From Intentions to Actions: A Model and Experimental Evidence of Inattentive Choice*

Dmitry Taubinsky †

First version: November, 2013
This version: February, 2014

Abstract

A growing body of evidence suggests that people's inattention may be a significant friction in domains of behavior ranging from medical compliance, to financial decisions, to residential energy use. In this paper, I present a psychologically grounded model of *inattentive choice* and investigate its implications for dynamic decisions. The model explains seemingly puzzling patterns of consumer behavior, makes novel predictions that I confirm in two experiments, and generates a rich set of market implications. Applied to repeated actions, the model provides an attention-based foundation for the formation of "good" habits in domains such as exercise or energy use. The model explains the recent evidence on the intertemporal spillover effects of temporary incentives, and makes testable predictions about when attention-focusing cues, such as reminders, will dampen or amplify the effects of incentives. Consistent with these predictions, the first experiment reported in this paper shows that while temporary interruptions to daily routines decrease subsequent performance of the behavior, reminders have the largest impact after an interruption. Applied to tasks that must be completed by a deadline, the model identifies when longer deadlines will make people less likely to complete a task. But additionally, the model leads to new comparative statics, tested in the second experiment reported in this paper, about how reminders can eliminate the potentially perverse effect of longer deadlines. Finally, I apply the model to study market interactions between sophisticated firms and inattentive consumers: the model predicts how firms will take advantage of consumer's inattention through sales strategies such as rebates, and also leads to a dynamic theory of how firms use reminder advertisements to steer the behavior of inattentive consumers.

*I am especially grateful to David Laibson, Michael Luca, Sendhil Mullainathan, and Alvin Roth for their encouragement and generous advice. For helpful comments, I thank Hunt Allcott, John Beshears, Pedro Bordalo, Tom Cunningham, Drew Fudenberg, Ed Glaeser, Judd Kessler, Antonio Rangel, Alex Rees-Jones, Todd Rogers, Josh Schwartzstein, Andrei Shleifer, Tomasz Strzalecki, and Eric Zwick. I also thank the staff at the Harvard Decision Science Lab, especially Gabe Mansur, for technical assistance with the experiments reported in this paper. Finally, I thank the Harvard Business School Doctoral Office for financial support of this project. The title of this paper is taken from Icek Ajzen's (1985) classic paper "From Intentions to Actions: A Theory of Planned Behavior."

†Department of Economics, Harvard University and HBS (taubinsk@fas.harvard.edu)

1 Introduction

The relationship between people's preferences and actions is characterized by a variety of frictions. Frictionless theories of intertemporal choice are hard to reconcile with people's poor adherence to medication regimens, lack of exercise, low savings rates, frequently late bill payments, and inefficient energy use. Incomplete information, systematically biased beliefs, and self-control problems are often invoked to explain what appear to be suboptimal choices, and have been formally modeled in a variety of economic settings.

This paper theoretically and experimentally investigates a different bias that may play an important role in intertemporal decision-making: people are *inattentive*, and will not take an action that is not on their minds. At one point in time, a person may have fully intended to take all of his daily medications, pay his bills on time, and apply for a rebate; at another point in time, those once salient intentions may no longer be contemplated, buried by daily distractions, worries, and new goals.

In the case of poor compliance with medication regimens, for example, chronic disease patients frequently report "forgetting" as the reason for why they do not take all of their prescribed daily medication (MacDonell et al., 2013; Vyankandondera et al., 2013). This is corroborated by evidence from cognitive psychology experiments, which show that attentional and memory failures measured in controlled laboratory environments are significantly correlated with poor medication adherence (Poquette et al., 2013; Zogg et al., 2012).

Another growing body of evidence builds on a basic implication of limited attention: cues directing people's attention to a particular action should make its execution more likely. In decisions affecting health outcomes, simple cues such as email or text message reminders have been shown to increase compliance with daily medication regimens (Vervloet et al., 2012), scheduling of check-ups and screening tests (DeFrank et al., 2009), exercise (Calzolari and Nardotto, 2012), and healthy food choices (Patrick et al., 2009). In financial decision making, simple cues have been shown to increase savings (Karlan et al., 2012a) and timely loan repayments (Cadena and Schoar, 2011). And various cues directing attention to electricity use have decreased energy consumption by increasing simple behaviors such as turning off energy-using appliances (Gilbert and Zivin, 2013; Jessoe and Rapson, forthcoming).

However, the full implications of inattention to choices and actions in dynamic decisions are poorly understood:[1] What makes a certain choice or action more or less likely to come to mind? When should reminder cues should be more or less effective, and how do they interact with financial incentives?[2] How do sophisticated, profit-maximizing firms respond to consumer inattention?

[1]In contrast, the implications of limited capacity to update expectations in dynamic environments have been much more studied in rational inattention models and other approaches.

[2]Note that models of rational inattention (e.g. Sims 2003), which I discuss in more detail in Section 2.6, do not naturally allow for simple cues such as reminders to affect choice, and thus do not shed light on these issues. The

This paper proposes a dynamic model of *inattentive choice* and theoretically and experimentally investigates its implications. The first ingredient of the model is that cues can affect behavior by reminding the person about the choice or action. The second ingredient is what research in psychology and other cognitive sciences calls *rehearsal* or *accessibility bias*: the idea that recent engagement with a certain choice or action will increase the likelihood that it rises to the top of the mind again. Two forms of rehearsal are key to the decisions studied in this paper: a weak form of rehearsal is when the person thinks about the action, a stronger form is when the person performs the action.

For repeated actions—such as taking medication, attending the gym, or turning off electricity-using appliances to cut energy use—rehearsal implies that these behaviors will be habit-forming. The more a person has performed these behaviors in the past, the more likely they are to be top of mind, and thus the more likely they are to be performed again. A consequence of rehearsal, therefore, is that temporary incentives have a ripple effect: for example, incentivizing people to make energy conserving actions a more integral part of their routines will increase the attention devoted to those actions in a way that will persist when those incentives are removed. I show that rehearsal also has implications for tasks that must be completed by a deadline; in particular, it can significantly diminish the option value of longer deadlines. A consumer who purchases a product with the intention of mailing in the accompanying $50 rebate will be likely to think about the rebate again several days later but, because of rehearsal, will be less likely to think about it again in 4 weeks. As a consequence, consumers who are naive about their future inattention can be hurt by longer deadlines.

These predictions are consistent with a broad set of empirical findings, and show that patterns of behavior that are sometimes attributed to other theories such as hyperbolic discounting (Laibson, 1997; O'Donoghue and Rabin, 1999) or habit-forming *preferences* (Becker and Murphy, 1988) may also be a consequence of time-varying attention. At the same time, the model is sharply distinguished from other theories by its predictions about how changes in cues can modify behavioral responses to changes in payoff parameters. For tasks with deadlines, the model predicts that a small set of reminders can alter the potentially perverse effect of longer deadlines. For repeated actions, the model makes predictions about how reminders can amplify or diminish the spillover effects of temporary shocks to peoples' routines. These novel predictions about the interaction of cues and payoff parameters are immediately translatable into new experimental designs, and are tested and confirmed in two experiments reported in this paper.

These results, as well as others, lead to four main contributions. First, I show that a simple and psychologically grounded model of inattentive choice provides a unifying explanation of what appear to be "non-standard" responses to changes in payoff parameters. Second, I show that the model

simple reason is that an uninformative reminder should not lower the costs associated with processing information about a particular option.

extends the economic analysis of incentives by providing a framework for studying the effects of cue provision with a theoretical precision that generates sharp, testable hypotheses. The model generates novel predictions for the previously unexamined question of how cues and incentives interact to influence behavior, and provides a needed organization of conditions under which cues should or should not have a significant impact. Third, I test the new predictions of the model in two real-effort experiments that employ 2×2 factorial designs to vary both payoff parameters and cues. These experiments provide the first direct evidence that both the deadline and the habit effects predicted by the rehearsal property of the model are, indeed, a consequence of time-varying attention. Fourth, I show that beyond its predictions for individual behavior, the inattention model can be fruitfully incorporated into a variety of market analyses. I illustrate the model's applicability by using it to theoretically study marketing tactics such as consumer rebates, and to complement existing theories of advertising by deriving optimal policies of *reminder advertising*.

Section 2 presents the formal model, which builds on the imperfect recall model in Mullainathan (2002), and related approaches by Ericson (2010) and Holman and Zaidi (2010). I consider a decision maker (DM) who each period must choose whether or not to take a certain action when it is available. But with some probability, the DM is *inattentive*, and therefore does not even consider taking the action. Closely following evidence from psychology, I assume that cues and rehearsal increase the probability of the DM being attentive. To complete the model and apply it to dynamic decisions, I further draw on the experimental literature to specify how people form beliefs about the possibility of their future inattentiveness. Because the evidence discussed in Section 2 suggests that, on average, people vastly underestimate the likelihood of future inattentiveness, I draw out the additional implications of this naivete by considering two different ways of completing the model: *sophisticated beliefs* that are correct about future (in)attentiveness or *naive beliefs* that assume full attentiveness in all future periods.[3]

Section 3 studies the behavior of an inattentive DM in a repeated action setting such as medication adherence, exercise, or residential energy use. Section 3.1 contains the theoretical results, while Section 3.2 reports an experiment testing the new predictions. Section 3.1 shows that because engaging in a behavior increases the probability that it will be attentively considered in the future, the model predicts behavioral patterns similar to those predicted by preference-based theories of habit formation (Becker and Murphy, 1988), and documented for exercise (Charness and Gneezy, 2009; Acland and Levy, 2011) and residential energy use (Jessoe and Rapson, forthcoming). But at the same time—and consistent with evidence from gym attendance (Calzolari and Nardotto, 2012) and energy use (Allcott and Rogers, 2012; Gilbert and Zivin, 2013)[4]—the model predicts that cues

[3] Appendix A.3 further generalizes these two illustrative extremes and proposes a model of partial naivete.

[4] The cues described in Allcott and Rogers (2012), however, may act through channels other than just attention, because they also included normative information.

such as reminders affect both current *and* future behavior.[5] Moreover, in addition to capturing both the habit effects and the cue effects in a single parsimonious framework, the model makes predictions about when increasing cues will amplify or diminish the impacts of incentives. Consistent with Jessoe and Rapson (forthcoming), the model predicts that people will be more responsive to incentives following a prior or concurrent increase in cues. At the same time, the model also predicts that the effects of current incentives on *future* behavior are decreasing in the strength of future cues. This second comparative static follows from core idea of the model that people do not need reminders for behaviors that are routine. That is, the model predicts that habits and reminders are substitutes.

The experiment reported in Section 3.2 examines the interaction between repeat performance and reminders. The task in this experiment resembles taking daily medication and involves completing a daily survey in return for \$1 per completed survey. The 2×2 design varies 1) whether this task is available each day for three weeks straight or whether it is available for only the first and third weeks, and 2) whether or not participants receive daily reminders in the third week. I find that even without week 3 reminders, participants who are assigned to perform the task for three weeks straight complete approximately 5 out of 7 surveys in week 3. In contrast, those participants who experience the week 2 interruption and don't receive week 3 reminders complete only about 2.5 surveys. But consistent with the prediction that habits and reminders are substitutes, this effect of the week 2 interruption is almost entirely eliminated when subjects receive reminders in week 3. Reminders have only a minor effect on subjects who do not experience the interruption, but nearly double the week 3 completion rate of subjects who do experience the interruption. This interaction between repeat performance and reminders provides strong evidence that repeat performance of a behavior increases the likelihood that it is attentively considered in the future.

Section 4 studies the behavior of an inattentive DM who must complete some task—e.g., pay a bill, schedule a medical appointment, redeem a rebate, return a product—by a certain deadline. Section 4.1 contains the theoretical results, while Section 4.2 reports an experiment testing the new predictions. The central theoretical results in Section 4.1 concern how extending the deadline affects completion rates and welfare. On the one hand, the model predicts that a longer deadline increases a *sophisticated* DM's welfare. On the other hand, the model predicts that a longer deadline can decrease a *naive* DM's welfare and completion probability *when the likelihood of being attentive to the task decays over time*. Intuitively, this is because the naive DM does not realize that he will be significantly less likely to think of the task at a later point in time, and thus puts off the task more than he should with a longer deadline. This result is consistent with experimental evidence on short vs. long deadlines for rebate redemption (Silk, 2004), gift certificate redemption

[5] This prediction resembles models of cue-triggered cravings (Laibson, 2001; Bernheim and Rangel, 2004). The scopes are very different, however. Laibson (2001) and Bernheim and Rangel (2004) focus on conditioned, visceral responses to immediate consumption opportunities, and apply the models mostly to addictions. Most of the repeated-action behaviors studied in this paper, in contrast, are "good" habits that typically involve temporally distant benefits that should not elicit any sort of craving.

(Shu and Gneezy, 2010), and product returns (Janakiraman and Ordóñez, 2012). The second central theoretical prediction for task completion helps differentiate the inattention mechanism from alternative explanations of deadline effects. The prediction is that *if* a longer deadline leads to lower completion rates *then* an appropriately placed reminder will have a significantly larger effect on decision makers facing the longer deadline, potentially reversing the "perverse" effect of the longer deadline.

Section 4.2 reports an experiment testing the new prediction about the interaction of deadlines and reminders. Participants in the experiment receive a cash reward for completing a 20 minute questionnaire by a certain deadline. The 2×2 design varies the deadline length between either 2 days or 21 days, and varies whether participants receive reminders during the last two days including the deadline. Replicating previous evidence, I find that among participants who do not receive reminders, the longer deadline *decreases* the probability of completing the task from 59% to 42%. However, I find that reminders increase completion rates by 15 percentage points for subjects facing the shorter deadline, and increase completion rates by a striking 31 percentage points for subjects facing the longer deadline. Thus, while existing results on deadline effects replicate in the absence of reminders, they are nearly eliminated by reminders—a prediction unique to the inattention mechanism.

Section 5 builds on the theoretical results in Sections 3 and 4 to formally explore some market implications of the model. Section 5.1 builds on Section 4 and proposes a formal model of consumers rebates. In contrast to previous theoretical work—which has focused on price discrimination and has modeled consumers' redemption decision as a static, one-period choice—I consider a dynamic model of the rebate redemption process and use the inattention model to analyze the previously ignored question of how firms choose the redemption deadlines. Different from existing theoretical work, but consistent with policymakers' and industry experts' claims, the model predicts that rebates are deceptively attractive to inattentive and naive consumers. Moreover, the model predicts that firms will use "intermediate length" deadlines to maximally exploit consumer mistakes. Finally, the model shows that consumer rebates can facilitate socially inefficient transactions, and provides a formal framework for evaluating various policy proposals to regulate rebate offers. Building on the insights developed in the rebate model, I also discuss implications for product return policies and automatic renewal billing.

Section 5.2 builds on Section 3 to study the optimal cue provision strategy of an organization interested in increasing consumers' likelihood of taking some repeated action. The simple advertising model developed in this section generates insights for a variety of different types of communications: A health care provider or insurer using SMS messages or phone calls to remind chronic disease patients to take their medications, an organization sending reports to remind consumers of ways to save energy, or a firm advertising a repeat-purchase product to make sure that it stays top of mind. I show that as in models of informative advertising, the returns to each additional message are decreasing in the intensity of previous communications. But in contrast to models of informative advertising, the inattention model predicts that the optimal intensity of reminder messages should not converge to zero in the long run, and that the optimal advertising strategy might involve cycles or intermittent messages. Moreover, when the behavioral rehearsal effect is sufficiently strong, the model also predicts that the optimal message intensity will be non-monotonic in consumers' preferences for choosing the action; that is, reminder messages will be most effective for behaviors that consumers want to take, but not so often that these actions become habitual. These results, as well as others, illustrate how the model complements existing theories of informative and persuasive advertising by providing new insights and foundations for what marketers refer to as reminder advertising.

Section 6 discusses some limitations of the analysis in this paper, as well as directions for future research.

2 Model and Evidence

Sections 2.1-2.3 set up the model, Section 2.4 further motivates the assumptions, Section 2.5 offers some further remarks on the model, and Section 2.6 briefly reviews the related economics literature.

2.1 The Decision Environment

I define the model for a general binary choice decision environment in which payoffs and the action space can depend arbitrarily on the past history of actions. This encompasses actions that must be taken only once over the course of T periods, such as paying a bill, and actions that are taken repeatedly every period, such as taking medication. Although not analyzed in this paper, the model can also be applied to actions that are taken "every so often"; e.g., actions that must be taken once every five periods.

There are T periods $t = 0, \ldots, T$, with $T < \infty$. Each period the decision maker (DM) makes a choice x_t from a choice-set $X_t(h_t) \in \{\{d\}, \{d, a\}\}$, where $h_t = (x_0, \ldots, x_{t-1})$ denotes the history of choices up to period t. The choice $x_t = a$ is an *active* choice—e.g., paying the bill or taking the medication. The choice $x = d$ is a passive default—e.g., not paying the bill or not taking the medication. When $X_t(h_t) = \{d\}$, there is no choice to be made in period t. In the bill example, for instance, this will be the case if the bill has already been paid in period $\tau < t$.

The DM's period t flow utility after choice x_t is given by $u_t(x_t, \xi_t, h_t)$, where the ξ_t are independent random draws from a distribution F with bounded support in $[\underline{\xi}, \bar{\xi}]$, and are realized prior to the DM taking his action. Unless otherwise stated, I will assume that F is atomless and fully supported on $[\underline{\xi}, \bar{\xi}]$. In this case, I will assume that F has a density function, which I will denote by

f.

Variation in ξ_t represents fluctuations in daily opportunity costs or variation in taste shocks. I assume that $u(d, \xi_t, h_t) = 0$ for all ξ_t and h_t, so that only the utility from the non-default choice fluctuates. The DM's period t utility from a sequence of realizations $u_t, u_{t+1}, \ldots$ is given by $U_t = \sum_{\tau \geq t} u_\tau$. Note that for simplicity, I assume no discounting, though allowing for an exponential discount factor would not change the results in any way. Appendix A.1, generalizes the model to infinite horizons and time discounting.

2.2 Attention Dynamics

Each period, the DM either thinks about a, denoted $\alpha_t = 1$, or does not think about a, denoted $\alpha_t = 0$. When $\alpha_t = 0$, he always chooses the default d. When $\alpha_t = 1$, the DM compares a to the default alternative and chooses the better option subject to the DM's beliefs about his own future actions. I will describe the state $\alpha_t = 1$ as *attentive* and the state $\alpha_t = 0$ as *inattentive*.

Each period, the DM also receives a set of attention cues $\Omega_t \sim G_t$, with strength $\sigma(\Omega_t) \in [0, 1]$. Cues increase the probability that the DM is attentive, and may include a host of events that direct attention to the choice of $x_t = a$, including advertisements, fliers, email or text message reminders, conversations with others, and various visual and auditory cues intended to serve as reminders. I will frequently use the shorthand of writing σ instead of $\sigma(\Omega_t)$, and I will let H_t denote the period t distribution of values of σ.

In period $t \geq 1$, the DM is more likely think about a if 1) he has thought about it in period $t-1$ (i.e. $\alpha_t = 1$); if 2) he actually engages in the behavior a in period $t-1$ (i.e., $x_{t-1} = a$); and if 3) he receives a salient set of cues in period t. I will refer to the effect of α_{t-1} on period t attentiveness as *mental rehearsal* and I will refer to the effect of x_{t-1} on period t attentiveness as *behavioral rehearsal*. Section 2.4 reviews the cognitive psychology literature that motivates these attention effects.

Formally, I assume that for $t \geq 1$, $Pr(\alpha_t = 1) = g(\alpha_{t-1}, x_{t-1}, \sigma(\Omega_t))$, where $(\alpha_t, x_t) \in \{(0, d), (1, d), (1, a)\}$ and g satisfies the following assumptions:[6]

A1 g is continuous and increasing in σ, and $g(0, d, 1) = 1$

A2 $g(1, d, 0) > 0$

A3 $g(0, d, \sigma) \leq g(1, d, \sigma) \leq g(1, a, \sigma)$ for all σ

A4 For any $\sigma_1 < \sigma_2$, $g(0, d, \sigma_2) - g(0, d, \sigma_1) \geq g(1, d, \sigma_2) - g(1, d, \sigma_1) \geq g(1, a, \sigma_2) - g(1, a, \sigma_1)$

The first assumption states that the stronger the set of cues the DM receives, the more likely he is to be attentive. Moreover, there exists a sufficiently strong set of cues that would make the DM fully

[6] Note that since the DM cannot choose $x_t = a$ when he is inattentive, the pair $(0, a)$ is not possible.

attentive. The second assumption states that if the DM was attentive last period, there is always a chance that he will be attentive this period. The third assumption formalizes the idea that rehearsal increases the likelihood of being attentive to a. The fourth assumption states that the effect of cues on period t attentiveness is less pronounced if behavioral and mental rehearsal already make a more top of mind.

The fourth assumption can be motivated with the following intuitive model of cues: Suppose that the DM receives a set of cues $\Omega = \{\omega^1, \ldots, \omega^K\}$ in period t. The effect of each additional cue ω_k is that it triggers thoughts of the action with probability $\sigma(\{\omega_k\})$, and thus decreases the probability of being inattentive by a factor $1 - \sigma(\{\omega_k\})$. Thus the set of cues Ω makes the DM attentive with probability

$$g(\alpha_{t-1}, x_{t-1}, \Omega) = 1 - [(1 - g(\alpha_{t-1}, x_{t-1}, 0))(1 - \sigma(\{\omega_1\})) \cdots (1 - \sigma(\{\omega_K\}))] \tag{1}$$

and can be said to have an aggregate strength of $1 - \prod_{i=1}^{K}(1 - \sigma(\{\omega_i\}))$. Clearly, specification (1) satisfies A1-A4 as long as $g(\alpha_{t-1}, x_{t-1}, 0)$ satisfies A3.

Assumptions A1-A4 will be in full force throughout the paper. Some results will also rely on strengthening A3 and A4 to require that rehearsal strictly matters:

A3′ $g(0, d, \sigma) < g(1, d, \sigma) < g(1, a, \sigma)$ for all $\sigma \in [0, 1)$.

A4′ For any $\sigma_1 < \sigma_2$, $g(0, d, \sigma_2) - g(0, d, \sigma_1) > g(1, d, \sigma_2) - g(1, d, \sigma_1) > g(1, a, \sigma_2) - g(1, a, \sigma_1)$.

Note that specification (1) satisfies A4′ when $g(\alpha_{t-1}, x_{t-1}, 0)$ satisfies A3′.

In period 0, $Pr(\alpha_0 = 1) = g(0, d, \sigma(\omega_0))$, with $\sigma(\omega_0)$ potentially very close to 1 if that's when the DM first learns about the behavior option. To further simplify notation, I will set $\gamma_t(\alpha_{t-1}, x_{t-1}) = \int g(\alpha_{t-1}, x_{t-1}, \sigma)dH_t(\sigma)$, and set $\gamma_0 = \int g(0, d, \sigma)dH_0(\sigma)$. That is, $\gamma_t(\alpha_{t-1}, x_{t-1})$ is simply the probability of being attentive in period t, as a function of period $t-1$ events.

2.3 Strategies and Beliefs

Sophisticated Decision Makers

A sophisticated DM correctly anticipates the dynamics of his (in)attention, and optimally chooses his actions with those dynamics in mind. His strategies can be computed by backward induction.

To formalize, let E_F denote the expectation taken with respect to F, and let (h_t, x_t) denote the period $t+1$ history that results when x_t is chosen after history h_t. Let $V_t^{\mathrm{s}}(h_t, \alpha_t)$ denoted the sophisticated DM's expected period t utility conditional period t history h_t and on whether or not he is attentive in period t. Let $\mathbf{x}^{\mathrm{s}}(h_t, \xi_t)$ denote the DM's period t strategy *conditional on being attentive.* The functions V_t^{s} and $\mathbf{x}^{\mathrm{s}}$ are defined recursively below. To ease notation, I sometimes suppress the arguments of $\mathbf{x}^{\mathrm{s}}(\cdot)$. For $t = T$,

$$V_T^{\mathrm{s}}(h_T, \alpha_t) = \alpha_t E_F u(\mathbf{x}^{\mathrm{s}}, \xi_T, h_T) \tag{2}$$

$$\mathbf{x}^{\mathrm{s}}(h_T, \xi_T) = \mathrm{argmax}_{x \in A(h_t)} u(x, \xi_T, h_T) \tag{3}$$

and for $t < T$,

$$V_t^{\mathrm{s}}(h_t, \alpha_t) = \alpha_t E_F \left[u(\mathbf{x}^{\mathrm{s}}, \xi_t, h_t) + V_{t+1}^{\mathrm{s}}((h_t, \mathbf{x}^{\mathrm{s}}), \gamma(1, \mathbf{x}^{\mathrm{s}})) \right] + (1 - \alpha_t) E_F V_{t+1}^{\mathrm{s}}((h_t, d), \gamma(0, d)) \tag{4}$$

$$\mathbf{x}^{\mathrm{s}}(h_t, \xi_t) = \mathrm{argmax}_{x \in A(h_t)} \{ u(x, \xi_t, h_t) + V_{t+1}^{\mathrm{s}}((h_t, x), \gamma(1, x)) \} \tag{5}$$

As formalized in equation (5), a sophisticated DM recognizes that he may be inattentive in the future and considers how cues and rehearsal affect the probability of being attentive in period $t + 1$.

Naive Decision Makers

As summarized in Section 2.4, many people appear to be naive about the possibility of future inattention. To investigate the role of naivete in a stark and illustrative way, in the body of the paper I make the extreme assumption that a naive DM is *fully naive*; that is, he thinks that he will be perfectly attentive in the future.[7] In appendix A.3, however, I introduce models of partial naivete. In general, a sufficiently—though not fully—naive DM would exhibit many of the same forms of misoptimization that a fully naive DM exhibits, but to a lesser extent.[8]

Conditional on being attentive, the naive DM makes the same choice that a fully attentive DM would make. And the utility that a naive DM *thinks* he will realize, in expectation, corresponds to the utility that a perfectly attentive DM *actually* realizes, in expectation.

Formally, let $\mathbf{x}_t^{\mathrm{n}}$ denote the naive DM's strategy, and let $\tilde{V}_t^{\mathrm{n}}$ denote the naive DM's expectation of his time t utility conditional on history h_t. Then analogous to equations (2)-(5),

$$\tilde{V}_T^{\mathrm{n}}(h_T) = E_F u(\mathbf{x}^{\mathrm{n}}, \xi_T, h_T) \tag{6}$$

$$\mathbf{x}_T^{\mathrm{n}}(\xi_T) = \mathrm{argmax}_{x \in A(h_T)} u(x, \xi_T, h_T) \tag{7}$$

and for $t < T$,

$$\tilde{V}_t^{\mathrm{n}}(h_t) = E_{F_t} \left[u(\mathbf{x}^{\mathrm{n}}, \xi_t, h_t) + \tilde{V}_{t+1}^{\mathrm{n}}((h_t, \mathbf{x}^{\mathrm{n}})) \right] \tag{8}$$

$$\mathbf{x}^{\mathrm{n}}(h_t, \xi_t) = \mathrm{argmax}_{x \in A(h_t)} \{ u(x, \xi_t, h_t) + \tilde{V}_{t+1}^{\mathrm{n}}((h_t, x)) \} \tag{9}$$

[7] This is analogous to O'Donoghue and Rabin's (1999) analysis of (fully) naive versus sophisticated hyperbolic discounters.

[8] In the decisions studied in Section 4, for example, the partially naive DM would still put off the task more than he should, but not as much as the fully naive DM. And as I discuss in appendix A.3, the results of Section 3 for naive DMs go through verbatim under milder forms of naivete.

The expected period t utility V_t^{n} that a naive DM actually realizes is determined analogously to equations (2) and (4), with $\mathbf{x}^{\mathrm{n}}$ in place of $\mathbf{x}^{\mathrm{s}}$ and with V_t^{n} in place of V_t^{s}.

2.4 Psychological foundations: Evidence and examples

In this section I discuss psychological evidence motivating the foundations of this model. Readers may skip this section without loss of continuity.

Attention dynamics

Rehearsal. A key difference between one-time actions such as redeeming a rebate and repeated actions such as taking medication is that repetition of a behavior option can increase the likelihood that it will come to mind again. The idea that future retrieval of information or intentions is made easier by rehearsing the retrieval has a long tradition of research in cognitive psychology (Atkinson and Shiffrin, 1969). In laboratory studies, Jacoby et al. (2001) coined the term "accessibility bias" to refer to the effect that repetition of a response has on the ease with which it comes to mind.[9] Recent work has proposed that repeated engagement in recycling behaviors (Tobias, 2009) or repeated purchasing of a product (Henderson et al., 2011) increases the attention devoted to those actions.

Other work (Sellen et al., 1997) suggests that even passive thoughts about one's intentions can increase the likelihood that they will be retrieved in the future. As Sellen et al. (1997) argue, however, mental rehearsal is not enough to permanently preserve intentions at the top of the mind in the absence of external cues. McBride et al. (in press), for example, gave study participants postcards and instructed participants to mail them back—in return for a $50 lottery—after randomly chosen delays of 1, 2, 5, 14, or 30 days. Corroborating the hypothesis that mental rehearsal does not by itself preserve goals, performance was significantly better by subjects randomized into the short delays.[10] Analogous results on task completion have been obtained for laboratory tasks with delays varying between 5 minutes and 60 minutes (McBride et al., 2011; Martin et al., 2011).[11]

Cues. Many types of cues may trigger thoughts of the behavior. Visual cues such noticing the pill bottle can trigger thoughts of the associated action; glancing at a billboard can refresh consideration

[9]In laboratory tests of accessibility bias (Jacoby et al., 2001; Hay and Jacoby, 1996), subjects first become accustomed to responding a certain way to stimulus words such as "knee." For some subjects the required response to "knee" is "bone" 75% of the time while for others it is "bend." The next day, subjects are asked to memorize a list of word pairs such as "knee/bone." In the quiz stage of day 2, subjects who become accustomed to responding to "knee" with "bone" in day 1, are both much faster and more accurate in their responses to "knee" in day 2 than are subjects who become accustomed to the response "bend."

[10]An average 65% return rate for delays of 1, 2, or 5 days versus an average 48% return rate for delays of 14 or 30 days.

[11]However, it is not clear that these laboratory tasks tap into the same psychological mechanisms as the naturalistic experiments. The typical "prospective memory" tasks require subjects to respond a certain way to a cue or prompt after a short time delay. Success in these tasks relies on short bursts of vigilance efforts that don't seem to have direct "real world" analogs. Most field behaviors involve unprompted actions, such as "canceling the gym membership at some point next week."

of the advertised product. Auditory cues can have similar effects. Charles et al. (2007) found that a combination of visual and audio reminders built into an electronic pill bottle had a substantial impact on adherence to inhaled corticosteroid therapy over the course of 24 weeks.

Firms and organizations often capture attention through various messages: reminder emails, SMS messages, or advertisements. Recent evidence from randomized controlled trials shows that reminder messages have large effects on health behaviors such as gym attendance (Calzolari and Nardotto, 2012), sunscreen use (Armgstrong et al., 2009), adherence to medication regimens (Krishna et al., 2009; Vervloet et al., 2012), obtaining immunizations (Szilagyi et al., 2000), scheduling/attending medical appointments (Altmann and Traxler, 2012), and even weight loss (Patrick et al., 2009).[12] In financial settings, Karlan et al. (2012a) find that reminders increase savings deposits, while Cadena and Schoar (2011) find that reminders increase timely loan repayments by small business owners.[13] In marketing, Nedungadi (1990) and Mitra and Lynch. (1995) find that non-persuasive product primes can increase the consideration and subsequent purchase of the product, without increasing the product's perceived value.

Finally, incidental events such as conversations with others can also constitute cues. Hearing a colleague complain about her referee report, for example, may trigger thoughts about one's own refereeing duties.

Cues are imperfect. Although cues can have powerful effects on behavior, most cues are still imperfect in the sense that a period t cue cannot guarantee attentiveness in period t. One reason for this imperfection is that emails, SMS messages, fliers, calendar reminders, and many other types of communications do not always reach the targeted person. There is no guarantee that the person will check his mail or phone, or be at his computer. The person may also choose to discard many of these communications before even reading them. Second, even within a time period of 1 day, a person will not necessarily receive the cue at a time when he is ready to act (Tobias, 2009). A reminder email to schedule a medical appointment might be read by the person when he is at work, but then subsequently forgotten several hours later when the person is ready to act. Third, indirect cues such as conversations with others are not guaranteed to trigger associated thoughts of the behavior. Fourth, a person with a limited capacity to process information may not meaningfully process all communications or stimuli. Section 6, for example, discusses how a person may become desensitized to certain cues.

