Emotional Inattention

Lukas Bolte* Collin Raymond*

February 7, 2022

PRELIMINARY DRAFT

Abstract

We introduce a model of emotional inattention. A decision-maker (DM) decides how much attention to allocate across decisions. Attention to a decision i) increases material payoffs, and ii) leads to an emotional response, increasing the weight of that decision in her total payoff. The cost of attention is thus endogenous and depends on the relative payoffs of the decisions. An emotionally inattentive DM exhibits the ostrich effect and avoids devoting attention to low-payoff decisions. She also exhibits excessively volatile levels of attention and can be caught in attention traps, where cognitive scarcity along with dynamic inconsistency leads her to be inattentive to welfare-improving actions. Standard interventions to improve decision-making, such as the provision of free information or penalties for mistakes, can in fact worsen decision-making. In a consumption-savings application, the DM is shown to react asymmetrically to income shocks immediately adjusting to good and at times ignoring bad ones. In a portfolio choice problem, the DM avoids assets that require re-adjustments to information (an attention premium), which can lead to an excessive premium for risky assets.

JEL Classification Codes: D81, D83

Keywords: Attention, emotional response, information

*Lukas Bolte; Department of Economics, Stanford University, Stanford, CA; lbolte@stanford.edu. Collin Raymond; Department of Economics, Purdue University, West Lafayette, IN; collinbraymond@gmail.com. We thank Douglas Bernheim, Gabriel Carroll and Muriel Niederle, and seminar participants at Stanford University for helpful comments. All errors are our own.

> *[. . .] we have at least two distinct roles for our minds at play, that of the information processing and reasoning machine by which we choose what to consume out of the array of things that our resources can be exchanged for, and that of the pleasure machine or consuming organ, the generator of direct consumer satisfaction.*
>
> Schelling (1988)

1 Introduction

An individual's mental state is not only affected by contemporaneous consumption, but also her anticipation of future or remembrance of past consumption. The latter, all consumption outside the "here and now," is not directly felt by the individual; for it to affect the current mental state, the individual thus needs to devote attention to it. Attention serves as the medium via which noncontemporaneous consumption leads to contemporaneous flow utility. Thus, in addition to attention improving future payoffs (e.g., by allowing information acquisition), the decision to allocate attention affects contemporaneous utility flows. For example, by not devoting attention to a negative future consumption (in other words ignoring it), an individual can minimize negative anticipatory feelings so that "out of sight" really is "out of mind." Instead, she may focus her attention on positive future consumption, or reminisce in the past, elevating current utility.

This implies that individuals may not devote attention to negative situations, and avoid information that improves the material payoffs associated with them (they "bury their head in the sand"). Evidence for such "ostrich effect" behavior is well-documented. In the domain of health, researcher have noted low rates of testing for serious medical illnesses such as Huntington's disease (Shouldson and Young, 2011; Oster et al., 2013). Ganguly and Tasoff (2017) experimentally document that the demand for medical testing for sexually transmitted diseases is decreasing as the expected health outcome worsens. Moreover, avoidance of attention to one's health is seen as a motivating factor when individuals (often) fail to follow medical recommendations, e.g., taking medicine or self-screening for symptoms (Becker and Mainman, 1975; Sherbourne et al., 1992; Lindberg and Wellisch, 2001; DiMattero et al., 2007).

Related evidence has also been found in the domain of personal finance. Karlsson et al. (2009) and Sicherman et al. (2015) show that investors' propensity to check their portfolios comoves with the market (both with market levels and changes). A down market (for most investors) implies low future consumption. By decreasing attention to their portfolio, investors can minimize the effect of the associated decrease in anticipatory utility. Strikingly, Sicherman et al. (2015) find a strong positive correlation between market returns and the frequency of investors logging in twice during a single weekend, when markets are closed and no new information is revealed.

In this paper, we introduce the emotional inattention model, at the heart of which is attention and its dual role as: 1) an aggregator of noncontemporaneous consumption, and 2) an input in improving material payoffs. We show how our model captures a novel form of attentional costs. It can not only match the previously mentioned data, but generates a number of new predictions about the effects of attentional scarcity, dynamic inconsistency, and intrinsic preferences for learning. We also show that it can help resolve a number of puzzles in consumption-saving and portfolio environments. Our approach generates distinct predictions compared to models where the costs of attention are exogenous (i.e., rational inattention), where anticipatory emotions are experienced independent of the allocation of attention. Our model also generates distinct predictions compared to a few related models that also think about attention in the context of anticipation (i.e., Tasoff and Madarasz (2009); Karlsson et al. (2009); Golman and Loewenstein (2018)).

The central starting point of the emotional inattention model is that attention is, at least partially, directed by the DM and not, e.g., entirely determined by exogenous stimuli, a premise referred to as "top-down" attention. While this assumption is standard in, for example, the rational inattention literature (as in Sims (2003), a seminal paper), it warrants some discussion. The aforementioned evidence of information, or more broadly attention, avoidance in some, and excessive attention in other situations, suggests attention as a choice, or equivalently be affected by the level of material payoffs. Furthermore, individuals seem to be aware of how their ability to freely allocate attention: Falk and Zimmermann (2016) find that information about an electrical shock is viewed differently depending on whether individuals have a distracting task available.

A more subtle, but equally important, premise is that the two consequences of attention—generating flow utility from the level of material payoffs and improving them—are inherently linked together. Alternatively, the DM could devote more targeted attention and, e.g., acquire materially beneficial information without contemplating the level of material payoffs. The aforementioned evidence suggests that, in some situations, individuals cannot completely compartmentalize the consequences of attention. This tight linkage is referred to as "bottom-up" attention.

In Section 2, we begin by sketching out the basic structure of our model. A DM faces two situations with (noncontemporaneous) consumption: one with some underlying decision-problem ("the decision") and a another trivial one with fixed material payoff ("the outside option"). In each of two periods, the DM decides how much attention to devote to the decision and how much to the outside option. In a final period, consumption takes places and the DM receives the material payoffs from the decision and the outside option. We assume that the material payoff from the decision is increasing in the amount of attention paid, but make no further substantive assumptions. Thus, we take no stand on why attention is a useful input in the underlying decision-problem; we believe it could happen for a variety of reasons and may vary by environment, including that attention allows an individual to think about the decision and the optimal action (e.g., via mental simulations of potential outcomes), or reduce the noise attached to their choice of action (as in random choice models), or to move away from a default action. The second key ingredient is that attention is the aggregator of noncontemporaneous consumption into flow utility: We operationalize this by assuming that attention to the decision or the outside option generates current flow utility proportional to the associated material payoffs.

We consider attention allocation over time for our model to capture important applications such as consumption-savings and portfolio allocation studied in Section 6. Furthermore, as we show, the DM is generally time-inconsistent which has important implications.

We frame the exposition with all consumption taking place after the DM allocates attention. We thus refer to the flow utility attention generates as anticipatory utility. However, in principle, some of the consumption could also take place in the past, e.g., attention to the outside option can correspond to reminiscing pleasant memories from the past.

We then study the implications of the model working backwards from the second period taking the initial allocation of attention as given (Section 3). In Section 3.1, we show that our model (not surprisingly) generates the asymmetric ostrich effect: Individuals do not attend to situations with low material payoffs (even if attention increases the payoff), but are happy to engage when when material payoffs are high. As eluded to previously, much empirical evidence from various domains is consistent with this prediction.

In Section 3.2, we formally show that the two consequences of attention, creating anticipatory utility and improving material payoffs, complement each other. For example, if a particular situation receives a lot of weight in the DM's flow utility of noncontemporaneous consumption, then increasing the material payoff is very beneficial; conversely, if the material payoff of a situation is high, then the DM wants it to take a large weight in her flow payoff from anticipatory utility. This complementarity can lead to volatile behavior, small changes in the environment, salience of a situation, or attentional bandwidth, can lead to drastic changes in attention and decision-making quality. While total payoff, i.e., including anticipatory utility, is continuous, we highlight a connection to poverty traps (Mullainathan and Shafir, 2013a): a threshold for attentional bandwidth can exist that governs whether individuals have high or low material payoffs. Furthermore, if utility from eventual consumption is discounted due to impatience while anticipatory utility is consumed contemporanesouly, and the individuals act too impatiently as suggested by Caplin and Leahy (2004), then such poverty traps can be costly.

In Section 4, we turn to understanding behavior in the first period. This allows us to focus on the dynamic implications of the emotional inattention. In Section 4.1, we note that the DM is generally time-inconsistent: The DM's future self devotes too little attention to the decision relative to the perspective of the current self; this is because the future self does not internalize the beneficial effect of future attention on current anticipatory utility. We show that this time-inconsistency has not bite, if the DM can achieve the optimum with commitment by devoting attention in only one period. However, when the DM's optimal attention strategy (with commitment) involves attention to the decision in both periods, then a lack of commitment can lead to her not devoting any attention and strictly worse total payoffs in both periods. This occurs in situations where it is beneficial for the future self to devote attention to the decision only when considering the current self's anticipatory utility. Without commitment, the future self then reduces her attention which in term worsen the anticipatory utility of the current self which may then also not devote attention. Indeed, there is a form of complementarity between attention devoted to the decision across periods. We show how this complementarity can lead, again, to attentional volatility where small changes in, e.g., attentional bandwidth, can lead lead to drastic changes in attention. In contrast to our results on attentional volatility in Section 3.2, this dynamic attentional volatility imply that small changes in attentional bandwidths can make both current and future self discontinuously worse off.

Section 4.2 discusses how emotional inattention generates intrinsic preferences for the timing of learning. In particular, an emotionally inattentive DM have an intrinsic preference for devoting attention (and so learning) earlier rather than later, even when there are no observable actions. This is because the DM anticipates in the future exhibiting an asymmetric ostrich effect—she will devote different amounts of attention depending on whether she learns the decision will generate a high or low material payoff. This generates a hidden action on the part of the DM, and so a preference for early learning. Such a preference for early learning is consistent with experimental evidence on intrinsic preferences for information (e.g., Masatlioglu et al., 2017 and Nielsen, 2020).

In Section 5 we take as a jumping off point the fact that increasing attention in the second period can be a good thing: It improves material payoffs (all else being equal), which, as eluded to earlier, may be optimal for the DM, and also helps overcome the commitment issues highlighted in Section 4.1. As we show, otherwise innocuous policies may in fact fail to achieve their goal if emotional inattention is taken into account.

We focus on two policies. First, we study the effect of increasing the instrumental return of attention by increasing the difference in material payoffs between good and bad actions in the underlying decision-problem. Without the effect of attention on anticipatory utility, such intervention increases the value of devoting attention to the decision; however, e.g., penalizing the DM for taking a bad action in the underlying decision-problem may sufficiently lower her expected material payoff so that she shies away and stops devoting attention. Our result imply that the DM will often avoid commitment contracts which impose penalties in the case of poor decisions: the future self's potential reduction in attention due to looming penalties leads the (sophisticated) present self to not impose such penalties in the first (although she would take up commitment devices that offered solely rewards). Second, we consider what happens when the DM is provided with free information (or more generally a technology that increase the material payoff of the decision given the DM's attention). While information provision improves the material payoff of the decision (with fixed attention), it can serve as a substitute for the DM's attention. In particular, we show that if the provided information and the DM's attention are strong enough substitutes, then the DM reduces her attention to a level so that her decision-making in fact worsens.

