Overcoming Temptation: 
Incentive Design For Intertemporal Choice

Michael C. Mozer†, Shruthi Sukumar⋄, Camden Elliott-Williams⋆, Shabnam Hakimi†, Adrian F. Ward‡

⋆ Department of Computer Science, University of Colorado, Boulder
⋄ Department of Electrical, Computer, & Energy Engineering, University of Colorado, Boulder
† Institute of Cognitive Science, University of Colorado, Boulder
‡ McCombs School of Business, University of Texas, Austin

Abstract

Individuals are often faced with temptations that can lead them astray from long-term goals. We’re interested in developing interventions that steer individuals toward making good initial decisions and then maintaining those decisions over time. In the realm of financial decision making, a particularly successful approach is the prize-linked savings account: individuals are incentivized to make deposits by tying deposits to a periodic lottery that awards bonuses to the savers. Although these lotteries have been very effective in motivating savers across the globe, they are a one-size-fits-all solution. We investigate whether customized bonuses can be more effective. We formalize a delayed-gratification task as a Markov decision problem and characterize individuals as rational agents subject to temporal discounting, costs associated with effort, and moment-to-moment fluctuations in willpower. Our theory is able to explain key behavioral findings in intertemporal choice. We created an online delayed-gratification game in which the player scores points by choosing a queue to wait in and patiently advancing to the front. Data collected from the game is fit to the model, and the instantiated model is then used to optimize predicted player performance over a space of incentives. We demonstrate that customized incentive structures can improve goal-directed decision making.

Should you go hiking today or work on that manuscript? Should you have a slice of cake or stick to your diet? Should you upgrade your flat-screen TV or contribute to your retirement account? Individuals are regularly faced with temptations that lead them astray from long-term goals. These temptations all reflect an underlying challenge in behavioral control that involves choosing between actions leading to small but immediate rewards and actions leading to large but delayed rewards. We introduce a formal model of this delayed gratification decision task, extending the Markov decision framework to incorporate the psychological notion of willpower, and using formal models to optimize behavior by designing incentives to assist individuals in achieving long-term goals.

Consider the serious predicament with retirement planning in the United States. Roughly 45% of working-age households have no retirement account assets. Median account balance for near-retirement households is $14,500. Even considering households’ net worth, 2/3 fall short of conservative savings targets based on age and income (Rhee & Boivie, 2015). Furthermore, 40% of every dollar contributed to the accounts of savers under age 55 simultaneously flows out of the retirement systems, not counting loans to oneself (Argento, Bryant, & Sabelhaus, 2015).

One technique that has been extremely successful in encouraging savings, primarily in Europe and the developing world but more recently in the US as well, is the prize linked savings account (PLSA) (Kearney, Tulino, Guryan, & Hurst, 2010). The idea is to pool a fraction of the interest from all depositors to fund a prize awarded by periodic lotteries. Just as ordinary lotteries entice individuals

to purchase tickets, the PLSA encourages individuals to save. Disregarding the fact that lotteries function in part because individuals overvalue low-probability gains (Kahneman & Tversky, 1979), the core of the approach is to offer savers the prospect of short-term payoffs in exchange for them committing to the long term. Although the account yields a lower interest rate to fund the lottery, the PLSA increases the net expected account balance due to greater commitment to participation.

The PLSA is a one-size-fits-all solution. A set of incentives that work well for one individual or one subpopulation may not be optimal for another. In this article, we investigate approaches to customizing incentives to an individual or a subpopulation with the aim of achieving greater adherence to long-term goals and ultimately, better long-term outcomes for the participants. Our approach involves: (1) building a model to characterize the behavior of an individual or group, (2) fitting the model with behavioral data, (3) using the model to determine an incentive structure that optimizes outcomes, and (4) validating the model by showing better outcomes with model-derived incentives than with alternative incentive structures.