[12] Over the course of 4 months, Patrick et al. (2009) found that individuals receiving SMS messages reminding them to purchase healthy foods, pack healthy snacks, drink water, and other such activities lost 1.97kg more than the control group. Haapala et al. (2012) found that over the course of 12 months, individuals receiving SMS reminders lost 3.4kg more than the control group ($N = 125$).

[13] See, however, Karlan et al. (2012b) who find that only personalized SMS reminder messages seem to affect repayments of loans in the Philippines.

Naivete

Experimental studies show that people overestimate the likelihood that they will take a certain action on a later date. Ericson (2011), for example, used an incentive compatible mechanism to elicit subjects' beliefs about the likelihood that on a later date, they will send an email to the experimenter to claim a $20 payment (that would subsequently be mailed to them). Quite strikingly, only 53% of the subjects sent the email, though the average incentive-compatible forecast was 76%. While the failure to carry out such a high reward/low cost task suggests significant inattention,[14] the large discrepancy between forecasts and actual behavior also suggests significant naivete. In similar tasks resembling rebate redemption, Silk (2004), Letzler and Tasoff (2013),[15] and Shu and Gneezy (2010) also find a large and robust discrepancy between forecasts and behavior.

Several psychological factors likely contribute to naivete about one's future inattention. One psychological channel is motivational: people like to hold favorable beliefs about the future or about their own traits (Camerer, 1997). A second psychological channel is cognitive: people overuse their current disposition to predict their future disposition, as in projection bias (Loewenstein et al., 2003). A person focusing on a certain task on Tuesday will simply have trouble imagining how he could forget about this task on Wednesday or Thursday.

2.5 Remarks

Note that in the baseline model proposed thus far, the agent has no control of the cues in his environment. Indeed, many types of cues are not set by the DM himself: cues coming from interested parties, incidental cues such as conversations with others, or random events (e.g, a new rebate opportunity reminds the DM of a current rebate that is about to expire). This starting point abstracts from the fact that people also use reminder technologies such as calendars, alarm clocks, "to do" lists, etc. In Appendix A.2, I consider a more general environment in which the DM first makes additional investments in cues, and then participates in the subgame corresponding to the baseline model defined in this section. In the appendix, I discuss the extent to which investment in cues modifies the results in the paper, and argue that the qualitative results remain largely unchanged. Although a *sophisticated* DM who can cheaply purchase a perfect cue technology and is always attentive to the possibility of this investment option can essentially eliminate his inattention, I argue in appendix A.2 that this is an extreme case.

[14] Ericson's (2011) calibrations show that alternative explanations such as time inconsistency or underestimation of effort cannot explain subjects' behavior given the high benefit-cost ratio.

[15] Letzler and Tasoff (2013) also found that a reminder sent approximately 2.5 weeks before the deadline reduced overconfidence by 7 percentage points by increasing redemption, though the difference was not statistically significant. Results in section 4, however, suggest that reminders sent far from the deadline should not be very effective for naive decision makers because these decision makers will delay completion of the task, and end up forgetting about the task again over the course of that delay.

2.6 Relation to economics work on limited attention

By studying inattention to *actions*, rather than *information*, my model differs from most existing economic models of inattention. The model in this paper complements models of limited attention in information processing (Sims, 2003; Gabaix and Laibson, 2005; Peng and Xiong, 2006; Reis, 2006; Schwartzstein, 2012; Woodford, 2012; Caplin and Dean, 2013b,a) and models of disproportionate focusing on some attributes over others (Gabaix and Laibson, 2006; Gabaix, 2012; Köszegi and Szeidl, 2013; Bordalo et al., 2013). The information processing models capture the idea that decision makers may have cognitive limitations in how much information they can use to evaluate an option; my model, in contrast, captures the idea that a decision maker may form a clear intention for how he would like to act in the future, but then fails to follow through on that intention because it is not top of mind. Like the information processing models, the focusing models also lead to biases in the *evaluation* of options, but not to the gap between intentions and future actions captured in this paper. Moreover, neither the information processing models nor the focusing models allow for uninformative cues or past actions to directly influence attentiveness. In Chun et al.'s (2011) terminology, existing economic models would be classified as models of *external attention*—the modulation of how external stimuli or new information is processed—while my model would be classified as a model of *internal attention*—the modulation of internally generated information such as intentions and goals.

Mullainathan (2002) studies limited recall of *information*. The idea that repetition and cues increase attentiveness parallels Mullainathan's modeling of the rehearsal and associativity properties of recall memory. Closely related to this paper, Holman and Zaidi (2010) formulate a model of "prospective memory" ("memory for action") for a decision environment in which there are no external cues and an action can be taken only once (and thus there is no scope for repetition to increase future attentiveness).[16] Their model is formally nested as a special case of the model in this paper. Also related is the model that Karlan et al. (2012a) develop for the specific context of consumption and savings with "lumpy expenditures." Karlan et al. (2012a) assume that consumers might be inattentive to certain aspects of a decision—namely, potential future expenditures—but that cues can direct attention to those aspects of a decision.

[16]Ericson (2010) also considers a model in which present-biased agents permanently forget about a task with some probability. "Forgetting," as applied to intentions and actions, might not be the right term, however. In contrast to a person focused on recalling a historical date but consciously failing, the person who absent-mindedly fails to take his medication is not aware of the memory retrieval failure while it is occurring—the intention has been stored in memory and is in tact, but the problem is that the person has simply not directed his attention to retrieving that intention from memory. Some psychologists use the term "prospective forgetting" in laboratory paradigms in which subjects fail to respond to a certain cue or prompt after short time delays. Applied to field behaviors, however, Dismukes (2012) points out that the term "is something of a misnomer, given that what it refers to involves the cognitive process of planning, attention, and task management as much at involves memory. After forming an intention, individuals often become engaged with various ongoing tasks and, in most everyday situations, cannot hold the deferred intention in a focal attention." As footnote 11 also points out, laboratory studies of what cognitive psychologists often call prospective forgetting don't necessarily capture the same mechanisms that operate in field behaviors. I suggest that a better description of the underlying psychology is inattention to previously formed intentions, choices, and actions.

3 Repeated actions

3.1 Theory

In this section, I investigate the model's implications for decisions that are made repeatedly and on a regular basis. These include health behaviors such as taking medication or making plans to exercise, actions affecting energy conservation such as turning of energy-using appliances and adjusting the thermostat at peak hours, and actions constituting parts of workplace routines.

Formally, the DM learns about the repeated action in period $t = 0$ and chooses $x_t \in \{d, a\}$ in each period $t = 1, \ldots, T$. The DM's utility from choosing $x_t = a$ in period t is $b_t + \xi_t$. Throughout this section, I will rely on the stronger assumptions A3′ and A4′, which state that rehearsal strictly increases next period's probability of attention, and that this diminishes the impact of cues. I will also assume that $b_t + \bar{\xi} > 0$ and $b_t + \underline{\xi} < 0$ for all t, so that $Pr(b_t + \xi > 0) > 0$ and $Pr(b_t + \xi < 0) > 0$ for all t. Finally, I will assume that for all $t \geq 1$, the DM will be inattentive with positive probability: $\gamma_t(1, a) < 1$ for all $t = 1, 2, \ldots T$.

A naive DM will choose $x_t = a$ if and only if doing so generates positive flow utility that period: $b_t + \xi_t \geq 0$.[17] A sophisticated DM's strategies are more complicated because he considers how his current actions impact the probability of future attentiveness through behavioral rehearsal. A sophisticated DM will always choose $x_t = a$ when $b_t + \xi_t \geq 0$, but he may also choose $x_t = a$ when $b_t + \xi_t < 0$.

I begin the analysis in the section by characterizing how changes in payoffs b_t impact behavior in period t, as well as in all subsequent and prior periods. I will let $Pr^{\mathrm{s}}(x_t = a)$ and $Pr^{\mathrm{n}}(x_t = a)$ denote respective probabilities that sophisticated and naive DMs choose $x_t = a$ in period t, from the period 0 perspective.[18]

The assumption that F is atomless and has a density function ensures that for both sophisticated and naive DMs, $Pr(x_t = a)$ is differentiable in $b_{t'}$ for *any* $t' \geq 1$ (Lemmas D1 and D2 in appendix D.1).

Proposition 1. *1. For $t \leq t'$,*

$$\frac{\partial}{\partial b_t} Pr^n(x_{t'} = a) > 0 \text{ and } \frac{\partial}{\partial b_t} Pr^s(x_{t'} = a) > 0.$$

[17] In the model of naivete in the body of the paper, naive DMs don't consider how performing a behavior affects future attentiveness because they assume that they will be perfectly attentive to the behavior in the future. More generally, however, overestimation and naivete about the effects of behavioral rehearsal don't need to be as tightly linked. The more general models in Appendix A.3 allow for a separation between a general overestimation of attention and a more specific naivete about behavioral rehearsal. All results in this section about naive DMs hold under the narrower assumption that the DM simply does not understand that behavioral rehearsal increases the probability of future attentiveness.

[18] That is, probabilities of choosing $x_t = a$ when conditioning on the null, period 0 history.

2. For all $1 \leq t'' < t$,

$$\frac{\partial}{\partial b_t} Pr^n(x_{t''} = a) = 0 \text{ and } \frac{\partial}{\partial b_t} Pr^s(x_{t''} = a) > 0.$$

3. For all $1 \leq t'' < t \leq t'$,

$$\frac{\partial^2}{\partial b_{t''} \partial b_t} Pr^n(x_{t'} = a) > 0 \text{ and } \frac{\partial^2}{\partial b_{t''} \partial b_t} Pr^s(x_{t'} = a) > 0.$$

Proposition 1 fully characterizes how changes in payoffs affect the behavior of sophisticated and naive DMs. Part 1 states that increasing the payoff to choosing $x_t = a$ increases the probability that both a naive and a sophisticated DM will choose $x_{t'} = a$ for $t' \geq t$. The intuition is that increasing b_t increases the probability that conditional on being attentive, both sophisticated and naive DMs find it worthwhile to choose $x_t = a$. Because of rehearsal, this then leads to higher probabilities of being attentive in periods $t + 1, t + 2, \ldots$, and thus higher likelihoods of choosing $x = a$ in those periods.

Part 2 states that increasing future payoffs to choosing $x = a$ also increases a sophisticated DM's motivation to invest in future attentiveness through behavioral rehearsal. Thus for $t'' < t$, a sophisticated DM's likelihood of choosing $x_{t''} = a$ is increasing in b_t.

Part 3 states that for $t'' < t < t'$, period t'' and period t payoffs will have complementary effects on period t' behavior. The intuition is simple: Part 1 shows that increasing $b_{t''}$ makes the DM more attentive in all future periods. In particular, this means that the DM will be more attentive in period t, which will increase his responsiveness to period t payoffs.

Parts 1 and 2 of Proposition 1 show that the model predicts behavioral patterns similar to what theories of rational, taste-based habit formation predict (Becker and Murphy, 1988). The mechanism is quite different, however. Becker and Murphy (1988) assume that repeated consumption generates a "habit stock" that enters the utility function and creates higher marginal utility from future consumption. Most of the applications of the Becker and Murphy (1988) framework have focused on "bad" habits such as addictions, in which past consumption increases the marginal utility from increasing future consumption, but also lowers the level of utility for any given choice of consumption. The attention model in this paper does not lead to, or apply to such bad habits. However, the model provides an alternative mechanism for the various "good" habits discussed at the beginning of this section. Temporarily increasing the returns to taking some action will increase the likelihood of taking that action is taken during the temporary increase in incentives; but, because of rehearsal, the model predicts that the person will then become more attentive to the action, and thus more likely to take it even when the additional incentives are no longer in place. Evidence for such spillover effects has been document in the case of energy use (Jessoe and Rapson, forthcoming) and exercising (Charness and Gneezy, 2009; Acland and Levy, 2011).[19]

[19] In the case of gym attendance, it is probably less likely that the person literally "forgets" to attend the gym after

What distinguishes the inattention model from models such as Becker and Murphy (1988) are its predictions about how cues will affect behavior and how they will modulate the spillover effects of incentives. I now turn to formally investigating these effects. Analogous to Proposition 1, I begin by characterizing how changes in the cue distribution H_t affect behavior in period t, as well as behavior after period t and before period t. Throughout, I will write $H_t^1 >_{FOSD} H_t^2$ if H_t^1 first order stochastically dominates H_t^2. When comparing outcomes under two different sequences of cue distributions $\mathbf{H}^1 = (H_1^1, \ldots, H_T^1)$ and $\mathbf{H}^2 = (H_1^2, \ldots, H_T^2)$, I will index the corresponding outcomes with the superscripts H^1 or H^2. I will set $\mu_t^i = \int \sigma dH_t^i$.

Proposition 2. *Consider two sequences of cue distributions* $\mathbf{H}^1 = (H_1^1, \ldots, H_T^1)$ *and* $\mathbf{H}^2 = (H_1^2, \ldots, H_T^2)$ *such that for some* t', $H_{t'}^2 >_{FOSD} H_{t'}^1$ *but* $H_\tau^1 = H_\tau^2$ *for* $\tau \neq t'$. *Then*

1. $Pr^{n,H^2}(x_t = a) > Pr^{n,H^1}(x_t = a)$ *for all* $t \geq t'$
2. *When* $t' > 1$, $Pr^{n,H^2}(x_t = a) = Pr^{n,H^1}(x_t = a)$ *for all* $t < t'$.
3. $Pr^{s,H^2}(x_t = a) > Pr^{s,H^1}(x_t = a)$ *for all* $t \geq t'$ *when a)* $t' = 1$ *or b)* $\mu_{t'}^2$ *is sufficiently close to 1.*
4. *When* $t' > 1$, $Pr^{s,H^2}(x_t = a) < Pr^{s,H^1}(x_t = a)$ *for all all* $t < t'$.
5. $Pr^{n,H^2}(x_t = a) - Pr^{n,H^1}(x_t = a)$ *is decreasing* $H_{t''}$ *for all* $t'' < t' \leq t$.

Part 1 states that adding cues unambiguously increase the likelihood that a naive DM chooses $x_t = a$ both at the time that the cues arrive and in all future periods. By assumption, a period t_2 cue increases the likelihood that the DM is attentive in period t_2. And because of rehearsal, this effect spills over into all periods following t_2. Part 2 simply says that naive DMs will not adjust their behavior in anticipation of future cues.

An analog to part 1 holds for sophisticated DMs when the change in cues occurs in period 1, as in part (a) of part 3. Generally, however, the impact on sophisticated DMs' behavior is more nuanced because of another channel through which cues affect behavior: sophisticated DMs realize that the more likely they are to be reminded of the behavior in the future, the less important it is to make it habitual. Thus, the anticipation of future cues crowds out sophisticated DMs' motive to invest in making the behavior more habitual, as formalized in part 4 of Proposition 2.[20] When the

making plans to do so. Rather, it is more likely that infrequent gym goers probably don't think very often about the possibility of attending the gym, and typically don't think to incorporate it into their daily plans and routines. This is consistent with the findings in Charness and Gneezy (2009) that temporary incentives have the biggest effects on irregular gym goers.

[20] In the baseline model presented here, sophisticated DMs can increase attentiveness to $x_t = a$ only through repeated choice of the action. In the more general model in Appendix A.2, sophisticated DMs could also increase attentiveness through investments in reminder technologies. Either way, when a sophisticated DM's investment choice is discrete, an additional cue in period t_2 can indirectly lower the probability of this DM's period t_2 attentiveness by crowding out investments in increasing the likelihood of being attentive in the future.

additional cues are sufficiently strong, however, as in condition (b) of part 3 of the proposition, the crowd out effect becomes relatively weak.

Finally, part 5 of Proposition 2 states that for naive DMs, temporally separated cues always have substitutable effects on behavior. The more cues the DM gets in period 1, for example, the smaller the impact of period 6 cues on period $t \geq 6$ behavior. Intuitively, this is because the more cues the DM gets in period 1, the more likely he is to be attentive in period 6 (by part 1 of the proposition); as a consequence, there is less scope for period 6 cues to make him more attentive that period. This result that temporally separated cues are substitutes is in contrast to part 3 of Proposition 1, which shows that increases in temporally separated *payoffs* have *complementary* effects on behavior.

Recent evidence on exercise (Calzolari and Nardotto, 2012) and energy use (Allcott and Rogers, 2012; Gilbert and Zivin, 2013) is consistent with parts 1 and 3 of Proposition 2, and complements the evidence on habit formation in those domains. Calzolari and Nardotto (2012) show that over the course of 6 months, weekly reminders to attend the gym increase attendance from an average of 8.19 visits per month to an average of 9.31 visits per month. Moreover, they find that the reminder intervention continues to have an effect on behavior for up to 3 months after it ends.

Gilbert and Zivin (2013) show that households reduce consumption by 0.6% to 1% following each electricity bill. Allcott and Rogers (2012) show that following each home energy report, consumers are more likely to engage in repeated actions such as turning off lights, unplugging unused electronics, and adjusting thermostats. Interestingly, Allcott and Rogers (2012) also find that the impact of each additional home energy report diminishes with time. While Allcott and Rogers (2012) propose that one possible reason for this is "desensitization" to previously encountered cues,[21] part 5 of Proposition 2 shows that this effect of diminishing marginal returns to additional cues can arise endogenously as a consequence of rehearsal. In the advertising setting examined in Section 5.2, I further extend the prediction that temporally separated cues will be substitutes, and examine the implications for optimal policies of reminder advertising.

In addition to their direct impact on behavior, cues also modify how behavior responds to incentives. The next two propositions characterize how 1) period t_1 cues change the response to period $t_2 \geq t_1$ incentives and how 2) period t_1 incentives change the response to period $t_2 > t_1$ cues.

Proposition 3. *Take $t_1 \leq t_2$ and consider two sequences of cue distributions $\mathbf{H}^1 = (H_1^1, \ldots, H_T^1)$ and $\mathbf{H}^2 = (H_1^2, \ldots, H_T^2)$ such that for some t_1, $H_{t_1}^2 >_{FOSD} H_{t_1}^1$, but $H_\tau^1 = H_\tau^2$ for $\tau \neq t_1$.*

1. For all $t \geq t_2$,

$$\frac{\partial Pr^{n,H^2}(x_t = a)}{\partial b_{t_2}} > \frac{\partial Pr^{n,H^1}(x_t = a)}{\partial b_{t_2}}.$$

[21]See Section 6 for a discussion of this effect.

2. *If $t_1 = 1$, then for all $t \geq 1$,*

$$\frac{\partial Pr^{s,H^2}(x_t = a)}{\partial b_{t_2}} > \frac{\partial Pr^{s,H^1}(x_t = a)}{\partial b_{t_2}}.$$

Part 1 of Proposition 3 shows that receiving additional cues prior to, or during period t, makes the naive DM's behavior more responsive to period t incentives. In particular, the DM is more responsive in period t and, because of rehearsal, this heightened responsiveness also spills over into periods $t+1, \dots T$. Thus, in the specific sense described in Proposition 3, cues *amplify* the effects of incentives.

Intuitively, the DM cannot respond to incentives for choosing $x_t = a$ unless he is attentive. Thus, when adding cues increases the likelihood of attentiveness, it also magnifies the response to incentives. For a naive DM, this leads to the straightforward prediction that increasing cues in some period t' magnifies the response to incentives in any period $t \geq t'$.

For a sophisticated DM, matters can be more complicated because, by part 4 of Proposition 2, an increase in period t cues crowds out the motivation to choose $x_{t'} = a$ in periods $t' < t$. In particular, the DM anticipating fewer cues in the future may also be more sensitive to changes in future payoffs due to the strategic rehearsal motive. When $t_2 = 1$, however, so that the strategic rehearsal motive is shutdown, the result also holds for sophisticated DMs. I have not yet been able to find more general conditions under which an analog to part 1 of Proposition 3 holds for sophisticated DMs.

The results of Proposition 3 are in line with recent evidence on residential electricity use. Jessoe and Rapson (forthcoming) study how energy use responds to temporary price increases, and find that supplementing the usual price change notifications with additional cues in the form of electricity meters makes people significantly more elastic to price changes. The intuitive explanation provided by Proposition 3 is that the electricity meters increase the likelihood of being attentive to energy use, and thus increase the fraction of people who take actions to reduce energy consumption during a temporarily high price. Appendix A.4 derives a formal corollary of Proposition 3 for energy use elasticities.

At the same time, there is also an important sense in which increasing cues can diminish the impact of incentives. In particular, the intertemporal spillover effect of period t_1 incentives becomes less and less pronounced as period $t_2 > t_1$ cues are increased:

Proposition 4. *Take $t_1 < t_2$ and consider two sequences of cue distributions $\mathbf{H}^1 = (H_1^1, \dots, H_T^1)$ and $\mathbf{H}^2 = (H_1^2, \dots, H_T^2)$ such that for some t_2, $H_{t_2}^2 >_{FOSD} H_{t_2}^1$, but $H_\tau^1 = H_\tau^2$ for $\tau \neq t_2$.*

1. *For all $t \geq t_2$,*

$$\frac{\partial Pr^{n,H^2}(x_t = a)}{\partial b_{t_1}} < \frac{\partial Pr^{n,H^1}(x_t = a)}{\partial b_{t_1}}.$$

2. For all $t \geq t_2$,

$$\frac{\partial Pr^{s,H^2}(x_t = a)}{\partial b_{t_1}} < \frac{\partial Pr^{s,H^1}(x_t = a)}{\partial b_{t_1}}$$

when μ_{t_2} is sufficiently close to 1.

Part 1 of Proposition 4 shows that for a naive DM, the effect of period t_1 incentives on behavior in periods $t \geq t_2$ becomes less and less pronounced as period t_2 cues are increased.

Part 2 establishes the same result for a sophisticated DM. The statement is weaker because the period t_2 cues can change how a sophisticated DM responds to incentives in period $t_1 < t_2$. In particular, it is possible that the sophisticated DM may become more responsive to period t_1 incentives as period t_2 cues are increased. Nonetheless, a sufficiently large increase in period t_2 cues will diminish the impact of period t_1 incentives even for a sophisticated DM. Intuitively, if period t_1 incentives impact future behavior by changing the likelihood of future attentiveness, then their effects will become negligible if future cues are sufficiently strong to guarantee almost perfect attentiveness.

Proposition 4 is particularly helpful for distinguishing taste-based theories of habit formation from the theory proposed in this paper. The simple but diagnostic prediction that reminders should have the largest effect when the behavior has not been recently performed is difficult to rationalize with any other existing theories, or even combinations of existing models. For example, if people had habit-forming preferences *and* were inattentive with some probability p that was independent of their past behavior (but could be bolstered with cues), then they would behave in a manner opposite to Proposition 4.[22]

Yet to my knowledge, this interaction between repetition of a behavior and subsequent reminders has not been investigated in economics, psychology, or related disciplines. Section 3.2 below reports an experiment investigating this relationship.

3.2 Experimental Evidence

Design and Procedures

To examine the interaction between action repetition and reminders, I conducted an online, real-effort experiment spanning three weeks of daily tasks. The daily task that subjects had to perform repeatedly over the course of 3 weeks was to complete a short, online daily survey. The repetitive nature of this short, simple task closely resembles taking daily medication or online browsing behaviors. Disguising this task as a "survey study" and not disclosing the true purpose of the experiment

[22] Intuitively, this is because for these types of decision makers, a period t reminder would only amplify the impact that past behavior would have on the preference for choosing $x_t = a$. The logic behind this comparative static is analogous to the logic behind Proposition 3.

made it possible to study how subjects *naturally* approach these kinds of real-world behaviors.[23]

The 2×2 factorial design varied whether or not subjects were given the opportunity to complete the task in the middle week, and whether or not subjects received reminders in the third week. The experiment consisted of three phases, depicted in Figure 1:

Registration Phase (Day 0) Subjects interested in participating in the experiment were directed to the study website, where they created an account for the experiment. Upon creating an account, they were randomized into the study conditions described below. They then received a short overview of the 3-week experiment and completed 5 minutes of demographic and lifestyle questions for which they received $3. They were then given detailed instructions explaining their "daily survey" task, which started the *day after* they signed up and spanned the next three weeks. Immediately upon completing the registration phase, subjects received an automatic email with a copy of their electronic consent form, and all study instructions (which included a link to the study website).

Daily Survey Phase (Days 1-21) The simple task subjects faced in days 1-21 was to log into the study website (using the email and password they registered for this experiment), and to report their current level of excitement, happiness, stress, and worry using a 1-5 Likert scale. Subjects received $1 for each day that they completed this daily survey. To ensure a time break between any two consecutive survey completions, the daily survey was not available between 12:00 a.m. and 4:59 a.m. each day. In all four conditions, subjects did not initially know if the survey would be available to them in week 2 of the study. They were told that it would be available in week 2 with 50% chance, and that they would be notified of this at 5:00 a.m. on the first day of week 2. The four conditions were as follows:

- *No interruption / No reminders.* Subjects in this condition received only 2 communications throughout the study:[24] An email on day 1 reminding them to start completing the daily survey, and an email on day 8 informing them that the survey would be available for the next 14 days.

- *Week 2 interruption / No reminders.* This condition was identical to the condition above, except that on day 8, these subjects received email notification that the daily survey would not be available to them for the next 7 days, but that it would be made available again for the last 7 days of the study.

- *No interruption / Week 3 reminders.* This condition was identical to the *No interruption / No reminders* condition, except that subjects received an email each day of week 3 stating

[23] This experiment involved absolutely no deception, however—only *incomplete disclosure*. Subjects were told all details of the experiment other than the true purpose of the study (and were not led to believe that the purpose of the study was anything other than what it is).

[24] All communications were emailed out at 5:00 a.m.

"This is a reminder that the daily survey is available to you between 5:00 a.m. and 11:59 p.m. today."

- *Week 2 interruption / Week 3 reminders.* This condition was identical to the *Week 2 interruption / No reminders* condition, except that subjects received an email each day of week 3 stating "This is a reminder that the daily survey is available to you between 5:00 a.m. and 11:59 p.m. today."

Post-task phase At the end of the three weeks of the daily survey, subjects completed a short closing questionnaire, for which they were paid an additional $3.[25] The closing questions asked about subjects' use of reminder devices to complete the daily survey, as well as subjects' daily routines (if any) for completing the daily survey. The questions are described and analyzed in more detail below, as well as in Appendix B.

In all conditions, subjects were informed on day 0 of all of the communications they would receive, and were told that they would receive no other communications.

All earnings were emailed to subjects in the form of an Amazon.com gift certificate after the completion of the study.[26]

Formally, the week 2 interruption can be modeled as creating substantially lower benefits from performing the action.[27] Notice that in contrast to the decisions studied in the theory section, however, whether or not the week-long interruption would occur was not known with certainty. A consequence of this feature is that conditional on the reminders treatment, week 1 behavior should be identical for subjects with a week 2 break and without a week 2 break. Controlling for week 1 behavior thus generates additional statistical power. Propositions B1 and B2 in Appendix B generalize Propositions 1 and 4 to the experimental setting studied here.

Results

The experiment was run through the Harvard Decision Science Laboratory (HDSL) during July and August 2013, and enrollment was limited to members of the HDSL subject pool.[28] A total

[25]Subjects who completed the daily survey on day 21 were immediately prompted to complete the 5 minute closing questionnaire. Subjects who did not complete the daily survey on day 21 were sent an email the next day inviting them to complete the closing questionnaire. All subjects had one week to complete the closing questionnaire, and were sent a reminder each day of that week to complete the questionnaire. If subjects did not complete the closing questions within one week, they did not receive the additional $3, and were emailed a receipt for their other earnings.

[26]For subjects who completed the closing survey and signed a study receipt, these payments were emailed within 24 hours. Subjects who never completed the closing survey were emailed a link to a receipt for their previous earnings 8 days after the end of the daily survey phase.

[27]So technically, participants very serious about keeping this activity on their mind can always log into the study site, and then pretend to answer the 4 feelings question. Alternatively, this week 2 break can also be modeled as creating such low benefits that no subject would ever want to complete the survey in week 2. A few subjects did log in to the study site on the first day of week 2, but this probably reflects confusion.

[28]All members of HDSL were eligible to participate in this experiment. After signing up through HDSL's SONA system, subjects were given a link to the study site, where they could initiate the registration process.

of 187 subjects completed the Registration phase. Of these 187, 7 subjects did not complete any daily surveys during the first week, and are thus excluded from all subsequent analysis. This exclusion is particularly reasonable because these 7 subjects also did not complete any surveys during the subsequent weeks, meaning that in my sample, not doing the survey in week 1 was perfectly predictive of dropping out of the experiment altogether. Recruitment took place over the course of 6 days: on July 15,16,17 subjects who signed up were randomized into the no reminders conditions (105/180 subjects), while on August 6,7,8 subjects who signed up were randomized into the reminders conditions (75/180) subjects.[29]

Figure 2 shows weekly averages for subjects in all 4 conditions, with error bars corresponding to 95% confidence intervals. Strikingly consistent with the habit effect identified in Proposition 1 (and extended to this experimental design in Proposition B2) subjects who did not receive reminders in week 3 completed significantly fewer surveys if they experienced a week 3 interruption. While uninterrupted subjects completed an average of 5.2 out of 7 surveys in week 3, subjects who experienced a week 2 interruption completed an average of only 2.3 surveys in week 3.

At the same time, the effect of the interruption is almost completely undone by the daily reminders in week 3. While week 3 reminders have almost no effect on uninterrupted subjects (a minor increase of 0.3 surveys), they increase the interrupted subjects' week 3 completion rates from an average of 2.3 to an average of 4.4 surveys.

Table 1 quantifies the effects in Figure 2 by estimating linear probability models of daily survey completion, with robust standard errors clustered at the subject level.[30] Specification (1) includes only the experimental conditions as the independent variables, while specifications (2) and (3) also include the average daily completion rate in week 1 (i.e., number of surveys completed in week 1 divided by 7) as a covariate. All three specifications show that a break between weeks 1 and 3 reduces the probability that a subject completes a survey on any given day of week 3 by about 35 percentage points. At the same time, the regressions show that most of this effect is mitigated with daily reminders in week 3: the interaction effect between experiencing a week 2 break and receiving week 3 reminders is about 25 percentage points, which is significant at the 5% level in specification (1) and significant at the 1% level in specifications (2) and (3).

In the first specification, the effect of the week 2 interruption on week 3 performance in the reminders conditions is still a marginally significant 13.9 percentage points ($p = 0.097$), but some of

[29]That the reminders and no reminders conditions were split up across time might seem worrisome because of time-dependent shocks. However, even if calendar date random effects did cause differences in levels, it is hard to think of a reason for why they should confound the estimate of the interaction effect between the week 2 interruption and week 3 reminders. But to address the concern that calendar date random effects might somehow be inflating the interaction effect between week 2 interruption and reminders, I show that the results are also robust to clustering at both the subject level and calendar date level. I also show that subjects' demographics do not vary across the four experimental conditions, and that all results are robust to demographic controls.

[30]Table B3 shows that the results are unchanged with robust two-way clustering at both the subject level and the calendar date level.

that is due to a slightly unbalanced randomization. As shown in Figure 2, uninterrupted subjects also had a slightly higher week 1 average. Specification (2) controls for the week 1 average under the assumption that a subject's week 1 behavior should not be affected by study condition. This is the right assumption for naive subjects who do not adjust their behavior in anticipation of future cues, but not the right assumption for sophisticated subjects. Specification (3) addresses this theoretical point by interacting the week 1 average with whether or not a subject was in a reminders condition. In both of these specifications, the impact of the interruption is estimated to be 8 to 9 percentage points, which is not statistically different from zero ($p = 0.205$ and $p = 0.274$, respectively). Appendix B2 shows that all of these results are unchanged when controlling for various demographic characteristics (Table B2), and when calendar date random effects are taken into account (Table B3).

The end of study questionnaire, which was completed by 172 of the 180 subjects, provides further evidence in support of the model.[31] In response to the question "On a scale of 1-5, with 1 being 'not at all' and 5 being 'very much,' to what extent did the following factors contribute to you not filling out the daily surveys" subjects gave an average rating of 3.9 for "Forgot", 1.8 for "Didn't have time," 1.16 for "Didn't feel like it," 1.64 for "Could not remember log in information," and 1.58 for "No internet access." The difference between "Forgot" and each of the other reasons is significant at $p < 0.001$ in pairwise t-tests and signed rank tests.