We study the implications of our model in economically important applications (Section 6). In Section 6.1, we apply our model to the consumption and saving decision of a DM. The DM decides on a consumption profile before she knows her income. Once she receives her (random) income, she can either stick with the consumption profile with future consumption now lower due to reduced savings, or devote devote attention to the decision and re-optimize her consumption. The DM responds to income shocks asymmetrically: she always reacts to high income realizations by increasing her consumption, while she may ignore low income realizations. Furthermore, she anticipates such asymmetry and chooses her initially planned consumption level accordingly, i.e., suited for exactly these low-income realizations for which she does not devote attention.

We next, in Section 6.2, study the portfolio choice of an emotionally inattentive DM. Here, we highlight the dynamic effects of emotional inattention on investment strategies. We show, using two extreme cases, that there are competing effects of emotional inattention on risk-taking. First, because emotional inattention generates option value of investing in the risky asset, the DM seems more risk loving compared to standard measures of risk preferences. The reason is that a varied payoff allows her to focus on good realizations and ignore bad ones, which makes her payoff more convex than that of a standard DM—the same mechanism that leads the DM to prefer early learning. Second, because risky assets may require more attention than safe assets due to a need to re-optimize after learning about potential returns, the DM, rationally anticipating their later ostrich behavior, will under-invest in the risky asset.

Section 7 discusses how emotional inattention differs from several other classes of models to which it is closely related. We consider in turn models of rational inattention, anticipatory emotions, and reference-dependence preferences. We then discuss how our approach compares to a small set of other papers which try to compare attention as well as utility and attention.

Section 8 concludes.

2 Model

A decision-maker (she, and henceforth DM) faces two future situations, both involving consumption. One situation has an underlying decision-problem determining consumption and is referred to as "the decision." The other situation is trivial in the sense that its consumption, and thus the material payoff associated with the situation, is fixed. This situation is referred to as "the outside option."

In period t, one of two periods, i.e., $t \in \{1, 2\}$, the DM allocates attention across the situations; in a final period, consumption takes place. Let $\alpha_t \in A_t \subseteq [0, 1]$ denote the share of attention (or just attention) devoted to the decision, with the implication that $1 - \alpha_t$ attention is devoted to the outside option. We assume that A_t is compact for all $t \in \{1, 2\}$. Via the set A_t, we can impose restrictions on which attention allocations are admissible. For example, salience is capture by imposing a minimum amount of attention devoted to, say, the decision, i.e., $A_t = [\underaccent{\bar}{C}_t, 1]$ with $\underaccent{\bar}{C}_t > 0$, whereas a limited cognitive bandwidth corresponds to $A_t = [0, \bar{C}_t]$ with $\bar{C}_t < 1$.

Attention has two effects. First, as is a standard feature of attention in many economic models, it improves decision-making. Specifically, by devoting attention to the decision, the associated level of consumption, i.e., its material payoff from the decision, increases (in expectation). As we do not explicitly model the underlying decision-problem of the decision, we allow for a variety of reasons of why attention improves the material payoff. For example, attention could allow the DM to think about the decision and optimal actions in the underlying problem (e.g., via mental simulations of potential outcomes), to reduce the noise attached to their choice of such actions (as in random choice models), to move away from a default action, or, perhaps most standard, to extract signals from an information source. In Section 6, we explicitly model two underlying decision-problems: a consumption-savings problem (Section 6.1), in which attention allows the DM to reoptimize her consumption path, and a portfolio choice problem (Section 6.2), where attention enables her to alter her portfolio.

The second effect of attention is that it aggregates noncontemporaneous consumption into contemporaneous flow utility. Specifically, in period t, no contemporaneous consumption takes place; yet, by allocating attention α_t, the DM receives anticipatory utility from the decision and the outside option, proportional to α_t and $1-\alpha_t$, respectively. The DM evaluates the anticipatory utility using the final material payoffs, i.e., material payoffs evaluated after accounting for attention as an input.

In period t, the DM places a weight τ_t on anticipatory utility and the weight on material payoffs in the final period is normalized to unity. These weights allow us to capture individuals who vary in how much they experience anticipatory utility, as well as other concerns like discounting (e.g., the individual may weight, from period-1's perspective, current anticipatory utility more strongly than future anticipatory utility).[1]

The timing is as follows. In period 1, the DM decides how much attention $\alpha_1 \in A_1$ to devote to the decision. As mentioned previously, attention improves decision-making. Additionally, attention may resolve some uncertainty pertaining to the material payoff from the decision even when there is no action involved, i.e., the DM simply learns whether consumption will be high or low. Thus, in period 2, the DM's attention allocation problem is parameterized by some random variable ϵ that encapsulates the resolved randomness. This variable is distributed according to some distribution F, that depends on attention in period 1. The DM then decides again how much attention, now denoted by $\alpha_2 \in A_2$, to devote to the decision, where α_2 may optimally depend on the resolved uncertainty, ϵ.

Given ϵ, we denote the material payoff from the decision when the DM devotes attention α_2 by $V^D(\alpha_2|\epsilon)$; note that benefits from period-1 attention are captured by ϵ. We assume that V^D is continuous in α_2.[2]

The fixed material payoff from the outside option is denoted by $\underline{U}$.

[1] Specifically, discounting of future payoffs are captured by fixing a time invariant weight τ on anticipatory utility and discount factor δ and letting $\tau_1 = \frac{1}{\delta}\tau, \tau_2 = \tau$. Similarly, our formulation captures discounting of anticipatory utility if consumption is far into the future (Loewenstein, 1987). We do, however, impose the restriction that the weights on anticipatory utility are time invariant.

[2] Continuity of the material payoff from the decision will ensure that the DM's optimization problem in periods 1 and 2 admits a solution. If A_1 and A_2 are finite, continuity is a vacuous assumption.

Figure 1 provides an overview of the weights, choices and utility in each period.

weights	τ_1	τ_2	1
choices	$\alpha_1 \in A_1$	$\alpha_2 \in A_2$	none
utility	$\alpha_1 E_{\epsilon \sim F(\alpha_1)}[V^D(\alpha_2(\epsilon)\vert\epsilon)]$ $+(1-\alpha_1)\underline{U}$	$\alpha_2 V^D(\alpha_2(\epsilon)\vert\epsilon)]$ $+(1-\alpha_2)\underline{U}$	$V^D(\alpha_2(\epsilon)\vert\epsilon)]$ $+\underline{U}$
time period	$t=1$	$t=2$	final period

Figure 1: Overview of the decisions and payoffs.

Lastly, in period t, the DM maximizes her weighted sum of current and future utilities. Thus, the objective of the DM is period 2 (for each realized ϵ) is given by

$$\begin{aligned} V_2(\alpha_2|\epsilon), &= \overbrace{([V^D(\alpha_2|\epsilon) + \underline{U}]}^{\text{Consumption Utility}} + \underbrace{\tau_2 \alpha_2 V^D(\alpha_2|\epsilon)}_{\text{Decision Anticipatory Utility}} + \overbrace{\tau_2(1-\alpha_2)\underline{U}}^{\text{Outside Option Anticipatory Utility}} \\ &= V^D(\alpha_2|\epsilon)[\tau_2\alpha_2 + 1] + \underline{U}[\tau_2(1-\alpha_2)+1]. \end{aligned} \tag{1}$$

A solution to this problem exists as A_2 is compact and V^D continuous; if the DM in period 2 is indifferent between two attention levels, we assume she chooses the one she prefers from period-1 perspective. Denote the solution to this problem as $\alpha_2(\epsilon)$. In period 1, the DM maximizes

$$V_1(\alpha_1, \alpha_2(\epsilon)) = E_{\epsilon \sim F(\alpha_1)}[V^D(\alpha_2(\epsilon)|\epsilon)\tau_1\alpha_1 + \underline{U}(1-\alpha_1)\tau_1 + V_2(\alpha_2(\epsilon)|\epsilon)]. \tag{2}$$

When $\tau_1 = \tau_2 = 0$, the model collapses to that with a "standard" DM who optimally devotes attention to maximize material payoff, here, to the decision; as $\tau \to \infty$ the DM becomes less and less worried about material payoffs.

We return to a particular parameterization which can be interpreted as resulting from a particular underlying decision-problem associated with the decision: the DM has a normally distributed prior about a state of the world that she tries to match. The DM receives a signal about the state the precision of which is increasing in attention. Let $\alpha = \alpha_1 + \alpha_2$, be the total attention paid over two periods. The material payoff also includes a constant term, l (the "level"), and is thus given by

$$V^D(\alpha) = -\gamma \frac{1}{\beta + \psi\alpha} + l. \tag{3}$$

3 Static analysis

We begin our analysis by focusing on behavior in period 2. As already noted, a solution to the DM's problem exists.

Lemma 1. *For every* ϵ, $\operatorname{argmax}_{\alpha_2 \in A_2} V_2(\alpha_2|\epsilon)$ *exists.*

All proofs are omitted from the main text and instead given in Appendix A.

As for every realized uncertainty ϵ the DM solves an essentially static problem, in order to declutter notation, we will drop the period-2 subscript, and explicit references to ϵ.

The optimal level of attention need not be unique. When we formulate results on the effects of parameter changes on attention, phrases such as "attention is increasing" are to mean comparative statics in the strong set order unless noted otherwise.

3.1 Ostrich effect

The first key behavior that our model generates is an "ostrich effect." The ostrich effect refers to individuals avoiding thinking about unpleasant news or situations by burying their figurative heads in the sand.[3]

In our model, the DM devotes attention to situations whose material payoff is high relative to the outside option. In other words, she does not devote attention, or ignores, situations with very low material payoffs—the ostrich effect. The proposition below formally documents this effect. Furthermore, a DM who puts greater weight on anticipatory utility, will devote more attention to situations with high material payoff, holding fixed the material benefit of doing so.

Fix some increasing function $\tilde{V}^D$ and let $V^D(\alpha_2) = \tilde{V}^D(\alpha) + l$, where l is some constant, the "level payoff from the decision."

Proposition 1 (Ostrich effect)**.** *Attention to the decision:*

1. *increases in the level payoff from the decision; and*

[3] Although the idea of ostriches burying their heads in the sand to avoid predators dates back to Roman times, they actually do not display this behavior. Instead, they put their heads into their nests (which are built on the ground) in order to check temperatures and rotate eggs.

2. decreases in the material payoff from the outside option.

In particular, for every $\tilde{V}^D$ that is differentiable at 0, there exists l such that the DM devotes no attention to the decision.

Furthermore, if $\max_{\alpha \in A} V^D(\alpha) < \underline{U}$, then attention to the decision decreases in the weight on anticipatory utility, τ.

3.2 Complementarity of the two consequences of attention and volatility

The ostrich effect highlights how our model predicts that the optimal amount of attention will vary with the material payoff attachd to the decision. We now turn to highlighting the fact that our model predicts attention can change dramatically with small changes in the environment.

The two consequences of attention are complementary: Higher levels of attention generate higher material payoffs, which increases the value of anticipatory utility. This implies that as long as the material payoff function from the decision as a function of attention, V^D, is not "too" concave, the total payoff function is a convex function of attention. Throughout the rest of this sub-section we will assume V^D is twice differentiable. Similar results hold without the assumption, but conditions are more complicated to state.

Lemma 2. *Suppose V^D is twice differentiable. V is strictly convex on $\alpha \in A$ if and only if*

$$\frac{2\tau}{\tau\alpha + 1} > -\frac{\frac{\partial^2}{\partial\alpha^2} V^D(\alpha)}{\frac{\partial}{\partial\alpha} V^D(\alpha)} \text{ for all } \alpha \in A.$$

Convexity in the total payoff function implies that the DM's optimal attention is a boundary point—she should either devote no or full attention to the decision. In these circumstances the DM may exhibit attentional volatility—small changes in the decision problem can lead to large observed changes in behavior and hence quality of decision-making.