1 Intertemporal Choice

Intertemporal choice involves decisions that produce gains and losses at different points in time. How an individual interprets delayed consequences influences the utility or value associated with a decision. When consequences are discounted with the passage of time, decision making is biased toward more immediate gains (and more distant losses). The delay discounting task is often used to study intertemporal choice (Green & Myerson, 2004). Individuals are asked to choose between two alternatives, e.g., $1 today versus $X in Y days. By identifying the X that yields subjective indifference for a given Y, one can estimate an individual’s discounting of future outcomes. Discount rates vary across individuals yet show trait-like stability over extended periods of time (Kirby, 2009).

This paradigm involves a one-shot hypothetical decision and reveals the intrinsic future value of an outcome. However, it does not reflect the way in which choices play out in the real world. Most tasks faced by people involve continually deciding between accepting a small immediate gain and a larger gain following a delay. For example, in the classic marshmallow test (Mischel & Ebbesen, 1970), children are seated at a table with a single marshmallow. They are allowed to eat the marshmallow, but if they wait while the experimenter steps out of the room, they will be offered a second marshmallow when the experimenter returns. In this delayed gratification task, children must decide at every instant whether to eat the marshmallow or wait for two marshmallows. In this situation, behavior depends not only on the hypothetical discounting of future rewards but on an individual’s willpower—their ability to maintain focus on the larger reward and not succumb to temptation before the experimenter returns. Defection at any moment eliminates the possibility of the larger reward.

The marshmallow test attracted interest not only because it turns out to be predictive of later life outcomes (Mischel, Shoda, & Rodriguez, 1989), but because it is analogous to many situations involving delayed gratification. Like the marshmallow test, some of these situations have an unspecified time horizon (e.g., exercise, waiting for an elevator, spending during retirement); others have a known horizon (e.g., saving for retirement, getting a college degree). Our work addresses the case of a known or assumed horizon.

2 Formalizing Delayed-Gratification Tasks as a Markov Decision Problem

In this section, we formalize a delayed-gratification task as a Markov decision problem, which we will refer to as the DGMDP. To the best of our knowledge, we are the first to take this straightforward modeling approach; the study of intertemporal choice has relied primarily on verbal, qualitative accounts. We assume time to be quantized into discrete steps and we focus on situations with a known or assumed time horizon, denoted τ. At any step, the agent may DEFECT and collect a small reward, or the agent may PERSIST to the next step, eventually collecting a large reward at step τ. We use μSS and μLL to denote the smaller sooner (SS) and larger later (LL) rewards. Figure 1 shows a finite-state representation of the task environment with terminal states LL and SS that correspond to resisting and succumbing to temptation, respectively, and states for each time step between the initial and final times, \( t \in \{1, 2, ..., \tau\} \). Rewards are associated with state transitions. The possibility of obtaining intermediate rewards during the delay period is annotated via \( \mu_{t-1} \equiv \{\mu_1, ..., \mu_{\tau-1}\} \), which we return to later. Although human studies of intertemporal choice support hyperbolic
discounting (Frederick, Loewenstein, & O’Donoghue 2002), we follow the modeling tradition of treating discounting as exponential, which is a reasonable approximation when the range of time scales is limited: rewards n steps ahead are devalued by a factor of $\gamma^n$, $0 \leq \gamma < 1$.

Given the DGMDP, an optimal decision sequence is trivially obtained by value iteration. However, this sequence is a poor characterization of human behavior. With no intermediate rewards ($\mu_{1,\tau-1} = 0$), it takes one of two forms: either the agent defects at $t = 1$ or the agent persists through $t = \tau$. In contrast, individuals will often persist some time and then defect, and when placed into the same situation repeatedly, behavior is nondeterministic. For example, replicability on the marshmallow test is quite modest ($\rho < 0.30$; Mischel, Shoda, & Peake 1988).

The discrepancy between human delayed-gratification behavior and the optimal decision-making framework might indicate an incompatibility. However, we prefer a bounded rationality perspective on human cognition according to which behavior is cast as optimal but subject to cognitive constraints. We claim two specific constraints.

