too small to translate into measurable health impacts? We now present an analysis of the effects of monitoring on exercise. Because we do not have pedometer walking data from the control group, we use a before-after design. We find that monitoring alone has limited impact on overall steps. Monitoring does however change the distribution of steps, increasing the share of days on which participants met the 10,000-step target but decreasing the steps taken on other days for a null effect on total exercise.

Our before-after design compares pedometer-measured walking in the monitoring group during the phase-in period (during which we had not given participants a walking goal and just told them to walk the same as they normally do) to their behavior during the intervention period. This strategy will be biased either in the presence of within-person time trends in walking, or if the phase-in period directly affects walking behavior. We control for year-month fixed effects to help address time trends, but the latter concern is more difficult, as the phase-in period likely did increase walking above normal, either because of Hawthorne effects or because participants received a pedometer and a step-reporting system, which are two of the elements of the monitoring treatment itself (the other three remaining that we can still evaluate are (a) a daily 10,000 step goal, (b) positive feedback for meeting the step goal through SMS messages and the step-reporting system, and (c) periodic walking summaries). Thus, we consider a pre-post comparison of walking in the monitoring group to be a lower bound of the monitoring program treatment effect.

One can visualize the variation used for our pre-post estimate in Figure A.2, Panels (a) and (b). Walking increases immediately during the intervention period for the monitoring group, although the effects decay over time.

We next estimate the pre-post monitoring effect controlling for date effects. In order to increase the precision of our estimated year-month fixed effects, we include the incentive group in the regression as well since that group is much larger. We thus estimate the following difference-in-differences regression using data from both the intervention and phase-in periods for the incentive and monitoring groups:

$$\begin{aligned} y_{it} = & \alpha + \beta_1 Intervention\,Period_{it} + \beta_2 incentives_i + \beta_3 (Intervention\,Period_{it} \times incentives_i) \\ & + \boldsymbol{X}_i' \gamma + \boldsymbol{\mu}_m + \varepsilon_{it}, \end{aligned} \tag{76}$$

where y_{it} are daily pedometer outcomes measured during both the phase-in and the intervention period, $Intervention\,Period_{it}$ is an indicator for whether individual i has been randomized into their contract at time t, $incentives_i$ is an indicator for whether i is in an incentive treatment group, $\boldsymbol{X}_i$ is a vector of individual-specific controls, and $\boldsymbol{\mu}_m$ is a vector of month fixed effects. The coefficient β_1—the coefficient of interest—is the pre-post difference in pedometer outcomes within the monitoring group (controlling for aggregate time effects).

Table J.1 presents the results. Column 3 shows that the monitoring group achieves the 10,000-step target on approximately 6% more days in the intervention period than in the phase-in period, an effect significant at the 1% level and equal to roughly 30% of the estimated impact of incentives. In contrast, the estimated effect on steps is very small in magnitude, varies across specifications, and is in fact sometimes negative (columns 4–6). Thus, the monitoring treatment, if anything, appears to do more to make walking consistent across days than it does to increase total steps.

Appendix Table J.1: Impacts of Monitoring (Pre-Post) and Incentives (Difference-In-Differences) on Exercise Outcomes

	Achieved 10K steps			Daily steps		
	(1)	(2)	(3)	(4)	(5)	(6)
Incentives	0.012	0.013	0.012	66.7	66.4	48.9
	[0.024]	[0.024]	[0.014]	[268.1]	[266.9]	[112.3]
Intervention period	0.057***	0.073***	0.064***	-130.4	108.0	-18.5
	[0.020]	[0.020]	[0.020]	[237.8]	[240.8]	[234.1]
Intervention period X Incentives	0.19***	0.19***	0.19***	1270.9***	1258.9***	1212.7***
	[0.021]	[0.021]	[0.021]	[248.6]	[249.2]	[243.4]
Year-month FEs	No	Yes	Yes	No	Yes	Yes
Individual controls	No	No	Yes	No	No	Yes
Monitoring phase-in mean	.24	.24	.24	6,904.8	6,904.8	6,904.8
# Individuals	2,604	2,604	2,604	2,604	2,604	2,604
Observations	221,214	221,214	221,214	221,214	221,214	221,214

Notes: This table shows coefficient estimates from regressions of the form specified in equation (76). The outcomes are from daily panel data from the pedometers. Standard errors, in brackets, are clustered at the individual level. Individual controls are the same as Table 2. The omitted category is Monitoring in the phase-in period. The coefficient in the second row, on $Intervention\,Period_{it}$, corresponds to the pre-post estimate of the Monitoring effect. Significance levels: * 10%, ** 5%, *** 1%.