We first demonstrate the result in the context of small changes in the (relative) level payoffs associated with the two situations. Note that the convexity of V is independent of of l.

Proposition 2 (Attentional volatility as material payoffs change)**.** *Let V_l denote the objective given level payoff l from the decision. If V_l is strictly convex, then there exists l^* such that for all $l < l^*$, the DM devotes the minimum attention to the decision, and for all $l > l^*$, the DM devotes the maximum attention.*

We next turn to investigating another source of attentional volatility: the fact that the DM may have bounds on the attention they can devote to the the decision. We presume that $A = [0, C]$; analogous results occur when there is a lower bound on attention. This bound potentially represents the fact that the DM must pay at least some attention to the outside option, perhaps due to salience, reminders outside their control, or outside distractors.[4]

It is relatively clear that decreasing C will weakly decrease the chosen level of attention to the decision. However, when the total payoff function is strictly convex and nonmonotone in attention, we may observe the DM exhibiting attentional volatility. In other words, a small change in the available attentional bandwidth leads to dramatic changes in attention allocation and subsequently in observable material payoffs.

Proposition 3 (Attentional volatility as the bandwidth changes)**.** *The optimal level of attention to the decision is weakly increasing in C. Moreover, suppose that V is strictly convex and nonmonotone with $V(1) > V(0)$. Then there exists a C^* such that for all $C \geq C^*$, the DM chooses attention level $\alpha = C$, and for all $C < C^*$, the DM chooses attention level $\alpha = 0$.*

In those situations, i.e., when the total payoff is strictly convex and nonmonotone, the optimal level of attention may jump from 0 to its highest admissable level, C as C increase. The conditions for this to happen are relatively mild. We already described when total payoff is strictly convex in attention.

Nonmonotonicity is given when small amounts of attention do not improve the material payoff from the decision by much, relative to the potential "cost" from a decrease in anticipatory utility. Thus, we require that $V^D(\alpha) < \underline{U}$, for some α, say $\alpha = 0$, so that the DM prefers anticipatory utility to come from the outside option, and τ to be large enough so that she puts sufficient weight on anticipatory utility.

We provide two examples to illustrate attentional volatility. First, consider the left panel in Figure 4, which shows the DM's total payoff (1) as a function of α. Even though V^D is concave, as it corresponds to a parameterization of (3), total payoff is strictly convex in α. Consequently, the DM's optimal level of attention is a boundary point. Small changes in the attentional bandwidth on α can thus cause attention devoted to the decision to drop to 0. In the figure, we can see that for bandwidths above the blue line, the DM would choose the maximum amount of attention to

[4]It could also represent a reduced form way of capturing third situation that also require attention. Let $\hat{\alpha}$ denote the attention devoted to said third situation. Then for any fixed level of attention $\hat{\alpha}$ devoted to the third situation, the DM allocates $C \equiv 1 - \hat{\alpha}$ between the decision and the outside option. Changes in C can then be understood as changing the parameters of the third situation so that the DM changes the amount of attention devoted to it.

devote to the decision, while for bandwidths less than below it, the DM would choose to devote no attention to the decision.

(a) V_2^D is given by (3) with $\beta = 10, \gamma = 2000, l = 87.5, \underline{U} = 10, \tau_2 = 0.8, \psi = 13$.

(b) V_2^D is given by (3) with $\beta = 10, \gamma = 2000, l = 87.5, \underline{U} = 20, \tau_2 = 0.8, \psi = 13$.

Figure 2

Similarly, shifts in the minimum of attention the DM needs to devote to the decision can cause her to switch from only devoting the minimum level of attention to devoting all her attention to it. The right panel of Figure 4 provides an example; we can see that for minimal bounds below the blue line, the DM would choose the minimal amount of attention to devote to the decision, while for minimal bounds above it, the DM would choose to devote full attention to the decision.

4 Dynamic analysis

We now turn to understanding behavior in the first period, and show how period 1 behavior is both directly affected by flow anticipatory utility in period 1, as well as the anticipated period 2 behavior.

The optimization problems faced in period 1 and in period 2 are related: in period 2, without commitment, the DM maximizes (2) with $\tau_1 = 0$. We first confirm that under mild conditions there exists a pair of attention levels that solves the dynamic optimization problem.

Lemma 3. *Suppose V^D is continuous in ϵ and α_2. There exists an $\alpha_1, \alpha_2(\cdot)$ that maximizes V.*[5]

In order to sharpen our results in this section, we will discuss two particular environments. First, we discuss environments where the material payoff of the payoffs do not depend on what is learned in period 1.

[5]Our continuity requirement is sufficient but not necessary for the existence of a solution, and in fact a solution may exist even with discontinuous payoffs.

Definition 1. *The environment is realization-independent if for all* α_1, α_2*,*

$$V^D(\alpha_2|\epsilon)$$

is constant across $\epsilon \in \text{support}(F(\alpha_1))$*. (With slight abuse of notation) we write* $V^D(\alpha_1, \alpha_2)$ *as the material payoff of the decision, taking* α_1 *as given in period 2.*

Examples of realization-independent environments are those where $F(\alpha_1)$ is degenerate, i.e., there is no uncertainty, where attention (deterministically) increases the available choices (e.g., allow the DM to deviate from a default). However, realization-independent environments can include uncertainty: When the DM tries to match a normally distributed state and receives a normally distributed signal as in equation (3), the variance of her posterior is independent of the signal realization and so is her material payoff.

A second special case is when the the DM has no action to take for the decision. Such a setting allows us to focus on the intrinsic psychological value of attention, as opposed to the instrumental value of attention (in terms of changing actions). In particular, there is no instrumental value of information if the DM cannot condition an observed action on the realization of the information.

Definition 2. *There is no instrumental value of attention if* $V^D(\alpha_2(\epsilon)|\epsilon)$ *is is constant in* $\alpha_2(\cdot)$ *and* $E_{\epsilon \sim F(\alpha_1)}[V^D(\alpha_2(\epsilon)|\epsilon)]$ *is constant in* α_1 *for every* $\alpha_2(\cdot)$*.*

An example of an environment in which there is no instrumental value of attention is one where the DM's material payoff is determined by a random state but she does not take any action or make any choice. By devoting attention to the decision, the DM may still learn about the material payoff.

4.1 Attention traps and time inconsistency

Because, as the DM moves forward through time, they are no longer concerned about the anticipatory utility of past selves and may be time inconsistent. In particular, without commitment, in period 2, she will choose some $\alpha_2^*(\epsilon)$ that maximizes $V_2(\cdot|\epsilon)$; with commitment, she maximizes (2), without this constraint. We thus refer to period-1 and period-2 selves.

In general, period-1 self prefers her period-2 self to devote more attention to the decision that her period-2 self does: An increase in period 2 attention leads to an increase in anticipatory utility in period 1, an effect that period-2 self does not internalize.[6]

Lemma 4. *Fix any α_1 and ϵ. The optimal α_2 chosen by period-2 self is less than that period-1 self would choose (i.e., with commitment).*

The time inconsistency, in particular the fact that period-2 self will devote lower levels of attention than preferred by period-1 self, will lead to period-1 self also altering her behavior. Because period-1 self anticipates that next period she will not devote as much as attention, relative to what period-1 self would commit to, the material return to devoting additional attention (the marginal value of and increase in α_1) changes, and the anticipatory utility from devoting attention worsens (because less attention will be devoted in period 2). Both of these effects can cause period-1 self to readjust her attention level (relative to commitment).

We can thus speak of two solution concepts. The first is the commitment solution: this is optimal α_1, α_2 combination that maximizes V_1. The second is the sub-game perfect equilibrium outcome (we refer to it as the equilibrium outcome) which derived via backwards induction. For each α_1, period-2 self chooses α_2 to maximize V_2. Anticipating period 2's behavior, period-1 self chooses α_1 to maximize V_1.

We now turn to understand how period-1 self readjusts her attention in the face of the commitment problem, when doing so is effective, and in what situations the commitment problem has severe consequences. In order to study time inconsistency, we assume that without commitment period-2 self, when indifferent, chooses the attention level that period-1 self prefers.

The following definitions define two special attentional strategies; where all attention is paid in a single period (either period 1 or period 2).

Definition 3. *An attention pair $(\alpha_1, \alpha_2(\cdot))$ has* frontloading *if $\alpha_2(\epsilon) = 0$ for all $\epsilon \in \text{support}(F(\alpha_1))$, and* backloading *if $\alpha_1 = 0$. If $(\alpha_1, \alpha_2(\cdot))$ further solves $\max_{(\alpha_1,\alpha_2(\cdot))} V_1(\alpha_1, \alpha_2(\cdot))$, we then say $(\alpha_1, \alpha_2(\cdot))$ is a* frontloading solution *or* backloading solution*, respectively.*

The next result points out that time inconsistency is only a problem for period-1 self when she cannot find solution where she devotes attention only in a single period. In other words, issues with dynamic inconsistency arise because the individual needs to coordinate attention levels across different periods.

[6]The DM may also want to adjust the timing of consumption to increase or decrease the time of anticipation. This can also lead to time inconsistency (or "reverse time inconsistency") but is studied elsewhere (Loewenstein, 1987).

Proposition 4. *Suppose V^D is continuously differentiable and with commitment always full attention in period-2 is not optimal. Period-1 self is strictly worse off without commitment than with commitment if and only if with commitment there does not exist a front- or backloading solution.*

Of course, period-1 self is worse off when the commitment outcome cannot be achieved. Because period-2 self places less weight on the material payoff of α_2 than period-1 self, she is better off when she can deviate from period-1 self's commitment solution, fixing the attention devoted by period-1 self. However, because period-1 self anticipates period-2 self's deviation from the commitment solution, α_1 will not be chosen at the commitment level. If period-1 self either increases, or does not reduce her attention too much, relative to the commitment level of attention, then period-2 self benefits from the lack of commitment. However, if period-1 self reduces her attention by too much, in anticipation of period-2 self's behavior, then period-2 self will be worse off relative to the commitment solution.

In order for both selves to be worse off relative to the commitment solutions, we need both period selves to pay less attention than they would in the commitment solution. The next proposition provides a striking example of this. In particular, it provides sufficient conditions so that we can find a situation where the unique equilibrium outcome is (0,0), while the commitment solution is to provide strictly positive amounts of attention in both periods.

We call such an example an "attention trap." This is because we demonstrate that if the DM had enough attention capacity in either period 1 or in period 2, they could achieve the commitment outcome by either frontloading or backloading their attention. In contrast, if attention is limited enough in both period 1 and in period 2, then frontloading or backloading is can no longer generate a high enough payoff to make either strategy optimal. Thus, the lack of having enough attention in any given period traps the individual in a low payoff (for both period's selves) equilibrium (i.e., paying no attention in either period). In contrast, with additional attentional capacity in either period, the DM could achieve the commitment payoff for both selves.

The conditions in order for this to occur are relatively mild. In addition to technical smoothness conditions, we require it to be the case that i) if the DM could devote enough attention in period 1, it would be worth thinking about the decision, and ii) the DM places sufficient weight on anticipatory utility in each period.

Proposition 5. *Suppose V^D is continuously differentiable with derivative bounded away from zero everywhere. Suppose $\underline{U} = V^D(1, 0)$. For any τ_1 large enough, there exists τ_2 and bounds on attention*

(C_1, C_2) *so that both selves prefer the commitment solution to the solution without commitment, which features no attention.*

Our attentional traps represent a dynamic extension of the attentional volatility we found in Section 3.2. Whenever the conditions of Proposition 5 are satisfied, we may observe threshold effects, where, if attention in different periods falls just below a particular level, then attention drops discontinuously in all periods. In other words, we can find a situation where, if the DM had a little more attention in both periods, their attentional strategy would discontinuously change; moving from (0,0) to an attentional strategy featuring strictly positive attentional values. However, our results are even stronger than the attentional volatility discussed in the previous section. In our single-period analysis, attentional volatility causes behavior (i.e., attention) and material payoffs to jump discontinouosly. However, overall utility still smoothly moves with changes in attentional capacity. In contrast, in the dynamic analysis, changes in attentional capacity can not only discontinuously change behavior, but also discontinuously change utilities.