1. Individuals exhibit moment-to-moment fluctuations in willpower based on factors such as sleep, hunger, mood, etc. Low willpower causes an immediate reward to seem more tempting, and high willpower, less tempting. We characterize willpower as a one-dimensional Gaussian process, $W = \{W_t\}$, with $w_1 \sim \text{Gaussian}(0, \sigma_1^2)$ and $w_t \sim \text{Gaussian}(w_{t-1}, \sigma^2)$. We suppose that willpower modulates an individual’s subjective value of defecting at step $t$:

$$Q(t, w; \text{DEFECT}) = \mu_{SS} - w,$$

where $Q(s; a)$ denotes the value associated with performing action $a$ in state $s$, and the state space consists of the discrete step $t$ and the continuous willpower level $w$.

2. Behavioral, economic, and neural accounts of decision making suggest that effort carries a cost, and that rewards are weighed against the effort required to obtain it (e.g., Kivetz 2003). This notion is incorporated into the model via an effort cost, $\mu_E$ associated with persevering:

$$Q(t, w; \text{PERSIST}) = \begin{cases} \mu_E + \mu_t + \gamma E_{W_{t+1}|W_t=w}V(t+1, w_{t+1}) & \text{for } t < \tau \\ \mu_{LL} & \text{for } t = \tau \end{cases}$$

where $V(t, w) \equiv \max_a Q(t, w; a)$.

To summarize, we have formalized the known-horizon delayed-gratification task as a Markov decision problem with parameters $\Theta_{task} \equiv \{\tau, \mu_{SS}, \mu_{LL}, \mu_{1,\tau-1}\}$, and a constrained rational agent parameterized by $\Theta_{agent} \equiv \{\gamma, \sigma_1, \sigma, \mu_E\}$. We now turn to solving the DGMDP and characterizing its properties.

### 2.1 Solving The Delayed-Gratification Markov Decision Problem (DGMDP)

The simple structure of the environment allows for a backward-induction solution to the Bellman equation (Equation 2). Although the continuous willpower variable precludes an analytical solution for $V(t, w)$, we construct a piecewise linear approximation (PLA) over $w$ for each step $t$. To justify the PLA, consider the shape of $V(t, w)$. With high willpower ($w \rightarrow \infty$), the agent almost certainly persists to the LL reward and the function asymptotes at the discounted $\mu_{LL}$. With low willpower ($w \rightarrow -\infty$), the agent almost certainly defects and the function approaches $\mu_{SS} - w$. Thus, both extrema of the value function are linear with known slope and intercept. At step $\tau$, these two linear segments exactly define the value function. At $t < \tau$, there is an intermediate range within which small fluctuations in willpower can influence the decision and the expectation in Equation 2 yields a
weighted mixture of the two extrema, which is well fit by a single linear segment—defined by its
slope $a_t$ and intercept $b_t$. With $V(t, w)$ expressed as a PLA, the expectation in Equation 2
becomes:

$$\mathbb{E}_{W_t|W_{t-1}=w}[V(t, w_t) = \Phi(z_t^-) (\mu_{SS} - w_t) + (\Phi(z_t^+) - \Phi(z_t^-)) (b_t + a_t w_t) + (1 - \Phi(z_t^+)) c_t + \sigma \phi(z_t^-) + \sigma a_t (\phi(z_t^+) - \phi(z_t^-)),$$  

where $\Phi(.)$ and $\phi(.)$ are the cdf and pdf of a standard normal distribution, respectively, and
the standardized segment boundaries are $z_t^- = \sigma^{-1}(t_{SS} - b_t)/(a_t + 1) - w_t$ and $z_t^+ = \sigma^{-1}(t_{LL} - b_t)/(a_t - w_t)$. The backup is seeded with $w_t^* = z_t^+ = \sigma^{-1}(t_{LL} - c_t)$ and $a_t = b_t = c_t = b_{t-1}$; $c_t - 1$—the value of steadfast persistence—is obtained by propagating the discounted reward for
existence: $c_1 - 1 = \mu_{SS} + \mu_t + \gamma c_t$.

Figure 2 shows the value as a function of willpower at each step of an eight step DGMDP with
an LL reward twice that of the SS reward, like the canonical marshmallow test. To ensure accuracy of the estimate and to eliminate an accumulation of estimation errors, we have also used a fine piecewise constant approximation in the intermediate region, yet the model output is almost identical.