During the registration phase, subjects were also asked to predict how many times they would complete the daily survey each week. They were asked four questions: 1) how many times they would complete the survey in week 1; 2) how many times they would complete the survey in week 2 if it was available then; 3) how many times they would complete the survey in week 3 if it was available in week 2; 4) how many times they would complete the survey in week 3 if it was not available in week 2. Although unincentivized and thus needing to be interpreted with caution, subjects' day 0 forecasts of their daily survey completion rates support the naive model. Of the 168/180 subjects who completed the predictions questions and gave answers within range,[32] the average forecast was between 6.72 and 6.9 surveys for each week. Subjects did not predict that having a week 2 interruption would affect their week 3 behavior, and their forecasts were insensitive to whether or not they were in the reminders or no reminders conditions. For all weeks and all conditions, the forecasts were significantly higher than actual behavior (paired t-tests $p < 0.001$ in all comparisons).

Of course, some of the initial confidence may reflect subjects' intentions to use reminder technologies, which, as analyzed in more detail in Appendix B, were employed by 30% of the subjects.

[31] It was completed by 102/105 subjects in the no reminders conditions, and by 71/75 subjects in the reminders conditions. Subjects were given one week to complete the end of study questionnaire, and received a daily reminder each day of the week to complete it.

[32] 98/105 in the no reminders conditions and 74/75 in the reminders conditions completed the questions. Of these, 2/98 subjects in the no reminders conditions and 2/74 in the reminders conditions gave out of range answers.

There is, however, no statistically significant interaction between forecasts and using reminder technologies. Thus the 70% of the subjects who used no reminder technologies appear to be quite naive. Appendix B also shows that subjects' week 3 performance was less affected by the week 2 interruption if they reported using a reminder technology such as a calendar (typically Google calendar), writing notes or asking others to remind them.[33] Moreover, these subjects were also less affected by the week 3 reminders.

4 Tasks With Deadlines

4.1 Theory

In this section I investigate the model's basic implications for decisions involving a single task that must be completed by a deadline T. The task might be making a savings deposit, applying for a rebate, paying a bill, canceling a subscription, scheduling an appointment, and so forth.

The DM learns about the task in period $t = 0$, and must complete the task at some time between period $t = 1$ and period $t = T$. The task can be completed only once, with $x_t = a$ corresponding to completing the task in period t, and with $x_t = d$ corresponding to not doing the task in period t. The DM obtains payoff $b_t + \xi_t$ if he completes the task in period t, and obtains a payoff of 0 if he never completes it. Throughout this section, I will consider a DM who is always attentive in period 0 when he first learns about the task: $\gamma_0 = 1$.

Throughout much of the analysis in this section, I will assume that the distribution F, deterministic payoffs $\mathbf{b} = (b_1, b_2, \dots)$, and cue distributions $\mathbf{H} = (H_1, H_2, \dots)$ are fixed, and that the choice of deadline T is applied to the (infinite) sequences $\mathbf{b}$ and $\mathbf{H}$, specifying that $X_t = \{d\}$ for all $t > T$. Thus for any two deadlines $T_1 < T_2$, a DM given a deadline T_2 faces the same payoff and cue distributions for the first T_1 periods as the DM facing a deadline T_1.

I assume that for an *inattentive* DM there is some $\bar{z} < 1$ such that $\gamma_t(1, d) \leq \bar{z}$ for all $t \geq 1$. A perfectly attentive DM will be defined as having $\gamma_t(1, d) = 1$ for all t.

The next proposition characterizes how inattention affects the probability of completing a task by some time $t \leq T$. In stating the results, let Q^{s} denote the random variable corresponding to the period in which a sophisticated DM completes the task, with $Q^{\mathrm{s}} = \infty$ if the task is never completed. Let Q^{n} and Q^{pa} be defined analogously for naive DMs and for perfectly attentive DMs, respectively.

Proposition 5. *1. $Pr(Q^n \leq t) < Pr(Q^{pa} \leq t)$ for all $t \leq T$.*

2. Suppose that $b_t = b$ for all t and that $b + \underline{\xi} \geq 0$. Then for each $t^\dagger \geq 1$, sequence of cue distributions $\mathbf{H}$, and atomless F, there is a sufficiently large T such that $Pr(Q^s \leq t) > Pr(Q^{pa} \leq t)$ for all $t \leq t^\dagger$, but such that $Pr(Q^s \leq T) < Pr(Q^{pa} \leq T)$.

[33] Of course, standard caveats about endogeneity apply.

3. *For each $t \geq 1$, there exists a $z^\dagger > 0$ such that $Pr(Q^s \leq t) < Pr(Q^{pa} \leq t)$ whenever $\gamma_{t'}(1,d) \leq z^\dagger$ for all $t' \leq t$.*

For naive DMs, the effect of inattention is straightforward: for any $t \leq T$, they are less likely to complete the task by period t than perfectly attentive DMs. This is because conditional on being attentive, they follow the same strategy as perfectly attentive DMs, but they never complete the task when inattentive. Part 1 of Proposition 5 formalizes this idea.

The behavior of inattentive but sophisticated DMs, however, is less straightforward to analyze because they may follow the strategy of "I should do it while it's on my mind." Part 2 of Proposition 5 shows that for sufficiently long deadlines and sufficiently high value tasks, inattentive but sophisticated DMs will actually be more likely to complete the task by some early date than perfectly attentive DMs. The intuition is that as the deadline becomes longer, the option value to waiting increases for perfectly attentive DMs, making it more and more attractive to put off the task until later. Inattentive but sophisticated DMs, however, will be afraid to put off the high value task to a later time because of the possibility of forgetting about it.

The last part of Proposition 5 is a partial analogue to the result for naive DMs. It is only a partial analogue because even for $t = T$, inattentive but sophisticated DMs may be more likely to complete the task due to the fact that they are less likely to put it off than perfectly attentive DMs. Part 3 shows, however, that if the sophisticated DM is sufficiently inattentive, then he will be less likely than the naive DM to complete the task by any period $t \leq T$.

I now turn to the question of how deadline length affects welfare and completion rates. In stating the formal results below, I will let Q^{s,T_i}, Q^{n,T_i}, Q^{pa,T_i} denote the random variables corresponding to when sophisticated, naive, and perfectly attentive DMs, respectively, complete the task (with $Q^{T_i} = \infty$ if the task is never completed) when facing deadline T_i. I will let V_0^{s,T_i}, V_0^{n,T_i}, V_0^{pa,T_i} denote the ex-ante expected utilities of sophisticated, naive, and perfectly attentive DMs, respectively, when facing a deadline T_i.

While many of the results will be stated for any sequence of payoffs $\mathbf{b}$, a leading case that will be considered throughout this section is the case of time-invariant payoffs: $b_t = b$ for all $t \leq T$. Most of the examples and applications discussed in this section involve a task that pays a fixed benefit b but involves a stochastic effort or opportunity cost ξ. A more general condition that subsumes this special case and that will be used in some of the results is the following:

Condition I(T_1, T_2) Given two deadlines T_1 and $T_2 = T_1 + \Delta > T_1$: i) $b_t \leq b_{t+\Delta}$ for all $t \leq T$ and ii) $b_t - b_{t-1} \geq b_{t+\Delta} - b_{t-1+\Delta}$ for all $t \leq T_1$.

In words, Condition I(T_1, T_2) states that the last T_1 periods of the longer deadline have a payoff profile that is no lower and no "steeper" than the payoff profile of the shorter deadline T_1.

Proposition 6 states results for perfectly attentive and for sophisticated DMs.

Proposition 6. *Consider two deadlines $T_1 < T_2$. Then*

1. $V_0^{s,T_1} < V_0^{s,T_2}$ *and* $V_0^{pa,T_1} < V_0^{pa,T_2}$

2. *Suppose that either Condition $I(T_1, T_2)$ holds or that $b_{T_2} + \underline{\xi} \geq 0$. Then $Pr(Q^{pa,T_1} \leq T_1) \leq Pr(Q^{pa,T_2} \leq T_2)$.*

3. *Suppose $b_t = b$ for all $t \leq T_2$ and that $\gamma_t(0, d) \geq \underline{z} > 0$ for all t. Then $Pr(Q^{s,T_1} \leq T_1) < Pr(Q^{s,T_2} \leq T_2)$ for sufficiently high b.*

Part 1 of Proposition 6 states that a sophisticated or fully attentive DM can only be made better off by longer deadlines. The straightforward intuition is that when given the longer deadline T_2, in the first T_1 periods the DM can still behave the same way he would behave if he was given the shorter deadline T_1.

Part 2 of Proposition 6 analyzes completion rates of DMs who are fully attentive. While longer deadlines unambiguously increase welfare of sophisticated DMs, the effect of deadlines on completion rates is less clear cut in certain special cases. When the task is not very pleasant or important *and* the distribution of task completion payoffs decreases with time, longer deadlines may lead to lower completion rates even for perfectly attentive DMs. As a simple example, suppose that $\xi \sim U[-1, 0]$ and that $b_1 = 1$ while $b_2 = 2/3$. Then when $T = 1$, the probability of task completion is 1 because $b + \xi_1$ is always positive. When $T = 2$, the option value of delaying to period 2 makes it suboptimal to complete the task in period 1 whenever ξ_1 is sufficiently close to zero. In period 2, however, the DM completes the task with probability less than 1. This then implies that the overall probability of task completion is less than 1 with a two period deadline.

Part 3 of Proposition 6 proves a result analogous to part 2 for inattentive but sophisticated DMs. When the task is sufficiently important to complete, sophisticated DMs will use the extra time to increase the probability of completing the task.

For naive DMs, however, I show below that extending a deadline can both reduce their welfare and lead to lower completion rates, even in situations in which longer deadlines lead to higher completion rates for sophisticated DMs.

Proposition 7. *Consider two deadlines $T_1 < T_2$. Then for any $\mathbf{b}$ and F, there exists a $\lambda > 0$ such that $V_0^{n,T_1} > V_0^{n,T_2}$ and $Pr(Q^{n,T_1} \leq T_1) > Pr(Q^{n,T_2} \leq T_2)$ if:*

i) $\gamma_t(1, d) \geq \gamma_{t+1}(1, d)$ *and* $\gamma_t(0, d) \geq \gamma_{t+1}(0, d)$ *for all* $t \geq 1$

ii) $\gamma_t(1, d) \leq \lambda$ *for all* $t > T_1$

ii) $\gamma_t(0, d) \leq \lambda\gamma_1(1, d)$ *for all* $t > T_1$

Proposition 7 shows that giving a naive DM more time to complete the task can make him worse off and less likely to complete the task if his probability of being attentive to the task decays sufficiently quickly over time, as guaranteed by the three conditions in the proposition. The simple intuition is that the longer the deadline, the higher is a naive DM's *perceived* option value of delaying the task and waiting for a more opportune time. What a naive DM doesn't realize, however, is that at a later date the task will probably not be on his mind.

Proposition 7 generalizes an insight developed by Holman and Zaidi (2010) for a setting that is formally equivalent to the case of no-cues, time-invariant payoff distributions, and an exponential attention decay curve that limits to zero. The intuition behind conditions (i)-(iii) in the proposition is as follows. Conditions (i) and (ii) ensure that if the DM is attentive in period $t = 1$, he will be attentive in period $t > T_1$ with low probability. Condition (iii) ensures that if the DM is not attentive in period $t = 1$, then his probability of being attentive in periods $t > T_1$ is low *relative* to his probability of being attentive in period $t = 1$. Combined, these three conditions ensure that the probability of being attentive in periods $t > T_1$ is substantially lower than the probability of being attentive in periods $t \leq T_1$. The naive DM's perception of periods' $t > T_1$ is substantially inflated, however. Consequently, he is much less likely to complete the task before period T_1 when given more time, while simultaneously not being very likely to complete the task after period T_1.

Importantly, the potentially perverse effect of longer deadlines is not generated simply by limited attention, but by the *decay* of attention over time. In fact, with time-invariant payoffs and a constant probability of attention ($\gamma_t(1, d) = \gamma_t(0, d) = \rho$ for all $t \geq 1$), longer deadlines would always leave a naive DM better off and lead to higher completion probabilities.

The conditions in Proposition 7 also provide insight into the mechanisms that lead to the decay of attention. When cue distributions don't change over time—$H_t = H$ for all $t \geq 1$—rehearsal always leads to decay of attention. To see this, suppose that for some $\bar{\rho} > \underline{\rho}$, $\gamma_t(1, d) = \bar{\rho}$ and $\gamma_t(0, d) = \underline{\rho}$ for all $t \geq 1$. Then the period t unconditional probabilities of being attentive, p_t, satisfy $p_{t+1} = p_t(\bar{\rho} - \underline{\rho}) + \underline{\rho}$. When $p_0 = 1$, the sequence $\{p_t\}$ can be shown to be decreasing; and Proposition 7 implies that there exists some $\lambda > 0$ such that T_2 leads to lower welfare and completion rates when $\bar{\rho} < \lambda$ and $\underline{\rho} < \lambda\bar{\rho}$.

Additionally, however, attention decay could also be a consequence of decreasing cues. Various paperwork lying on the table, conversations with others, or emails at the top of the inbox are examples of cues that are likely to be present when the DM first learns of the task but that will dissipate with time. Even in the extreme case in which γ_t is not affected by mental rehearsal, the cue distributions could still be such that the probability of being attentive in period $t \geq 1$ is given by, e.g., $\gamma_t(1, d) = \gamma_t(0, d) = k\bar{\rho}^t + \underline{\rho}$. Proposition 7 then again implies that completion rates will be lower under the longer deadline T_2 when $\bar{\rho}$ is low and when $\underline{\rho}$ is low *relative* to $\bar{\rho}$.

Empirically, an inverse relationships between deadline length and completion rates was first documented by Shafir and Tversky (1992), and subsequently replicated and extended by Shu and Gneezy (2010), Silk (2004), and Janakiraman and Ordóñez (2012). Shafir and Tversky (1992) offered students \$5 to complete a long questionnaire by a given date. One group ($N = 56$) was given a 5 days to complete the questionnaire while the other group ($N = 58$) was given 3 weeks. The rates of return were 60% for the short deadline group and 42% for the long deadline group.[34] Silk (2004) replicated this deadline effect in a setting closely resembling consumer rebates. Shu and Gneezy (2010) replicated this effect with gift certificates for "immediately enjoyable experiences," such as pastries and movies. And Janakiraman and Ordóñez (2012) replicated this effect in a setting closely resembling product returns.[35] Further supporting naive inattention, Silk (2004) and Shu and Gneezy (2010) also elicited beliefs, and found that while longer deadlines led to lower completion rates, participants predicted higher completion rates with a longer deadline.

The model provides a formal explanation for the evidence above, but also makes predictions about when longer deadlines *should not* lead to lower completion rates. The first prediction is that when the probability of being attentive is bounded away from zero—either because there is always a small chance of encountering a cue, or because $g(0, d, \sigma)$ is intrinsically bounded away from zero for all σ—the probability of task completion will approach 1 as deadlines become very long. Longer deadlines can unambiguously decrease completion rates only when $\gamma_t(0, d) = 0$ for all t—so that the probability of being attentive decays rapidly toward zero.[36]

Proposition 8. *Suppose that there is some $b^* > -\bar{\xi}$ such that $b_t \to b^*$.*

1. *If there is some $\underline{z} > 0$ such that $\gamma_t(0, d) \geq \underline{z}$ for all t, then $\lim_{T\to\infty} Pr(Q^{s,\ T} \leq T) = \lim_{T\to\infty} Pr(Q^{n,\ T} \leq T) = 1$.*

2. *If $b_t = b$ and if $\gamma_t(0, d) = 0$ for all t, then $\lim_{T\to\infty} Pr(Q^{n,\ T} \leq T) = 0$.*

Combined, Propositions 7 and 8 imply that completion rates will be lowest for "intermediate length" deadlines when cues are rare and/or weak. Figure 3 illustrates this "U-shaped" pattern. In both panels $b_t + \xi_t \sim U[0, 1]$ and the unconditional probability of $\alpha_t = 1$ is given by $0.8(0.75)^t + 0.2$. In the left panel, this decay is generated purely through the mental rehearsal property, while in the right panel this decay is generated purely by changes in cues.[37]

[34] Shafir and Tversky (1992) also had a "no definite deadline condition," but did not provide details on how they assessed completion rates in that condition. Without a definite deadline, even perfectly sophisticated and attentive agents might generate low completion rates by any finite date.

[35] Bertrand et al. (2010) examine how interest rate reductions that last 2, 4, or 6 weeks affect the demand for small, short-term loans by the working poor population in South Africa. Bertrand et al. (2010) find that consumers are significantly more likely to take out a loan with a longer deadline. This result, however, is not in conflict with the experimental evidence discussed in this paragraph. The reason is that a consumer is, mechanically, more likely to generate a need for a loan over the course of 6 weeks rather than over the course of 2 weeks. Conditional on generating a need within 2 weeks, consumers may be less likely to take out a loan from the lender when the interest perk lasts 6 weeks rather than 2; however, this effect is trumped by the fact that far fewer people will generate a need over the course of only two weeks.

[36] The convergence to zero completion probability is a knife-edge result, however, because if the DM discounted future payoffs by some $\delta < 1$, then his probability of task completion would be bounded away from zero. Still, longer deadlines could keep decreasing the completion probability even when $\delta < 1$.

[37] There are two caveats to this prediction, however. First, even if $\gamma_t(0, d)$ is bounded away from zero, then the

Unfortunately, Proposition 8 is difficult to test because it does not provide guidance on *when* the non-monotonicity should occur, if it does. A more diagnostic and easily testable prediction of the model is that even relatively small changes in cues can alter the effects of a longer deadline. The rough idea is that *if* longer deadlines decrease task completion due to the decay of attention over time, *then* increasing cues several periods before the deadline to stop the decay can have a very large effect on the completion probability of a DM facing the longer deadline. When attention decay is substantial enough to lead to perverse deadline effects, a DM who faces a long deadline T_2 and has not yet completed the task by period $T_2 - 1$ will respond very strongly to an increase in period $T_2 - 1$ cues; this DM simply won't be attentive unless he gets a cue.

To formally state comparative statics about adding cues under two different deadlines T_1 and $T_2 = T_1 + \Delta > T_1$, I will begin with *initial* attention probabilities $\gamma_t^0(\alpha, x)$ and consider increasing them to $\gamma_t^{\kappa_t^i}(\alpha, x) = \gamma_t^0(\alpha, x) + (1 - \gamma_t^0(\alpha, x))\kappa_t^i$, where $i \in \{1, 2\}$ denotes the resulting probability for each of two deadlines T_1 and T_2. Note that as in previous analysis, I assume that the cue distributions in periods $t \leq T_1$ are initially identical in both of the deadlines conditions. But in contrast to the previous analysis, I now allow the modified cue distributions to differ between deadlines. This set-up makes it possible to compare how, for example, sending a reminder on the last or second-to-last period affects completion rates under short versus long deadlines.

The parametrization of $\gamma_t^{\kappa_t^i}(\alpha, x)$ is natural and fits into the simple model of cues proposed in the discussion surrounding equation (1): Suppose that the DM is initially inattentive with probability $1 - \gamma$. Now consider sending reminders whose effect is that a DM who would have otherwise been inattentive is now attentive with probability κ. The reminders then increase the probability of being attentive to $\gamma + (1 - \gamma)\kappa$. In the results below, I make the additional restriction that $\kappa_t^1 = \kappa_{t+\Delta}^2$ for $t \leq T_1$, which allows me to examine how adding the same types and same number of cues to both the short and long deadline environments will affect completion rates and welfare.

Proposition 9. *Consider two deadlines T_1, $T_2 = T_1 + \Delta > T_1$. Then for any $\{\gamma_t^{\kappa_t^i}(1, d)\}_{t \geq 1}$, $\{\gamma_t^{\kappa_t^i}(0, d)\}_{t \geq 1}$ with the additional restriction that $\kappa_t^1 = \kappa_{t+\Delta}^2$:*

1. *Suppose that either Condition $I(T_1, T_2)$ holds or that $b_{T_2} + \underline{\xi} \geq 0$. Then there is a $\bar{\kappa} < 1$ such that $Pr(Q^{n,T_1} \leq T_1) < Pr(Q^{n,T_2} \leq T_2)$ and $Pr(Q^{s,T_1} \leq T_1) < Pr(Q^{s,T_2} \leq T_2)$ if $\kappa_t^1 = \kappa_{t+\Delta}^2 \geq \bar{\kappa}$ for $t \leq T_1$.*

2. *Suppose that $b_{T_2} + \underline{\xi} \geq 0$. Then $Pr(Q^{n,T_1} \leq T_1) - Pr(Q^{n,T_2} \leq T_2)$ is strictly decreasing in $\kappa_{T_1}^1 = \kappa_{T_2}^2$.*

possibility that the DM might, for example, lose the necessary paperwork to mail in his rebate would, effectively, lead to a task completion decay curve that limits to zero. Second, if the DM is sufficiently present-biased (Laibson, 1997; O'Donoghue and Rabin, 1999) and naive about the present bias, and if the variation in ξ is sufficiently small, then he will always put off the task until the last period. Combined with decaying, this would lead longer deadlines to unambiguously decrease completion rates. (Relatedly, Ericson (2010) considers a model in which present-biased agents permanently forget about the task with some probability ρ each period. Part 2 of Proposition 8 shows that even without present bias, longer deadlines can be uniformly bad in such a permanent forgetting model.)

3. *Suppose that Condition $I(T_1, T_2)$ holds and the initial cue distributions are time invariant, $\gamma_t^0 = \gamma^0$ for all $t \geq 1$. Then for each $t \leq T_1$, $Pr(Q^{n,T_1} \leq T_1) - Pr(Q^{n,T_2} \leq T_2)$ is strictly decreasing in $\kappa_t^1 = \kappa_{t+\Delta}^2$ while it is positive.*

Proposition 9 motivates a simple and testable hypothesis for situations in which longer deadlines lead to lower completion rates, and in which payoffs from task completion don't change with time, as in the settings studied by Shafir and Tversky (1992), Shu and Gneezy (2010), Silk (2004), and Janakiraman and Ordóñez (2012). Suppose, for example, that study participants are more likely to complete a task when given a 2-day deadline than when given a 3-week deadline. Part 1 of Proposition 9 implies that if the participants receive a sufficiently strong set of cues during the last two days in the 3-week condition and during both days of the 2-day condition, then the longer deadline should no longer lead to lower completion rates.

The second part of the proposition considers the case in which doing the task in the last period of the long deadline always generates positive flow utility that period. In this case, adding cues, however weak, in the last period of both the short and long deadline conditions should diminish the difference in completion rates between the short and long deadline conditions.

The last part of the proposition shows that under the additional assumption that the initial cue distributions do not vary from period to period, adding additional cues (however weak) any number of periods before the deadline has a bigger impact on a naive DM's completion probability when the deadline is long.

Attention decay is not only a sufficient condition, but also a necessary condition for reminder provision to have a larger effect in the long deadline condition. If the probability of being attentive to the task does not decline with time then, because more decisions makers will have completed the task by period $T_2 - 1$ then by period $T_1 - 1$, a reminder in period $T_i - 1$ will actually have a *smaller* effect in the longer deadline condition. And even when the likelihood of being attentive does decay with time, the "dropout effect" described in the previous sentence is still a countervailing force that pushes cues to have a smaller net effect in the long deadline condition. The conditions described in Proposition 9 ensure that the "decay effect" dominates the "dropout effect."

Results such as Proposition 9 motivate a straightforward way to diagnose when certain patterns of behavior are caused by inattention: increasing attention through various types of cues should diminish those behavioral patterns. Although other biases such as procrastination may also play a role people's ability to complete tasks, tests such as those suggested by Proposition 9 can help assess *how much* of a certain behavioral pattern is caused by inattention.[38][39]

4.2 Experimental Evidence

Design and Procedures

Motivated by Proposition 9, I conducted an online, real-effort experiment to measure how much of the long deadline effect (previously documented in other studies) is due to the decay of attention over time. In addition to varying deadline length as in previous work, the new experiment also varied reminder provision to test the comparative static derived in Proposition 9. The experiment consisted of two phases described below.

Registration Phase (Day 0) Potential subjects received an invitation email to complete a 10-20 minute survey in which they would have to choose between hypothetical gambles, and for which they would receive a $10 Amazon.com gift card. Those who were interested in the opportunity followed the link to the study site, where they created an account for the experiment. Subjects then answered a few demographics questions, were randomized into one of the four experimental conditions described below, and read instructions for completing the risk survey. Instructions (which included a link to the study site) were also automatically emailed to the subjects upon completion of the registration phase.

Task Completion Phase (Days 1−deadline) The risk survey became available to subjects at 8:00 a.m. the day after registration. The four conditions for task completion were as follows:

1. *Short Deadline / No Reminders.* Subjects in this condition faced a deadline of 11:59 p.m. on day 2 and received no reminders. The study instructions clarified that they would receive no reminders, stating "You will receive no further communications from us over the course of the study."

2. *Long Deadline / No Reminders.* Subjects in this condition faced a deadline of 11:59 p.m. on

[38] Short-run impatience (Laibson, 1997; O'Donoghue and Rabin, 1999) can lead to suboptimally low completion rates but, like the perfectly attentive and time consistent model, it does not predict that increasing the (finite) deadline can decrease completion rates when the distributions of completion payoffs do not change over time. The simple reason is that even if a longer deadline T_2 induces more procrastination, upon reaching the last T_1 periods of the longer deadline, the present-biased DM now plays the same game that he would play with the shorter deadline. Moreover, a fully attentive but present-biased DM would not exhibit suboptimally low completion rates for immediately pleasurable tasks as, arguably, in the experiments of Shu and Gneezy (2010).

There are, however, other deadline effects that are explained by present-biased preferences but not by the inattention model. Ariely and Wertenbroch (2002) study tasks that can only be completed over several days, such as term papers, and find that students self-impose shorter deadlines. This is consistent with students wanting to commit their short-run self to complete the task in a timely manner, but cannot be explained by inattention.

[39] Another theory that would be consistent with Proposition 7 but not Proposition 9 is a combination of declining motivation and projection bias (Loewenstein et al., 2003). For example, people may be excited when they first decide to complete some task, but their motivation to do it decays each day. If they misforecast this decay in motivation, then logic similar to that of Proposition 7 would predict that longer deadlines could lead to lower completion rates. However, reminders would then not reverse this trend, contrary to Proposition 9.

day 21. As in the previous condition, they received no reminders on any of the days, and the study instructions clarified that they would not.

3. *Short Deadline / Reminders.* Subjects in this condition faced a deadline of 11:59 p.m. on day 2, and received reminders on days 1 and 2. Subjects were informed that they would be getting reminders each day. The reminders are described below in further detail.

4. *Long Deadline / Reminders.* Subjects in this condition faced a deadline of 11:59 p.m. on day 21, and received reminders on days 20 and 21. Subjects were informed that they would be getting reminders on days 20 and 21.

The timing and content of the reminders were as follows. Reminders were only sent to those subjects in the reminders conditions who had not yet completed the survey. Of those subjects, email reminders were sent to everyone. Additionally, during the registration phase, subjects randomized into the reminders conditions were also given the option to consent to receive text message reminders. All subjects in the reminders conditions who had not yet completed the survey received an email reminder at 8 a.m. on each of the two reminder days. Subjects were sent a second set of reminders at either 10 a.m., 11 a.m., 12 p.m., 1 p.m., or 2 p.m. that consisted of an email and, for those choosing to receive text messages, a text message. Two-thirds of all subjects in the reminders condition received only *basic* email reminders. These emails mentioned the deadline date and nothing else (see Appendix C for text). One-third of the subjects received a *basic+info* email reminder instead of a basic email reminder at 8 a.m. of each reminder day. The purpose of the augmented message was to examine whether information loss played a role in task completion. The email therefore included information about the $10 reward and when it would be emailed, the length of the survey, and the URL of the study (see Appendix C). The language of the reminder messages was intentionally simple to minimize demand effects.

After completing the risk survey, subjects were prompted to sign a receipt for the $10 gift card, which was always emailed the day after the deadline.[40]

[40]Note that the reward was never immediately available. Nevertheless, one potential concern with this aspect of the procedures is that if subjects have a high exponential daily discount factor, then the discounted value of the reward for subjects on day 1 in the long deadline conditions is lower than the discounted value of the reward for subjects on day 1 in the short deadline conditions. However, all results about completion rates would continue to hold even with exponential discounting—the simple intuition is that the discounted value of task completion on days 20 and 21 in the long deadline conditions is equal to the discounted value of task completion on days 1 and 2 in the short deadline conditions. At the same time, to the extent that small amounts of money are fungible, subjects should not treat the $10 on day 22 any differently than the $10 on day 3. For small stakes monetary rewards, Andreoni and Sprenger (2012) experimentally estimate an annual discount rate of 0.3, which translates into a 3-week discount rate of less than 1.7%, or a discount of 0.17 from $10.00.

Results

A total of 403 subjects (93% students; 88.6% Harvard undergraduates) were recruited for the experiment in August and September. The email was sent to all 12 undergraduate residences at Harvard[41] and generated a total of 34 unique dates on which subjects signed up for the study. Of subjects in the reminders condition, 64.5% chose to receive text messages. Appendix C.2 verifies that the randomization was successful, finding no significant differences between any of the demographic variables across conditions.

Table 2 summarizes the completion rates by condition; and table 3 estimates a linear probability model of task completion, with robust standard errors clustered by both start date and undergraduate residence. Consistent with previous evidence and the prediction of Proposition 7, a longer deadline decreases completion rates from 59.4% to 41.6% in the no reminders conditions ($p < 0.01$). Consistent with Proposition 9, however, reminders close the gap between the short and long deadlines. With reminders, the longer deadline reduces completion rates by an insignificant 2.3%.

The key prediction that is confirmed by the data is that *if* longer deadlines decrease completion rates, *then*, because of how rapidly the likelihood of being attentive must decay, reminders have a larger impact on decision makers facing a longer deadline. Table 3 shows that while subjects given the *short* deadline are 14.9 percentage points more likely to complete the task when they receive reminders, subjects facing the *long* deadline are 30.5 percentage points more likely to complete the task when they receive reminders. The difference between differences is a marginally significant 15.6 percentage points ($p < 0.1$).

Table C7 in Appendix C shows that these results are robust to both demographic controls as well as day of week controls.[42] Table C8 in Appendix C analyzes whether different types of reminders had different effects. The content of the reminder had no additional effect on completion rates, suggesting that information loss is not an important factor. The table also shows that the 64.5% of the subjects who agreed to receive SMS messages did not behave any differently from the subjects who did not. This fact, of course, is difficult to interpret because of the usual endogeneity caveats.

Figure 4 provides further evidence for the decaying attention dynamics proposed by the model. For each of the four experimental conditions, the figure shows the fraction of subjects completing the task on a given day. Consistent with the theoretical prediction that longer deadlines increase the perceived option value of delaying task completion, subjects in the long deadlines conditions are

[41] Almost all sophomores, juniors and seniors at Harvard live in one of the 12 undergraduate "houses." A recruitment email was sent to each house only once. Recruitment occurred over the course of 5 weeks, with 2 or 3 houses receiving a recruitment email each week.

[42] Recruitment was staggered across different days of the week, so that each day of the week served as a start date for at least 33 and no more than 77 subjects in the study. Table C6 shows that deadline day of week appears to have no impact on probability of completing the task: the hypothesis that there are no day of week effects cannot be rejected for either the short deadline or the long deadline conditions ($p = 0.18$ for short deadline; $p = 0.26$) for long deadline).

significantly less likely to complete the task on the first day that it becomes available. Consistent with decaying attention, however, subjects who don't complete the task in the first few days are very unlikely to ever complete it if they don't receive reminders. Quite strikingly, only 2 out of the 101 subjects in the *long deadline / no reminders* condition completed the task between days 11 and 21. Subjects in the *long deadline / reminders* condition exhibited a similar pattern of behavior during days 1-19. When they received reminders on days 20 and 21, however, more than 50% of the subjects who had not yet completed the task ended up completing it.[43]

The results of this experiment thus replicate existing studies on deadline effects but, by showing how a small set of reminders can drastically change the outcomes, provide strong evidence that these deadline effects are generated by the decaying attention mechanism postulated in this paper.

5 Inattention in the Market

In this section, I explore how the kinds of behavioral patterns analyzed in Sections 3 and 4 might play out in various strategic interactions between sophisticated firms and inattentive consumers. Section 5.1 builds on the task completion results from Section 4 to study consumer rebates and related applications, while Section 5.2 builds on the repeated action results from Section 3 to study a model of reminder advertising.

5.1 Rebates and Related Applications

In this section I illustrate the applicability of the inattention model by embedding it in a simple model of consumer rebates. Rebates are a commonly used marketing tactic that requires consumers to 1) make a purchase at some up-front price p and 2) to submit a (partial) refund request, along with proof of purchase, by mail or internet by some deadline T. Edwards (2007) estimates the annual volume of rebate offers to range from \$4 to \$10 billion. Silk and Janiszewski (2008) estimate redemption rates of 10% to 30% for consumer rebate offers between \$20 and \$100, and redemption rates of 40% for bigger ticket consumer electronics.