Proposition 6. *Suppose the conditions of Proposition 5 hold and V^D is not too concave in α_2, and V^D has not too decreasing differences in (α_1, α_2). For any τ_1 large enough, there exists τ_2 and values (C_1, C_2) so that for all $\delta > 0$, i) attention bounds $(C_1 - \delta, C_2 - \delta)$ lead to the unique equilibrium solution (0,0); ii) attention bounds $(C_1 + \delta, C_2 + \delta)$ lead to a unique equilibrium solution of $(C_1 + \delta, C_2 + \delta)$, and iii) utility payoffs are discontinuous for both selves.*

Attention traps can occur in less extreme ways, in situations in which the lack of commitment does not lead to a drop in attention all the way to no attention. Consider a realization-independent environment with $\tau_1 = \tau_2 = 0.1$, $\underline{U}$ normalized to zero and

$$V^D(\alpha_1, \alpha_2) = l + \sqrt{\alpha_1} + \sqrt{\alpha_2},$$

with $l = -7$. Period-1 self's objective, $V_1(\alpha_1, \alpha_2)$, is maximized by equalizing attention across periods; and plotted as a function of attention in Figure 3(a).

Consider period-1 self's total payoff under commitment. Evidently, it is concave and maximized at $\alpha_1 = \alpha_2 = 1$ $(= \alpha^c)$. In the absence of commitment, Lemma 4 suggests that period-2 self may want to deviate from this prescribed level of attention devoted to the decision; indeed, Figure 3(b) shows that period-2 self's total payoff, $V_2(\alpha_2|\alpha^c)$, is maximized at a lower level of attention, $\alpha_2(\alpha^c) < \alpha^c$. Figure 3(c) plots period-2 self's optimal level of attention devoted to the decision as a

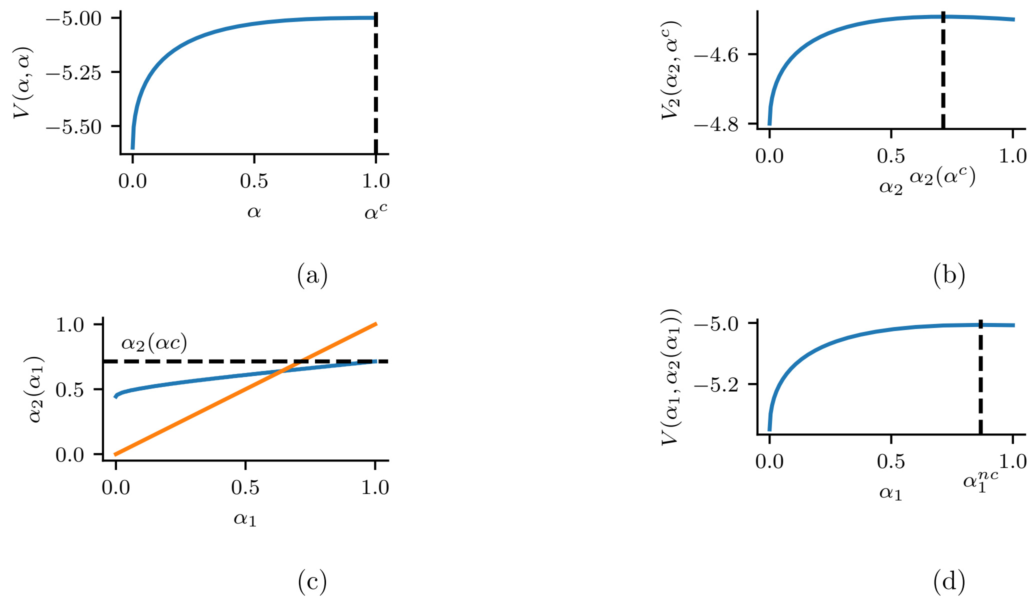

Figure 3

function of period-1 self's attention (blue line). Again, we can see that $\alpha_2(\alpha^c) < \alpha^c$ as $(\alpha^c, \alpha_2(\alpha^c))$ lies below the 45-degree line (orange line). Period-1 self takes period-2 self's optimal response into account and maximizes her total payoff as a function of α_1 as depicted in Figure 3(d). Without commitment, period-1 self then chooses a lower level of attention, α_1^{nc}, than with commitment. However, she does not stop devoting any attention.

Period-1 self is obviously worse off without commitment. Here, period-2 self is also worse off: the loss in total payoff due to the decrease in attention by period-1 self outweighs the benefit from flexibly choosing own attention. In particular, period-2 self's payoff with commitment is -4.5 and ≈ -4.55 without. In other words, the lack of commitment makes both selves worse off.

An emotional inattentive DM thus demands commitment against falling into attention traps, or more generally being time-inconsistent; e.g., she could set automatic reminders that make the decision salient to her at a future date.

Thus, our dynamic extension can shed some light on issues surrounding poverty traps. In particular, even tasks that can ultimately generate large returns in helping individuals escape poverty may not be undertaken, because they require attention spread out over many periods. Due to dynamic inconsistency, a DM may simply not take on these tasks at all. These results consistent with the threshold evidence of Balboni et al. (2021) for scarcity in physical assets. Building on Mullainathan and Shafir (2013b)'s point, it is not just scarce cognitive resources that act as a constraint, but the inability to coordinate the use of those resources over time. Thus, either simplifying tasks so they can be accomplished in one period (thus eliminating the need for temporal coordination as in Proposition 4), assisting a DM with commitment devices, or providing them with enough resources to accomplish a task quickly, can all potentially serve to help individuals escape poverty.

4.2 Early vs late attention

It is well known that when individuals have non-linear anticipatory utility they can have "non-standard" preferences towards information. For example, a classic question is to what extent individuals have preferences over information even when they cannot take any actions (e.g., Kreps and Porteus, 1978; Epstein and Zin, 1989). Experimental evidence has been strongly consistent with the hypothesis that individuals have a preference for earlier resolution of information (see Masatlioglu et al., 2017 and Nielsen, 2020 for recent papers exploring this).

In our setting, we can ask an equivalent question. Because information is endogenously generated via attention, we can ask when it is the case that a DM wants to acquire information earlier, rather than later, even when the information does not influence any action they are taking?[7]

In order to focus on intrinsic reasons for informational preferences we consider environments where there is no instrumental value of information. Our environment not only rules out information being instrumentally valuable after period 1 but also instrumentally valuable after period 2. This eliminates a classic motivation for the timing of learning in dynamic models: with costly experimentation, there is a benefit to sequential learning (i.e., learning some in period 1, and then conditioning learning in period 2 on the outcome of information in period 1) — e.g., in the setting of Moscarini and Smith (2001). In these settings, the DM would benefit from devoting some attention early (in period 1), with the option of devoting further attention late (in period 2).

Instead, our setting allows us to focus on a second, distinct form of informational preferences: The DM can not only condition her attention devoted to the decision in period 2 on how much she learned in period 1, but also on how it shifted her belief about the eventual payoff up or down: If the payoff is likely to be high, the DM continues to devote attention to the decision; otherwise, she reduces attention. In particular, upon observing different values for the material payoff in period 1,

[7] A distinct question asks about the timing of consumption — does the DM prefer consumption to occur in period 1 or period 2. If anticipatory utility drives the emotional response, then the DM may prefer to delay consumption of pleasurable goods while consuming unpleasant goods immediately as to vary the duration of anticipatory feelings (see Loewenstein (1987)).

the DM may allocate her attention in period 2 depending on such values because of the emotional focus: she may devote more attention to the decision if the material payoff turns out to be high.

Proposition 7.

- *For every* $\underline{U}$ *and* l*, there exists* V^D *with no instrumental value of attention and expected payoff equal to* l*, so that the DM devotes all her attention to the decision in period 1,* $\alpha_1 = 1$*.*
- *If* $V^D(0|\epsilon)$ *is not a constant with probability* $\delta > 0$ *when* $\epsilon \sim F(\alpha_1')$ *for some* α_1'*, and there is no instrumental value of attention, then there exists* $\underline{U}$ *greater than the expected payoff from the decision such that* $\alpha_1 > 0$*.*

Classic models of anticipatory utility generate a preference for early resolution of information via an assumption that utility is convex in anticipatory utility (e.g., Kreps and Porteus, 1978; Caplin and Leahy, 2001). In contrast, anticipatory utility enters total utility in a linear fashion in our model. Despite this, we have a preference for early attention. This is because a DM has hidden actions (e.g., Ergin and Sarver, 2015). Because of this, information generates option value for the DM, and so the payoff function is convex in information, even without observable actions.

Allowing for instrumental value of attention does not change the substance of the proposition. For example, suppose the DM can acquire information, say a single signal, in either period by devoting at least $\underline{\alpha}$ of attention to the decision of (she can always devote more attention and attention in the period without information). Then the DM prefers the signal to be available in period 1.

5 Interventions

As the previous propositions makes clear, emotionally inattentive individuals engage in strategies often ignore information that could be materially beneficial, and which could potentially increase payoffs to all periods' selves.

Such issues raise the question of whether there are potential interventions that could take place. These may arise from policymakers who want to improve the material payoffs of individuals. Or they could emerge because past selves may try to motivate their future selves to pay attention via a commitment device.[8]

[8]Because our individuals are dynamically inconsistent, they may have a desire to change future selves actions. Similarly, a welfare-based justification can be made if consumption utility from the tasks is discounted due to impatience and anticipatory utility consumed contemporaneously. In this case, a benevolent policymaker should be more patient (Caplin and Leahy, 2004), i.e., put more weight on the material payoff.

As our results show, some natural interventions may, perhaps unexpectedly, not work, i.e., the treated population may end up with less total information. The reason for these surprising results is the complementarity between the two roles of attention (increasing material payoffs and causing an emotional response) in determining the effect of policies. In particular, although the response of a standard DM to some intervention is a function of its effect on the curvature of the material payoff function only, an emotionally inattentive DM also reacts to changes in the level of material payoff.

Increasing the value of information

The first intervention we consider is whether changing the payoff function in period 2 may help period-2 self acquire more information. For example, the DM, in the past, could have taken some action that changes how their payoff depends on the information acquired today. Or it could be that a policymaker increased the reward of making good decisions or the penalty of making poor ones. To keep the analysis simple, we suppose that such interventions occur at the end of period 1, after anticipatory utility in that period has been realized.

For a standard DM, increased rewards for good decisions or the penalties for poor decisions should generate equivalent effects: The return of information (attention) increases so that she devotes more attention to the decision. To an emotionally inattentive DM, however, these two “feel” rather different: increasing rewards raises the overall material payoff of the decision, and the anticipatory utility from devoting attention increases; increasing penalties has the opposite effect. From our discussions of the ostrich effect in Section 3.1, we expect the DM’s attention response to differ.