Using the value function, we can characterize the agent’s behavior in the DGMDP via the likelihood of deflection at various steps. With $D$ denoting the deflection step, we have the hazard probability

$$h_t = P(D = t|D \geq t) = P(W_t < w_t^*|W_1 \geq w_1^*, \ldots, W_{t-1} \geq w_{t-1}^*),$$  

where $w^*$ is the willpower threshold that yields action indifference, $Q(t, w^*; \text{DEFECT}) = Q(t, w^*; \text{PERSIST})$. To represent the posterior distribution over willpower at each non-deflection step, we initially used a particle filter but found a computationally more efficient and stable solution
with quantile-based samples. We approximate the $W_t$ prior and $\Delta W$ with discrete, equal probability $q$-quantiles. We reject values for which deflection occurs, and then propagate $W_{t+1} = W_t + \Delta W$ which results in up to $q^2$ samples, which we thin back to $q$-quantiles at each step. Using $q = 1000$ produces nearly identical results to selecting a much higher density of samples. Figure 2b shows hazard functions for low and high LL reward values.

### 2.2 Behavioral Phenomena Explained

We consider the solution of the DGMDP as a theory of human behavior. It is meant to explain
both an individual’s initial choice (“Should I open a retirement account?”) as well as the temporal
dynamics of sustaining that choice (“Should I withdraw the funds to buy a car?”) Nearly all previous
case conceptualizations of DG tasks have focused on the eventual outcome not the time course. One
exception is the work of McGuire and Kable [2013] who frame failure to postpone gratification as a rational, utility-maximizing strategy when the time at which future outcomes materialize is uncertain.

Our theory is complementary in providing a rational account in the known time horizon situation.

Our theory explains two key phenomena in the literature. First, failure on a DG task is sensitive
to the relative magnitudes of the SS and LL rewards (Mischel [1974]. Figure 2b) presents hazard functions for two reward magnitudes. The probability of obtaining the LL reward is greater with $\mu_{LL}/\mu_{SS} = 3$ than with $\mu_{LL}/\mu_{SS} = 2$. Figure 2 can also accommodate the finding that environmental reliability and trust in the experimenter affect outcomes in the marshmallow test (Kidd, Palmeri, & Aslin [2012]: in unreliable or nonstationary environments, the expected LL reward is lower than the advertised reward, and the DGMDP is based on reward expectations. Second, a reanalysis of data
from a population of children performing the marshmallow task shows a declining hazard rate over the task period of 7 minutes (McGuire & Kable [2013]). The rapid initial drop in the empirical curve looks remarkably like the curves in Figure 2b. One might interpret this phenomenon as a finish-line effect: the closer one gets to a goal, the greater is the commitment to achieve the goal. However, the model suggests that this behavior arises not from abstract psychological constructs but because of correlations in willpower over time: if an individual starts down the path to an LL reward, the individual’s willpower at that point must be high. The posterior willpower distributions reflect the elimination of individuals with low momentary willpower, which contributes to the declining hazard rate. Also contributing is the exponential increase in value of the discounted LL reward as the agent advances through the DGM. McGuire and Kable [2013] explain the empirical hazard function via a combination of uncertainty in the time horizon and time-fluctuating discount rates. Our theory shows that these strong assumptions are not necessary, and our theory can address situations with a well delineated horizon such as retirement saving. Additionally, our theory aims to move beyond population data and explain the granular dynamical behavior of an individual.

3 Optimizing Incentives

With a computational theory of the DG task in hand, we now explore a mechanism-design approach (Nisan & Ronen [1999]) aimed at steering individuals toward improved long-term outcomes. We ask whether we can provide incentives to rational value-maximizing agents that will increase their expected reward subject to constraints on the incentives.