Previous work on consumer rebates has focused on their potential use for price discrimination (Narasimhan, 1984; Banks and Moorthy, 1999; Lu and Moorthy, 2007),[44] and has modeled consumers' redemption decision as a static, one-period choice. The analysis here differs from previous

[43] This striking pattern of behavior also shows that reminder provision is not likely to have generated a demand effect. One potential hypothesis is that reminders induce a demand effect by signaling to subjects that the experimenter is taking the task seriously. Note, however, that subjects knew from the beginning whether or not they were getting reminders. The extremely low completion rates in days 11-19 in the long deadline/ reminders condition are not consistent with this signaling demand story.

[44] Gerstner and Hess (1991) propose that rebates may arise due to channel distribution issues between manufacturers, retailers and consumers. Chen et al. (2005) propose that rebates may function as "state-contingent discounts" that redistribute money to states with high marginal utility from income. More similar to the analysis here is Gilpatric's (2009) recent work on offering rebates to time-inconsistent consumers.

work in two crucial ways. First, the analysis is the first, to my knowledge, to investigate the question of how firms choose optimal redemption deadlines in a dynamic model of consumers' redemption decisions. Second, I investigate a setting in which firms may offer rebates to naive and inattentive consumers even when there is no incentive to price discriminate among them. While price discrimination is surely a potential motivation as well, the intentionally simple setting in this section allows me to formally capture and evaluate policymakers' and industry leaders' claims that consumer rebates are deceptively attractive to consumers.[45]

The monopolistic seller produces a certain product at a constant marginal cost c, and in period 0 chooses an offer $\mathcal{P} = (p, T, r)$ consisting of an upfront price p, a rebate $r \geq 0$ and a redemption deadline T chosen from a set of possible deadlines $\mathcal{T}$. For simplicity, I assume that $\mathcal{T}$ if finite. The set $\mathcal{T}$ may be bounded from below because of a minimum deadline length requirement as in, e.g., New York state.[46], and bounded from above because of similar restrictions[47] or because of logistical and credibility issues.

I will often focus on the monopolist's optimal offer conditional on being restricted to a particular deadline T.[48] This is a helpful intermediate step in the analysis that also facilitates the evaluation of various policy proposals to restrict deadline choice. The optimal policy for a particular deadline T is denoted $\mathcal{P}_T^* = (p_T^*, r_T^*)$, and the corresponding monopolist's profits are denoted π_T^*. The optimal policy $\mathcal{P}$ is computed by choosing the deadline $T \in \mathcal{T}$ that maximizes π_T^* and then setting p and r according to $\mathcal{P}_T^*$.

There is a unit mass of consumers who are interested in the product and value it at $v > 0$. They make a purchase decision in period 0. In each period $t = 1, \ldots T$, consumers who have purchased the product can incur an i.i.d. hassle cost $\xi_t \sim F$ (with $\xi_t \leq 0$) and mail in the rebate. Consumers who never attempt to redeem the rebate derive utility $v - p$ from the transaction. Consumers who mail in the rebate in period t derive expected utility $v - p + \theta r + \xi_t$ from the transaction, where $\theta < 1$. The discount factor θ arises from the possibility of filling out the form incorrectly and thus not receiving the rebate (Edwards, 2007, 2009) or the possibility of losing the check once it arrives in the mail (Edwards, 2007, 2009).[49] Accordingly, the monopolist's profit is given by $\pi = p - c - \theta \mu r$, where μ is the probability that the consumer attempts to redeem over the course of the T periods.

Following the recent behavioral industrial organization literature on two-part pricing (Heidhues

[45]Preliminary results show that even when price discrimination is possible, making rebates maximally deceptive is more profitable than trying to price discriminate among consumers.

[46]See Edwards (2009) for a summary of the law.

[47]To my knowledge, no such restrictions exist. However, the theoretical analysis in this section will consider the possibility of setting an upper bound.

[48]Proposition 10 shows that an optimal offer always exists. An optimum could fail to exist if it was the case, for example, that the monopolist's profits increase without bound for an appropriate combination of p and r approaching infinity.

[49]The results in this section would be unchanged if instead I assumed that 1) there is some probability that consumers lose the rebate form or the necessary information they need to claim the rebate (e.g., barcodes) or if 2) consumers are partially naive but not fully naive. What matters for the analysis in this section is that all consumers, including the naive ones, discount the value of the rebate because of a mistake they may make.

et al., 2012b,a; Grubb, 2012) I assume a price floor: there is some $\underline{p}$ such that any offer must satisfy $p - r \geq \underline{p}$. Following (Heidhues et al., 2012b, 2011), this can be motivated by the existence of market-savvy arbitrageurs who have no interest in the actual product, but who will exploit a firm offering easy money. Appendix A.5 walks-through this microfoundation.[50] For smaller value items, $\underline{p} = 0$ is consistent with, and helps explain the fact that "free after rebate" deals are extremely common, while deals in which rebates are even slightly above up-front prices are non-existent.[51] All results will be proven for any $\underline{p} < c$, however.[52]

I begin by characterizing when rebates will be offered.[53]

Proposition 10. *For each T, there exists an optimal policy $\mathcal{P}_T^*$, and consumers always buy in equilibrium if $v \geq c$. If consumers are sophisticated about their inattention then $p_T^* = v$ and $r_T^* = 0$. If consumers are naive, then for any set of attention probabilities $\gamma_t(\alpha, x)$, there is a $v^\dagger > 0$ such that*

1. *If $v < v^\dagger$ then $r_T^* = 0$ in any equilibrium in which consumers buy the product.*

2. *If $v > v^\dagger$ then $p_T^* > v$ and $r_T^* > 0$ and consumers buy the product.*

According to Proposition 10, rebates will not be offered to sophisticated consumers. The simple intuition is that the monopolistic seller extracts all surplus generated by a transaction with a sophisticated consumer. Offering a rebate, however, reduces total surplus because it forces consumers to incur effort costs to redeem it.

With naive consumers, the monopolist's profits are no longer determined solely by the transaction surplus; the profits are given by the sum of the actual transaction surplus *and* consumers' misperceptions of the value of the deal. Thus a rebate will be optimal when it increases misperceptions more than it decreases the actual surplus. Note now that the introduction of a rebate cannot decrease true surplus by more than $E_F \xi_t$, the expected redemption effort cost in a given period. In contrast, holding T fixed, the misperceptions of naive consumers will strictly increase in r. But as shown in Appendix D.3, the post-rebate price floor constrains the rebate value to be no higher than some continuous function $\bar{n}(v)$ that is strictly increasing in v and satisfies $\bar{n}(0) = 0$. Thus a high v is necessary (and sufficient) to guarantee that the monopolist can set a rebate that will generate sufficiently high misperceptions.

[50] And as I also discuss in the appendix, all results hold under the plausible assumption a small fraction of the consumers who derive value from the product happen to be market-savvy arbitrageurs. As Heidhues et al. (2011) point out, these arbitrageurs are equivalent to Gabaix and Laibson's (2006) *sophisticates*.

[51] See, for example, http://www.dealigg.com/free.php. Of course, current rebate offers still don't keep out all arbitrageurs. Woodruff (2012), for example, describes profit-making schemes involving the resale of free or nearly free after rebate offers.

[52] $\underline{p} = 0$ may not be a good assumption for items such as expensive consumer electronics that can be resold in a secondary market. This suggests that a more general model could allow the price floor to be a function of v. All results in hold for a piece-wise linear formulation of the form $\underline{p}(v) = \max(0, a + bv)$, where $b \in (0, 1)$.

[53] As before, I assume that there is some $\bar{z} < 1$ such that $\gamma_t(1, d) \leq \bar{z}$ for all t.

Because rebates are used to create deceptively appealing offers, expected consumer utility is negative in any equilibrium with a rebate. The next proposition shows that rebates may also lead to socially wasteful transactions in the sense that a product that costs c to produce is sold to consumers who derive utility $v < c$ from the product.

Proposition 11. *Suppose consumers are naive. Then for each $L > 0$ and T there exists $c^\dagger$ such that if $v = c - L$ and $c \geq c^\dagger$ then $p_T^*, r_T^* > 0$ and consumers purchase the product.*

The intuition behind proposition 11 is again derived from the fact that the monopolist's profits are given by the sum of the actual transaction surplus and consumers' misperceptions of the value of the deal. Thus even when $v < c$ and the transaction surplus is negative, a rebate can still be profitable by making the deal appear deceptively attractive to consumers. As with Proposition 10, the price-floor requires the product to be a sufficiently big ticket item to create scope for sufficiently deceptive rebates.

I now characterize how the monopolist will choose a redemption deadline and how that will impact consumers.

Proposition 12. *Let $R(r, T)$ denote the probability that a consumer mails in a rebate of size r given a deadline T and suppose that consumers are naive.*

1. *If $T_1 < T_2$ and $R(r_{T_1}^*, T_1) < R(r_{T_1}^*, T_2)$ then $\pi_{T_2}^* > \pi_{T_1}^*$.*
2. *Assume there is some $\underline{z} > 0$ such that $\gamma_t(0, d) \geq \underline{z}$ for all t. Then holding all other parameters constant, there is a $T^\dagger$ such that $r_T^* = 0$ for all $T \geq T^\dagger$.*

The main message of Proposition 12 is that if consumers are sufficiently inattentive, firms will choose deadlines that are of "intermediate length." This basic message follows from Proposition 7, which shows that longer deadlines can lead to lower task completion rates, and from Proposition 8, which shows that long enough deadlines will actually lead to higher completion rates than "intermediate length" deadlines. Although a systematic analysis of redemption deadlines is lacking, the "medium length" deadline prediction seems consistent with industry practice of setting deadlines that typically range between 15 and 60 days.[54][55]

The model's prediction that rebates may be used to create deceptively attractive deals is echoed by industry experts and policymakers. As bluntly explained by a VP of an electronics retailer, "Manufacturers love rebates because the redemption rates are close to none. It's just human nature that we go after them, and they get people into stores, but when it comes time to collect, few

[54]See, however, the caveats in footnote 37

[55]Rebates for products offered by the retailer Staples, for example, typically have redemption deadlines of either 30 or 60 days (see https://www.stapleseasyrebates.com/promocenter/staples/promo_search.html). Silk and Janiszewski (2008) analyze a random online sample and find that most deadlines are either 15 or 30 days.

people follow through."[56] This prediction is in contrast to the more traditional views that rebates are used to price discriminate among rational consumers. And even more starkly in contrast to the rational price discrimination framework is the prediction that rebates can facilitate the sale of socially wasteful products.

While some have explored the possibility of banning rebates on the grounds that they are deceptive (Lynch and Zauberman, 2006; Sovern, 2006), the model suggests that there are less paternalistic interventions that can increase consumer welfare or market efficiency, without interfering with the potential of using rebates for purposes of efficiency-enhancing price discrimination. One such possibility, for example, is to require very short or very long deadlines. Legislators in the states of California, New York, Texas, and North Carolina have imposed minimum deadline lengths (Edwards, 2007), though it is unclear whether those 2 or 4 week minimums are sufficiently high. Note, however, that such deadline restrictions have less clear-cut effects on social efficiency. This is because by virtue of attempting to minimize redemption probability, a monopolist's choice of deadline also minimizes consumers' socially wasteful redemption effort.[57]

Lastly, note that the inattention model also suggests that because of attention decay, exploitative firms may want to create *deferred rebate programs* that require consumers to wait before they can claim the rebate. Some companies have, in fact, offered 54-week deferred rebates of up to $10,000 to consumers purchasing big ticket items such as pools or automobiles, but have subsequently faced lawsuits for deceptive and exploitative practices.[58] Following incidents of massive consumer complaints, North Carolina has banned rebates with a deferral period of more than 6 months, in addition to creating other restrictions.[59] I speculate that deferred rebates are not more common because firms fear that such offers are so transparently exploitative that they would tarnish the firms' reputations on the one hand, and—as in the case of North Carolina—invite more severe government oversight and intervention on the other hand.

5.1.1 Related Applications

There are a number of other transactions between firms and consumers that share some key similarities with consumer rebates. Namely, the feature that, at a later date, the consumer can take some action x by a deadline T, which generates some (endogenously set) benefit $b - \xi_t$ to the consumer, at a (endogenously set) cost c to the firm. A product return policy with a deadline T is

[56] Catherine Greenman, "The Trouble With Rebates." The New York Times. September 16, 1999. Accessed August 15, 2013, http://www.nytimes.com/1999/09/16/technology/the-trouble-with-rebates.html?pagewanted=all&src=pm

[57] A second possibility is to minimize inattention bias by requiring firms to send consumers reminders to redeem their rebates. A third possibility is to attempt to debias naive consumers through salient disclosure of low redemption rates (Edwards, 2007), or by encouraging them to set their own reminders to minimize the probability of forgetting.

[58] Jamie Boll, "PSI: Big rebate offer leads to big disappointment." Feb 08, 2011. Accessed 8/25/2013, http://www.wbtv.com/story/13983724/psi-big-rebate-offer-leads-to-big-disappointment. See also the Pennsylvania Attorney General Press release:http://www.attorneygeneral.gov/press.aspx?id=835

[59] See the North Carolina Department of Justice on rebates: http://www.ncdoj.gov/Consumer/Purchases-and-Contracts/Rebates.aspx

one such example. Another set of examples is contracts with automatic renewal. DellaVigna and Malmendier (2004) list examples such as free trial offers that automatically transition into paid services, automatically renewing contracts in the health club industry, and automatically renewing newspaper subscriptions. In all of these examples, naive and inattentive consumers will overvalue the deal because they will overestimate the likelihood of returning the product or canceling their membership.

As DellaVigna and Malmendier (2004) show, naive hyperbolic discounting can also lead to overvaluation, and can explain why firms increase switching costs by, for example, stipulating that consumers can only cancel their gym membership in person. The analysis in this paper provides an alternative explanation: by increasing switching costs, firms increase the option value of delay, which makes naive and inattentive consumers more likely to put off the task until later and subsequently lose track of it. Intuitively, if canceling a gym membership involved only one click online, a consumer who first decided to cancel the membership would likely do it right away. The requirement to cancel the membership in person, however, makes the consumer put off the task until later and subsequently forget about it.

The inattention model can thus explain why firms make some actions very easy—like the "1-Click" buying feature—while making other actions more effortful. But additionally, the inattention model also makes predictions about which actions firms will remind consumers to take, and which they won't. Firms exploiting naive inattention would remind consumers to, for example, follow through on limited time offers, but would not remind consumers to, for example, consider canceling their subscription after the expiration of the free-trial phase. These predictions seem to be consistent with casual observation, and future work should provide systematic evidence for these new predictions.

5.2 Optimal Cue Provision by an Interested Party

Economic theories of advertising typically assume that marketing communications either provide information (Stigler, 1961; Butters, 1977) or enter directly into consumers' utility functions (Becker and Murphy, 1993). Both the informative and the persuasive (or complementary) functions of advertising are important, particularly when the product is first introduced. However, marketers also emphasize a third function of marketing communications: after a product has been introduced, advertising must keep the product at the top of the consumer's mind.

This section uses the inattention model to formalize a model of reminder communications by a sophisticated firm or organization. To keep things simple, I focus on an organization whose objective is to direct consumer behavior *solely* through the reminder communications. For example, a health care provider or insurer might use SMS messages or phone calls to remind patients with chronic diseases to take their medications,[60] or an organization might send reports reminding consumers of ways to save energy (Allcott and Rogers, 2012). The insights developed in this simple model could also be applied to understand the optimal policy of a firm that uses reminder advertising to make sure that its repeat purchase product stays a top of mind consideration.

While recent work in economics has investigated advertising to inattentive consumers in a static, one-period setting (e.g., Falkinger 2007, 2008; Eliaz and Spiegler 2011a,b),[61] the inattention model proposed in this paper provides foundations for extending these analyses to dynamic, multi-period environments. As this section shows, some of the most distinctive predictions of reminder communications are manifested only in dynamic settings.

Formally, there is a unit mass of homogeneous, inattentive consumers who choose $x_t \in \{d, a\}$ in each period $t \geq 1$. Each period, the organization chooses between sending a communication, denoted $m_t = 1$, or not sending a communication, denoted $m_t = 0$. A period t communication costs c to send and reaches a consumer with probability $w \in (0, 1)$. For simplicity, I assume that when a consumer is reached he is attentive with probability 1 regardless of previous history. Otherwise, there are no other cues and the DM is attentive with probability $g(\alpha, x, 0)$. For simplicity, I assume that $g(0, d, 0) = 0$, and that the probability of being attentive in period 0 is 0.

The organization's objective function is to choose $m = (m_1, m_2, \dots)$ to maximize $E_0 \sum_{t=1}^{T} \delta_o^t (\phi_t - m_t c)$, where ϕ_t is the fraction of consumers choosing $x_t = a$ and $\delta_o < 1$ is the organization's discount factor. Formally, I will assume that the organization commits to a strategy in period 0. However, the assumptions in this section will guarantee that the consumers will not alter their strategies in anticipation of future cues; thus, the results would be identical under the assumption that the organization can revise its strategy each period.

Note that initially uninformed but otherwise attentive consumers, in the sense of Stigler (1961) or Butters (1977), are nested as a special case of the framework: $g_0(0, d, 0) = 0$ and $g(1, d, 0) = 1$. That is, these consumers will not consider $x_t = a$ until they are contacted by the organization, but will always consider $x_t = a$ after the first time. The general framework proposed here thus nests a simple variation of informative advertising as a special case.

To fully draw out the long-run implications for the optimal messaging strategy, I extend the baseline model and set $T = \infty$, and assume that future payoffs are discounted by an exponential discount factor δ_{DM} (see Appendix A.1 for a formal extension of the inattention model to infinite horizons). Each consumer's period t payoff from choosing $x_t = a$ is given by $b_H > 0$ with probability $\ell \in (0, 1)$ and by $b_L < 0$ with probability $1 - \ell$. These taste variations are independently distributed

[60] Providers and insurers may purchase these services from independent third party firms specializing in population health and patient communication; e.g., Healthways, Phytel, Optum, IncentOne, Staywell.

[61] See also Hefti (2013) for a model building on Falkinger (2007, 2008). See Spiegler (forthcoming) for a model of competitive framing. Most related to the analysis here, see Villas-Boas (1993) for a simple model of reminder advertising in a duopoly.

across the unit mass of consumers each period.[62] To further simplify the model, I assume that b_L is sufficiently low so that even sophisticated consumers always choose $x_t = d$ when b_L is realized.[63]

Proposition 13 establishes necessary and sufficient conditions for repeated communication to be optimal in the long run.

Proposition 13. *Let m^* be an optimal messaging strategy and set $\psi = \ell g(1, a, 0) + (1 - \ell)g(1, d, 0)$.*

1. *If*
$$\frac{w\ell}{1 - \delta_o \psi} \leq c$$
then $m_t^* = 0$ for all t.

2. *If $\psi = 1$ then there exists a $t^\dagger < \infty$ such that $m_t^* = 1$ if and only if $t < t^\dagger$.*

3. *If*
$$\frac{w\ell}{1 - \delta_o \psi} > c \tag{10}$$
and $g(1, a, 0) < 1$, then for any $t > 0$ there exists a $t' \geq t$ such that $m_{t'}^* = 1$.

Part 1 establishes conditions under which sending communications to the consumers is not optimal in any period. The quantity $\frac{w\ell}{1-\delta_o\psi}$ is an upper bound on how much a single message can increase the organization's discounted payoff, so that sending a message is never optimal when this upper bound is not greater than c. As intuition would suggest, the upper bound is increasing in ℓ, the consumers' likelihood of wanting to choose $x_t = a$ conditional on being attentive, and increasing in w, the efficacy of the organization's messaging attempt.

Note that in the absence of rehearsal effects, $\psi = 0$, and the upper bound reduces to $w\ell$. With rehearsal effects, $\psi > 0$, and thus a single period t message has a larger effect: in addition to making DMs more attentive in period t, the message also makes DMs more attentive in periods $t + 1, t + 2, \ldots$.

Part 2 establishes conditions under which sending communications to the consumers cannot be optimal in the "long-run," though may be optimal in the short run. The condition that $\psi = 1$ simply ensures that the probability of being attentive does not decay with time. Importantly, this condition holds for consumers with $g_0(0, d, 0) = 0$ and $g(1, d, 0) = 1$—which can be interpreted as initially uninformed but otherwise attentive consumers.

Part 3 completes the proposition by showing that when neither the condition in part 1 nor the condition in part 2 holds, the organization's optimal strategy will involve messaging in the long run.

[62] To map this into the formal framework, assume that $b_t \equiv b$ for all t, and suppose that ξ_t takes on the values ξ_H or $\xi_L < \xi_H$. So $b_H = b + \xi_H$ while $b_L = b + \xi_L$. Note that in contrast to the assumptions in earlier sections, the distribution F is not atomless here.

[63] A sufficient condition for this is that $b_L < -\gamma b_H/(1 - \psi)$, where $\psi = \ell\gamma(1, a, 0) + (1 - \ell)\gamma(1, d, 0)$. Note that while the parameter ℓ is a useful comparative statics parameter corresponding to the consumers' strength of preference for choosing $x_t = a$, the variation in taste shocks does create additional complexity arising from the fact that sophisticated consumers' choice may be shaped by their beliefs about the organization's future messaging strategy.

Proposition 13 illustrates a key difference between optimal communications with inattentive versus attentive but uninformed consumers: repeated communication is optimal in the long run if and only if consumers are inattentive *and* their attention decays with time in the absence of communications. In fact, parts 1 and 3 imply that when $\psi < 1$, so that attention decays over time, the organization's optimal advertising strategy is to either not send messages at all, or to never stop sending messages.[64]

But while the inattention model implies that the returns to additional communication do not converge to zero in the long run, like models of informative advertising it also predicts that the returns to each additional communication are diminishing. A version of this intuition was already established in part 5 of Proposition 2, which showed that temporally separated cues are substitutes. In the simple setting here, it is similarly true that for $t' < t$, the returns to choosing $m_t = 1$ are decreasing in $m_{t'}$ because a choice of $m_{t'} = 1$ decreases the fraction of consumers who would otherwise be inattentive in period t. Such intermittent communications create an "action and backsliding effect": for $t > t'$, the probability of $x_{t'} = a$ is high following $m_{t'} = 1$, then decays while $m_{t'+1}, m_{t'+2}, \cdots = 0$, and then increases substantially again when $m_t = 1$.[65] Both the diminishing returns prediction and the action and backsliding prediction are key determinants of the optimal advertising policy and both have received empirical support. Allcott and Rogers (2012), for example, find that the effect of each additional report about energy-saving behaviors declines with time. At the same time, they also find a significant action and backsliding effect.

Lemma 1 formalizes the conditions under which choosing $m_t = 1$ is each period is not optimal. For the remainder of the section, I will use $\mathbf{1} = (1, 1, \dots)$ to denote the strategy of sending a message each period, and likewise use $\mathbf{0}$ to denote the strategy of never sending a message.

Lemma 1. *$m = \mathbf{1}$ is an optimal messaging strategy if and only if*

$$\frac{w(1-\psi)\ell}{(1-\psi(1-w))(1-\delta_o\psi(1-w))} \geq c. \tag{11}$$

Combined, Proposition 13 and Lemma 1 begin to provide insights into how consumers' preferences for choosing $x_t = a$, parametrized here by ℓ, affect the organization's returns to choosing $m_t = 1$. When consumers are attentive but initially uninformed, the returns to sending an additional message are increasing in ℓ, for the simple reason that conditional on considering $x_t = a$, more consumers will choose it. Similarly, when consumers are inattentive but rehearsal has no effect,

[64] Note that models of informative advertising can sustain long-run advertising policies in a market by assuming overlapping generations of consumers, but these models also predict that an organization or firm will eventually want to limit its communications to zero with any one particular household or consumer. This distinction is particularly important for communications such as those encouraging health-improving or energy conservation behaviors, which are targeted to specific households and can be sent on a frequent and regular basis.

[65] Agarwal et al. (2013) first coined the phrase "action and backsliding" to describe this kind of pattern of behavior with respect to credit card fees. Their finding that behavior responds most strongly to cues in the form of late fees, however, is not directly explained by the model in this paper.

$\psi = \ell g(1,a,0) + (1-\ell)g(1,d,0) = 0$ and thus condition (10) and condition (11) are equivalent. This then implies that there is some $\ell^\dagger$ such that the optimal strategy is $m = \mathbf{1}$ for $\ell > \ell^\dagger$ and $m = \mathbf{0}$ for $\ell < \ell^\dagger$. Lemma 2 summarizes these results:

Lemma 2. *1. If $g(1,d,0) = 1$ then the total number of messages is increasing in ℓ.*

2. If $g(1,a,0) = 0$ then the optimal strategy is $m = \mathbf{0}$ if $w\ell < c$ and is $m = \mathbf{1}$ if $w\ell > c$.

Behavioral rehearsal adds an additional nuance to the relationship between preferences and the optimal messaging strategy. On the one hand, the higher the ℓ, the more likely consumers are to choose $x_t = a$ conditional on being attentive. On the other hand, higher ℓ also implies that consumers are more likely to be attentive due to rehearsal. The combination of these two effects implies that the optimal communication intensity can be non-monotonic in ℓ:

Proposition 14. *Fix $g(1,d,0) < 1$ and $c < 1/4 - g(1,d,0)/4$. Then there exist high enough $g(1,a,0) < 1$ and $w < 1$ such that the optimal messaging strategy $m^* = (m_1^*, m_2^*, \dots)$ is non-monotonic in ℓ: for some $\ell_1 < \ell_2 < \ell_3$,*

i) If $\ell \in (\ell_1, \ell_2)$ then condition (10) holds but condition (11) does not hold

ii) If $\ell \in (\ell_2, \ell_3)$ then condition (11) holds

iii) If $\ell \in (\ell_3, 1)$ then condition (10) holds but condition (11) does not hold

The key implication of Proposition 14 is that when the effect of behavioral rehearsal is sufficiently strong, cue provision has the biggest impact on consumers who want to choose $x_t = a$ "often enough" but not necessarily all of the time. Intuitively, consumers who want to choose $x_t = a$ all of the time develop a strong enough habit that will not be significantly affected by cues. In the daily action experiment, for example, participants who completed the survey for three weeks straight were not significantly affected by week 3 cues. On the other hand, consumers who rarely want to choose $x_t = a$ will also not be affected by additional cues for the simple reason that they will not want to choose a conditional on being attentive. These results generate implications not only for how the organization should design its optimal messaging strategy for a given population of consumers, but also for how the organization should attempt to target its messages to different subgroups within a population.[66]

Although the analysis in this section focuses on a simple setting in which the organization only chooses cues, an important question for future theoretical and empirical work is how to optimally

[66] The model also generates insights about how to optimally tailor communications to a given consumer over time. The results in Section 3 show that all else equal, reminders should be most effective following an interruption to the decision maker's routine. An organization that can effectively monitor each consumer's behavior should thus increase the intensity of communications following what appear to be random breaks in the repeated behavior. Such targeting should be especially affective when consumers are naive, because naive consumers do not change their behavior in anticipation of future cues.

combine incentives and cues. The substitutes and complements predictions in Propositions 3 and 4 suggest that when both incentives and cues part of a program, they should be used simultaneously rather than one after the other. This is because cues amplify the effects of contemporaneous incentives, but crowd-out the spillover effects of past incentives.

6 Concluding Remarks

6.1 Recap

This paper proposes a model of inattentive choice, focusing on two factors determining attention: cues and rehearsal.

Applied to repeated actions, the model provides an attention-based theory of “good” habits and provides a unifying explanation of many recent empirical findings in domains of behavior ranging from exercise to residential energy use. The model is distinguished from theories of habit-forming *preferences* (Becker and Murphy, 1988) by its predictions about how 1) cues affect current and future behavior 2) how cues can amplify the effects of temporary shocks to routines and 3) how cues can also diminish the effects of temporary shocks to routines. Consistent with these predictions, the first experiment in this paper uses a three-week, real-effort task to test and confirm the distinguishing prediction that repeat performance and reminders are substitutes.

Applied to tasks that must be completed by a deadline, the model identifies when the likelihood of being attentive to the task will decay with time, which then leads naive decision makers to be hurt by longer deadlines. But at the same time, the model also generates new comparative statics about how changes in cues can break the time-decay of attention and therefore eliminate the “perverse” effect of longer deadlines. The second experiment reported in this paper replicates the finding that longer deadlines can lead to lower completion rates but, consistent with the new comparative statics about cue effects, shows that a small set of reminders can eliminate this effect.

Finally, I apply the model to study market interactions between sophisticated firms and inattentive consumers. Building on the results about tasks with deadlines, I show how firms will use sales tactics such as consumer rebates and automatic renewal billing to facilitate deceptively attractive and socially inefficient transactions with consumers. Building on the results about repeated actions, I use the model to study a firm or organization’s optimal policy of *reminder advertising* to inattentive consumers who must take some repeated action. In contrast to models of informative advertising, the optimal messaging intensity does not converge to zero in the long run and may be non-monotonic in consumers’ preferences for taking the action.

Taken together, the theoretical and experimental results show that 1) time-varying attention can explain a broad range of economically important behaviors that are not easily explained by other theories; 2) the effects of payoff-irrelevant cues on economic behaviors can be predicted and studied with theoretical and empirical precision; 3) theoretically motivated variations in cues can be used to empirically distinguish between time-varying attention and other potential drivers of behavior; and 4) consumer inattention may be an important determinant of firms' sales and marketing strategies.

Industry players', policymakers' and academics' growing interest in combining incentives with cue-based interventions to encourage behaviors ranging from preventive health measures to energy conservation—as well as the need to understand the medium- and long-run effects of these (sometimes short-run) programs—makes the developments in this paper especially pertinent. The theoretical model and experimental frameworks developed in this paper can provide useful structure for future empirical work on these questions.

6.2 Extensions

A stark and not fully realistic characteristic of the model in this paper is the assumption that all else equal, inattention does not vary with stakes. Some models of bounded rationality (e.g., Mullainathan 2002) make a similar assumption, while others (e.g., Sims 2003) propose that a limited cognitive resource is allocated optimally toward the highest stake decisions.[67] In this paper, higher stakes may make sophisticated agents more motivated to engage in rehearsal or, as in Appendix A.2, to invest in reminder technologies. An intuition that is absent from the model, however, is that even in the absence of all cues, a person would be more attentive to higher stakes tasks. At the same time, assuming that attention allocation is completely "rational" in the sense of Sims (2003) is discordant with the fact that many people are naive about how attentive they will be in the future. A completely "rational" approach to attention also seems to be at odds with the notion that sometimes people just pay attention to the "wrong" things. One avenue for future theoretical and experimental work is to investigate how the types of inattentive choices modeled in this paper might depend on the stakes of the decision.

A second way in which the model in this paper could be extended is by incorporating the possibility that some types of contextual cues derive their strength from repeated pairing with a certain action, as in the work of Laibson (2001) and Bernheim and Rangel (2004) on addiction. Interventions to improve medication adherence, for example, recommend associating the action of taking medicine with an event-based cue such as breakfast (Insel et al., 2013). Endogenously formed associations between context and actions may further help explain why some routines, such as brushing one's teeth every evening, are *so* stable. Relatedly, people may also become desensitized to cues that they receive often but typically don't pair with the action (Rankin et al., 2009). The sensitization and desensitization considerations may play an important role in the design of optimal communication strategies such as the ones analyzed in Section 5.2, and limit the extent to which

[67] Psychologists recognize that attention has both a "goal-driven"/"top-down" component and a "stimulus-driven"/"bottom-up" component (Yantis, 1998)

communications can be used to increase task completion, as in the examples studied in Section 4. The experimental design described in Section 3.2 could be extended to study desensitization to reminders, as well as the sensitization to event-based cues.

Finally, future work should broaden the choice set beyond the simple decisions studied in this paper. A key concept to theoretically and experimentally explore in this more general framework is the hypothesis that making some actions more top of mind might crowd out attention to other actions. Consider, for example, a driver who must choose between left, right, or straight at an intersection, and suppose that left is the correct choice on the driver's usual route to work. In the rare and atypical instances in which the driver must, instead, go straight, he may still go with the routine choice of left when he is inattentive. This example illustrates the intuition that rehearsal of one action may make it very mentally accessible, but at the expense of making other actions less so.[68] Extending the model to larger choice sets and, with that, formalizing the nuances of how attention is allocated to the different alternatives is a modeling challenge that may shed light on the formation and power of defaults, and may generate new insights about how firms compete through advertising policies.

[68] At the same time, there seem to be some *a priori* default choices such as *not* taking the daily medication or *not* turning off the lights whose consideration is much less likely to be crowded out by the increased accessibility of the alternative(s). My speculation is that the option of *not* performing an action is salient whenever execution of the action requires a certain degree of cognitive monitoring.