For simplicity, we drop the period-2 subscripts. In order to formalize these notions we can write the material payoff of the decision equivalently as

$$V(\alpha) = V^D(0) + \underbrace{[V^D(\alpha) - V^D(0)]}_{\text{reward}}, \quad \text{or} \quad V(\alpha) = V^D(1) - \underbrace{[V^D(1) - V^D(\alpha)]}_{\text{penalty}}.$$

More generally, fix l (the *level*) and $\tilde{V}^D(\alpha)$, and consider $V^D(\alpha) = l + \gamma \tilde{V}^D(\alpha)$ for $\gamma > 0$. If $\tilde{V}^D(\alpha) \geq 0$, increasing γ corresponds to increasing the reward, and if $\tilde{V}^D(\alpha) \leq 0$, increasing γ corresponds to increasing the penalty.

Increasing the reward will always lead to an increase in attention. Note that in both cases, the return of the value of information increases with γ so that a standard DM would devote more attention. However, the changes in material payoff differ: for a fixed level of attention, the material payoff increases when increasing the reward, and decreases when increasing the penalty making the effects on the emotional response drastically different.

Proposition 8. *If $\tilde{V}^D \geq 0$ (reward), then attention devoted to the decision is increasing in γ.*

If $\tilde{V}^D \leq 0$ (penalty), then attention devoted to the decision is decreasing (increasing) in γ if

$$\tilde{V}^D(\alpha)[(1-\tau)+\tau\alpha] \tag{4}$$

is decreasing (increasing) in α.

Thus, the effect of increasing penalties on attention depends on how $\tilde{V}^D(\alpha)[(1-\tau)+\tau\alpha]$ changes with attention. The following result provides some sufficient conditions for penalties to either decrease attention, or increase attention.

Proposition 9. (4) *is decreasing in α if $\tilde{V}^D$ is semi-differentiable and at least one of the two following conditions is satisfied*

1. *$\tilde{V}^D(1)$ is low enough and $\tau > 0$;*
2. *τ is large enough ($\tau < 1$) and $\tilde{V}^D(1) < 0$.*

In contrast, (4) *is increasing in α if τ small enough and there exists $C > 0$ so that for all α', α with $\alpha' > \alpha$,*

$$\frac{\tilde{V}^D(\alpha') - \tilde{V}^D(\alpha)}{\alpha' - \alpha} \geq C.$$

Some evidence, such as Böheim et al. (2019) and Ariely et al. (2009) indicates that individuals "choke," i.e., perform worse, when the stakes go up. This is clearly at odds with standard rational models. However, it can be consistent with our model, where increased stakes increase the downside of getting a decision wrong, and so reduce attention (ostrich effect), and overall performance. Of course, key to our mechanism is that the increased stakes reduce the worst payoffs, not just increase the best payoffs. If only the latter were true, then we would expect increased performance. Thus, our model may also be able to rationalize why we observe choking in some environments but not others.

Our model also suggests that the effects of penalties and rewards depend on whether the individual can reallocate her attention. A laboratory experiment consisting of a decision problem where subjects are treated with either increased penalties or rewards may provide dramatically different results depending on whether attention is held constant or not.

Providing Information

Another way to try to try to increase the total amount of information acquired is for more information to be acquired prior to period 2. Of course if the period 1 does this prior to receiving anticipatory utility, this imposes an emotional cost. But, as we will show, even if additional information could be acquired at 0 informational cost, it still may not increase the total amount of information acquired. Understanding the effects of such actions are often important because policymakers often try to force individuals to acquire information, typically in order to improve material payoffs. We can study how their later endogenous information acquisition reacts to such policies. We find that in the presence of emotional inattention, providing additional information prior to period 2 does not necessarily increase the total amount of information acquired — i.e. the sum of information acquired prior to period 2 and in period 2. Thus, acquiring additional information prior to period 2 can lead to less total information at the end of period 2.

Consider the following policy: before the beginning of period 2, the DM is provided with some free information, β, that she is forced to attend to (this could be e.g., because the DM acquired it in the previous period, or it was forced on them by a social planner). This information is equivalent to what the DM would acquire by devoting β attention to the decision. The information increases the weight on the material payoff of decision in the DM's anticipatory utility by $\beta' < \beta$, that is less attention is required to attain a certain level of material payoff.

We focus on situations where payoffs depend on the sum of attention in period 2, and the previously acquired free information: i.e., attention α given provided information β leads to material payoff of $V^D(\alpha + \beta)$ (similar, albeit more complex statements can be formulated for more general payoff functions).[9] Furthermore, the weight of the anticipatory utility of the decision given attention α is given by $\beta' + \alpha$ (and $\alpha \in [0, 1 - \beta']$) for some $\beta' < \beta$, i.e., the extracting the "free" information requires some attention.

There are two mechanical reasons why such information policy information may lead to higher material payoff: 1) the DM's maximum material payoff potentially increases (see Section 3.2 for a more general discussion of extreme attention allocations), and 2) the DM may be forced to attend

[9] For example, if β represents previously acquired information in period 1, then this is equivalent to V^D is a function of the sum of attentions in two periods.

to more information than she initially does. However, in the absence of these two reasons, total information, $\alpha + \beta$, and thus material payoff decreases.

Proposition 10. *Denote the optimal level of attention without the information policy by α^* and with the information policy by $\alpha^*(\beta)$ and suppose both are unique.*

If $\beta < \alpha^*$ and $\alpha^*(\beta) + \beta \leq 1$, then

$$\alpha^*(\beta) + \beta \leq \alpha^*.$$

The key driver behind the aforementioned result is a complementarity between increasing the material payoff and the anticipatory utility. Here, suppose the DM responds to the free information by reducing attention to exactly offset β. In this case, the cost of devoting attention is the same as the material payoff of the decision is unchanged. However, this occurs at a lower level of attention (reduced by exactly $\beta - \beta'$), and thus a lower weight on the material payoff of the decision. Hence, the DM may in fact further decrease attention.

Our result is thus a cautionary tale. Information provision is not an innocuous intervention that at worst will not have any effect on material payoffs. Instead, informational intervention can decrease material payoffs.

6 Applications

6.1 A consumption-savings model

We study the consumption and savings behavior of a DM in the presence of emotional inattention. There are two consumption periods, $t = 1, 2$. At the beginning of period 1, the DM needs to decide how much to consume in period 1 which is denoted by c_d (for 'default'). *After* making this decision, she receives income y which is distributed according to some continuous distribution F.[10] She can then either devote attention to the income, revise her consumption, and consume $c_1^*(y)$ in period 1 and $y - c_1^*(y)$ in period 2, or she can devote attention to an outside option that gives a payoff of $\underline{U}$ in which case she consumes c_d in period 1 and $y - c_d$ in period 2. The DM is risk-averse and equipped with a CARA within-period utility function denote by u. Let y_{inf} denote the infimum of the support of y. c_d is restricted to be at most y_{inf} so that the DM need not devote attention to consume in period 2. Given her risk-aversion, $c_1^*(y) = y/2$.

[10] Continuity is not a substantive assumption but simplifies the statement of the results.

Her total payoff if she devotes attention is given by

$$2u(\frac{y}{2}) + \underline{U}[1-\tau],$$

and if she does not devote attention

$$(u(c_d) + u(y - c_d))[1-\tau] + \underline{U}.$$

For a given y, she pays attention to her consumption if and only if the former exceeds the latter.

Let y_{sup} denote the supremum of the support of y and suppose that $2u(\frac{y_{\text{sup}}}{2}) > \underline{U}$ so that the DM pays attention for some income shocks.

Proposition 11. *Given any c_d, there exist $\underline{y}, \bar{y}$, so that the DM devotes attention if $y < \underline{y}$ or $y > \bar{y}$, does not devote attention if $y \in (\underline{y}, \bar{y})$ and is indifferent otherwise.*

When $\underline{U}$ is large enough, then for any c_d, $\underline{y} = -\infty$, i.e., the DM never pays attention for low-income realization, but may otherwise be finite. In contrast, $\bar{y}$ is always finite.

When $\underline{y} = \infty$, the DM responds *asymmetrically* to income shocks: low-income shocks are ignored and she does not adjust her planned income, whereas she pays attention to positive income shocks and increases from her planned consumption.

The DM anticipates this and has a precautionary savings motive. Let c_{na} (for 'no adjustment') denote her optimal level of consumption in period 1 if the DM could not readjust her consumption after observing y. c_{na} solves:

$$u'(c_{\text{na}}) = E[u'(y - c_{\text{na}})],$$

where u' is the first derivative of u.

When $\underline{y} = -\infty$ and the DM can adjust consumption by devoting attention to it, the DM chooses a (tentative) consumption level c_d that is at most c_{na}. c_d solves:

$$\int_{y=-\infty}^{\bar{y}} u'(c_d)dF = \int_{y=-\infty}^{\bar{y}} u'(y - c_d)dF.$$

Thus, an emotionally inattentive DM may exhibit several potential biases simultaneously. First, ex-ante pessimism, because she wants to ensure that she does not lose too much utility in states she fails to pay attention to. Second, she exhibits a default effect — for some states of the world the DM fails to adjust her plan. Third, she exhibits ex-post optimism for the worst states — because she fails to pay attention for the worst realizations and overconsumes when times are bad. (She may also overconsume for income realizations just below $\bar{y}$.)

Within this setting, we can consider the effects of changing τ (while remaining in the regime with $\underline{y} = -\infty$). We focus on situations where $\tau > 0$ because when $\tau = 0$ the solution for c_1 is set-valued.

Lemma 5. *Suppose $\underline{y} = -\infty$ and consider an increase in $\tau \in (0, 1)$. Then both $\bar{y}$ and c_1 increase.*

In other words, a DM who puts more weight on anticipatory utility will avoid paying attention for more income realizations. Consequently, she will choose a consumption level that is more suitable for those higher-income realizations.

6.2 Portfolio choice

We next study the effects of emotional inattention on the DM's portfolio choice in an asset allocation problem. As previously mentioned, it is well known that individuals avoid checking their investment accounts when markets are down (Karlsson et al., 2009; Sicherman et al., 2015). We extend our results to show how anticipation of such behavior can have important impacts on investment behavior. The asset allocation problem is kept as simple as possible. Formally, it is an example of the dynamic model introduced in Section 2: here, the material payoff of the decision is the consumption value of the DM's final wealth, and attention is required to (re)optimize her portfolio.

In period 1, the DM makes an initial portfolio allocation (which requires $\alpha_1 = 1$); specifically, she decides how much of her wealth W_1 to invest in a risky asset, $I_1 \in [0, W_1]$, with the remainder, $W_1 - I_1$, saved via a safe asset. The investment in the risky asset generates a return so that her total wealth, after the first period, is given by $W_2 = W_1 + I_1 r_1$, where $r_1 \geq -1$ is the return of the risky asset in period 1. The DM also receives information about future returns, the return of the risky asset in the second period. This information is denoted by $\mathcal{N}_1$ and may be correlated with the return of the risky asset in the first period, i.e., $(r_1, \mathcal{N}_1) \equiv \epsilon \sim F_1$.

Given the return of the risky asset r_1 and the information about future returns $\mathcal{N}_1$, the DM then decides whether to devote attention to her portfolio and reoptimize ($\alpha_2 = 1$) or to not devote attention and leave her portfolio as is (in which case $\alpha_2 = 0$ is optimal). Specifically, in the latter case, her investment in the risky asset is given by $I_2 = I_1(1 + r_1)$. When reoptimizing, she chooses $I_2 \in [0, W_2]$. The return $r_2 \sim F_2(\mathcal{N}_1)$, with $r_2 \geq -1$, is realized, and the DM consumes

$W_{\text{final}} = W_1 + I_1 r_1 + I_2 r_2$, which, in utils, she values as $u(W_{\text{final}})$.

To avoid trivialities, both the standard and the emotionally inattentive DM, always devote attention in the first period, $\alpha_1 = 1$. For simplicity, we further set $\tau_1 = 0$ and denote $\tau_2 = \tau$ so that the commitment and no-commitment solutions coincide.