We focus on an investment scenario that is analogous to a prize-linked savings account (PLSA). Suppose an individual has $x$ dollars which they can deposit into a bank account earning interest at rate $r$, compounded annually. At the start of each year, they decide whether to continue saving (PERSIST) or to withdraw and spend their entire savings with interest (DEFECT). Our goal is to assist them in maximizing the profit they reap over $\tau - 1$ years from their initial investment. Our mechanism is a promised schedule of distributions from the bank. We will refer to these distributions as bonuses, even though they come from an individual’s savings much as the lotteries of the PLSA do. With $\mu_t$ denoting the bonus awarded in year $t$ and $\mu_{t;\tau-1}$ denoting the set of scheduled bonuses, our goal as mechanism designers is to identify the schedule that maximizes the expected net accumulation from an individual’s investment:

$$\mu^*_{t;\tau-1} = \arg\max_{\mu_{1;\tau-1}} \sum_{t=1}^{\tau} P(D = t|\mu_{1;\tau-1}) \left[ b_t + \sum_{t'=1}^{t-1} \mu_{t'} \right], \quad (6)$$

where $b_t$ is the amount banked at the start of year $t$, with $b_1 = x$ and $b_{t+1} = (1+r)(b_t - \mu_t)$, and $D$ is the year of defection, where $D = 1$ represents immediate defection and $D = \tau$ represents the account reaching maturity. Defection probabilities are obtained from the theory (Equation 5).

To illustrate this approach, we conducted a simulation with $\gamma \in [0.55, 0.95]$, $\tau = 10$, $r = 0.1$, and $x = 100$, comparing an agent’s expected accumulation without bonuses and with optimal bonuses (Figure 3). Optimization is via direct search using the simplex algorithm over unconstrained variables $p_t = \logit(\mu_t/b_t)$, representing the proportion of the bank being distributed as a bonus. With high discounting, this particular simulation yields a modest (~ 10%) improvement in an individual’s expected accumulation by providing bonuses in year 1 and then in the final year, although in other simulations we have found that the benefit of bonuses increases greatly when penalties are applied for early withdrawals. Bonuses are recommended only when the gain from encouraging persistence beats the loss of interest on an awarded bonus. With low discounting, the model optimization recommends no bonuses. We now turn to human behavioral studies in which this framework is applied.
4 Experiments

We created a simple online delayed-gratification game in which players score points by waiting in a queue, much as diners score delicious foods by waiting their turn at a restaurant buffet (Figure 4a). The upper queue is short, having only one position, and delivers a 100 point reward when the player is serviced. The lower queue is long, having $\tau$ positions, and delivers a $100\tau\rho$ point reward when the player is serviced. The reward-rate ratio, $\rho$, is either 1.25 or 1.50 in our experiments. The player starts in a vestibule (right side of screen) and selects a queue with the up and down arrow keys. The game updates at 2000 msec intervals, at which point the player’s request is processed and the queues advance (from right to left). Upon entering the short queue, the player is immediately serviced. Upon entering the long queue, the player immediately advances to the next-to-last position as the queue shuffles forward. With every tick of the game clock, the player may hit the left-arrow key to advance in the long queue or the up-arrow key to defect to the short queue. If the player takes no action, the simulated participants behind the player jump past. When the player defects to the short queue, the player is immediately serviced. When points are awarded, the screen flashes the points and a cash register sound is played, and the player returns to the vestibule and a new episode begins.

The game is structured such that the episodes are independent of one another. Specifically, the reward rate (points per action) for the short and long queues is always 100 and $100\rho$, respectively. Thus, the game has no long term dependencies. However, because the game requires time-constrained score maximization, the one-shot DGMDP (Figure 1a) is not a suitable model of the task. Reward rate is what matters in the game. Consequently, we model each episode with a variant of the DGMDP (Figure 1b) in which a defection is treated as if the player continues to defect until $\tau$ steps are reached, each step delivering the small reward. The vestibule in Figure 4a corresponds to state 1 in Figure 1b and lower queue position closest to the service desk to state $\tau$. Note the left-to-right reversal of the two Figures, which has often confused the authors of this article.

Participants were recruited to play the game for five minutes via Amazon Mechanical Turk. In our analyses of player behavior, we remove the first and last thirty seconds of play. At the start, players are learning the game actions; at the end, players may not have sufficient time to traverse the long queue and defection is the optimal strategy. Participants are paid $0.80 to play and are awarded a score-based bonus. The long-queue length $\tau$ is uniformly drawn from $\{4, 6, 8, 10, 12, 14\}$.