Tables

Table 1: Probability of completing a survey on any given day in week 3

Pr(complete survey)	(1)	(2)	(3)
Interruption	–0.416***	–0.398***	–0.401***
	(0.061)	(0.054)	(0.054)
Reminders	0.036	0.028	–0.121
	(0.067)	(0.051)	(0.137)
Interruption*Reminders	0.267**	0.305***	0.320***
	(0.108)	(0.089)	(0.091)
Week1Avg		0.625***	0.527***
		(0.078)	(0.093)
Week1Avg*Reminders			0.226
			(0.162)
Adj. R^2	0.139	0.289	0.293
Observations	1260	1260	1260

Notes: This table estimates a linear probability model of completing the daily survey in week 3 of the study. The variable "Interruption" equals 1 if the daily survey was available in week 3 and equals 0 otherwise. The variable "Reminders" equals 1 if subject received daily reminders in week 3 and equals 0 otherwise. The variable "Week1avg" denotes the fraction of surveys completed in week 1. Robust standard errors clustered at subject level. All regressions include controls for day of week and day in study. *$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

Table 2: Fraction of subjects completing task, by experimental condition

	No reminders	Reminders on last two days
2 days	59.4%	74.3%
3 weeks	41.6%	72.0%

Table 3: Probability of completing task, by experimental condition

Pr(Complete)	(1)
LongDeadline	–0.178***
	(0.067)
Reminders	0.149**
	(0.072)
LongDeadline*Reminders	0.156*
	(0.081)
Adj. R^2	0.074
Observations	403

Notes: This table estimates a linear probability of task completion, by condition. The variable "LongDeadline" equals 1 if subjects had 3 weeks to complete the task, and equals 0 if subjects had 2 days to complete the task. The variable "Reminders" equals 1 if subjects received two days of reminders—days 1 and 2 for subjects with a 2-day deadline and days 20 and 21 for subjects with a 3-week deadline. Robust standard errors are computed by specifying both start date and undergraduate residence as the cluster groups, following the multiway clustering method suggested by Cameron et al. (2011). *$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

Figures

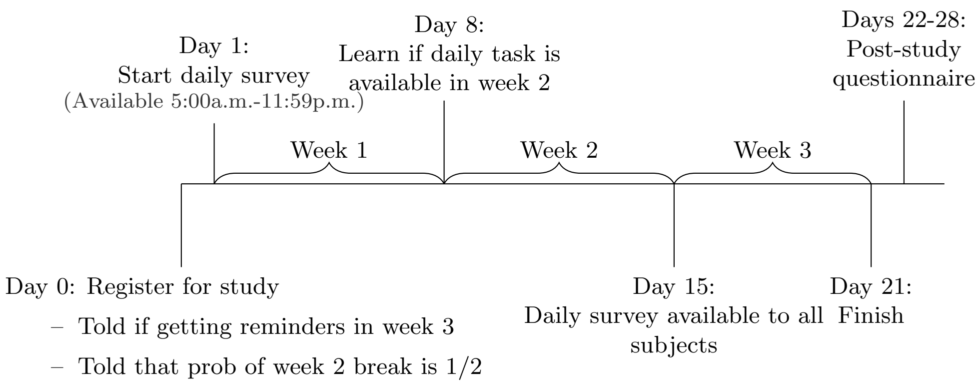


Figure 1: Timeline for repeated action experiment

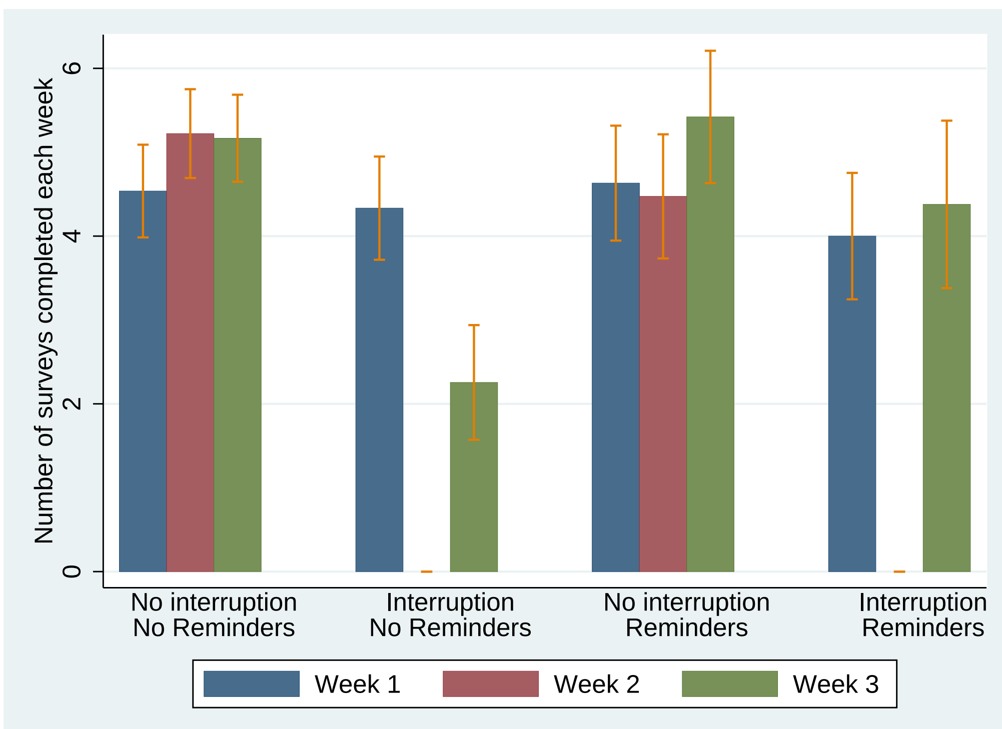


Figure 2: Surveys completed per week, by experimental condition

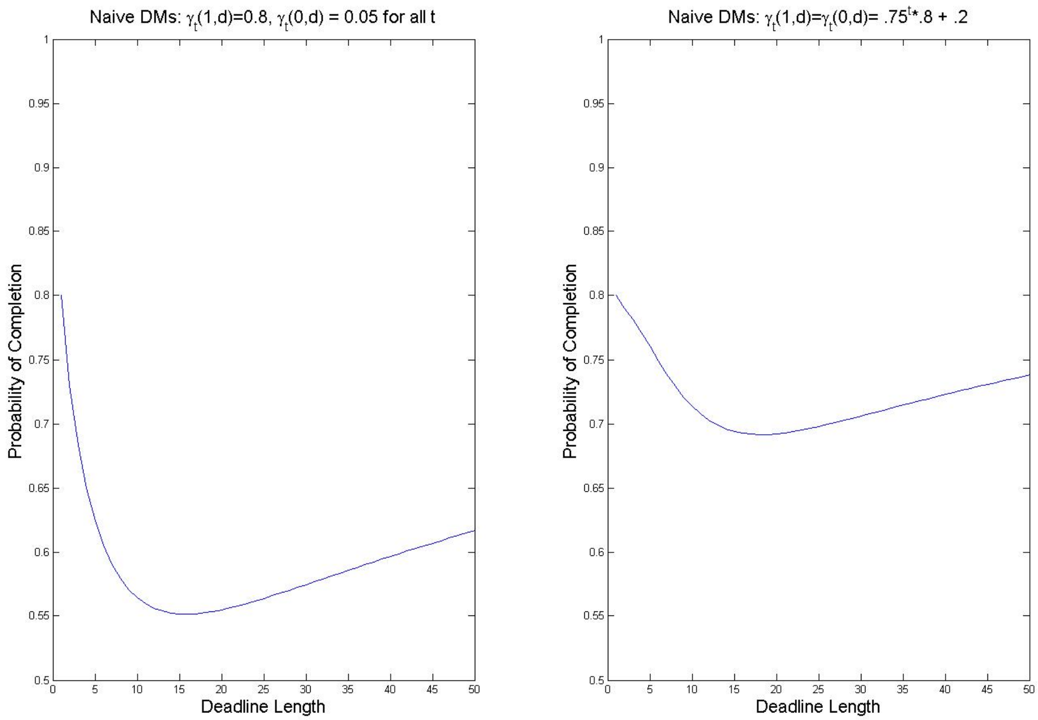
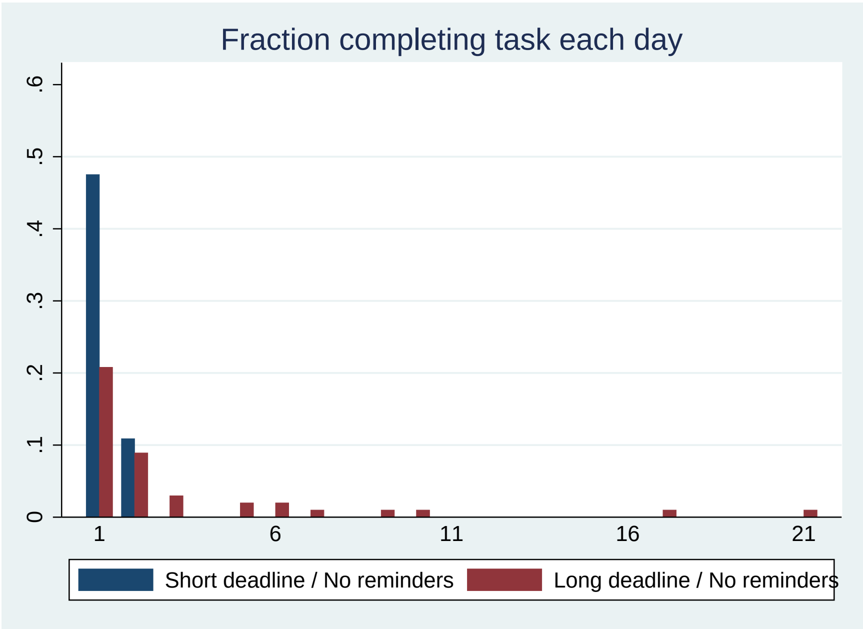
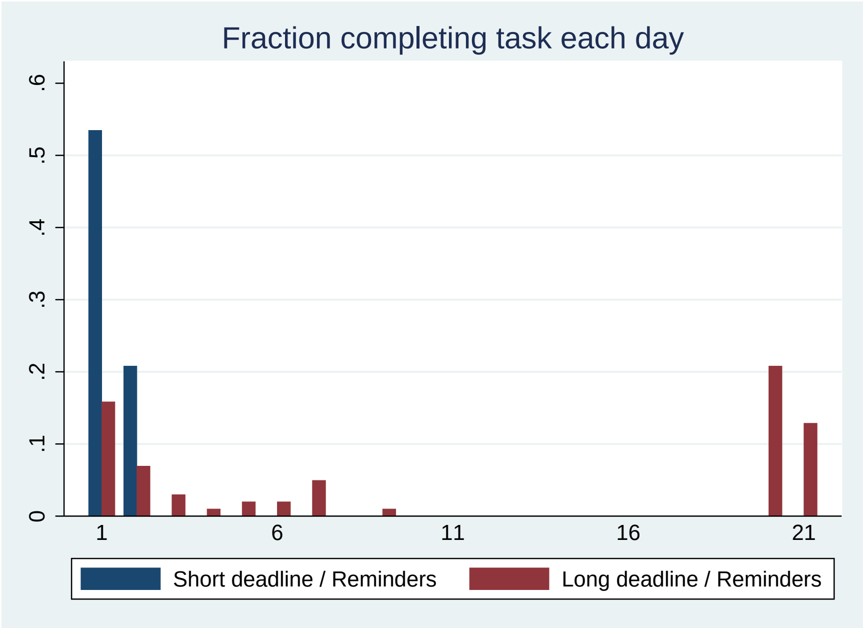


Figure 3: "U-shaped" curve



(a) No reminders



(b) Reminders

Figure 4: Fraction of subjects completing task on each day, by condition

Appendices

Note: These appendices may still contain minor typos

A Extensions and Additional Examples

A.1 Infinite Horizon

Let $\delta < 1$ be the DM's discount factor, and suppose now that T is possibly infinite. For simplicity, I will assume that period t flow payoffs are independent of the history t. Let H_t be the set of all possible period t histories, and set $\mathbf{H} \equiv \bigcup H_t$. Let $\mathbf{x}^{\mathrm{s}} : \mathbf{H} \times \mathbb{R} \to \{d, a\}$ denote a sophisticated DM's plan, which maps each pair (h_t, ξ_t) to an action in $A(h_t)$, *conditional on the DM being attentive.* Let $z_t = (x_t, \xi_t)$ denote the period t outcome, and let $u_t(z_t)$ denote the period t flow payoff corresponding to that outcome. Finally, let $\mathcal{F}(\mathbf{x}^{\mathrm{s}})$ be the distribution over the possible outcomes $z = (z_0, z_1, \dots)$ induced by strategy $\mathbf{x}^{\mathrm{s}}$. Note that the product measure $\mathcal{F}(\mathbf{x}^{\mathrm{s}})$ takes into account how the likelihood of being attentive evolves over time as a function of past events.

The sophisticated DM chooses the strategy $\mathbf{x}^{\mathrm{s}}$ that maximizes

$$\sum_{t=0}^{\infty} \delta^t u(z) d\mathcal{F}^{\mathrm{s}}(\mathbf{x}^{\mathrm{s}}).$$

The naive DM chooses the strategy $\mathbf{x}^{\mathrm{n}}$ that corresponds to the strategy of a perfectly attentive DM. In particular, let $\mathcal{F}^{\mathrm{pa}}(\mathbf{x}^{\mathrm{n}})$ denote the product measure over outcomes that would be induced if a perfectly attentive DM used strategy $\mathbf{x}^{\mathrm{n}}$. The naive DM chooses the strategy $\mathbf{x}^{\mathrm{n}}$ that maximizes

$$\sum_{t=0}^{\infty} \delta^t u(z) d\mathcal{F}^{\mathrm{pa}}(\mathbf{x}^{\mathrm{n}}).$$

His utility, however, is given by

$$\sum_{t=0}^{\infty} \delta^t u(z) d\mathcal{F}^{\mathrm{n}}(\mathbf{x}^{\mathrm{n}})$$

where $\mathcal{F}^{\mathrm{n}}(\mathbf{x}^{\mathrm{n}})$ is the actual product measure over a naive DMs' outcomes that $\mathbf{x}^{\mathrm{n}}$ induces.

A.2 Endogenous Cue Generation

Does setting cues eliminate inattention for sophisticates?

The possibility to manipulate cues does not eliminate the DM's inattention problem for a number of reasons. As discussed in Section 2.4, cues are imperfect. Second, cues can be costly. While setting a single electronic calendar reminder is cheap, purchasing an electronic pill bottle with an array of audio and visual reminders is more costly. Even electronic calendar reminders, however, can be expensive. Because one calendar reminder is highly imperfect in the sense that it can be forgotten minutes later, ensuring perfect attentiveness over the course of even a 1 week period might require hundreds, if not thousands of reminders—which carries a high nuisance cost. Third, an inattentive

DM will not only be inattentive about the primary behavior: he will also be inattentive about setting additional cues. Especially when combined with naivete, these issues substantially reduce the possibility that a DM might eliminate his inattention through reminder technologies.

Formal model

I now explore an augmented model in which the DM can set cues in period 0. I begin by focusing on sophisticated DMs. As in section 4, suppose that the DM starts out with initial attention probabilities $\gamma_t^0(\alpha, x)$ but can modify them to $\gamma_t^{\kappa_t}(\alpha, x)$ at cost $C(\kappa)$, where $\kappa = (\kappa_1, \ldots, \kappa_T)$. For these DMs, let $V_0^{\mathrm{s}}(\kappa)$ be the period 0 utility as a function of modified attention probabilities, with the V_t^{s} still defined as in Section 2.3. Different from Section 2.3, I now allow the vector κ to be set by the DM in period 0. The period 0 optimization is now

$$\max_{\kappa \in [0,1]^T} \{V_0^{\mathrm{s}}(\kappa) - C(\kappa)\}, \tag{12}$$

Notice that all of the analysis in the paper is just a subgame of this more general framework. Notice also that this general framework still incorporates the idea that some cues are generated exogenously, either by an interested party or as incidental events such as conversations with others.

A basic prediction of this more general framework is that as stakes become larger, the DM invests more in increasing his attention. This would not modify most of the results, however. I now walk through each of the paper's results to check their robustness. To consider analogs to the results about creating additional cues, as in several propositions in the paper, assume that the additions are applied to the initial cue distributions H_t, but that the DM's cost function $C(\kappa)$ does not change.

An assumption I will rely on throughout the analysis is that the DM cannot guarantee perfect attentiveness:

Assumption B For $\kappa = (\kappa_1, \ldots, \kappa_T)$, $\lim_{\kappa_t \to 1} C(\kappa) = \infty$ for each t.

Prop 1 Holds under Assumption B (which is needed to guarantee that there is scope for "external" cues to affect inattention). The logic is slightly more involved, though, because increasing $b_{t'}$ will now increase both investments in reminders, as well as investments through behavioral rehearsal.

Prop 2 Part 4 goes through exactly. The higher the period t' cues, the less the DM invests in future attentiveness both through rehearsal and through reminder technologies. Thus the likelihood of him taking an action in period $t < t'$ decreases. The logic of 2b goes through under assumption B: A sufficiently large increase in period t cues cannot be crowded out by lack of investment through rehearsal or own reminder technologies. Finally, whether or not 3a goes through under assumption B depends on the curvature of C and on the function $g(x, \alpha, \sigma)$. It is possible that increasing period 1 cues crowds out investment in future cues more than one-for-one. However, in the simple model of multiplicative cues introduced in (1), investment in cues will be crowded out less than one-for-one when C is continuous and convex. This ensures that increasing H_1 always increases the likelihood of attentiveness in all future periods. Part 5 will hold under these same assumptions.

Prop 3 Whether or not part 2 goes through under assumption B depends on the curvature of C on the function $g(x, \alpha, \sigma)$. As argued above, in the simple model of multiplicative cues introduced in (1), investment in cues will be crowded out less than one-for-one when C is continuous and convex. This ensures that increasing H_1 always increases the likelihood of attentiveness in all future periods.

Prop 4 The logic goes through under assumption B. Making the DM close to attentive as in part (b) makes him close to non-responsive to rehearsal effects (or investments in reminder technologies) from prior periods.

Prop 5 Part (2) of Proposition 5 will hold under assumption B. As discussed in Section 2.5, no cue is perfect because there is always a chance that it will be ignored or missed. Part 3 of Proposition 5 may not necessarily hold for very high benefit tasks, since the DM will set many reminders for those tasks. However, this claim should still hold for low benefit tasks for which the DM will not invest as much in reminders.

Prop 7 Parts 1 and 2 clearly hold verbatim, since the more attentive the DM is, the more advantageous the longer deadline. Part 3 will also hold but possibly through a different mechanism: the more important the task, the more attentive the DM becomes to it through endogenous cue setting, thus approaching the behavior of the perfectly attentive DM.

Prop 8 Holds verbatim, since the probability of being attentive is still bounded away from 0.

Prop 9 The analog to part 2 will still hold, in the sense that if the $\gamma_t^0(\alpha, x)$ are sufficiently high to make the DM close to attentive, then longer deadlines should generate higher completion rates.

Props 13,14 The major new question that the more general framework raises is the extent to which third-party cue-provision crowds out decision makers' personal cue provision. As long as crowd-out is not one-for-one, however, the impact of third-party cues should not go to zero in the long run. The DM's own personal provision may also be non-monotonic in ℓ, thus dampening the extent to which the third party's optimal cue intensity may be non-monotonic in ℓ.

A.3 Modeling Partial Naivete

At the most general level, prediction mistakes can be incorporated by supposing that the DM's forecasts are given by $\hat{g}(\alpha, x, \sigma)$ that satisfies assumptions A1-A4, but that doesn't necessarily correspond to the true g. A simple and parametric way of modeling overconfidence is to set

$$\hat{g}(\alpha, x, \sigma) = \chi_o + (1 - \chi_o)g(\alpha, x, \sigma) \tag{13}$$

Then $\chi_o = 0$ corresponds to a fully sophisticated DM, $\chi_o = 1$ corresponds to a fully naive DM, and $\chi \in (0, 1)$ corresponds to a partially overconfident DM.

The DM may be naive in more nuanced ways, however. For example, the DM may recognize that he can be inattentive in the future, but be fully naive about the role that behavioral rehearsal plays in increasing accessibility. This can be captured by setting $\hat{g}(\alpha, a, \sigma) = \hat{g}(\alpha, d, \sigma) = \chi_r g(\alpha, d, \sigma) + (1 - \chi_r)g(\alpha, a, \sigma)$. In this formulation of naivete, the DM will always overestimate his future attentiveness for one time tasks such as the ones studied in Section 4. In the repeated action environments studied in Section 3, this formulation will lead the DM to be completely ignorant of the effect that past behavior will have on his future behavior. This ignorance of the behavioral rehearsal effect will sometimes lead the DM to overestimate and sometimes underestimate the probability of choosing $x_t = a$ in the future.

Even more generally, the partial overconfidence in equation (13) can be combined with the naivete about rehearsal described in the previous paragraph.

A.4 Response elasticities for section 3.1

Let $D_t^i(b_t) = e + \Big(1 - Pr^{H^i}(a_t = 1)\Big)$ be the expected total consumption of energy in period t, conditional on sequence of cue distributions $\mathbf{H}^i$. Here, $e \geq 0$ is a baseline electricity use from other decisions. Then $\frac{\partial}{\partial b_t} D_t^i(b_t) = -\frac{\partial}{\partial b_t} Pr^{H^i}(x_t = a)$.

Let s_t be the naive or sophisticated DM's threshold rule in period t, so that $x_t = a$ if and only if $\xi_t \geq s_t$. Then for $i = 1, 2$,

$$\frac{\frac{\partial Pr^{H^i}(x_t=a)}{\partial b_t}}{Pr^{H^i}(x_t = a)} = \frac{\partial}{\partial b_t}(1 - F(s_t))$$

is constant in i. To simlify notation, set $k_t == \frac{\partial}{\partial b_t}(1 - F(s_t))$. Then

$$\begin{aligned}
& -\frac{\frac{\partial}{\partial b_t} D_t^1(b_t)}{D_t(b_t)} < -\frac{\frac{\partial}{\partial b_t} D_t^2(b_t)}{D_t(b_t)} \\
& \Leftrightarrow -\frac{k_t Pr^{H^1}(x_t = a)}{e + (1 - Pr^{H^1}(x_t = a))} < -\frac{k_t Pr^{H^2}(x_t = a)}{e + (1 - Pr^{H^2}(x_t = a))} \\
& \Leftrightarrow -Pr^{H^1}(x_t = a)\Big(e + 1 - Pr^{H^2}(x_t = a)\Big) < -Pr^{H^2}(x_t = a)\Big(e + 1 - Pr^{H^1}(x_t = a)\Big) \\
& \Leftrightarrow -Pr^{H^1}(x_t = a)(e + 1) < -Pr^{H^2}(x_t = a)(e + 1) \\
& \Leftrightarrow Pr^{H^1}(x_t = a) > Pr^{H^2}(x_t = a)
\end{aligned}$$

Thus, when cues increase the probability of $x_t = a$, they also increase the demand response elasticity.

A.5 Microfoundations for the price floor assumption

An arbitrageur who derives no intrinsic value from the product gets a total payoff of $\hat{v}_a - p + r$ from the offer, where $\hat{v}_a$ corresponds to the scrap value of the product, net of potential inconvenience costs of going through with the deal. I assume that $\hat{v}_a < c$. In contrast to the inattentive consumers, an arbitrageur is perfectly attentive, and always mails in the form correctly. Thus, conditional on purchasing a product with a rebate, he obtains the rebate with probability 1. This is in contrast to naive consumers who derive utility v from one unit of the product but derive 0 utility from further purchases.

Proposition A1. *Conditional on any deadline T, the profit maximizing choice of p_T and r_T is such that $p_T - r_T \geq \hat{v}_a$.*

Proof. If $p_T - r_T < \hat{v}_a$ then arbitrageurs will have infinite demand for the product. However, the profit from selling to each arbitrageur will be $p_T - r_T - c < \hat{v}_a - c < 0$. □

Note that it is possible there are some arbitrageurs who are intrinsically interested in the product, and derive utility $v - p + r > \hat{v}_a - p + r$ from the first unit. The analysis in the body of the paper is a limit case of an economy in which the fraction ϵ of such arbitrageurs approaches 0.

B Daily Action Experiment: Supplementary Material

B.1 Theoretical Extensions

To formally model the decision environment in experiment 2, suppose now that the payoff to choosing $a_t = 1$ is $b_t + \xi_t + \zeta_t$, where b_t and $\xi_t \sim F$ are as in Section 3. The new source of variation is the correlated shock ζ_t. Specifically, there are t_1 and t_2 such that

- $\zeta_t = \bar{z}$ for all $t \leq t_1$ and $t > t_2$
- With probability p, $\zeta_t = \underline{\zeta}$ for all $t_1 < t \leq t_2$, while with probability $1 - p$, $\zeta_t = \bar{\zeta}$ for all $t_1 < t \leq t_2$

Proposition B1. *For all $t > t_2$, $t_1 < \tau \leq t_2$: $Pr^s(a_t = 1)$ and $Pr^n(a_t = 1)$ are increasing in the realization ζ_τ.*

Proposition B2. *Consider two different sequences of cue distributions $\mathbf{H}^1 = (H_1^1, \ldots, H_T^1)$ and $\mathbf{H}^2 = (H_1^2, \ldots, H_T^2)$ such that $H_\tau^1 = H_\tau^2$ for $\tau \leq t_2$, but such that $H_\tau^2 \geq H_\tau^1$ for $\tau > t_2$, with strict equality for at least one such τ.*

1. *For all $t > t_2$, $t_1 < \tau \leq t_2$: $Pr^{n,H^2}(a_t = 1) - Pr^n(a_t = 1)$ is decreasing in ζ_τ.*
2. *If $\underline{\zeta}$ is sufficiently low then for all $t > t_2$ and $t_1 < \tau \leq t_2$, $Pr^{s,H^2}(a_t = 1) - Pr^{s,H^1}(a_t = 1)$ is decreasing in ζ_τ.*
3. *For each $\mathbf{H}^1$, there is a $\mu^\dagger < 1$ such that if $\mu_\tau^2 \geq \mu^\dagger$ then for all $t > t_2$ and $t_1 < \tau \leq t_2$, $Pr^{s,H^2}(a_t = 1) - Pr^{s,H^1}(a_t = 1)$ is decreasing in ζ_τ.*

B.2 Additional Experimental Details

Table B1 shows the age, sex and student breakdown by arm. Chi-squared tests from multinomial a probit regression show that the randomization was succesful and that none of these demographic variables are statistically different by arm: $p = 0.20$, $p = 0.41$, $p = 0.29$ separately and $p = 0.40$ jointly. The reminders and no reminders groups are also comparable: chi-square $p = 0.40$, $p = 0.23$, $p = 0.10$ separately and $p = 0.28$ jointly. And similarly for interrupted versus non-interrupted groups: chi-square $p = 0.04$, $p = 0.29$, $p = 0.25$ separately and $p = 0.15$ jointly.

Table B1: Demographics by Experimental Condition

Condition	NB-NR	B-NR	NB-R	B-NR	Overall
Mean Age	31.3	28.1	31.2	28.4	29.6
% Male	26.0%	37.3%	23.7%	27.0%	30.5%
% Students	57.0%	58.8%	47.4%	46.0%	52.6%
Observations	54	51	38	37	180

*Notes:*Condition 1 = No interruption/ No reminders; Condition 2 = Week 2 interruption / No reminders; Condition 3 = No interruption / Reminders; Condition 4 = Week 2 Interruption / Reminders.

B.3 Robustness to Demographic Controls

One subject listed an out of range answer ("Beau") in the age category, and is thus excluded from analysis. Table B2 shows that adding demographic controls does not at all alter the results of Table 1. The demographic controls also do not appear to explain any additional variance in week 3 completion rates.

Table B2: Replication of Table 1 with demographic controls.

Pr(complete survey)	(1)	(2)	(3)
Interruption	–0.420***	–0.419***	–0.421***
	(0.063)	(0.054)	(0.055)
Reminders	0.034	0.032	–0.134
	(0.065)	(0.050)	(0.136)
Interruption*Reminders	0.270**	0.316***	0.326***
	(0.115)	(0.094)	(0.095)
Week1Avg		0.635***	0.527***
		(0.081)	(0.097)
Week1Avg*Reminders			0.251
			(0.163)
Adj. R^2	0.147	0.289	0.294
Observations	1253	1253	1253

Notes: This table estimates a linear model of the probability of completing the daily survey in week 3 of the study, controlling for the demographic variables gathered in the registration phase of the study. The variable "Interruption" equals 1 if the daily survey was available in week 3 and equals 0 otherwise. The variable "Reminders" equals 1 if subject received daily reminders in week 3 and equals 0 otherwise. The variable "Week1avg" denotes the fraction of surveys completed in week 1. Robust standard errors clustered at subject level. All regressions include controls for: day of week, day in study, age, age^2, sex, race, and whether or not subject is a student. $^*p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

B.4 Robustness to calendar date random effects

Table B3 estimates a linear probability model with robust standard errors clustered at the subject level and calendar date level, following Cameron et al. (2011). Unfortunately, when clustering at the calendar date level, including dummy variables for day of the week and for day in study causes the variance-covariance matrix to be highly singular; thuse these dummies are omitted. Instead, I just include day in study as a covariate.

Table B3: Replication of Table 1 with clustering at the calendar date level

Pr(complete survey)	(1)	(2)	(3)
Interruption	–0.416***	–0.398***	–0.401***
	(0.065)	(0.058)	(0.058)
Reminders	0.036	0.028	–0.121
	(0.068)	(0.056)	(0.136)
Interruption*Reminders	0.267**	0.305***	0.320***
	(0.108)	(0.092)	(0.095)
Week1Avg		0.625***	0.527***
		(0.078)	(0.081)
Week1Avg*Reminders			0.226
			(0.152)
Adj. R^2	0.143	0.292	0.297
Observations	1260	1260	1260

Notes: This table estimates a linear probability model of completing the daily survey in week 3 of the study. The variable "Interruption" equals 1 if the daily survey was available in week 3 and equals 0 otherwise. The variable "Reminders" equals 1 if subject received daily reminders in week 3 and equals 0 otherwise. The variable "Week1avg" denotes the fraction of surveys completed in week 1. Robust standard errors clustered at the subject level and calendar date level, following Cameron et al. (2011). $^*p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

B.5 Analysis of Post-Experimental Survey Results

Out of the 172 subjects completing the post-experimental questionnaire, 51 subjects, or 29.7% reported using a reminder technology. Subjects were coded as using a memory aid if they reported 1) using a calendar 2) using notes/diaries/daily planners/Google recurring tasks etc 3) asking others to remind them 4) leaving a tab with the study site open on the computer 5) making the study site their homepage or 6) automating their own daily reminders. Subjects were not coded as using a reminder technology if they reported a routine such as doing the survey each morning, saying that they put it on a "mental to do list" or saying that they bookmarked the survey or starred the study overview email.

Interestingly, subjects were not less likely to use their own reminder technology if they were told that they would not be getting reminders in week 3 (Fisher's exact test $p = 0.40$). There is also no difference in reminder technology use by week 2 Interruption (Fisher's exact test $p = 0.17$).

Table B4: Effect of using own reminder technology, by experimental condition.

Pr(complete survey)	(1)	(2)	(3)	(4)
Interruption	–0.432***	–0.438***	–0.458***	–0.463***
	(0.058)	(0.060)	(0.056)	(0.058)
Reminders	0.051	–0.095	0.046	–0.110
	(0.049)	(0.136)	(0.049)	(0.136)
Interruption*Reminders	0.338***	0.346***	0.362***	0.361***
	(0.093)	(0.093)	(0.100)	(0.097)
Own_Technology	0.142***	0.137***	0.143***	0.137***
	(0.046)	(0.045)	(0.047)	(0.046)
Interruption*Own_Technology	0.255***	0.274***	0.256***	0.276***
	(0.091)	(0.092)	(0.091)	(0.091)
Interruption*Own_Tech*Reminders	–0.208	–0.209	–0.185	–0.182
	(0.165)	(0.168)	(0.163)	(0.165)
Week1Avg	0.553***	0.460***	0.566***	0.467***
	(0.078)	(0.090)	(0.080)	(0.091)
Week1Avg*Reminders		0.222		0.236
		(0.162)		(0.163)
Adj. R^2	0.319	0.323	0.321	0.325
Observations	1204	1204	1197	1197

Notes: This table estimates a linear model of the probability of completing the daily survey in week 3 of the study. The variable "Interruption" equals 1 if the daily survey was available in week 3 and equals 0 otherwise. The variable "Reminders" equals 1 if subject received daily reminders in week 3 and equals 0 otherwise. The variable "Week1avg" denotes the fraction of surveys completed in week 1. The variable "Own_Technology" equals 1 if subject reported having used a reminder technology, and equals 0 otherwise. Robust standard errors clustered at subject level. All regressions include controls for day of week and day in study. Regressions (3) and (4) also include controls for age, age^2, sex, race, and whether or not subject is a student. *$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

C Task Completion Experiment: Supplementary Material

C.1 Reminders Text

Basic Email

Dear Participant,

This is a reminder that the deadline for the risk questionnaire is [[date]], 11:59pm EST.