Starting with period 2, let $V_2^D(\mathcal{N}_1, W_2, I_2) = E_{r_2 \sim F(\mathcal{N}_1)}[u(W_2 + I_2 r_2)]$ and $I_2^* \in \operatorname{argmax}_{I_2 \in [0, W_2]} V_2^D(\mathcal{N}_1, W_2, I_2)$. In period 2, if the DM learned information $\mathcal{N}_1$, has wealth W_2, and without reoptimizing invests I_2, she devotes attention, that is $\alpha_2 = 1$, if

$$V_2^D(\mathcal{N}_1, W_2, I_2^*) + \underline{U}[1-\tau] > V_2^D(\mathcal{N}_1, W_2, I_2)[1-\tau] + \underline{U}. \tag{5}$$

As $\tau_1 = 0$, period-1 self then chooses I_1 as to maximize

$$E_{(r_1, \mathcal{N}_1) \sim F_1}[\max\{V_2^D(\mathcal{N}_1, W_2, I_2^*) + \underline{U}[1-\tau], V_2^D(\mathcal{N}_1, W_2, I_2)[1-\tau] + \underline{U}\}].$$

Our formal results highlight two competing effects. First, the DM likes risk if there is no need to reoptimize her portfolio. Essentially, the risky asset allows the DM to condition her attention in period 2 on whether the return in period 1: if the return is high, the DM increases the weight on consumption by devoting attention to her portfolio, otherwise, she shifts weight to the outside option. Not devoting attention in when the return in period 1 is low may pose an issue if there are material benefits from reoptimizing her portfolio. When there is "no need" to reoptimize, however, the emotionally inattentive DM always invests more in the risky asset than the standard DM.[11]

Proposition 12. *Suppose $r_2 = 0$ (so that the risky and safe assets have the same return in period 2) for all $\mathcal{N}_1$ and u is strictly concave and strictly increasing. Then I_1 is increasing in τ.*

The previous result demonstrates (using an extreme case) that emotional inattention generates excessive preferences for risk. In order to highlight the second effect, we focus on an alternative extreme case. We suppose that the DM may need to reoptimize, for example as she learns about the likely return of the risky asset in period 2. To isolate the effects of potentially needing to reoptimize, we set $r_1 \geq 0$ so that the DM is not intrinsically risk-seeking due to emotional attention as in Proposition 12.

In general, the DM reoptimizes if (5) holds. To simplify the already cluttered analysis, we

[11] Making the risky asset riskless in period 2 gets rid of the need to reoptimize. Alternatively, the need to reoptimize can be eliminated if the DM's asset allocation automatically readjusts via some technology. In other words, the DM can costlessly (when measured in attention) reoptimize.

suppose that there are two realizations of $\mathcal{N}_1$, in particular, $\mathcal{N}_1 \in \{\mathcal{N}_G, \mathcal{N}_B\}$, where $F_2(\mathcal{N}_G)$ first-order stochastically dominates $F_2(\mathcal{N}_B)$. In words, $\mathcal{N}_G$ signifies good news about the risky payoff, whereas $\mathcal{N}_B$ implies low returns. To avoid trivialities, we assume that the good news are such that (5) holds for all $I_2 = I_2^*$, i.e., when the news are good, the DM's expected material payoff from the consumption task is larger than that of the outside option. Further, when the news are bad, there are some levels of initial investment in the risky asset, I_1, so that the DM does not devote attention (and thus does not reoptimize), i.e., (5) does not hold.

We then compare the emotionally inattentive DM, $\tau > 0$, with the standard DM, $\tau = 0$.

Proposition 13. *In period 1 the standard DM's (weakly) optimal level of investment is $I_1 = W_1$, while the emotionally inattentive DM strictly prefers to chose an investment in the risky asset that is optimal given bad news, $I_1 \in \operatorname{argmax}_{I_1 \in [0,W_1]} V_2^D(\mathcal{N}_B, W_1, I_1)$. In period 2 the standard DM always (without loss) devotes attention to the portfolio problem and reoptimizes, while the emotionally inattentive DM reoptimizes only if the news are good.*

The proof follows straight from the simplifying assumptions and is omitted.

Proposition 13 states that the emotionally inattentive DM chooses an investment level in period 1 that is optimal precisely for when the news are bad, e.g., the market is down, and she avoids her portfolio problem. On the other hand, she happily reoptimizes if the news are good. Thus, she is generally willing to forego some investment benefits and distort her initial investment in order to have a portfolio that works well in poor market conditions. This implies that the emotionally inattentive DM under-invests in the risky asset. Combining these two results, we can see that emotional inattention can be a force for either excessive risk-taking or excessive risk avoidance. In general, she will tradeoff the optimal current portfolio with the optimal future portfolio in the case in which she does not want to reoptimize.

7 Discussion

Several other broad frameworks have been used to try and explain seemingly anomalous behaviors of individuals not devoting attention and, e.g., avoid materially beneficial information. We now turn to comparing our model of emotional inattention to four alternative frameworks: i) rational inattention where acquisition has exogenous costs unrelated to anticipatory feelings, ii) anticipatory utility, where there is no explicit cost of information acquisition but individuals have non-linear payoffs from beliefs about future payoffs, iii) news utility and reference-dependent preferences, where individuals gain flow utility not by how likely they think outcomes are, but how those likelihoods have changed recently, and iv) a small set of papers that have tried to understand attention and anticipation simultaneously. A key distinction between our approach and approaches i) - iii) is that our model incorporates a very particular form of "income effects." An emotionally inattentive DM pays more attention (and acquires more information) when the value of the decision problem increases (i.e., the height increases), and when the value of the outside task decreases. In models i)-iii) changes in the height of the payoff attached to the decision and to the outside task should have equivalent effects.[12]

Models of rational inattention (e.g., Sims (2003)) also attach a cost to attention. But, they differ from emotional attention in that the cost of attention is an exogenous function that depends only on the informativeness of the signal and the prior over the states (it, e.g., captures the mental effort of thinking). In contrast, in emotional inattention the cost of attention is endogenous — it depends on future payoffs, which themselves depend on the amount of attention that is paid. This endogeneity drives many of our important results; including that we can have attentional volatility even when the payoff function is concave Moreover, monetary and cognitive costs (stemming from a limited capacity for processing information), do not seem sufficient in many important situations to justify individuals' behavior, e.g., genetic tests for Hunting's disease cost no more than $300 (Oster et al., 2013). Furthermore, attention and information avoidance varies with the level of future consumption (Karlsson et al. (2009); Sicherman et al. (2015) for market returns, and Ganguly and Tasoff (2017) for expected health outcome) with no (obvious) corresponding change in costs or benefits. These examples suggest that a "cost" is missing from the consideration; indeed, our approach can be seen as modeling costly attention, but where the costs, or benefits, are endogenous and determined by the effect of attention on flow utility. In dynamic settings, existing models of rational inattention suppose dynamic consistency and that individuals will not have any intrinsic preference for the resolution of uncertainty via attention. This is in contrast to our stark results on dynamic inconsistency and a preference for paying attention earlier.

The key conceptual distinction between emotional inattention and models of anticipatory utility (e.g., Caplin and Leahy, 2001 and Dillenberger and Raymond (2020)) is that in anticipatory models a DM gains (or loses) anticipatory utility in accordance with her beliefs about future utility regardless of how much they focus on a decision. In contrast, an emotional inattentive DM can

[12]In fact, most models of rational inattention, with the exception of Liu et al. (2019), make the cost of attention additively separable from the benefits. This rules out "income effects" in the cost of information acquisition.

control the extent to which she gains or loses anticipatory utility by directing her attention. Thus, differences in beliefs induced by different information are integral in these anticipatory models in a way that isn't necessarily true in emotional inattention. But many behaviors, such as low adherence to medical recommendations (Becker and Mainman, 1975; Sherbourne et al., 1992; Lindberg and Wellisch, 2001; DiMattero et al., 2007), and double logins when markets are closed (Sicherman et al., 2015) do not feature shifts in beliefs, and so anticipatory models would fail to explain them. The relevance of distractors as a way to manage attention, as in Falk and Zimmermann (2016), is also not explained by these models. In contrast, they can be matched by emotional inattention. In order to generate an ostrich effect, models of anticipatory utility must be have intrinsic preferences for later information when utility levels are low (see Kreps and Porteus, 1978 and Dillenberger and Raymond, 2020 for a discussion of how the shape of the utility function over beliefs affects preferences towards learning). An emotionally inattentive DM instead always prefers earlier learning if they have to pay any attention at when there is no instrumental value to information.

Distinct from models of anticipatory utility, where flow utility depends on the beliefs about future outcomes, there are also models of changing beliefs, news utility, and reference dependence, where utility depends on changes in beliefs due to recent information (e.g., Kőszegi and Rabin, 2009). In these models (unlike our model) flow utility from emotions is 0 if and only if nothing has been learned. It is known that these models could generate ostrich-type behavior. (Olafsson and Pagel, 2017) show that that reference-dependence implies individuals will want to avoid paying attention to their bank accounts and that this attention aversion will decrease when wealth is large. However, as Olafsson and Pagel (2017) explain their model "cannot rationalize an increase in attention at a fully expected income payment or a jump in the probability of logging in when balances turn from negative to positive." In contrast, attentional volatility can both lead to jumps in attention, and can rationalize paying attention to even fully anticipated changes.[13]

Both anticipatory and reference-dependent models deliver a distinct prediction from emotional inattention in a realization independent environment (where the realization of ϵ does not change payoff expectations) when all states are equally likely. They predict individuals should exhibit no intrinsic preferences for information (because utility doesn't depend on the realized signal); in contrast, we point out that an emotionally inattentive DM may have a strict preference to devote some attention earlier rather than later.

[13]The asymmetric ostrich effects generated by reference-dependent models are "second-order" — they emerge because of decreases in the concavity over money at higher levels of wealth. In emotional inattention such shifts are first-order — they occur because of the changing the height of payoffs in the decision problem.

There are a few other papers that directly think about the dual role of attention (as we do).[14] Karlsson et al. (2009) propose a model where individuals can control attention, and it amplifies the emotional impact of information, modeling attention as a discrete choice.[15] They suppose individuals gain flow utility not from anticipatory emotions but rather via gain-loss utility changes from change in expected future outcomes from the information. Paying attention increases the relative impact of gain-loss utility, and also speeds up the adjustment of the reference point. This is important, as it implies that Karlsson et al. (2009) cannot capture the complementary roles of attention as (i) improving decisions and (ii) generating anticipatory utility, it cannot generate the "feedback loops" which underlie our results on attentional volatility and attentional traps. Moreover, although models such as Karlsson et al. (2009) can generate ostrich type effects, they are sensitive to the structure of model, and intuitive relaxations of any of the assumptions cause the assymetry in attention between bad and good states of the world to disappear.[16]

Tasoff and Madarasz (2009)'s model is closest to our own. A DM faces a decision problem with multiple dimensions (analogous to our different decisions), and receives anticipatory utility as a function of utility in that dimensions and the attention that is paid to that dimension. Like us, they model attention as a fixed resource. Unlike us, they assume that attention is a discrete outcome, and that attention is directed at a particular dimension when information about that dimension is acquired.

Formally, their Proposition 2 (and Corollary 1) draws on the same intuition as our results on the ostrich effect: fixing the instrumental value of information, bad situations attract little attention, while good situations do. Their Proposition 3 implies that there will be "default" effects, that individuals will not always adjust their decisions, and that this happens more often after bad signals than good, similar to what we find in our applications on consumption-savings and portfolio problems.