4.1 Experiment 1: Varying Reward Magnitude

In Experiment 1, we manipulated the reward-rate ratio. Twenty different participants were tested for each $\rho \in \{1.25, 1.50\}$. Figure 5a shows the reward accumulation by individual participants in the two conditions as a function of time within the session. The two dashed black lines represent the reward that would be obtained by deterministically performing the SS or LL action at each tick of the game clock. (Participants are not required to act every tick, but they are warned after 7 sec and rejected after 14 sec if they fail to act.) The traces show that some participants had a strong preference for the short queue, others had a nearly perfect preference for the long queue, and still others alternated between strategies. The variability in strategy over time within an individual suggests that they did not simply lock into a fixed, deterministic action sequence.

For each participant, each queue length, and each of the $\tau$ positions in a queue, we compute the fraction of episodes in which the participant defects at the given position. We average these proportions across participants and then compute empirical hazard curves. Figure 5b shows hazard curves for each of the
Figure 5: (a) Game points accumulated by individual participants over time in Experiment 1. (b) Hazard curves in Experiment 1 for 6 line lengths and two reward-rate ratios. Human data shown with asterisks and dashed lines, model fits with circles and solid lines. (c) Hazard curves for Experiment 2, with only one free model parameter. (d) Hazard curves for Experiment 3, with no free model parameters. (e) Model-predicted optimal bonus sequences—early (yellow) and late (blue) bonuses for weak and strong participants, respectively. (f) Average reward rate in Experiment 4 for weak and strong subpopulations and three bonus conditions. Error bars are ±1 SEM, corrected for between-subject variance (Masson & Loftus, 2003).

six queue lengths and the two $\rho$ conditions. The $\rho = 1.50$ curves are lighter and are offset slightly to the left relative to the $\rho = 1.25$ curves to make the pair more discriminable. The Figure presents both human data—asterisks connected by dotted lines—and simulation results—circles connected by solid lines. Focusing on the human data for the moment, initial-defection rates rise slightly with queue length and are greater for $\rho = 1.25$ than for $\rho = 1.50$. We thus see robust evidence that participants are sensitive to game conditions.

To model the data, we set the DGMDP parameters ($\Theta_{\text{agent}}$) based on the game configuration. We obtain least-squares fits to the four agent parameters ($\Theta_{\text{agent}}$): discount rate $\gamma = 0.957$, initial and delta willpower spreads $\sigma_1 = 81.3$, and $\sigma = 21.3$, and effort cost $\mu_E = -52.1$. The latter three parameters can be interpreted using the scale of the SS reward, $\mu_{\text{SS}} = 100$ points. Although the model appears to fit the pattern of data quite well, the model has four parameters and the data can essentially be characterized by four qualitative features: the mean rate of initial defection, the modulation of the initial-defection rate based on queue length and on $\rho$, and the curvature of the hazard function. The model parameters have no direct relationship to these features of the curves, but the model is flexible enough to fit many empirical curves. Consequently, we are cautious in making claims for the model’s validity based solely on the fit to Experiment 1. We note, however, that we investigated a variant of the model in which willpower is uncorrelated across steps, and it produces qualitatively the wrong prediction: it yields curves whose hazard probability depends only on the steps to the LL reward. In contrast, the curves of the correlated-willpower account depend primarily on the distance from the initial state, $t$, but secondarily on distance to the LL reward, $\tau - t$.

4.2 Experiments 2 and 3: Modulating Effort

To obtain additional support for the theory, we modified the queue-waiting game such that the player must work harder and experiences more frustration in reaching the front of the long queue. By increasing the required effort, we may test whether model parameters fit to Experiment 1 will also fit new data, changing only the effort parameter, $\mu_E$. To increase the required effort, the long queue advanced only every other clock tick in an apparently random fashion. Nonetheless, the player must press the advance key every tick to move with the queue, thus requiring exactly two keystrokes for
each action in the game FSM (Figure 4). The game clock in Experiment 2 updated every 1000 msec, twice the rate as in Experiment 1, and thus the overall timing was unchanged. We tested only the reward-rate ratio $\rho = 1.50$.