Thank you.
-The study team
Questions? Send an email to RiskQuestionnaires@gmail.com and we will get back to you

Augmented Email

Dear Participant,

Thank you for signing up for the risk attitudes study. This is a reminder that to receive the $10 Amazon gift card on [[day after deadline]], you must complete the 10-20 minute risk questionnaire by [[deadline]], 11:59pm EST.

To access the questionnaire, please click on the link below and log in with the email and password with which you registered: [[Study URL]]

If you forgot your password and need help resetting it, you can reply to this email.

Thank you.
-The study team
Questions? Send an email to RiskQuestionnaires@gmail.com and we will get back to you

Text Message

This is a reminder that the deadlie for the risk questionnaire is [[date]].

C.2 Demographics

Tabe C5 shows demographics by experimental condition. The last row displays year in college for subjects who are students. Freshmen are coded as 1, Sophomores as 2, etc. Multinomial probit regressions do not reject the null hypothesis that student status, sex, race, and year in college are equally distributed across all 4 conditions ($p > 0.38$ for all variables in separate regressions; $p = 0.85$ in a joint test of significance).

Table C5: Demographics by Experimental Condition

Condition	1	2	3	4	Overall
% Harvard students	87.1%	88.1%	89.1%	90.0%	88.6%
% Male	38.6	37.3%	23.7%	27.0%	41.7%
% White	57.0%	58.8%	47.4%	46.0%	51.6%
% Asian	29.7%	30.7%	32.7%	27.0%	30.0%
% Hispanic	6.9%	5.9%	7.9%	11.0%	7.9%
% African American	6.9%	4.0%	5.9%	2.0%	4.7%
Year (if student)	2.96	3.04	2.89	2.75	2.91
Observations	101	101	101	100	403

Notes: Condition 1 = Short deadline/ No reminders; Condition 2 = Long deadline / No reminders; Condition 3 = Short deadline / Reminders; Condition 4 = Long deadline / Reminders.

C.3 Day of week effects

Table C6: Day of week effects

Pr(complete)	(1)	(2)
reminder	0.150**	0.303***
	(0.068)	(0.066)
Sun	–0.037	–0.080
	(0.138)	(0.104)
Mon	–0.125*	–0.135*
	(0.075)	(0.071)
Tues	0.071	–0.086
	(0.117)	(0.090)
Wed	–0.000	–0.002
	(0.077)	(0.134)
Thurs	–0.131	–0.193**
	(0.121)	(0.098)
Fri	–0.053	0.003
	(0.098)	(0.092)
Adj. R^2	0.047	0.119
Observations	202	201
Day effects F-test	$p = 0.183$	$p = 0.257$

Notes: This table estimates a linear probability model of completing the task, by experimental condition, to test for day of week effects. Column (1) check whether the day of week on which the deadline falls impacts completion rates for the short deadline conditions. Column2 (2) checks whether the day of week on which the deadline falls impacts completion rates for the long deadline conditions. The F tests checks the joint significance of day of week effects for each regression.

C.4 Robustness

Table C7: Robustness to day of week and demographics

Pr(complete)	(1)	(2)	(3)
LongDeadline	–0.177***	–0.171***	–0.170***
	(0.067)	(0.057)	(0.057)
Reminders	0.151**	0.150**	0.151*
	(0.075)	(0.076)	(0.078)
LongDeadline*Reminders	0.163*	0.154*	0.163*
	(0.089)	(0.083)	(0.091)
Adj. R^2	0.087	0.089	0.102
Observations	403	403	403
Demographic Controls	Yes	No	Yes
Day of Week Controls	No	Yes	Yes

Notes: This table estimates a linear probability model of completing the task, by experimental condition. The variable "LongDeadline" equals 1 if subjects had 3 weeks to complete the task, and equals 0 if subjects had 2 days to complete the task. The variable "Reminders" equals 1 if subjects received two days of reminders—days 1 and 2 for subjects with a 2-day deadline and days 20 and 21 for subjects with a 3-week deadline. Robust standard errors are computed by specifying both start date and undergraduate residence as the cluster groups, following the multiway clustering method suggested by Cameron et al. (2011). Demographic controls include whether or not subject is a Harvard undergraduate, sex, and race (White, Asian, Hispanic, African American, or Other). Day of week controls include dummies for which day of the week the deadline falls on. *$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

C.5 Reminder Types

Table C8: Different types of reminders don't have differential effects

	(1)	(2)	(3)	(4)
LongDeadline	–0.178***	–0.178***	–0.178***	–0.178***
	(0.067)	(0.067)	(0.067)	(0.061)
Reminders	0.158**	0.163**	0.165	0.164*
	(0.074)	(0.078)	(0.106)	(0.088)
LongDeadline*Reminders	0.156*	0.146	0.154*	0.157*
	(0.081)	(0.091)	(0.082)	(0.086)
Info	–0.031	–0.047		
	(0.052)	(0.112)		
Info*LongDeadline		0.033		
		(0.176)		
SMSmessage			–0.025	–0.022
			(0.073)	(0.069)
SMS*LongDeadline				–0.005
				(0.116)
Adj. R^2	0.074	0.074	0.074	0.060
Observations	403	403	403	403

Notes: This table estimates a linear probability model of completing the task, by experimental condition. The variable "LongDeadline" equals 1 if subjects had 3 weeks to complete the task, and equals 0 if subjects had 2 days to complete the task. The variable "Reminders" equals 1 if subjects received two days of reminders—days 1 and 2 for subjects with a 2-day deadline and days 20 and 21 for subjects with a 3-week deadline. The variable "Info" equals 1 if the email message contained information in addition to the deadline. The variable SMSMessage equals 1 if subject request to receive SMS message reminders in addition to the emails. Robust standard errors are computed by specifying both start date and undergraduate residence as the cluster groups, following the multiway clustering method suggested by Cameron et al. (2011).

D Proofs of Mathematical Results

Note: The exposition in these proofs is still a little rough and omits minor computations

D.1 Proofs for Section 3

As before, I will let s_t^{s} and s_t^{n} denote the sophisticated and naive DMs' threshold strategies, respectively: the DM chooses $x_t = a$ if and only if $\xi_t \geq s_t$. I will let $\mathbf{s}^{\mathrm{s}} = (s_1^{\mathrm{s}}, \ldots, s_T^{\mathrm{s}})$ denote the sophisticated DM's vector of threshold strategies, and I will let $\mathbf{s}^{\mathrm{n}}$ denote the naive DM's vector of threshold strategies.

Since past of actions does not affect payoffs in this setting, I will use $V_t^{\mathrm{s}}(\alpha_t)$ and $V_t^{\mathrm{n}}(\alpha_t)$ to denote the sophisticated and naive DMs' expected period t utility conditional on being attentive or not that period.

I will let $Pr(\alpha_t = 1)$ denote the probability of being attentive in period t, from the period 0 perspective.

Preliminary results

I begin by showing that for both sophisticated and naive DMs, $Pr(x_t = a)$ is differentiable in $b_{t'}$ for *any* $t' \geq \tau$.

Lemma D1. *$Pr^n(x_t = a)$ is differentiable in $b_{t'}$ for any $t' \geq \tau$.*

Proof. Note that this is clearly true for $t < t'$, since a naive DM does not adjust his current behavior in anticipation of future benefits.

Now suppose that $t = t'$. Because $Pr^{\mathrm{n}}(x_t = a) = Pr(b_t + \xi_t \geq 0)Pr^{\mathrm{n}}(\alpha_t = 1)$, it follows that $Pr^{\mathrm{n}}(x_t = a) = Pr(b_t + \xi_t \geq 0)Pr^{\mathrm{n}}(\alpha_t = 1)$. But $Pr(b_t + \xi_t \geq 0) = 1 - F(-b_t)$ is of course differentiable in b_t by assumption.

Next suppose that $t > t'$. For $t > t'$,

$$\begin{aligned} Pr^{\mathrm{n}}(\alpha_{t+1} = 1) &= \gamma_{t+1}(0, d) + Pr^{\mathrm{n}}(\alpha_t = 1)(\gamma_{t+1}(1, d) - \gamma_{t+1}(0, d)) \\ &+ Pr^{\mathrm{n}}(\alpha_t = 1)Pr(b_t + \xi_t \geq 0)(\gamma_{t+1}(1, a) - \gamma_{t+1}(1, d)). \end{aligned}$$

Thus $Pr^{\mathrm{n}}(\alpha_{t+1} = 1)$ is differentiable in $b_{t'}$ if $Pr^{\mathrm{n}}(\alpha_t)$ and $Pr(b_t + \xi_t \geq 0)$ are differentiable in $b_{t'}$. A simple induction thus shows that $Pr^{\mathrm{n}}(\alpha_t = 1)$ is differentiable in $b_{t'}$ for all $t > t'$. And because $Pr^{\mathrm{n}}(x_t = a) = Pr(b_t + \xi_t \geq 0)Pr^{\mathrm{n}}(\alpha_t = 1)$, it then follows that $Pr^{\mathrm{n}}(x_t = a)$ is differentiable in $b_{t'}$ for all $t > t'$. □

Lemma D2. *$Pr^s(x_t = a)$ is differentiable in $b_{t'}$ for any $t' \geq \tau$.*

Proof. First, suppose that $t = t'$. Then $V_t^{\mathrm{s}}(0)$, $V_{t+1}^{\mathrm{s}}(1)$, and $V_{t+1}^{\mathrm{s}}(0)$ are all differentiable in $b_{t'}$

because they are not functions of $b_{t'}$. By definition,

$$
\begin{aligned}
s_t^{\mathrm{s}} &= -b_t - \left[\gamma_{t+1}(1,a) - \gamma_{t+1}(1,d)\right] \Delta V_{t+1}^{\mathrm{s}} && (14)\\
V_t^{\mathrm{s}}(0) &= V_{t+1}^{\mathrm{s}}(0) + \gamma_{t+1}(0,d)\Delta V_{t+1}^{\mathrm{s}} && (15)\\
V_t^{\mathrm{s}}(1) &= V_{t+1}^{\mathrm{s}}(0) + \gamma_{t+1}(1,d)\Delta V_{t+1}^{\mathrm{s}} + \int_{\xi_t \geq s_t^{\mathrm{s}}} \left[(b_t + \xi_t) + (\gamma_{t+1}(1,a) - \gamma_{t+1}(1,d))\, \Delta V_{t+1}^{\mathrm{s}}\right] dF && (16)
\end{aligned}
$$

where $\Delta V_{t+1}^{\mathrm{s}} = V_{t+1}^{\mathrm{s}}(1) - V_{t+1}^{\mathrm{s}}(0)$. Now suppose that $t = t'$. Then $V_t^{\mathrm{s}}(0)$, $V_{t+1}^{\mathrm{s}}(1)$, and $V_{t+1}^{\mathrm{s}}(0)$ are all differentiable in $b_{t'}$ because they are not functions of $b_{t'}$, and thus equations (14)-(16) show that s_t^{s} and $V_t^{\mathrm{s}}(1)$ are differentiable in b_t. By recursive reasoning, equations (14)-(16) more generally show that s_t^{s} and $V_t^{\mathrm{s}}(\alpha_t)$ are differentiable in $b_{t'}$ for $t \leq t'$. Moreover, s_t^{s} is differentiable in $b_{t'}$ for $t > t'$ for the simple reason that s_t^{s} is not a function of $b_{t'}$ for $t > t'$.

Next, note that

$$
\begin{aligned}
Pr^{\mathrm{s}}(\alpha_{t+1} = 1) &= \gamma_{t+1}(0,d) + Pr^{\mathrm{s}}(\alpha_t = 1)(\gamma_{t+1}(1,d) - \gamma_{t+1}(0,d)) \\
&+ Pr^{\mathrm{s}}(\alpha_t = 1) Pr(\xi_t \geq s_t)(\gamma_{t+1}(1,a) - \gamma_{t+1}(1,d)). && (17)
\end{aligned}
$$

from which it follows that $Pr^{\mathrm{s}}(\alpha_{t+1} = 1)$ is differentiable in $b_{t'}$ if $Pr^{\mathrm{s}}(\alpha_t = 1)$ is differentiable in $b_{t'}$. Similarly, (17) also shows that $Pr^{\mathrm{s}}(\alpha_t = 1)$ is differentiable in $b_{t'}$ if $Pr^{\mathrm{s}}(\alpha_{t+1} = 1)$ is differentiable in $b_{t'}$. But because $Pr(\alpha_t)$ is differentiable in $b_{t'}$ for $t = t'$, recursive reasoning shows that $Pr(\alpha_t)$ is differentiable in $b_{t'}$ for all t.

Finally, because $Pr^{\mathrm{n}}(x_t = a) = Pr(\xi_t \geq s_t^{\mathrm{s}}) Pr^{\mathrm{n}}(\alpha_t = 1)$, it then follows that $Pr^{\mathrm{n}}(x_t = a)$ is differentiable in $b_{t'}$ for all t. □

Proofs of Propositions

Proof of Proposition 1, parts 1 and 2. A naive DM chooses $x_{t'} = a$ if and only if $b_{t'} + \xi_{t'} \geq 0$. So an increase in $b_{t'}$ clearly increases the likelihood of choosing $x_{t'} = a$. Because of the behavioral rehearsal property, this increases the probability of being attentive in period $t+1$, and subsequently choosing $x_{t'+1} = a$. A simple induction shows that for all periods $t > t'$ the DM will thus be more likely to be attentive, and therefore choose $x_t = a$.

I now turn to sophisticated DMs. I first consider how $\mathbf{s}^{\mathrm{s}}$ changes as $b_{t'}$ changes. Conditional on being attentive in some period $t > t'$, a change in $b_{t'}$ does not affect the DM's payoff in any period $\tau \geq t$, or the likelihood of the DM being attentive in any period $\tau > t$. Thus s_t^{s} remains fixed for $t > t'$.

Similarly, $V_t^{\mathrm{s}}(\alpha_t)$ remains fixed for any $t > t'$. The DM's optimal strategy in any period t takes into account period t flow utility, as well as how it affects future expected utility through changes in future expected attentiveness,

$$
s_t^{\mathrm{s}} = -b_t - \left[\gamma_{t+1}(1,a) - \gamma_{t+1}(1,d)\right] \Delta V_{t+1}^{\mathrm{s}} \tag{18}
$$

where $\Delta V_{t+1}^{\mathrm{s}} = V_{t+1}^{\mathrm{s}}(1) - V_{t+1}^{\mathrm{s}}(0)$. But since V_t^{s} is constant in $b_{t'}$ for all $t > t'$, it follows that $s_{t'}^{\mathrm{s}}$ is decreasing in $b_{t'}$.

Next, for s_t^{s} defined as in (18),

$$\begin{aligned} V_t^{\mathrm{s}}(1) - V_t^{\mathrm{s}}(0) &= V_{t+1}^{\mathrm{s}}(0) + \gamma_{t+1}(1,d)\Delta V_{t+1}^{\mathrm{s}} + \int_{\xi_t \geq s_t^{\mathrm{s}}} \left[(b_t + \xi_t) + (\gamma_{t+1}(1,a) - \gamma_{t+1}(1,d))\, \Delta V_{t+1}^{\mathrm{s}}\right] dF \\ &- \left[V_{t+1}^{\mathrm{s}}(0) + \gamma_{t+1}(0,d)\Delta V_{t+1}^{\mathrm{s}}\right] \\ &= \int_{\xi_t \geq s_t^{\mathrm{s}}} \left[(b_t + \xi_t) + (\gamma_{t+1}(1,a) - \gamma_{t+1}(1,d))\, \Delta V_{t+1}^{\mathrm{s}}\right] dF \\ &+ (\gamma_{t+1}(1,d) - \gamma_{t+1}(0,d))\, \Delta V_{t+1}^{\mathrm{s}}. \end{aligned}$$

Differentiating $V_t^{\mathrm{s}}(1) - V_t^{\mathrm{s}}(0)$ with respect to b_t, shows that

$$\begin{aligned} \frac{\partial}{\partial b_t}\left(V_t^{\mathrm{s}}(1) - V_t^{\mathrm{s}}(0)\right) &= \int_{\xi_t \geq s_t^{\mathrm{s}}} (1) dF + \left[(b_t + s_t^{\mathrm{s}}) + (\gamma_{t+1}(1,a) - \gamma_{t+1}(1,d))\, \Delta V_{t+1}^{\mathrm{s}}\right] f(s_t^{\mathrm{s}}) \\ &= 1 - F(s_t^{\mathrm{s}}) > 0 \end{aligned} \tag{19}$$

where (18) is used to obtain the second line. This shows that $\Delta V_{t'}^{\mathrm{s}}$ is increasing in $b_{t'}$.

Next, differentiating $V_t^{\mathrm{s}}(1) - V_t^{\mathrm{s}}(0)$ with respect to $\Delta V_{t+1}^{\mathrm{s}}$, shows that

$$\begin{aligned} \frac{d}{d\Delta V_{t+1}^{\mathrm{s}}}\left(V_t^{\mathrm{s}}(1) - V_t^{\mathrm{s}}(0)\right) &= \int_{\xi_t \geq s_t^{\mathrm{s}}} (\gamma_{t+1}(1,a) - \gamma_{t+1}(1,d))\, dF + \left[(b_t + s_t^{\mathrm{s}}) + (\gamma_{t+1}(1,a) - \gamma_{t+1}(1,d))\, \Delta V_{t+1}^{\mathrm{s}}\right] f(s_t^{\mathrm{s}}) \\ &= 1 - F\left(\gamma_{t+1}(1,a) - \gamma_{t+1}(1,d)\right) > 0 \end{aligned} \tag{20}$$

Thus, since $\Delta V_{t'}^{\mathrm{s}}$ is increasing in $b_{t'}$, a straightforward induction shows that ΔV_t^{s} is increasing in $b_{t'}$ for $t < t'$. Equation (18) thus shows that s_t^{s} is decreasing in $b_{t'}$ for all $t \leq t'$.

Equation (17) shows that $Pr^{\mathrm{s}}(\alpha_{t+1} = 1)$ is increasing in $Pr^{\mathrm{s}}(\alpha_t = 1)$ and $Pr(\xi_t \geq s_t)$. And because $Pr^{\mathrm{s}}(\alpha_1 = 1)$ is constant in $b_{t'}$, recursively applying equation (17) shows that $Pr^{\mathrm{s}}(\alpha_{t'} = 1)$ is increasing in b_t for all t'. □

Proof of Proposition 1, part 3 (naive DM). I first consider naive DMs. Note that $Pr^{\mathrm{n},H^i}(x_t = a) = Pr(b_t + \xi_t \geq 0) Pr^{\mathrm{n},H^i}(\alpha_t = 1)$. Thus

$$\frac{dPr^{\mathrm{n},H^i}(x_t = a)}{db_{t''}} = \frac{dPr(b_t + \xi_t \geq 0)}{db_{t''}} Pr^{\mathrm{n},H^i}(\alpha_t = 1).$$

So the result follows for $t' = t$ if $Pr^{\mathrm{n}}(\alpha_t = 1)$ is increasing $b_{t''}$ This is true, as shown in the proof of Proposition 1, part 1.

Now

$$\begin{aligned} Pr^{\mathrm{n}}(\alpha_{t+1} = 1) &= \gamma_{t+1}(0,d) + Pr^{\mathrm{n}}(\alpha_t = 1)(\gamma_{t+1}(1,d) - \gamma_{t+1}(0,d)) \\ &+ Pr^{\mathrm{n}}(\alpha_t = 1) Pr^{\mathrm{n}}(b_t + \xi_t \geq 0)(\gamma_{t+1}(1,a) - \gamma_{t+1}(1,d)), \end{aligned} \tag{21}$$

which shows that
$$\frac{\partial Pr^{\mathrm{n}}(\alpha_{t+1} = 1)}{\partial b_t} = Pr^{\mathrm{n}}(\alpha_t = 1) \frac{\partial Pr^{\mathrm{n}}(b_t + \xi_t \geq 0)}{\partial b_t}$$

is increasing in $Pr^{\mathrm{n}}(\alpha_t = 1)$. For $t' > t$, similar computations show that

$$\begin{aligned}\frac{\partial Pr^{\mathrm{n}}(\alpha_{t'+1}=1)}{\partial b_t} &= \frac{\partial Pr^{\mathrm{n}}(\alpha_{t'}=1)}{\partial b_t}(\gamma_{t+1}(1,d)-\gamma_{t+1}(0,d)) \\ &+ \frac{\partial Pr^{\mathrm{n}}(\alpha_{t'}=1)}{\partial b_t} Pr^{\mathrm{n}}(b_{t'}+\xi_{t'}\geq 0)(\gamma_{t'+1}(1,a)-\gamma_{t'+1}(1,d))\end{aligned}$$

A simple proof by induction then shows that $\frac{dPr^{\mathrm{n}}(\alpha_{t'}=1)}{db_t}$ is increasing in $Pr^{\mathrm{n}}(\alpha_t = 1)$ for all $t' > t$. But since $Pr^{\mathrm{n}}(\alpha_t = 1)$ is increasing in $b_{t''}$, the statement of the proposition follows. □

Proof of Proposition 1, part 3 (sophisticated DM). Equation (19) shows that

$$\frac{\partial}{\partial b_{t''}}\left(V^{\mathrm{s}}_{t''}(1)-V^{\mathrm{s}}_{t''}(0)\right) = 1 - F(s^{\mathrm{s}}_{t''}) \tag{22}$$

from which it follows that

$$\frac{\partial^2}{\partial b_{t''}\partial b_t}\left(V^{\mathrm{s}}_{t''}(1)-V^{\mathrm{s}}_{t''}(0)\right) = -f(s^{\mathrm{s}}_{t''})\frac{\partial}{\partial b_t}s^{\mathrm{s}}_{t''} > 0 \tag{23}$$

because $\frac{\partial}{\partial b_t}s^{\mathrm{s}}_{t''} < 0$, as established in part 2. Equation (20) combined with a simple induction now shows that

$$\frac{\partial^2}{\partial b_{t''}\partial b_t}\left(V^{\mathrm{s}}_{\tau}(1)-V^{\mathrm{s}}_{\tau}(0)\right) > 0 \tag{24}$$

for all $\tau \leq t''$. Equation (18) now shows that

$$\frac{\partial^2}{\partial b_{t''}\partial b_t}s^{\mathrm{s}}_{\tau} < 0 \tag{25}$$

for all $\tau < t''$ and that

$$\frac{\partial^2}{\partial b_{t''}\partial b_t}s^{\mathrm{s}}_{t''} = 0 \tag{26}$$

for $\tau \geq t''$.

Now equation (17), implies that for any $t' \geq \tau$,

$$\begin{aligned}\frac{\partial^2}{\partial b_{t''}\partial b_t}Pr^{\mathrm{s}}(\alpha_{t'+1}) &= \frac{\partial^2}{\partial b_{t''}\partial b_t}Pr^{\mathrm{s}}(\alpha_{t'})(\gamma_{t'+1}(1,d)-\gamma_{t'+1}(0,d)) &(27)\\ &+ \Bigg[\left(\frac{\partial^2}{\partial b_{t''}\partial b_t}Pr^{\mathrm{s}}(\xi_{t'}\geq s_{t'})\right)Pr^{\mathrm{s}}(\alpha_{t'}) &(28)\\ &+ Pr^{\mathrm{s}}(\xi_{t'}\geq s_{t'})\left(\frac{\partial^2}{\partial b_{t''}\partial b_t}Pr^{\mathrm{s}}(\alpha_{t'})\right) &(29)\\ &+ \left(\frac{\partial}{\partial b_{t''}}Pr^{\mathrm{s}}(\xi_{t'}\geq s_{t'})\right)\left(\frac{\partial}{\partial b_t}Pr^{\mathrm{s}}(\alpha_{t'})\right) &(30)\\ &+ \left(\frac{\partial}{\partial b_t}Pr^{\mathrm{s}}(\xi_{t'}\geq s_{t'})\right)\left(\frac{\partial}{\partial b_{t''}}Pr^{\mathrm{s}}(\alpha_{t'})\right)\Bigg] &(31)\end{aligned}$$

from which it follows that $\frac{\partial^2}{\partial b_{t''}\partial b_t}Pr^{\mathrm{s}}(\alpha_{t'+1}) > 0$ if $\frac{\partial^2}{\partial b_{t''}\partial b_t}Pr^{\mathrm{s}}(\alpha_{t'}) \geq 0$. But since $\frac{\partial^2}{\partial b_{t''}\partial b_t}Pr^{\mathrm{s}}(\alpha_1) \geq 0$, recursively using the computations above, as well as equations (25) and (26) shows that $\frac{\partial^2}{\partial b_{t''}\partial b_t}Pr^{\mathrm{s}}(\alpha_{t'}) > 0$ for all $t' \leq t$. Reasoning analogous to part 1 now shows that $\frac{\partial^2}{\partial b_{t''}\partial b_t}Pr^{\mathrm{s}}(\alpha_{t'}) > 0$ for all $t' > t$ as well. □

Proof of Proposition 2, parts 1,2. Part 2 is obvious because naive DMs just choose $x_t = a$ if $b_t + \xi_t \geq 0$. Part 1 follows because the DM will be more attentive in period t_2, which, as in the proof of Proposition 1 (part 1), will make him more likely to be attentive in all future periods. □

Proof of Proposition 2, part 3. That the result holds under condition (a) is obvious, and follows from the same logic as for the naive case.

Consider now condition (b). Assumptions A1 and A2 imply that there exists an ϵ such that $\int g(\alpha, x, \sigma)dH^1_{t'} < (1-\epsilon)g(\alpha, x, 1-\epsilon)$. But now if $\mu^2_{t'} > 1-\epsilon^2$, then under $H^2_{t'}$, $\sigma > 1-\epsilon$ with probability at least $1-\epsilon$. Thus for a high enough $\mu^2_{t'}$, $Pr^{\mathrm{s},H^2}(\alpha_{t'} = 1) > Pr^{\mathrm{s},H^1}(\alpha_{t'} = 1)$. From this, it follows that the probability of attentiveness under H^2 is higher in all period $t > t'$ as well. Moreover, as in the proof of Proposition 1, the sophisticated DM's strategies don't change for $t \geq t'$, from which the results follow. □

Proof of Proposition 2, part 4. Let $\gamma^i_t(\alpha_{t-1}, x_{t-1})$ denote the probability of being attentive in period t under cue distribution $\mathbf{H}^i_t$. As shown in the proof of part 2 of Proposition 1,

$$V^{\mathrm{s},i}_t(1) - V^{\mathrm{s},i}_t(0) = \int_{\xi_t \geq s^{\mathrm{s}}_t} \left[(b_t + \xi_t) + \left(\gamma^i_{t+1}(1,a) - \gamma^i_{t+1}(1,d)\right)\Delta V^{\mathrm{s}}_{t+1}\right] dF + \left(\gamma^i_{t+1}(1,d) - \gamma^i_{t+1}(0,d)\right)\Delta V^{\mathrm{s}}_{t+1}$$

Consider now $t = t'-1$. By assumption A4, $\gamma^2_{t+1}(1,a) - \gamma^2_{t+1}(1,d) < \gamma^1_{t+1}(1,a) - \gamma^1_{t+1}(1,d)$ and $\gamma^2_{t+1}(1,d) - \gamma^2_{t+1}(0,d) < \gamma^1_{t+1}(1,d) - \gamma^1_{t+1}(0,d)$. It thus follows that $V^{\mathrm{s},2}_{t_2-1}(1) - V^{\mathrm{s},2}_{t_2-1}(0) < V^{\mathrm{s},1}_{t_2-1}(1) - V^{\mathrm{s},1}(0)$. Now the proof of part 2 of Proposition 1 shows that $V^{\mathrm{s}}_t(1) - V^{\mathrm{s}}_t(0)$ is increasing in $\Delta V^{\mathrm{s}}_{t+1}$. From this it follows that $V^{\mathrm{s},2}(1) - V^{\mathrm{s},2}(0) < V^{\mathrm{s},1}(1) - V^{\mathrm{s},1}(0)$ for all $t = t_1+1, t_1+1, \ldots t_2$. As shown in equation (18), this implies that s^{s}_t is higher under $\mathbf{H}^2$; which means that the DM will be less likely to do the task under $\mathbf{H}^2$. □

Proof of Proposition 2, part 5. By definition,

$$Pr^{\mathrm{n},H^2}(\alpha_{t_2} = 1) - Pr^{\mathrm{n},H^1}(\alpha_{t_2} = 1) \tag{32}$$
$$= \left(\gamma^2_{t_2}(0,d) - \gamma^1_{t_2}(0,d)\right) + Pr^{\mathrm{n}}(\alpha_{t_2-1} = 1)\left[(\gamma^2_{t_2}(1,d) - \gamma^1_{t_2}(1,d)) - (\gamma^2_{t_2}(0,d) - \gamma^1_{t_2}(0,d))\right] \tag{33}$$
$$+ Pr^{\mathrm{n}}(\alpha_{t_2-1} = 1)Pr(b_{t_2-1} + \xi_{t_2-1} \geq 0)\left[(\gamma^2_{t_2}(1,a) - \gamma^1_{t_2}(1,a)) - (\gamma^2_{t_2}(1,d) - \gamma^1_{t_2}(1,d))\right]. \tag{34}$$

But assumption A4$'$ imlies that that $(\gamma^2_{t_2}(1,d) - \gamma^1_{t_2}(1,d)) - (\gamma^2_{t_2}(0,d) - \gamma^1_{t_2}(0,d)) < 0$ and that $(\gamma^2_{t_2}(1,a) - \gamma^1_{t_2}(1,a)) - (\gamma^2_{t_2}(1,d) - \gamma^1_{t_2}(1,d)) < 0$. It thus follows that $Pr^{\mathrm{n},H^2}(\alpha_{t_2} = 1) - Pr^{\mathrm{n},H^2}(\alpha_{t_2} = 1)$ is decreasing in $Pr^{\mathrm{n}}(\alpha_{t_2-1} = 1)$. But since $Pr^{\mathrm{n}}(\alpha_{t_2-1} = 1)$ is increasing in $H_{t'}$ by part 1, it thus follows that $Pr^{\mathrm{n},H^2}(\alpha_{t_2} = 1) - Pr^{\mathrm{n},H^2}(\alpha_{t_2} = 1)$ is decreasing in $H_{t'}$.

Finally, note that for $t > t_2$,

$$
\begin{aligned}
& Pr^{\mathrm{n},H^2}(\alpha_t = 1) - Pr^{\mathrm{n},H^1}(\alpha_t = 1) && (35)\\
&= \left(Pr^{\mathrm{n},H^2}(\alpha_t = 1) - Pr^{\mathrm{n},H^1}(\alpha_t = 1)\right)(\gamma_{t+1}(1,d) - \gamma_{t+1}(0,d)) && (36)\\
&= \left(Pr^{\mathrm{n},H^2}(\alpha_t = 1) - Pr^{\mathrm{n},H^1}(\alpha_t = 1)\right) Pr(b_t + \xi_t \geq 0)\,(\gamma_{t+1}(1,a) - \gamma_{t+1}(1,d)) && (37)
\end{aligned}
$$

is increasing in $Pr^{\mathrm{n},H^2}(\alpha_t = 1) - Pr^{\mathrm{n},H^1}(\alpha_t = 1)$. A simple proof by induction thus shows that $Pr^{\mathrm{n},H^2}(\alpha_t = 1) - Pr^{\mathrm{n},H^2}(\alpha_t = 1)$ is decreasing in $H_{t'}$ for all $t \geq t_2$. □

Proof of Proposition 3, part 1. Note that $Pr^{\mathrm{n},H^i}(x_t = a) = Pr(b_t + \xi_t \geq 0)Pr^{\mathrm{n},H^i}(\alpha_t = 1)$. Thus

$$\frac{\partial Pr^{\mathrm{n},H^i}(x_t = a)}{\partial b_{t_2}} = \frac{\partial Pr(b_t + \xi_t \geq 0)}{\partial b_{t_2}} Pr^{\mathrm{n},H^i}(\alpha_t = 1).$$

So the result follows for $t = t_2$ if $Pr^{\mathrm{n},H^2}(\alpha_{t_2} = 1) > Pr^{\mathrm{n},H^1}(\alpha_{t_2} = 1)$. This is true, as shown in the proof of Proposition 2, part 1.