[14] The earliest reference we know of is Loewenstein (1987). Although he focuses his formal analysis on anticipation alone, he briefly discusses an extension to his model which allows for the vividness of an experience (which can be related to the amount of attention an experience attracts).

[15] Golman and Loewenstein (2018) and Golman et al. (2021) provide models that incorporate attention, as well as anticipatory utility, but model attention provision as automatic, rather than as a choice variable. Because attention is exogenous, their analysis excludes most of the features and results that we emphasize.

[16] The implications of the Karlsson et al. (2009) model are generated in part because the first-period gains or losses (relative to the reference point) are not affected by attention, they are just scaled up or down by attention. Only the second-period reference point is affected. If instead both the realized gains/losses in the first period, and the reference point in the second period are equally affected by attention, then their major finding, of the asymmetric ostrich effect, disappears, and instead, a symmetric ostrich effect emerges where all information is avoided. Workaround, such as relying on the concavity of utility over wealth to generate asymmetries, as in Olafsson and Pagel (2017), only generates second-order asymmetries as noted.

However, there are also important differences. First, they model the acquisition of information as causing attention. In other words, while in our model the choice of attention can determine the information structure, in theirs the choice of information structure determines (in expectation) the attention. Thus, for example, they do not focus on the situation where attention can be directed even when no information has come to light, a key driving force for our results regarding intrinsic preferences for the timing of learning (and so they do not have any analogue of our preference for learning early). This distinction also motivates very different applications than us (their focus is on a monopolist manipulating information provided to — and thus the attention of — consumers). Second, although their model extensively covers how there may be a consumption-attention complementarity, they do not focus on the importance of this in generating attentional volatility.[17] Third, although they mention the issue of time inconsistency, their analysis doesn't focus on the effects of this, while we highlight the importance of dynamic inconsistency in generating attentional traps.

8 Conclusion

This paper introduces a model of emotional inattention in which attention serves a dual role leading to instrumental value but also prompting an emotional response. The emotional response from attention then immediately predicts the well-documented ostrich effect of individuals not attending to negative situations (to avoid the negative emotional response). A key driver of our results is the complementarity between the instrumental and emotional consequences of attention. We study preferences over the timing of uncertainty and highlight potentially perverse consequences of policies designed to increase attention. We also show how dynamic inconsistency and volatile attention levels can lead to attention traps. We relate some of our predictions to existing empirical evidence and leave others to be tested in the future. We show the implications of emotional inattention for consumption-savings decisions and portfolio choice, again providing novel explanations and predictions, but also highlighting the flexibility of our modeling approach showing its ripeness for applied work.

Our approach has made several substantive assumptions about the sophistication of our DM, all of which could be relaxed. First, it could be that the DM supposes in period 1 that their commitment plans will be followed through on. In this case, rather than having the total unravelling

[17]E.g., their Proposition 4 discusses a form of complementary between consumption and attention, but does not explore the effects on attentional volatility.

of attention that we provide examples of, we will observe the DM engaging in a "lack of follow through" where she initially pay attention in period 1, but then under-invests in period 2 relative to expectations. Second, it could be that the actual τ_2 used in period 2 is different than the τ_2 used in period 1. If period 1 overestimates τ_2 relative to the truth, this accentuates the dynamic inconsistency problems we highlighted (since period-1 self's behavior is based on their anticipation of period-2 self's behavior). If period-1 self underestimates the actual τ_2, we again see a lack of follow through. A third way in which a DM could be naive is that she in period 1 rather than receiving anticipatory utility from what she expect to be their final payoffs (which includes both period 1 and period 2 attention), only base their anticipatory utility on the expected payoff given her period 1 attention. This implies that the anticipatory utility for period 1 weakly falls, making it harder to sustain attention, an increasing the chance that we see an attention trap.

We have also made important assumptions about the structure of our environment. First, our model assumes that there is an exogenously given way of dividing up the world into a decision and into an outside task. One might imagine that individuals have a choice of how to bundle tasks — it could be that whenever a person thinks of task A they must think of B, or that they can think of them separately. Our model naturally extends to an addition "period 0" where the DM decides whether to bundle the decision and the outside task together (so that paying attention to one requires paying attention to both) or separately. Although the DM may want to bundle occasionally, to encourage her future selves to think about both problems, the DM will still want to keep particularly bad decisions mentally distinct from the outside task. A related issue is that we assumed that payoffs are separable across the decision and the outside task, which may not be the case. Moreover, our model, like many models of inattention which are based on optimization, raises questions about infinite regress. One might ask whether our DM needs to think about both the decision and the outside task in order to first understand the parameters of the problem facing them, and if so, how that impacts utility.

References

Ariely, Dan, Uri Gneezy, George Loewenstein, and Nina Mazar, "Large stakes and big mistakes," *The Review of Economic Studies*, 2009, *76* (2), 451–469.

Balboni, Clare A, Oriana Bandiera, Robin Burgess, Maitreesh Ghatak, and Anton Heil, "Why do people stay poor?," Technical Report, National Bureau of Economic Research 2021.

Becker, Marshall H and Lois A Mainman, "Sociobehavioral determinants of compliance with health and medical care recommendations," *Medical Care*, 1975, *1* (13), 10–24.

Böheim, René, Dominik Grübl, and Mario Lackner, "Choking under pressure–Evidence of the causal effect of audience size on performance," *Journal of Economic Behavior & Organization*, 2019, *168*, 76–93.

Caplin, Andrew and John Leahy, "Psychological Expected Utility Theory and Anticipatory Feelings*," *The Quarterly Journal of Economics*, 02 2001, *116* (1), 55–79.

_ and _ , "The Social Discount Rate," *Journal of Political Economy*, 2004, *112* (6), 1257–1268.

Dillenberger, David and Collin Raymond, "Additive-Belief-Based Preferences," Working Paper 20-020, SSRN July 2020.

DiMattero, Robin M, Kelly B Haskard, and Summer L Williams, "Health beliefs, disease severity, and patient adherence: a meta-analysis," *Medical Care*, 2007, *6* (45), 521–8.

Epstein, Larry G and Stanley E Zin, "Substitution, Risk Aversion, and the Temporal Behavior of Consumption," *Econometrica*, 1989, *57* (4), 937–969.

Ergin, Haluk and Todd Sarver, "Hidden actions and preferences for timing of resolution of uncertainty," *Theoretical Economics*, 2015, *10* (2), 489–541.

Falk, Armin and Florian Zimmermann, "Beliefs and Utility: Experimental Evidence on Preferences for Information," Working Paper 10172, Institute for the Study of Labor August 2016.

Ganguly, Ananda and Joshua Tasoff, "Fantasy and Dread: The Demand for Information and the Consumption Utility of the Future," *Management Science*, 2017, *63* (12), 4037–4060.

Golman, Russell and George Loewenstein, "Information gaps: A theory of preferences regarding the presence and absence of information," *Decision*, 2018, *5* (3), 143–164.

_ , Nikolos Gurney, and George Loewenstein, "Information gaps for risk and ambiguity," *Psychological Review*, 2021, *128* (1), 86–103.

Karlsson, Niklas, George Loewenstein, and Duane Seppi, "The ostrich effect: Selective attention to information," *Journal of Risk and Uncertainty*, 2009, *38* (2), 95–115.

Kreps, David M. and Evan L. Porteus, "Temporal Resolution of Uncertainty and Dynamic Choice Theory," *Econometrica*, 1978, *46* (1), 185–200.

Kőszegi, Botond and Matthew Rabin, "Reference-Dependent Consumption Plans," *American Economic Review*, June 2009, *99* (3), 909–36.

Lindberg, Nangel M. and David Wellisch, "Anxiety and compliance among women at risk for breast cancer," *Annals of Behavioral Medicine*, 2001, *4* (23), 298–303.

Liu, Ce, Christopher Chambers, John Rehbeck et al., "Costly Information Acquisition," Technical Report 2019.

Loewenstein, George, "Anticipation and the Valuation of Delayed Consumption," *The Economic Journal*, 1987, *97* (387), 666–684.

Masatlioglu, Yusufcan, A Yesim Orhun, and Collin Raymond, "Intrinsic information preferences and skewness," *Ross School of Business Paper*, 2017.

Moscarini, Giuseppe and Lones Smith, "The optimal level of experimentation," *Econometrica*, 2001, *69* (6), 1629–1644.

Mullainathan, Sendhil and Eldar Shafir, *Scarcity: why having too little means so much*, first edition ed., New York: Times Books, Henry Holt and Company, 2013.

_ **and** _ , *Scarcity: Why having too little means so much*, Macmillan, 2013.

Nielsen, Kirby, "Preferences for the resolution of uncertainty and the timing of information," *Journal of Economic Theory*, 2020, *189*, 105090.

Olafsson, Arna and Michaela Pagel, "The Ostrich in Us: Selective Attention to Financial Accounts, Income, Spending, and Liquidity," Working Paper 23945, National Bureau of Economic Research October 2017.

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

Schelling, T. C., "THE MIND AS A CONSUMING ORGAN," 1988.

Sherbourne, Cathy Donald, Ron D. Hays, Lynn Ordway, M. Robin DiMatteo, and Richard L. Kravitz, "Antecedents of adherence to medical recommendations: Results from the medical outcomes study," *Journal of Behavioral Medicine*, 1992, (15), 447–468.

Shouldson, Ira and Anne Young, "Milestones in huntington disease," *Movement Disorders*, 2011, *6* (26), 1127–33.

Sicherman, Nachum, George Loewenstein, Duane J. Seppi, and Stephen P. Utkus, "Financial Attention," *The Review of Financial Studies*, 11 2015, *29* (4), 863–897.

Sims, Christopher A, "Implications of rational inattention," *Journal of monetary Economics*, 2003, *50* (3), 665–690.

Tasoff, Joshua and Kristof Madarasz, "A Model of Attention and Anticipation," Working Paper, SSRN November 2009.

A Proofs omitted from the main body

Proof of Lemma 3. As $V^D(\cdot|\epsilon)$ is continuous, so is $V_2(\cdot|\epsilon)$. As A_2 is compact, maximizing $V_2(\cdot|\epsilon)$ over A_2 admits a solution. □

Proof of Proposition 1. The first claims follows as the objective has increasing differences in (α, l) and $(\alpha, -\underline{U})$.

For the third claim, note that if $\tilde{V}^D$ is differentiable at 0, so is V, and we have

$$\frac{\partial}{\partial\alpha}V(0) = \frac{\partial}{\partial\alpha}V^D(0) - \tau[\underline{U} - (V^D(0) + l)],$$

which is negative for large enough l as $\tau > 0$.