Figure 5 shows hazard curves for Experiment 2. Using Experiment 1 parameter settings for $\gamma$, $\sigma_1$, and $\sigma$, we fit only the effort parameter, obtaining $\mu_1 = -99.7$, which is fortuitously twice the value obtained in Experiment 1. Model fits are superimposed over the human data. To further test the theory’s predictive power, we froze all four parameters and ran an Experiment 3 in which we introduced a smattering of 50 and 75 point bonuses along the path to the LL reward, adjusting the front-of-queue reward such that the reward-rate ratio $\rho = 1.50$ was attained when traversing the entire queue (see example in Figure 5c). Using the fully constrained model from Experiment 2, the fit obtained for Experiment 3 was quite good (Figure 5d). The model might underpredict long-queue initial defections, but it captures the curvature of the hazard functions due to the presence of bonuses.

4.3 Experiment 4: Customized Bonuses

In Experiment 4, we tested the effect of bonuses customized to a subpopulation. To situate this Experiment, we reviewed the Experiment 2 data to examine inter-participant variability. We stratified the 50 participants in Experiment 2 based on their mean reward rate per action. This measure reflects quality of choices and does not penalize individuals who are slow. With a median split, the weak and strong groups have average reward rates of 103 and 132, respectively. Theoretically, rates range from 0 (always switching between lines and never advancing) to 100 (deterministically selecting the short queue) to 150 (deterministically selecting the long queue). We fit the hazard curves of each group to a customized $\gamma$, leaving unchanged the other parameters previously tuned to the population. We obtained excellent fits to the distinctive hazard functions with $\gamma_{\text{strong}} = 0.999$ and $\gamma_{\text{weak}} = 0.875$.

We then optimized bonuses for each group for various line lengths. Instead of using the investment framework (Section 3) for determining bonus-based rewards—which would have led to variable reward rates and thus dependencies among game episodes—we searched over a bonus space consisting of all arrangements of four 50-point bonuses, allowing multiple bonuses at the same queue position. We subtracted 200 points from the LL reward, maintaining a reward-rate ratio of $\rho = 1.50$ for completing the long queue. We constrained the search such that no mid-queue defection strategy would lead to $\rho > 1$. A brute-force optimization yields bonuses early in the queue for the weak group, and bonuses late in the queue for the strong group (Figure 5e).

Experiment 4 tested participants on three line lengths—6, 10, and 14—and three bonus conditions—early, late, and no bonuses. (The no-bonus case was as in Experiment 2.) The 54 participants who completed Experiment 4 were median split into a weak and a strong group based on their reward rate on no-bonus episodes only. Consistent with the model-based optimization, the weak group performs better on early bonuses and the strong group on late bonuses (the yellow and blue bars in Figure 5f). Importantly, there is a $2 \times 2$ interaction between group and early versus late bonus ($F(1, 51) = 11.82, p = .001$) indicating a differential effect of bonuses on the two groups. Figure 5f also shows model predictions based the parameterization determined from Experiment 2. The model has a perfect rank correlation with the data, and correctly predicts that both bonus conditions will facilitate performance, despite the objectively equal reward rate in the bonus and no-bonus conditions. That bonuses should improve performance is nontrivial: the persistence induced by the bonuses must overcome the tendency to defect because the LL reward is lower (as we observed in Experiment 1 with $\rho = 1.25$ versus $\rho = 1.50$).

5 Discussion

Our primary contribution is a formal theoretical framework for modeling the dynamics of intertemporal choice. We obtained support for the theory by demonstrating that it explains key qualitative phenomena (Section 2.2) and predicts quantitative outcomes from a series of behavioral experiments (Section 4). The theory allows us to design incentive mechanisms that steer individuals toward better outcomes (Section 3), and we showed that this idea works in practice for customizing bonuses to subpopulations playing our queue-waiting game. Because the theory has just four free parameters, we have hope for the feasibility of fitting the model to the data of an individual. And with such fits comes the potential for maximally effective, truly individualized solutions.
Acknowledgments

This research was supported by NSF grants SES-1461535, SBE-0542013, SMA-1041755, and seed summer funding from the Institute of Cognitive Science at the University of Colorado. We thank Ian Smith and Brett Israelson for design and coding of the experiments.

References