Now for the naive DM,

$$
\begin{aligned}
Pr(\alpha_{t+1} = 1) &= \gamma_{t+1}(0,d) + Pr(\alpha_t = 1)(\gamma_{t+1}(1,d) - \gamma_{t+1}(0,d)) \\
&+ Pr(\alpha_t = 1)Pr(b_t + \xi_t \geq 0)(\gamma_{t+1}(1,a) - \gamma_{t+1}(1,d)), \quad (38)
\end{aligned}
$$

which shows that $\frac{\partial Pr(x_{t+1}=a)}{\partial b_{t_2}}$ is increasing in $Pr(\alpha_t = 1)$ when $t = t_2$. A simple proof by induction then shows that $\frac{\partial Pr(x_{t+1}=a)}{\partial b_{t_2}}$ is increasing in $Pr(\alpha_t = 1)$ for $t \geq t_2$. Since, $Pr^{\mathrm{n},H^2}(\alpha_{t_2} = 1) > Pr^{\mathrm{n},H^1}(\alpha_{t_2} = 1)$, the statement of the proposition follows. □

Proof of Proposition 3, part 2. When looking at period 1 behavior, there is no scope for cues to crowd out behavioral rehearsal as in Proposition 2, part 4. Moreover, period t_1 cues do not affect the sophisticated DM's strategy in periods $t \geq t_1$. Thus the rest of the argument follows analogously to the argument for part 1. □

Proof of Proposition 4, part 1. As shown in equation (38), the impact of period t rehearsal on period $t+1$ attentiveness is increasing in $(\gamma_{t+1}(1,d) - \gamma_{t+1}(0,d))$ and $(\gamma_{t+1}(1,a) - \gamma_{t+1}(1,d))$. Assumption A4 guarantees that $g(1,d,\sigma) - g(0,d,\sigma)$ are decreasing in σ, from which it follows that

$$(\gamma_{t+1}(1,d) - \gamma_{t+1}(0,d)) = \int (g(1,d,\sigma) - g(0,d,\sigma))dH_{t+1}$$

is decreasing in H_{t+1} (in the FOSD order). Similarly, $(\gamma_{t+1}(1,a) - \gamma_{t+1}(1,d))$ are decreasing H_{t+1}. Thus the impact of rehearsal in periods $t < t_2$ on the likelihood of being attentive in period $t = t_2$ is smaller under $\mathbf{H}^2$. □

Proof of Proposition 4, part 2. Assumptions A1 and A2 guarantee that $\gamma_{t_2}(0,d) \to 1$ as $\mu^2_{t_2} \to 1$. This implies that the impact of period $t < t_2$ rehearsal on period t_2 attentiveness can be made arbitrarily small when $\mu^2_{t_2}$ is sufficiently high. As a consequence, the impact of period $t < t_2$

rehearsal on period $t' > t_2$ attentiveness can be made arbitrarily small when $\mu^2_{t_2}$ is sufficiently high. □

D.2 Proofs for Section 4

Throughout, I will use the following additional notation. I will use $s_t^{\text{pa},T}$ o denote the sophisticated DM's period t threshold rule, given a deadline T. That is, the sophisticated DM chooses $x_t = a$ if and only if $\xi_t \geq s_t^{\text{pa},T}$. I will define $s_t^{\text{s},T}$ and $s_t^{\text{n},T}$ similarly. When considering results under two different deadlines T_1, and T_2, I will use the superscript T_i to index the respective strategies s_t^{pa,T_i}, s_t^{s,T_i}, s_t^{n,T_i}. I will let $\mathbf{d}_t$ denote the $1 \times t$ vector $(d, \ldots, d)$. Then $V_t^{\text{pa}}(\mathbf{d}_{t-1})$ will denote the perfectly attentive DM's period t utility, conditional on not having yet completed the task.

Preliminary Results

I begin with several lemmas that will be used in the proofs of the propositions.

Lemma D3. *Suppose that $b_t \equiv b$ for all t. Then $V_0^{pa,T} \to b + \bar{\xi}$ as a $T \to \infty$.*

Proof. Suppose that the DM follows the following strategy: For some $\epsilon > 0$, choose $x_t = a$ if and only if $b + \xi_t > b + \bar{\xi} - \epsilon/2$. Then conditional on completing the task, the DM's utility is at least $b + \bar{\xi} - e/2$. And his probability of completing the task is at least $p_{\epsilon,T} = 1 - \big(1 - F(\bar{\xi} - \epsilon/2)\big)^T$. Clearly, $p_{\epsilon,T} \to 1$ as $T \to \infty$. Thus for any $\epsilon > 0$,

$$\lim_{T\to\infty} V_0^{\text{pa},T} > b + \bar{\xi} - \epsilon. \tag{39}$$

But since equation (39) holds for any $\epsilon > 0$, it therefore follows that $V_0^{\text{pa},T} \to b + \bar{\xi}$ as a $T \to \infty$. □

Lemma D4. *Suppose that $b_t \equiv b$ for all t. Then $Pr(Q^{pa,T} \leq T) \to 1$ as a $T \to \infty$.*

Proof. Suppose, by way of contradiction, that $1 - Pr(Q^{\text{pa,T}} \leq T) > \eta$ for some $\eta > 0$. This would then imply that $V_0^{\text{pa},T} < (b + \bar{\xi})(1 - \eta)$ for all T. This is in direct contradiction to Lemma D3. □

Lemma D5. *Suppose that $b_t \equiv b$ for all t. Then for any fixed t, $V_t^{pa,T}(\mathbf{d}_{t-1}) \to b + \bar{\xi}$ as a $T \to \infty$.*

Proof. Follows identically to the proof of Lemma D3. □

Lemma D6. *Fix some $t \geq 1$. Then $\lim_{T\to\infty} s_t^{pa,T} = \bar{\xi}$.*

Proof. By definition, $b + s^{\text{pa},T} = V_{t+1}^{\text{pa},T}$. By Lemma D5, $V_{t+1}^{\text{pa},T}(\mathbf{d}_{t-1}) \to b + \bar{\xi}$ as $T \to \infty$. Thus $s^{\text{pa},T} \to \bar{\xi}$ as $T \to \infty$. □

Lemma D7. *For some $t^\dagger \leq T$, set $b_t = b'_t + \eta$ for $t \leq t^\dagger$, and $b_t = b'_t$ for $t > t^\dagger$. Then $Pr(Q^{pa} \leq T)$ is increasing in η.*

Proof. I will let $V_t^{\text{pa}}(\mathbf{d}_{t-1};\eta)$ denote the expected period t utility as a function of η. For $t > t^\dagger$, $V_t^{\text{pa}}(\mathbf{d}_{t-1};\eta)$ does not depend on η. For $t \leq t^\dagger$, I will now show that $V_t^{\text{pa}}(\mathbf{d}_{t-1};\eta) - V_t^{\text{pa}}(\mathbf{d}_{t-1};0) \leq \eta$ Note that the threshold rule is given by $s_t^{\text{s}}(\eta) = \max\left(V_{t+1}^{\text{pa}}(\mathbf{d}_{t-1}) - b_t' - \eta, \underline{\xi}\right)$. Setting $V_{T+1}(\mathbf{d}_T) = 0$, we now have:

$$
\begin{aligned}
V_t^{\text{pa}}(\mathbf{d}_{t-1};\eta) &= \int_{s_t^{\text{s}}(\eta)}^{\bar{xi}} (b_t' + \eta - \xi_t)dF + F(s_t^{\text{s}}(\eta))V_{t+1}^{\text{pa}}(\mathbf{d}_{t-1};\eta) \\
&= \int_{s_t^{\text{s}}(0)}^{s_t^{\text{s}}(\eta)} (b_t' + \eta - \xi_t)dF + \int_{s_t^{\text{s}}(0)}^{\bar{xi}} (b_t' + \eta - \xi_t)dF + F(s_t^{\text{s}}(\eta))V_{t+1}^{\text{pa}}(\mathbf{d}_{t-1};\eta) \\
&= \int_{s_t^{\text{s}}(0)}^{s_t^{\text{s}}(\eta)} \left[(b_t' + \eta - \xi_t) - V_{t+1}^{\text{pa}}(\mathbf{d}_{t-1})\right] dF + \int_{s_t^{\text{s}}(0)}^{\bar{xi}} (b_t' + \eta - \xi_t)dF + F(s_t^{\text{s}}(0))V_{t+1}^{\text{pa}}(\mathbf{d}_{t-1};\eta) \\
&= (1 - F(s_t^{\text{s}}(\eta)))\eta + \int_{s_t^{\text{s}}(0)}^{s_t^{\text{s}}(\eta)} \left[(b_t' - \xi_t) - V_{t+1}^{\text{pa}}(\mathbf{d}_{t-1})\right] dF + \int_{s_t^{\text{s}}(0)}^{\bar{xi}} (b_t' - \xi_t)dF + F(s_t^{\text{s}}(0))V_{t+1}^{\text{pa}}(\mathbf{d}_{t-1};\eta) \\
&\leq (1 - F(s_t^{\text{s}}(\eta)))\eta + \int_{s_t^{\text{s}}(0)}^{\bar{xi}} (b_t' - \xi_t)dF + F(s_t^{\text{s}}(0))V_{t+1}^{\text{pa}}(\mathbf{d}_{t-1};\eta) \\
&= (1 - F(s_t^{\text{s}}(\eta)))\eta + F(s_t^{\text{s}}(0))\left[V_{t+1}^{\text{pa}}(\mathbf{d}_{t-1};\eta) - V_{t+1}^{\text{pa}}(\mathbf{d}_{t-1};0)\right] + V_t^{\text{pa}}(\mathbf{d}_{t-1};0).
\end{aligned}
$$

Now clearly, $V_{T+1}^{\text{pa}}(\mathbf{d}_T;\eta) - V_{T+1}^{\text{pa}}(\mathbf{d}_T;0) < \eta$, and thus a simple induction shows that $V_t^{\text{pa}}(\mathbf{d}_{t-1};\eta) - V_t^{\text{pa}}(\mathbf{d}_{t-1};0) \leq \eta$. □

Proof of Propositions

Proof of Proposition 5. Part 1. Obvious. Conditional on being attentive, the naive DM follows the same strategy as the perfectly attentive DM. However, the naive DM is less likely to do the task each period because he is inattentive with some positive probability.

Part 2. Clearly, $Pr(Q^{\text{s}} \leq T) < Pr(Q^{\text{pa}} \leq T)$ when $b + \underline{\xi} \geq 0$: by assumption, the perfectly attentive DM will complete the task with probability 1, whereas the inattentive DM will complete the task with probability less than one. Now by assumption, the inattentive DM is attentive with probability no greater than $\bar{z} < 1$ in all periods, and thus $V_2^{\text{s}} \leq \bar{z}(b + \bar{\xi})$ for any value of T. This means that s_1^{s} is bounded away from $\bar{\xi}$ for all T or, equivalently, that $Pr^{\text{s}}(x_1 = a)$ is bounded away from 0 for all T. On the other hand, Lemma D6 implies that for any $t^\dagger$, $\Pr(Q^{\text{pa}} \leq t^\dagger) \to 0$ as $T \to \infty$.

Part 3. Obvious. Sufficiently low probability of being attentive can make completion probabilities arbitrarily low. □

Proof of Proposition 6. Part 1. Obvious

Part 2. An even stronger result is true. Under condition $I(T_1, T_2)$, the probability of completing the task between periods $t = \Delta + 1$ and $t = T_2$, *conditional on not having completed the task by period* $\Delta + 1$, is higher than the probability of completing the task when given the short deadline. Formally, $Pr(Q^{\text{pa},T_1} \leq T_1) < Pr(Q^{\text{pa},T_2} \leq T_2 | Q^{\text{pa},T_2} > \Delta)$.

The short deadline game can be transformed into the T_1 period subgame of the long deadline through the following series of operations:

- Increase the payoff in all periods by $b_{T_2} - b_{T_1}$.
- Increase the payoff all but the last period by $(b_{T_2-1} - b_{T_1-1}) - (b_{T_2} - b_{T_1})$
- ...
- Increase the payoff in the first period by $(b_{\Delta+1} - b_1) - (b_{\Delta+2} - b_2)$.

Lemma D7 implies that each operation increase the likelihood of completing the task, which establishes the result.

Part 3. Note that $V_t^{s,T_i} \le \bar{z}(b + \bar{\xi})$ for all t. By definition, $s_t^{s,T_i} = \max(V_{t+1}^{s,T_i} - b, \underline{\xi})$, from which it follows that $s_t^{s,T_i} = \max\left(\bar{z}\bar{\xi} - (1-\bar{z})b, \underline{\xi}, -b\right)$. However, $\bar{z}\bar{\xi} - (1-\bar{z})b \to -\infty$ as $b \to \infty$, from which it follows that $s_t^{s,T_i} \to \underline{\xi}$ as $b \to \infty$.

Thus for $t \le T_1$, $s_t^{s,T_1}/s_t^{s,\overline{T}_2} \to 1$ as $b \to \infty$. Moreover, the assumption that $\gamma_t(1,d)$ is bounded away from 1 implies that $\lim_{b\to\infty} Pr(Q^{s,T_i} \le T_i) < 1$ each $i = 1, 2$. Thus it follows that

$$\begin{aligned}
\lim_{b\to\infty} 1 - Pr(Q^{s,T_2} \le T_2) &= \lim_{b\to\infty} Pr(T_1 + 1 \le Q^{s,T_2} \le T_2)\left(1 - Pr(Q^{s,T_2} \le T_1)\right) \\
&< \lim_{b\to\infty} \left(1 - Pr(Q^{s,T_2} \le T_1)\right),
\end{aligned}$$

which completes the proof. □

Proof of Proposition 7. Clearly, $V_t^{pa,T_2} > V_t^{pa,T_1}$ for all $t \le T_1$. Therefore, $s_t^{n,T_1} < s_t^{n,T_2}$ for all $t \le T_1$. This implies that $Pr(2 \le Q^{n,T_2} \le T_1) < Pr(2 \le Q^{n,T_1} \le T_1)$. Noting that $\gamma_1(1,d)$ is the probability of being attentive in period 1 under both deadlines,

$$\begin{aligned}
Pr(Q^{n,T_1} \le T_1) - Pr(Q^{n,T_2} \le T_1) &= \left[1 - \gamma_1(1,d)(1 - F(s^{s,T_2}))\right]\left[1 - Pr(2 \le Q^{n,T_2} \le T_1 | Q^{n,T_2} > 1)\right] \\
&- \left[1 - \gamma_1(1,d)(1 - F(s_1^{s,T_1}))\right]\left[1 - Pr(2 \le Q^{n,T_1} \le T_1 | Q^{n,T_1} > 1)\right] \\
&> \left[1 - \gamma_1(1,d)(1 - F(s_1^{s,T_2}))\right]\left[1 - Pr(2 \le Q^{n,T_2} \le T_1 | Q^{n,T_2} > 1)\right] \\
&- \left[1 - \gamma_1(1,d)(1 - F(s_1^{s,T_1}))\right]\left[1 - Pr(2 \le Q^{n,T_2} \le T_1 | Q^{n,T_2} > 1)\right] \\
&= \left[\gamma_1(1,d)\left(F(s_1^{s,T_1}) - F(s_1^{s,T_2})\right)\right]\left[1 - Pr(2 \le Q^{n,T_2} \le T_1 | Q^{n,T_2} > 1)\right] \\
&> \left[\gamma_1(1,d)\left(F(s_1^{s,T_1}) - F(s_1^{s,T_2})\right)\right]\left[1 - Pr(2 \le Q^{pa,T_2} \le T_1 | Q^{pa,T_2} > 1)\right] \\
&= \left[\gamma_1(1,d)\left(F(s_1^{s,T_1}) - F(s_1^{s,T_2})\right)\right]\prod_{t=2}^{T_1} F(s_t^{n,T_2}),
\end{aligned} \tag{40}$$

As before, let $\mathbf{d}_t$ denoting the $1 \times t$ vector $(d, \ldots, d)$. Let $Pr(\alpha_t | h_t = \mathbf{d}_{t-1}, \alpha_1, T_i)$ denote the probability that the naive DM is attentive or inattentive in period t, conditional on: 1) not having completed the task by period t, 2) facing the deadline T_i, and 3) whether or not he was attentive in period 1.

Condition (i) in the statement of the proposition implies that $Pr(\alpha_{T_1} = 1 | \mathbf{d}_{T_1-1}, \alpha_1 = 1, T_2) \le$

$\gamma_1(1,d)$. Conditions (ii) and(iii) thus implies that for $t > T_1$

$$\begin{aligned} Pr(\alpha_t = 1|\mathbf{d}_{t-1}, \alpha_1 = 1, T_2) &\leq \lambda Pr(\alpha_{t-1} = 1|\mathbf{d}_{t-2}, \alpha_1 = 1, T_2) + \lambda\gamma_1(d,1)Pr(\alpha_{t-1} = 0|\mathbf{d}_{-1t}, \alpha_1 = 1, T_2) \\ &< \lambda Pr(\alpha_{t-1} = 1|\mathbf{d}_{T_1-1}, \alpha_1 = 1, T_2) + \lambda\gamma_1(d,1). \end{aligned} \quad (41)$$

Equation (41) thus shows that if $Pr(\alpha_{t-1} = 1|\mathbf{d}_{t-2}, \alpha_1 = 1, T_2) \leq \gamma_1(d,1)$ and $\lambda < 1/2$, then $Pr(\alpha_t = 1|\mathbf{d}_{t-1}, \alpha_1 = 1, T_2) < \gamma_1(d,1)$ and $Pr(\alpha_t = 1|\mathbf{d}_{t-1}, \alpha_1 = 1, T_2) < 2\lambda\gamma_1(d,1)$. A simple induction thus shows that if $\lambda < 1/2$, then $Pr(\alpha_t = 1|\mathbf{d}_{t-1}, \alpha_1 = 1, T_2) < 2\lambda\gamma_1(d,1)$ for all $t > T_1$.

Thus the probability of being attentive in any period $t > T_1$ conditional on not having completed the task by that period is bounded above by $2\lambda\gamma_1(d,1)$ when $\lambda < 1/2$. And overall, this implies that for $2\lambda(T_2 - T_1) < 1$, the probability of completing the task after period T_1, conditional on not completing it by period T_2, is bounded above by $2\lambda\gamma_1(d,1)(T_2 - T_1)$. Thus by equation (40),

$$\begin{aligned} Pr(Q^{\mathrm{n},T_2} \leq T_2) - Pr(Q^{\mathrm{n},T_1} \leq T_1) &= Pr(Q^{\mathrm{n},T_2} \leq T_1) + Pr(T_1 < Q^{\mathrm{n},T_2} \leq T_2 | Q^{\mathrm{n},T_2} > T_1) - Pr(Q^{\mathrm{n},T_1} \leq T_1) \\ &< 2\lambda\gamma_1(d,1)(T_2 - T_1) - \left[\gamma_1(1,d)\left(F(s_1^{\mathrm{s},T_1}) - F(s_1^{\mathrm{s},T_2})\right)\right]\prod_{t=2}^{T_1} F(s_t^{\mathrm{n},T_2}) \end{aligned}$$

is negative when λ is sufficiently small such that

$$2\lambda(T_2 - T_1) < \left(F(s_1^{\mathrm{s},T_1}) - F(s_1^{\mathrm{s},T_2})\right)\prod_{t=2}^{T_1} F(s_t^{\mathrm{n},T_2}).$$

□

Proof of Proposition 8, part 1. Let f be the density function of F, and let M denote the maximum f on $[\underline{\xi}, \bar{\xi}]$. By assumption $\iota > 0$.

Step 1. I will begin with the simpler case in which $b_t \equiv b$ for all t. For any $\epsilon > 0$, Lemma D4 implies that there is a sufficiently high T_ϵ such that $V_0^{\mathrm{pa},T} > b + \bar{\xi} - \epsilon^2$. This then implies that the perfectly attentive DM obtains a payoff that's at least $b + \bar{\xi} - \epsilon$ with probability at least $1 - \epsilon$. Thus with probability at least $1 - \epsilon$, the DM completes the task by using a threshold strategy $s_t^{\mathrm{s}} > \bar{\xi} - \epsilon$. Let $t^\dagger$ be such that $s_t^{\mathrm{s}} > \bar{\xi} - \epsilon$ whenever $t \geq t^\dagger$. Thus

$$\prod_{t=t^\dagger}^{T} F(s_t^{\mathrm{s}}) < \epsilon$$

or

$$\sum_{t=t^\dagger}^{T} \log F(s_t^{\mathrm{s}}) < \log(\epsilon). \quad (42)$$

Now a Taylor expansion shows that

$$\begin{aligned}
\sum_{t=t^\dagger}^{T} \log F(s_t^{\text{s}}) &= -\sum_{t=t^\dagger}^{T}\left[\sum_{i=1}^{\infty}(1-F(s_t^{\text{s}}))^i/i!\right] \\
&> -\sum_{t=t^\dagger}^{T}\left[\sum_{i=1}^{\infty}(1-F(s_t^{\text{s}}))^i\right] \\
&= -\sum_{t=t^\dagger}^{T}\frac{1-F(s_t^{\text{s}})}{F(s_t^{\text{s}})} \\
&> -\sum_{t=t^\dagger}^{T}\frac{1-F(s_t^{\text{s}})}{F(\bar{\xi}-\epsilon)}. \qquad (43)
\end{aligned}$$

Now for this same deadline T_ϵ, the probability that a naive DM never completes the task satisfies

$$1 - Pr(Q^{\text{n}} \leq T_\epsilon) < \prod_{t=t^\dagger}^{T}\left[1-\underline{z}(1-F(s_t^{\text{s}}))\right]$$

Taking logs and Taylor expanding,

$$\begin{aligned}
\log\left(1-Pr(Q^{\text{n}} \leq T_\epsilon)\right) &< \sum_{t=t^\dagger}^{T}\log\left[1-\underline{z}(1-F(s_t^{\text{s}}))\right] \\
&= -\underline{z}\sum_{t=t^\dagger}^{T}\left[\sum_{i=1}^{\infty}(1-F(s_t^{\text{s}}))^i/i!\right] \\
&< -\underline{z}\sum_{t=t^\dagger}^{T}(1-F(s_t^{\text{s}})) \\
&< -\underline{z}F(\bar{\xi}-\epsilon)\log(\epsilon), \qquad (44)
\end{aligned}$$

where the last line follows from equations (42) and (43).

But now the expression in Equation (44) approaches $-\infty$ as $\epsilon \to 0$, from which it follows that $Pr(Q^{\text{n}} \leq T_\epsilon) \to 1$ as $\epsilon \to 0$.

Step 2 (sketch) I now prove the more general statement in the proposition. Let Q_*^{n} denote the stopping time corresponding to the model in which period t payoffs are $b^* + \xi_t$. Step 1 is easily generalized to show that for each $\epsilon > 0$, there is a Δ_ϵ such that if $T_{t^\dagger,\epsilon} = t^\dagger + \Delta_\epsilon$ $Pr(Q_*^{\text{n}} \leq T_{t^\dagger,\epsilon}|Q_*^{\text{n}} > t^\dagger) > 1-\epsilon/2$ for all $t^\dagger$. The reason is as follows. For the sophisticated DM, equations (42) and (43) generalize immediately to the Δ_ϵ period subgame. For the naive DM, equation (44) also carries over to the Δ_ϵ subgame because all that matters is that the probability of being attentive is bounded from below by $\underline{z}$.

Next, it is easy to show that the threshold strategis s_t^{n} are continuous in the payoffs $b_1, \ldots b_T$. Thus by continuity, $Pr(Q_*^{\text{n}} \leq T_{t^\dagger,\epsilon}|Q_*^{\text{n}} > t^\dagger) \to Pr(Q^{\text{n}} \leq T_{t^\dagger,\epsilon}|Q^{\text{n}} > t^\dagger)$ as $t^\dagger \to \infty$.

Now choose some $\epsilon > 0$. And let Δ_ϵ be the value such that $Pr(Q_*^{\text{n}} \leq T_{t^\dagger,\epsilon}|Q_*^{\text{n}} > t^\dagger) > 1-\epsilon/2$ for all $t^\dagger$ and $T_{t^\dagger,\epsilon} = t^\dagger + \Delta_\epsilon$. Now fixing Δ_ϵ, we can find a high enough $t^\dagger$ be such that $Pr(Q^{\text{n}} \leq T_{t^\dagger,\epsilon}|Q^{\text{n}} >$

$t^\dagger) > Pr(Q^{\rm n}_* \leq T_{t^\dagger,\epsilon}|Q^{\rm n}_* > t^\dagger) - \epsilon/2$. Combined, this shows that $Pr(Q^{\rm n} \leq T_{t^\dagger,\epsilon}|Q^{\rm n} > t^\dagger) > 1 - \epsilon$. □

Proof of Proposition 8, part 2. Let $Pr(\alpha_t = 1|\mathbf{d}_{t-1})$ denote the probability that the naive DM is attentive in period t, conditional on not yet having completed the task. Note that

$$\begin{aligned} Pr(\alpha_t = 1|\mathbf{d}_{t-1}) &= \frac{\gamma_t(1,d)Pr(\alpha_{t-1} = 1|\mathbf{d}_{t-2})F(s^{\rm n}_{t-1})}{1 - Pr(\alpha_{t-1} = 1|\mathbf{d}_{t-2})F(s^{\rm n}_{t-1})} \\ &< \gamma_t(1,d)Pr(\alpha_{t-1} = 1|\mathbf{d}_{t-2})F(s^{\rm n}_{t-1}) \\ &\leq \bar{z}Pr(\alpha_{t-1} = 1|\mathbf{d}_{t-2})F(s^{\rm n}_{t-1}). \end{aligned}$$

Thus

$$Pr(\alpha_t = 1|\mathbf{d}_{t-1}) < \bar{z}^t. \tag{45}$$

Now for any $t^\dagger$ and $\epsilon > 0$, Lemma D6 implies that there exists a deadline length $T_{t^\dagger}$ such that $Pr\left(Q^{{\rm n},T_{t^\dagger}} \leq t^\dagger\right) < \epsilon$. This implies that the probability of ever completing the task is given by

$$\begin{aligned} Pr\left(Q^{{\rm n},T_{t^\dagger}} \leq T_{t^\dagger}\right) &= \epsilon + \sum_{t=t^\dagger+1}^{T_{t^\dagger}} Pr(\alpha_t = 1|\mathbf{d}_{t-1}) \\ &< \epsilon + \sum_{t=t^\dagger+1}^{T_{t^\dagger}} \bar{z}^t \\ &< \epsilon + \frac{\bar{z}^{t^\dagger}}{1-\bar{z}^{t^\dagger}} \end{aligned}$$

But $\frac{\bar{z}^{t^\dagger}}{1-\bar{z}^{t^\dagger}} \to 0$ as $t^\dagger \to \infty$, which shows that $Pr\left(Q^{{\rm n},T_{t^\dagger}}\right)$ can be made arbitrarily small with a large enough $T_{t^\dagger}$ □

Proof of Proposition 9, part 1. Part 1. By assumption A2, as $\kappa^1_t \to 1$ for all $t \leq T_1$, $Pr(Q^{{\rm n},T_1} \leq T_1) \to Pr(Q^{{\rm pa},T_1} \leq T_1)$. Similarly, $Pr(Q^{{\rm n},T_2} \leq T_2|Q^{{\rm n},T_2} > \Delta) \to Pr(Q^{{\rm pa},T_2} \leq T_2|Q^{{\rm pa},T_2} > \Delta)$ as $\kappa^2_{\Delta+t} \to 1$ for all $t \leq T_1$. But now the proof of part 2 of Proposition 6 shows that $Pr(Q^{{\rm pa},T_2} \leq T_2|Q^{{\rm pa},T_2} > \Delta) \geq Pr(Q^{{\rm pa},T_1} \leq T_1)$. Thus

$$Pr(Q^{{\rm n},T_2} \leq T_2) = Pr(Q^{{\rm n},T_2} \leq \Delta) + \left(1 - Pr(Q^{{\rm n},T_2} \leq T_2)\right) Pr(Q^{{\rm n},T_2} \leq T_2|Q^{{\rm n},T_2} > \Delta) > Pr(Q^{{\rm n},T_1} \leq T_1)$$

as $\kappa^1_t \to 1$ for all $t \leq T_1$.

The proof for sophisticated DMs follows identically. □

Proof of Proposition 9, part 2. To simplify notation, set $q^i = Pr(Q^{{\rm n},T_i} \leq T_i - 1)$, and set $p^i = Pr(\alpha_{T_i} = 1|\mathbf{d}_{T_i-1})$. That is, q^i is the probability that a naive DM completes the task by period $T_i - 1$ when facing the deadline T_i, whereas p^i is the probability of being attentive in period T_i, conditional on not having completed the task by that time.

The effect of adding a strength $\kappa \equiv \kappa^1_{T_1} = \kappa^2_{T_2}$ cue is that p^i is transformed to $p^i(\kappa) = p^i + (1-p^i)\kappa$.

Suppose that

$$\left[q^1 + (1-q^1)p^1(\kappa)Pr(b_{T_1} + \xi_{T_1} \geq 0)\right] - \left[q^2 + (1-q^2)p^2(\kappa)Pr(b_{T_2} + \xi_{T_2} \geq 0)\right] > 0 \tag{46}$$

when $\kappa = 0$.

Differentiating the expression in (46) with respect κ, and noting that $(Pr(b_{T_1} + \xi_{T_1} \geq 0) \leq Pr(b_{T_2} + \xi_{T_2} \geq 0)$ by assumption, yields

$$
\begin{aligned}
& (1-q^1)(1-p^1)Pr(b_{T_1}+\xi_{T_1} \geq 0) - (1-q^2)(1-p^2)Pr(b_{T_2}+\xi_{T_2} \geq 0) && (47)\\
= \; & (1-q^2)p^2 Pr(b_{T_2}+\xi_{T_2} \geq 0) - (1-q^1)p^1 Pr(b_{T_1}+\xi_{T_1} \geq 0) && (48)\\
- \; & (1-q^2)Pr(b_{T_2}+\xi_{T_2} \geq 0) + (1-q^1)Pr(b_{T_1}+\xi_{T_1} \geq 0) && (49)\\
< \; & q^1 - q^2 - (1-q^2)Pr(b_{T_2}+\xi_{T_2} \geq 0) + (1-q^1)Pr(b_{T_1}+\xi_{T_1} \geq 0) && (50)\\
< \; & (q^1 - q^2)\left(1 - Pr(b_{T_2}+\xi_{T_2} \geq 0)\right) && (51)\\
= \; & 0
\end{aligned}
$$

where equation (50) is obtained by substituting the inequality (46) into the experession in lines (48), (49). □

Proof of Proposition 9, part 3. To simplify notation, set $q_t^i = Pr(Q^{\mathrm{n},T_i} \leq t)$, and set $p_t^i = Pr(\alpha_t = 1|\mathbf{d}_{T_i-1})$. That is, q_t^i is the probability that a naive DM completes the task by period t when facing the deadline T_i, whereas p^i is the probability of being attentive in period t, conditional on not having completed the task by that time. Let $P_t^i(\alpha_t) = Pr(Q^{\mathrm{n},T_i} \leq T_i|\alpha_t, Q^{\mathrm{n},T_i} > t-1)$ denote the probability of completing the task conditional on i) facing the deadline T_i, ii) period t attentiveness α_t, and iii) not having completed the task before period t.

The effect of adding a strength $\kappa \equiv \kappa_t^1 = \kappa_{t+\Delta}^2$ cue in periods t and $t+\Delta$, respectively, is that p_t^1 is transformed to $p_t^1(\kappa) = p_t^1 + (1-p_t^1)\kappa$ and $p_{t+\Delta}^2$ is transformed to $p_{t+\Delta}^2(\kappa) = p_{t+\Delta}^1 + (1-p_{t+\Delta}^1)\kappa$

Suppose that $Pr(Q^{\mathrm{n},T_1} \leq T_1) - Pr(Q^{\mathrm{n},T_2} \leq T_2) > 0$ when $\kappa = \kappa_t^1 = \kappa_{t+\Delta}^2 = 0$. Then, equivalently,

$$
\begin{aligned}
& [q_{t-1}^1 + (1-q_{t-1}^1)p_t^1(\kappa)P_t^1(1) + (1-q_{t-1}^1)(1-p_t^1(\kappa))P_t^1(0)] && (52)\\
& -[q_{t-1+\Delta}^2 + (1-q_{t-1+\Delta}^2)p_{t+\Delta}^2(\kappa)P_{t+\Delta}^2(1) + (1-q_{t-1+\Delta}^2)(1-p_{t+\Delta}^2(\kappa))P_{t+\Delta}^2(0)] > 0 && (53)
\end{aligned}
$$

when $\kappa = 0$. Equivalently,

$$
\begin{aligned}
& [q_{t-1}^1 + (1-q_{t-1}^1)p_t^1(\kappa)(P_t^1(1) - P_t^1(0)) + (1-q_{t-1}^1)P_t^1(0)] && (54)\\
& -[q_{t-1+\Delta}^2 + (1-q_{t-1+\Delta}^2)p_{t+\Delta}^2(\kappa)(P_{t+\Delta}^2(1) - P_{t+\Delta}^2(0)) + (1-q_{t-1+\Delta}^2)P_{t+\Delta}^2(0)] > 0 && (55)
\end{aligned}
$$

Differentiating the expression in lines (54)-(55) with respect to κ yields

$$
[(1-q_{t-1}^1)(1-p_t^1(\kappa))(P_t^1(1) - P_t^1(0))] - [(1-q_{t-1+\Delta}^2)(1-p_{t+\Delta}^2(\kappa))(P_{t+\Delta}^2(1) - P_{t+\Delta}^2(0))] \quad (56)
$$

By assumption, $\gamma_t^1 = \gamma_{t+\Delta}^2$ for all $t \leq T_1$. Combining this with reasoning similar to that in Part 2 of Proposition 6 shows that Condition $I(T_1, T_2)$ guarantees that

$$
\begin{aligned}
& P_{t+\Delta}^2(1) \geq P_t^1(1) && (57)\\
& P_{t+\Delta}^2(0) \geq P_t^1(0) && (58)\\
& P_{t+\Delta}^2(1) - P_{t+\Delta}^2(0) \geq P_t^1(1) - P_t^1(0) && (59)
\end{aligned}
$$

The assumption that $\gamma_t^1 = \gamma_{t+\Delta}^2$ for all $t \leq T_1$ also implies that $p_t^1 > p_{t+\Delta}^2$. Thus, if $q_{t-1}^1 > q_{t-1+\Delta}^2$, then, using (59), the expression in equation (56) is negative.

Otherwise, when $q_{t-1}^1 < q_{t-1+\Delta}^2$, combining the inequality in lines (54)-(55) with inequality (59) shows that

$$\begin{aligned}
& [(1-q_{t-1}^1)(1-p_t^1(\kappa))(P_t^1(1)-P_t^1(0))] - [(1-q_{t-1+\Delta}^2)(1-p_{t+\Delta}^2(\kappa))(P_{t+\Delta}^2(1)-P_{t+\Delta}^2(0))] \\
< \; & q_{t-1}^1 - q_{t-1+\Delta}^2 + (1-q_{t-1}^1)P_t^1(1) - (1-q_{t-1+\Delta}^2)P_{t+\Delta}^2(1) \quad (60) \\
\leq \; & q_{t-1}^1 - q_{t-1+\Delta}^2 + (1-q_{t-1}^1)P_{t+\Delta}^2(1) - (1-q_{t-1+\Delta}^2)P_{t+\Delta}^2(1) \quad (61) \\
= \; & (q_{t-1}^1 - q_{t-1+\Delta}^2)(1-P_{t+\Delta}^2(1)) \quad (62) \\
< \; & 0. \quad (63)
\end{aligned}$$

In the computations above, simple algebra shows that (60) is a consequence of the inequality in lines (54)-(55), and (61) is a consequence of (57). □

D.3 Proofs for Section 5.1

Proof of Proposition 10. As argued in the text, it is clearly not optimal to create rebates for sophisticated consumers.

Consider now naive decision makers. Suppose that when $v = v_1$, the optimal rebate policy is $p_T^*, r_T^* > 0$. This implies that $p_T^* - \theta\mu_T r_T^* > v_1$, where μ_T is the probability that the DM applies for the rebate; otherwise the profit maximizing offer would set $p_T = v_1$ and $r_T = 0$. When v is raised by some amount $\delta > 0$ to $v = v_1 + \delta$, it still follows that $p_T^* + \delta - \theta\mu_T r_T^* > v_1 + \delta$.

Note that a change in v will not change the DM's redemption probability conditional on purchasing, or the perceived value of the rebate. Thus if a positive rebate is better than no rebate for $v = v_1$, simply increasing the upfront price by δ when v is increased to $v_1 + \delta$ increases the profits by δ, and is likewise preferred to no rebate.

I now show that a rebate will be offered for a high enough v. For all $r_T > -\underline{\xi}$, the DM's redemption strategy will not depend on r. So for $r_T > -\underline{\xi}$, let $\nu_{e,T}$ denote the DM's expected redemption effort. Note, also, that for $r_T > -\underline{\xi}$, the DM's *perceived* probability of applying for the rebate is 1. Then for $r_T > -\underline{\xi}$, the DM's expected payoff conditional on offer p_T, r_T is $v - p_T + \theta(r - \nu_{e,T})$. Profit maximization implies that $p_T = v + \theta(r - \nu_{e,T})$, and the price floor assumption thus requires that

$$v + \theta(r - \nu_{e,T}) - r \geq \underline{p},$$

from which it follows that

$$r \leq \frac{v - \theta\nu_{e,T} - p}{1-\theta} \equiv n_T(v).$$

Note that $n_T(v)$ grows without bound as v approaches infinity.

Now let $\theta\mu_T$ denote the DM's actual probability of applying for the rebate when $r + \underline{\xi} > 0$. Note that μ_T does not depend on r in the region $r > -\underline{\xi}$. Note, also, that $\mu_T < 1$ by assumption that the DM is inattentive. Now when $n_T(v) > -\underline{\xi}$ is high enough and generates profit $v + \theta(r - \nu_{e,T}) - r \geq \underline{p}$, the monopolist's profit from setting $r_T = n_T(v)$ and $p_T = v + \theta(r - \nu_{e,T})$ will be

$$\begin{aligned}\pi_T &= v + \theta r - \theta\nu_{e,T} - \theta\mu_T r - c \\ &= v + \theta(1-\mu_T)n_T(v) - \theta\nu_{e,T} - c.\end{aligned}$$

But for high enough $n_T(v)$, it is clear that $v - c + \theta(1-\mu_T)n_T(v) - \theta\nu_{e,T} > v - c$. □

Proof of Proposition 11. As shown in the proof of Proposition 10, for a high enough $v = c - L$ the profit from offering a rebate $r = n_T(v)$ will be given by

$$\theta(1-\mu_T)n_T(v) - \theta\nu_{e,T} - L. \tag{64}$$

As before, μ_T is constant in r in the region $r = n_T(v) > -\underline{\xi}$. Moreover, $_T(v) > -\underline{\xi}$ for high enough v. Plainly, then, expression (64) is positive for a high enough $n_T(v)$. □

Proof of Proposition 12. Part 1 On the one hand, a longer deadline always increase the DM's *perceived* value of the rebate. On the other hand, if a longer deadline increases the redemption probability, then it also increases the cost associated with a rebate.

Part 2 As show in Proposition 8, for any $r > 0$, redemption probability will approach 1 as $T \to \infty$. But profits when $r > 0$ are

$$v + \theta(1-\mu_T)r - \theta\nu_{e,T} - c$$

which approach $v - c - \theta\nu_{e,T} < v - c$ as $\mu_T \to 1$. □

D.4 Proofs for Section 5.2

Proof of Proposition 13, part 1. Suppose that a fraction ϕ_t of consumers are attentive in period t. When a message is not sent in period $t+1$, rehearsal implies that the fraction of consumers who will be attentive that period will be

$$\phi_{t+1} = \phi_t\psi \tag{65}$$

, where $\psi = \ell g(1, a, 0) + (1-\ell)g(0, d, 0)$.

Now since all consumers are initially inattentive, the payoff from sending no messages is 0. Consider now the payoff from sending a single message in period 1, and no other messages. In period 1, a fraction w of consumers will be attentive. Equation (65) then implies that the fraction of consumers attentive in period t will be $\phi_t = \psi^{t-1}w$. The total payoffs, therefore, are

$$\sum_{t=1}^{\infty} \delta_o^t \psi^{t-1} w\ell - \delta c = \delta\frac{w\ell}{1-\delta_o\psi} - \delta c \tag{66}$$

Now the quantity in equation (66) is the highest payoff a single message can ever generate: If some consumers would already be attentive in the period that the message is sent, then the payoff from sending the message would be smaller. If messages are to be sent in the future, then the incremental payoff from sending the message in the current period is also diminished. □

Proof of Proposition 13, part 2. The fraction of attentive consumers after some number n of message will be $1-(1-w)^n$. Thus, the payoff of sending message $n+1$ in some period t will be

$$\sum_{\tau=t}^{\infty}[(1-w)^n-(1-w)^{n+1}]\delta_o^\tau-\delta_o^t c\ell=\delta_o^t[\left((1-w)^n-(1-w)^{n+1}\right)\frac{\ell}{1-\delta_o}-c] \tag{67}$$

Now $(1-w)^n-(1-w)^{n+1}\to 0$ as $n\to\infty$, and thus the quantity in equation (67) approaches 0 as $n\to 0$. Thus there exists some $n^\dagger$ such that sending more than $n>n^\dagger$ messages is not optimal.

The proof of the claim is completed by noting that the expression in (67) is positive if and only if $(1-w)^n-(1-w)^{n+1}\frac{\ell}{1-\delta_o}-c$. But in this case, the expression in (67) is clearly decreasing in t: sending a message earlier is always preferred to sending a message later. Thus the $n^\dagger$ message will be sent in the first $n^\dagger$ periods. □

Proof of Proposition 13, part 3. Suppose, by way of contradiction, that there is some $t^\dagger$ such that $m_{t^\dagger}=1$ but $m_t=0$ for all $t>t^\dagger$. Then the fraction of consumers attentive in period $t>t^\dagger$ will be $\phi_{t^\dagger}\psi^{t-t^\dagger}$.

If a message is sent in period $t>t^\dagger$, but not in any future periods, then the fraction of consumers attentive in period t will be $\phi_{t^\dagger}\psi^{t-t^\dagger}+(1-\phi_{t^\dagger}\psi^{t-t^\dagger})w$. Rehearsal implies that the fraction of consumers attentive in period $\tau>t$ will be

$$[\phi_{t^\dagger}\psi^{t-t^\dagger}+(1-\phi_{t^\dagger}\psi^{t-t^\dagger})w]\psi^{\tau-t}=\phi_{t^\dagger}\psi^{\tau-t^\dagger}+(1-\phi_{t^\dagger}\psi^{t-t^\dagger})w\psi^{\tau-t} \tag{68}$$

The net benefits of sending a message in period $t>t^\dagger$ are therefore

$$\sum_{\tau=t}^{\infty}(1-\phi_{t^\dagger}\psi^{t-t^\dagger})w\ell\psi^{\tau-t}-\delta^t c=\sum_{\tau=t}^{\infty}\delta^t\left[(1-\phi_{t^\dagger}\psi^{t-t^\dagger})\frac{w\ell}{1-\delta_o}-c\right] \tag{69}$$

Since $(1-\phi_{t^\dagger}\psi^{t-t^\dagger})\to 1$ as $t\to\infty$, it follows that the expression in equation (69) is positive for high enough t. □

Proof of Lemma 1. When $m_t=1$ each period, the fraction of attentive consumers in period $t+1$ is given by

$$\phi_{t+1}=(1-\phi_t)w+\phi_t w+\phi_t(1-w)\psi=w(1-\phi_t\psi)+\phi_t\psi. \tag{70}$$

Since $\phi_0=0$, simple algebra shows that the sequence $\{\phi_{t+1}\}$ defined in (70) is strictly increasing and converges to $\phi^*=\frac{w}{1-\psi(1-w)}$.

Now set $\phi_t=\phi^*-\epsilon_t$. By definition, $\epsilon_t\to 0$ as $t\to\infty$. Consider the payoffs from choosing $m_{t+1}=1$ rather than $m_{t+1}=0$. When in period $t+1$, this increases the fraction of attentive consumers by

$$\begin{aligned}\Delta_{t+1} &= [(1-(\phi^*-\epsilon_t)\psi)w+(\phi^*-\epsilon_t)\psi]-[(\phi^*-\epsilon_t)\psi]\\ &= w-w\psi(\phi^*-\epsilon_t)\\ &= \frac{w(1-\psi)}{1-\psi(1-w)}+w\psi\epsilon_t\end{aligned} \tag{71}$$

Now if $m_\tau = 1$ for all $\tau \geq t+2$, then the period $t+1$ increase translates into an additional $\Delta_{t+1}(1-w)\psi$ attentive consumers period $t+2$. Intuitively, this is because of those extra consumers who have been made attentive by the period $t+1$ message, $(1-w)\psi$ will not receive the period $t+2$ message and but stay attentive because of rehearsal. Thus, the period $t+1$ message increase the fraction of attentive consumers by $\Delta_{t+1}(1-w)\psi$ in period $t+2$. Following this logic, the period $t+1$ message increase the fraction of consumers attentive in period $\tau > t+1$ by $\Delta_{t+1}((1-w)\psi)^{\tau-t+1}$.

Thus the payoff from sending the message in period $t+1$ is

$$\ell \sum_{\tau=t+1}^{\infty} \Delta_{t+1}((1-w)\psi)^{\tau-t+1}\delta_o^\tau - c\delta^{t+1} = \delta^{t+1}\left[\frac{w(1-\psi)}{1-\psi(1-w)} + w\psi\epsilon_t - c\right] \tag{72}$$

Since ϵ_t approaches zero as $t \to \infty$, the quantity in equation (72) is positive for all t if and only if $\frac{w(1-\psi)}{1-\psi(1-w)} \geq c$. □

Proof of Lemma 2. Obvious. □

Proof of Proposition 14. When $\ell = 1/2$ and $w = 1$,

$$\frac{w(1-\psi)\ell}{(1-\psi(1-w))(1-\delta_o\psi(1-w))} = \frac{1}{2}\left(1-g(1,a,0)/2 - g(1,d,0)/2\right) > (1-g(1,d,0))/2.$$

Since the experession above is continuous in ℓ and w for all $g(1,a,0)$, there is some $w' < 1$ such that condition (11) holds in an open neighborhood of $\ell = 1/2$ for $w > w'$.

When $\ell = 1$ and $w = 1$, ψ approaches 1 as $g(1,a,0)$ approaches 1, and thus the expression in (11) approaches 0 as $g(1,a,0)$ approaches 1. Continuity implies that there are some $\bar{g} < 1$ and $w'' < 1$ such that condition (11) does not hold in an open neighborhood of $\ell = 1$ when $g(1,a,0) > \bar{g}$ and $w > w''$.

When $w > 2c$, condition (10) holds for all $\ell \geq 1/2$.

Thus choosing $g(1,a,0) > \bar{g}$ and $w > \max\{w', w'', 2c\}$ guarantees that there exist ℓ_2 and ℓ_3 such that condition (11) holds when $\ell \in (\ell_2, \ell_3)$, condition (11) does not hold when $\ell \in (\ell_3, 1)$, but condition (10) does hold when $\ell \in (\ell_3, 1)$.

Now fix $g(1,a,0) > \bar{g}$ and $w > \max\{w', w'', 2c\}$, and let ℓ_2 be the highest value of ℓ smaller than 1/2 for which condition(11) holds. Since

$$\frac{w\ell}{1-\delta_o\psi} > \frac{w(1-\psi)\ell}{(1-\psi(1-w))(1-\delta_o\psi(1-w))}$$

there exists some $\ell_1 < \ell_2$ such that condition (10) holds but condition (11) does not hold for $\ell \in (\ell_1, \ell_2)$. □

References

Acland, Dan and Matthew Levy, "Habit Formation, Naivete, and Projection Bias in Gym Attendance," *working paper*, 2011.

Agarwal, Sumit, John C. Driscoll, Xavie Gabaix, and David Laibson, "Learning in the Credit Card Market," *working paper*, 2013.

Allcott, Hunt and Todd Rogers, "The Short-Run and Long-Run Effects of Behavioral Interventions: Experimental Evidence from Energy Conservation," *NBER Working Paper 18492*, 2012.

Altmann, Steffen and Christian Traxler, "Nudges at the Dentist," *working paper*, 2012.

Andreoni, James and Charles Sprenger, "Estimating Time Preferences from Convex Budgets," *American Economic Review*, 2012, *102* (7), 3333–3356.

Ariely, Dan and Klaus Wertenbroch, "Procrastination, Deadlines, and Performance: Self-Control by Precommitment," *Psychological Science*, 2002, *13* (3), 219–224.

Armgstrong, April W., Alice J. Watson, Maryanne Makredes, Jason E. Frangos, Alexandra Kimball, and Joseph C. Kvedar, "Text-Message Reminders to Improve Sunscreen Use," *Archives of Dermatology*, 2009, *145* (11), 1230–1236.

Atkinson, Richard C. and Richard M. Shiffrin, "Storage and Retrieval Processes in Long-Term Memory," *Psychological Science*, 1969, *76* (2), 179–193.

Banks, Jefrey and Sridhar Moorthy, "A Model of Price Promotions with Consumer Search," *International Journal of Industrial Organization*, 1999, *17*, 371–398.

Becker, Gary S. and Kevin M. Murphy, "A Theory of Rational Addiction," *The Journal of Political Economy*, 1988, *96* (4), 675–700.

_ **and** _ , "A Simple Theory of Advertising as a Good or Bad," *Quarterly Journal of Economics*, 1993, *108* (4), 941–964.

Bernheim, Douglas B. and Antonio Rangel, "Addiction and Cue-Triggered Decision Processes," *The American Economic Review*, 2004, *94* (5), 1558–1590.

Bertrand, Marianne, Dean Karlan, Sendhil Mullainathan, Eldar Shafir, and Jonathan Zinman, "What's Advertising Content Worth? Evidence from a Consumer Credit Marketing Field Experiment," *Quarterly Journal of Economics*, 2010, *125* (1), 1–45.

Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer, "Salience and Consumer Choice," *working paper*, 2013.

Butters, Gerard R., "Equilibrium Distributions of Sales and Advertising Policies," *The Review of Economic Studies*, 1977, *44* (3), 465–491.

Cadena, Ximena and Antoinette Schoar, "Remembering to Pay? Reminders vs. Financial Incentives for Loan Repayments," *NBER working paper #17020*, 2011.

Calzolari, Giacomo and Mattia Nardotto, "Nudging with Information: A Randomized Field Experiment," *working paper*, 2012.

Camerer, Colin, "Progress in Behavioral Game Theory," *Journal of Economic Perspectives*, 1997, *11*, 167–188.

Cameron, Colin, Jonah Gelbach, and Douglas Miller, "Robust Inference with Multi-way Clustering," *Journal of Business and Economic Statistics*, 2011, *29* (2), 238–249.

Caplin, Andrew and Mark Dean, "The Behavioral Implications of Rational Inattention with Shannon Entropy," *working paper*, 2013.

_ **and** _ , "Rational Inattention and State Dependent Stochastic Choice," *working paper*, 2013.

Charles, Thomas, Dean Quinn, Mark Weatherall, Sarah Aldington, Richard Beasley, and Shaun Holt, "An Audiovisual Reminder Function Improves Adherence with Inhaled Corticosteroid Therapy in Asthma," *Journal of Allergy and Clinical Immunology*, 2007, *119* (4), 811–816.

Charness, Gary and Uri Gneezy, "Incentives to Exercise," *Econometrica*, 2009, *77* (3), 909–931.

Chen, Yuxin, Sridhar Moorthy, and Z. John Zhang, "Research Note–Price Discrimination After the Purchase: Rebates as State-Dependent Discounts," *Management Science*, 2005, *51* (7), 1131–1140.

Chun, Marvin M., Julie D. Golomb, and Nicholas B. Turk-Browne, "A Taxonomy of External and Internal Attention," *Annual Review of Psychology*, 2011, *62*, 73–101.

DeFrank, Jessica T., Barbara K. Rimer, Jennifer M. Gierisch, J. Michael Bowling, David Farrell, and Celette S. Skinner, "Impact of Mailed and Automated Telephone Reminders on Receipt of Repeat Mammograms," *American Journal of Preventive Medicine*, 2009, *36* (6), 259–467.

DellaVigna, Stefano and Ulrike Malmendier, "Contract Design and Self-Control: Theory and Evidence," *Quarterly Journal of Economics*, 2004, *CXIX* (2), 353–402.

Dismukes, R. Key, "Prospective Memory in Workplace and Everyday Situations," *Current Directions in Psychological Science*, 2012, *21* (4), 215–220.

Edwards, Matthew A., "The Law, Marketing and Behavioral Economics of Consumer Rebates," *Stanford Journal of Law, Business, and Finance*, 2007, *12* (362), 362–424.

— , "The Rebate Rip-Off: New Yorks Legislative Responses to Common Consumer Rebate Complaints," *Pace Law Review*, 2009, *29* (3), 471–478.

Eliaz, Kfir and Ran Spiegler, "Consideration Sets and Competitive Marketing," *Theoretical Economics*, 2011, *6*, 127–155.

— **and** — , "On the Strategic Use of Attention Grabbers," *Review of Economic Studies*, 2011, *78* (1), 235–262.

Ericson, Keith M. Marzilli, "On the Interaction of Memory and Procrastination," 2010, p. working paper.

— , "Forgetting we Forget: Overconfidence and Memory," *Journal of the European Economic Association*, 2011, *9* (1), 43–60.

Falkinger, Josef, "Attention Economies," *Journal of Economic Theory*, 2007, *133*, 266–294.

— , "Limited Attention as a Scarce Resource in Information-Rich Economies," *The Economic Journal*, 2008, *118*, 1596–1620.

Gabaix, Xavier, "A Sparsity-Based Model of Bounded Rationality, with Application to Basic Consumer and Equilibrium Theory," *working paper*, 2012.

— **and David Laibson**, "Bounded Rationality and Directed Cognition," *working paper*, 2005.

— **and** — , "Shrouded Attributes, Consumer Myopia and Information Supression in Competitive Markets," *Quarterly Journal of Economics*, 2006, *121* (2), 505–540.

Gerstner, Eitan and James D. Hess, "A Theory of Channel Price Promotions," *The American Economic Review*, 1991, *81* (4), 872–886.

Gilbert, Ben and Joshua S. Graff Zivin, "Dynamic Salience with Intermittent Billing: Evidence from Smart Electricity Meters," *NBER Working Paper No. 19510*, 2013.

Gilpatric, Scott M., "Slippage in Rebate Programs and Present-Biased Preferences," *Marketing Science*, 2009, *28* (2), 229–238.

Grubb, Mchael D., "Consumer Inattention and Bill-Shock Regulation," *Working Paper*, 2012.

Haapala, Irja, Noel C. Barengo, Simon Biggs, Leena Surakka, and Pirjo Manninen, "Weight Loss by Mobile Phone: A 1-year Effectiveness Study," *Public Health Nutrition*, 2012, *12*, 2382–2391.

Hay, Janine F. and Larry L. Jacoby, "Separating Habit and Recollection: Memory Slips, Process Dissociations, and Probability Matching," *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 1996, *22* (6), 1323–1335.

Hefti, Andreas M., "Attention and Competition," *working paper*, 2013.

Heidhues, Paul, Botond Köszegi, and Takeshi Murooka, "Deception and Consumer Protection in Competitive Markets," in "Anthology on Pros and Cons of Consumer Protection" 2011.

_ , _ , and _ , "Exploitative Innovation," *working paper*, 2012.

_ , _ , and _ , "Inferior Products and Profitable Deception," *working paper*, 2012.

Henderson, Conor M., Joshua T. Beck, and Robert W. Palmatier, "Review of the Theoretical Underpinnings of Loyalty Programs," *Journal of Consumer Psychology*, 2011, *21*, 256–276.

Holman, Jeff and Farhan Zaidi, "The Economics of Prospective Memory," *working paper*, 2010.

Insel, Kathleen C., Gilles O. Einstein, Daniel G. Morrow, and Joseph T. Hepworth, "A Multifaceted Prospective Memory Intervention to Improve Medication Adherence: Design of a Randomized Control Trial," *Contemporary Clinical Trials*, 2013, *34*, 45–52.

Jacoby, Larry L., James A. Debner, and Janine F. Hay, "Proactive Interference, Accessibility Bias, and Process Dissociations: Valid Subjective Reports of Memory," *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 2001, *27* (3), 686–700.

Janakiraman, Narayan and Lisa Ordóñez, "Effect of Effort and Deadlines on Consumer Product Returns," *Journal of Consumer Psychology*, 2012, *22*, 260–271.

Jessoe, Katrina and David Rapson, "Knowledge is (Less) Power: Experimental Evidence from Residential Energy Use," *American Economic Review*, forthcoming.

Karlan, Dean, Margaret McConnell, Sendhil Mullainathan, and Jonathan Zinman, "Getting to the Top of Mind: How Reminders Increase Saving," *working paper*, 2012.

_ , Melanie Morten, and Jonathan Zinman, "A personal touch: Text messaging for loan repayment," 2012, p. working paper.

Köszegi, Botond and Adam Szeidl, "A Model of Focusing in Economic Choice," *Quarterly Journal of Economics*, 2013, *128* (1), 53–107.

Krishna, Santosh, Suzanne Austin Boren, and E. Andrew Balas, "Healthcare via Cell Phones: A Systematic Review," *Telemedicine and e-Health*, 2009, *15* (3), 231–240.

Laibson, David, "Golden Eggs and Hyperbolic Discounting," *Quarterly Journal of Economics*, 1997, *112* (2), 443–477.

_ , "A Cue-Theory of Consumption," *Quarterly Journal of Economics*, 2001, *116* (1), 81–119.

Letzler, Robert and Joshua Tasoff, "Everyone Believes in Redemption: Nudges and Overoptimism in Costly Task Completion," *working paper*, 2013.

Loewenstein, George, Ted O'Donoghue, and Matthew Rabin, "Projection Bias in Predicting Future Utility," *Quarterly Journal of Economics*, 2003, *118* (4), 1209–1248.

Lu, Qiang and Sridhar Moorthy, "Coupons Versus Rebates," *Marketing Science*, 2007, *26* (1), 67–82.

Lynch, John G. and Gal Zauberman, "When do you Want it? Time, Decisions, and Public Policy," *American Marketing Association*, 2006, *25* (1), 67–78.

MacDonell, Karen, Sylvie Naar-King, Heather Huszti, and Marvin Belzer, "Barriers to Medication Adherence in Behaviorally and Perinatally Infected Youth Living with HIV," *AIDS Behavior*, 2013, *17* (1), 86–93.

Martin, Benjamin A., Noelle L. Brown, and Jason L. Hicks, "Ongoing Task Delays Affect Prospective Memory More Powerfully Than Filler Task Delays," *Canadian Journal of Experimental Psychology*, 2011, *65* (1), 48–56.

McBride, Dawn M., Jaclyn K. Beckner, and Drew H. Abney, "Effects of Delay of Prospective Memory Cues in an Ongoing Task on Prospective Memory Task Performance," *Memory & Cognition*, 2011, *39* (7), 1222–1231.

_ , Jennifer H. Coane, Jason Drwal, and Stephanie-Ann M. LaRose, "Differential Effects of Delay on Time-based Prospective Memory," *Aging, Neuropsychology, and Cognition*, in press.

Mitra, Anusree and John G. Lynch., "Toward a Reconciliation of Market Power and Information Theories of Advertising Effects on Price Elasticity," *Journal of Consumer Research*, 1995, *21* (4), 644–659.

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

Narasimhan, Chakrvarthi, "A Price Discrimination Theory of Coupons," *Marketing Science*, 1984, *3* (2), 128–147.

Nedungadi, Prakash, "Recall and Consumer Consideration Sets: Influencing Choice without Altering Brand Evaluations," *Journal of Consumer Research*, 1990, *17* (3), 263–276.

O'Donoghue, Ted and Matthew Rabin, "Doing it Now or Later," *American Economic Review*, 1999, *89* (1), 103–124.

Patrick, Kevin, Fred Raab, Marc A. Adams, Lindsay Dillon, Marian Zabinski, Cheryl L. Rock, William G. Griswold, and Gregory J. Norman, "A Text Message-Based Intervention for Weight Loss: Randomized Controlled Trial," *Journal of Medical Internet Research*, 2009, *11.*

Peng, Lin and Wei Xiong, "Investor attention, overcondence and category learning," *Journal of Financial Economics*, 2006, *80*, 563–602.

Poquette, Amelia J., David J. Moore, Ben Gouaux, Erin E. Morgan, Igor Grant, and Steven P. Woods, "Prospective Memory and Antiretroviral Medication Non-Adherence in HIV: An Analysis of Ongoing Task Delay Length Using the Memory for Intentions Screening Test," *Journal of the International Neuropsychological Society*, 2013, *19*, 155–161.

Rankin, Catharine H., ..., and Richard F. Thompson, "Habituation Revisited: An Updated and Revised Description of the Behavioral Characteristics of Habituation," *Neurobiology of Learning and Memory*, 2009, *92*, 135–138.

Reis, Ricardo, "Inattentive Consumers," *Journal of Monetary Economics*, 2006, *53*, 1761–1800.

Schwartzstein, Josh, "Selective Attention and Learning," *working paper*, 2012.

Sellen, A.J., G. Louie, J.E. Harris, and A.J. Wilkins, "What Brings Intentions to Mind?," *Memory*, 1997, *5* (4), 483–507.

Shafir, Eldar and Amost Tversky, "Choice Under Conflict: The Dynamics of Deferred Decision," *Psychological Science*, 1992, *3*, 358–361.

Shu, Suzanne B. and Ayelet Gneezy, "Procrastination of Enjoyable Experiences," *Journal of Marketing Research*, 2010, *47*, 933–944.

Silk, Tim and Chris Janiszewski, "Managing Mail-in Rebate Promotions," 2008.

Silk, Timothy, "Examining Purchase and non-Redemption of Mail-in Rebates: The Impact of Offer Variableson Consumers' Subjective and Objective Probability of Redeeming," 2004.

Sims, Christopher A., "Implications of Rational Inattention," *Journal of Monetary Economics*, 2003, *50* (3), 665–690.

Sovern, Jeff, "Toward a New Model of Consumer Protection: The Problem of Inflated Transaction Costs," *William and Mary Law Review*, 2006, *47* (5), 1635–1709.

Spiegler, Ran, "Competitive Framing," *American Economic Journal: Microeconomics*, forthcoming.

Stigler, George J., "The Economics of Information," *Journal of Political Economy*, 1961, *69* (3), 213–225.

Szilagyi, Peter C., Clayton Bordley, Julie C. Vann, Ann Chelminski, Ronald M. Kraus, Peter A. Margolis, and Lance E. Rodewald, "Effect of Patient Reminder/Recall Interventions on Immunization Rates," *Journal of the Americal Medical Association*, 2000, *284* (14), 1820–1827.

Tobias, Robert, "Changing Behavior by Memory Aids: A Social Psychological Model of Prospective Memory and Habit Development Tested With Dynamic Field Data," *Psychological Review*, 2009, *116*, 408–438.

Vervloet, Marcia, Annemiek J. Linn, Julia C. M. van Weert, Dinny H. de Bakker, Marcel L. Bouvy, and Liset van Dijk, "The Effictiveness of Interventions Using Electronic Reminders to Improve Adherence to Chronic Medication: A systematic Review of the Literature," *Journal of the Americal Medical Informatics Association*, 2012, *19*, 696–704.

Villas-Boas, J. Miguel, "Predicting Advertising Pulsing Policies in an Oligopoly: A Model and Empirical Test," *Marketing Science*, 1993, *12* (1), 88–102.

Vyankandondera, Joseph, Kirstin Mitchell, Brenda Asiimwe-Kateera, Kimberly Boer, Philippe Mutwa, Jean-Paul Balinda, Masja van Straten, Peter Reiss, and Janneke van de Wijgert, "Antiretroviral therapy drug adherence in Rwanda: Perspectives from patients and healthcare workers using a mixed-methods approach," *AIDS care*, 2013, *25* (12), 1504–1512.

Woodford, Michael, "Inattentive Valuation and Reference-dependent Choice," 2012.

Woodruff, Mandi, "Genius PhD Student Paid His Tuition With Mail-In-Rebates," September 2012. http://www.businessinsider.com/a-genius-phd-student-paid-his-tuition-almost-entirely-with-mail-in-rebates-2012-9.

Yantis, Steven, "Control of Visual Attention," in H. Pashler, ed., *Attention*, The Guilford Press, 1998, pp. 223–156.

Zogg, Jennifer B., Steven P. Woods, John A. Sauceda, John S. Wiebe, and Jane M. Simoni, "The Role of Prospective Memory in Medication Adherence: A Review of an Emergin Literature," *Journal of Behavioral Medicine*, 2012, *35* (1), 47–62.