□

Proof of Lemma 2. As

$$\frac{\partial^2}{\partial\alpha^2}V(\alpha) = \frac{\partial^2}{\partial\alpha^2}V^D(\alpha)[\tau\alpha + 1] + 2\tau_2\frac{\partial}{\partial\alpha}V^D(\alpha),$$

the condition in the proposition implies that the second derivative is strictly positive, i.e., the total payoff is strictly convex. □

Proof of Proposition 2. Let $\underline{\alpha} \equiv \min_{\alpha\in A}\alpha, \bar{\alpha} \equiv \max_{\alpha\in A}\alpha$. Note that $V_l(\alpha)$ is continuous in l, $V_l(\underline{\alpha}) > V_l(\bar{\alpha})$ for low l, and $V_l(\underline{\alpha}) < V_l(\bar{\alpha})$ for large l. Thus, there exists l^* so that $V_{l^*}(\underline{\alpha}) = V_{l^*}(\bar{\alpha})$. Lastly, note that V_l has strictly increasing differences in (α, l), and the result follows. □

Proof of Proposition 4. First, suppose with commitment there does not exists a front- or backloading solution. As V^D is continuous, V^1 is continuous and with commitment period-1 self's optimal payoff is obtained. We proceed by showing that no solution remains a solution when commitment is dropped, and then note that without commitment no sequence of attention levels, where period-2 attention is chosen optimally without commitment, approaches a solution with commitment. To this end, take any solution from the set of solutions with commitment, (α_1, α_2), and consider

$$\frac{\partial V^1}{\partial \alpha_1}(\alpha_1, \alpha_2) = \frac{\partial V^D}{\partial \alpha_1}(\alpha_1, \alpha_2)[\tau_1\alpha_1 + \tau_2\alpha_2 + 1] - \tau_1(\underline{U} - V^D(\alpha_1, \alpha_2)). \tag{6}$$

If (6) is strictly positive, then it must be that $\alpha_2 = 1$, which is ruled out by assumption. If (6) is strictly positive, then $\alpha_2 = 0$: $\alpha_1 = 0$ is ruled out by assumption and $\alpha_1 > 0$ implies (α_1, α_2) is a frontloading solution, a contradiction. Thus, consider the case where (6) equals 0. As with commitment full attention in period 2 is not optimal, it must be that $\underline{U} > V^D(\alpha_1, \alpha_2)$ and $\alpha_2 < 1$, and as (6) equals 0, it must be that $\frac{\partial V^D}{\partial \alpha_1}(\alpha_1, \alpha_2) > 0$. If $\alpha_1 = 0$, $(\alpha_1, \alpha_2) >$ is a backloading solution, thus, consider α_1. It is than clear that

$$\frac{\partial V^2}{\partial \alpha_1}(\alpha_1, \alpha_2) = \frac{\partial V^D}{\partial \alpha_1}(\alpha_1, \alpha_2)[\tau_2\alpha_2 + 1] - \tau_1(\underline{U} - V^D(\alpha_1, \alpha_2)) < 0,$$

so that either $\alpha_2 = 0$, in which case (α_1, α_2) is a frontloading solution, or without commitment (α_1, α_2) is not a solution. Thus, no solution remains a solution when commitment is dropped. Lastly, take any sequence of attention levels where period-2 attention is chosen optimally without commitment and suppose its limit is (α_1, α_2). As V^D is continuously differentiable so is V^1 and the preceeding discussion implies that the sequence does not approach (α_1, α_2). Thus, period-1 self is strictly worse off when commitment is dropped.

Now, suppose with commitment that there exists a front- or backloading solution, (α_1, α_2). Suppose $\alpha_2 = 0$, i.e., it is a frontloading solution. V_1, fixing α_1, has increasing differences in α_2 and τ_1, and equals V_2 for $\tau_1 = 0$. Thus, period-2 self, if anything, prefers to reduce α_2 which is already at 0 so that (α_1, α_2) remains a solution when commitment is dropped. Suppose $\alpha_1 = 0$. In this case, V_1, holding α_1 fixed, and V_2 coincide. Thus, again, (α_1, α_2) remains a solution. $\square$

Proof of Proposition 5. To prove the result, we make use of some claims.

Claim 1. *For τ_1, τ_2 large enough, $(0, 0)$ is a strict local maximum of V_1.*

Proof of Claim 1. We show that for τ_1, τ_2 large enough, the derivatives of V_1 at $(0, 0)$ with respect to α_1 and α_2 are strictly negative. We have

$$\frac{\partial V_1}{\partial \alpha_1}(0, 0) = \frac{\partial V^D}{\partial \alpha_1}(0, 0) - \tau_1(\underline{U} - V^D(0, 0)), \quad \text{and}$$
$$\frac{\partial V_1}{\partial \alpha_2}(0, 0) = \frac{\partial V^D}{\partial \alpha}(0, 0) - \tau_2(\underline{U} - V^D(0, 0)).$$

As V^D is strictly increasing, $V^D(1,0) > V^D(0,0)$ and so as $\underline{U} = V^D(1,0)$, $^O - V^D(0,0) > 0$. Thus, for

$$\tau_1 > \underline{\tau}_1 \equiv \frac{\frac{\partial V^D}{\partial \alpha_1}(0,0)}{\underline{U} - V^D(0,0)} \quad \text{and} \quad \tau_2 > \underline{\tau}_2 \equiv \frac{\frac{\partial V^D}{\partial \alpha_2}(0,0)}{\underline{U} - V^D(0,0)},$$

$(0,0)$ is a strict local maximum. □

Claim 2. *Let $A_1 \equiv \{\alpha_1 : V_1(\alpha_1, 0) = V_1(0,0), \alpha_1 > 0\}$. For $\tau_1 > \underline{\tau}_1$, where $\underline{\tau}_1$ is defined in the proof of Claim 1, we have:*

i) $\underline{\alpha}_1 \equiv \min_{\alpha_1 \in A_1} \alpha_1$ *exists,*

ii) $\underline{\alpha}_1(\tau_1) < 1$,

iii) $\underline{\alpha}_1(\tau_1)$ *is strictly increasing in τ_1, and*

iv) $\lim_{\tau_1 \to \infty} \underline{\alpha}_1(\tau_1) = 1$.

Proof of Claim 2. The first subclaim follows as A_1 is compact and non-empty because i) $\tau_1 > \underline{\tau}_1$ implies $\frac{\partial V_1}{\partial \alpha_1}(0,0) < 0$, ii) $\underline{U} = V^D(1,0)$ implies $V_1(1,0) > V_1(0,0)$, and iii) by continuity of V_1.

For the second subclaim, note again that $V_1(1,0) > V_1(0,0)$ and by existence of $\underline{\alpha}_1(\tau_1)$.

The third subclaim follows as $V_1(\alpha_1,0) - V_1(0,0) = (V^D(\alpha_1,0) - V^D(0,0)) - (\underline{U} - V^D(\alpha_1,0))\tau_1\alpha_1$ is weakly decreasing in τ_1 and strictly so for $\alpha_1 \in (0,0)$. In particular, for $\tau_1' > \tau$ and any $\alpha_1 \in (0, \underline{\alpha}_1(\tau_1)]$, $(V^D(\alpha_1,0) - V^D(0,0)) - (\underline{U} - V^D(\alpha_1,0))\tau_1'\alpha_1 < 0$.

By the second and third subclaims, it must be that $\alpha_1(\tau_1)$ converges as $\tau_1 \to \infty$. Suppose $\alpha_1(\tau_1) \to \underline{\alpha}_1 < 1$, as $\tau_1 \to \infty$. As V^D is strictly increasing, this implies $\underline{U} - V^D(\underline{\alpha}_1, 0) > 0$ so that $\underline{U} - V^D(\underline{\alpha}_1,0))\tau_1\underline{\alpha}_1 - (V^D(\underline{\alpha}_1,0) - V^D(0,0))$ becomes negative as $\tau_1 \to \infty$. Furthermore, $V_1(\alpha_1,0) - V_1(0,0)$ applied to $\alpha_1 = \underline{\alpha}_1(\tau_1)$ equals 0 for all τ_1. As it is also continuous, however, the function value of the limit (which is negative) must equal to limit of the function value (which is 0), i.e., we have reached a contradiction and the final subclaim follows. □

Given τ_1 large enough, we first show that we can find $\underline{\alpha}_1$ and τ_2 so that three conditions hold:

i) τ_1, τ_2 are large enough so that by Claim 1, $(0,0)$ is a strict local maximum,

ii) period-1 self strictly prefers to increase α_2 from 0 if $\alpha_1 = \underline{\alpha}_1$:

$$\frac{\partial V_1}{\partial \alpha_2}(\underline{\alpha}_1, 0) = \frac{\partial V^D}{\partial \alpha_2}(\underline{\alpha}_1, 0)[\tau_1\underline{\alpha}_1 + 1] - \tau_2(\underline{U} - V^D(\underline{\alpha}_1, 0)) > 0$$

iii) period-2 self strictly prefers not to increase α_2 from 0 if $\alpha_1 = \underline{\alpha}_1$:

$$\frac{\partial V_2}{\partial \alpha_2}(\underline{\alpha}_1, 0) = \frac{\partial V^D}{\partial \alpha_2}(\underline{\alpha}_1, 0) - \tau_2(\underline{U} - V^D(\underline{\alpha}_1, 0)) < 0.$$

By part iv) of Claim 2, $\lim_{\tau_1 \to \infty}\{\underline{U} - V^D(\underline{\alpha}_1(\tau_1), 0)\} = 0$, and, as V^D is continuously differentiable with derivative bounded away from zero, we thus have

$$\lim_{\tau_1 \to \infty} \frac{\frac{\partial V^D}{\partial \alpha_2}(\underline{\alpha}_1(\tau_1), 0)}{\underline{U} - V^D(\underline{\alpha}_1(\tau_1), 0)} = \infty.$$

Thus, there exists $\bar{\tau}_1 > \underline{\tau}_1$, so that for every $\tau_1 > \bar{\tau}_1$, we have

$$\underline{\tau}_2 < \frac{\frac{\partial V^D}{\partial \alpha_2}(\underline{\alpha}_1(\tau_1), 0)}{\underline{U} - V^D(\underline{\alpha}_1(\tau_1), 0)} \tag{7}$$

implying that if condition iii) holds for some τ_2, then $\tau_2 > \underline{\tau}_2$.

Pick τ_2', τ_2'' so that ii) and iii) hold with equality, respectively: $\frac{\partial V^D}{\partial \alpha_2}(\underline{\alpha}_1, 0)[\tau_1 \underline{\alpha}_1 + 1] - \tau_2'(\underline{U} - V^D(\underline{\alpha}_1, 0)) = 0$, $\frac{\partial V^D}{\partial \alpha_2}(\underline{\alpha}_1, 0) - \tau_2'(\underline{U} - V^D(\underline{\alpha}_1, 0)) = 0$. By Part ii) of Claim 2, $\underline{\alpha}_1 < 1$ and as V^D is strictly increasing, $\underline{U} - V^D(\underline{\alpha}_1, 0) > 0$. Thus, $\tau_2' < \tau_2''$ and for, e.g., $\tau_2 = (\tau_2' + \tau_2'')/2$, both (ii)) and iii) are satisfied.

In other words, we have found the parameters we set out to find.

To briefly sketch out the rest of the construction, note that as α_2 increases at $(\underline{\alpha}_1, 0)$, by condition ii), the resulting point is strictly preferred by period-1 self to $(0, 0)$. However, given attention $\underline{\alpha}_1$ in period 1, period 2 strictly prefers devoting no attention to devoting a (small) positive amount. Hence, with commitment, period-1 self can choose a point she strictly prefers to devoting no attention while without commitment, the outcome is not devoting any attention.

To show that period-2 self is also strictly worse off without commitment, we make use of the following claim.

Claim 3. *If period-1 self is indifferent between* $(0, 0)$ *and* (α_1, α_2) *and* $\alpha_1 > 0$*, then there exists* $\delta > 0$ *so that period-2 self strictly prefers* $(\alpha_1 + x, \alpha_2 + y)$ *over* $(0, 0)$ *for all* $(x, y) \in [-\delta, \delta]^2$.

Proof of Claim 3. Note that $\frac{\partial V_1}{\partial \tau_2}(0, 0) = \underline{U} > \alpha_1 V^D(\alpha_1, \alpha_2) + (1 - \alpha_1)\underline{U} = \frac{\partial V_1}{\partial \tau_2}(\alpha_1, \alpha_2)$ (using that

$V^D(\alpha_1, \alpha_2) < \underline{U}$ as otherwise period-1 self cannot be indifferent). This, together with the continuity of V^D, proves the claim. $\square$

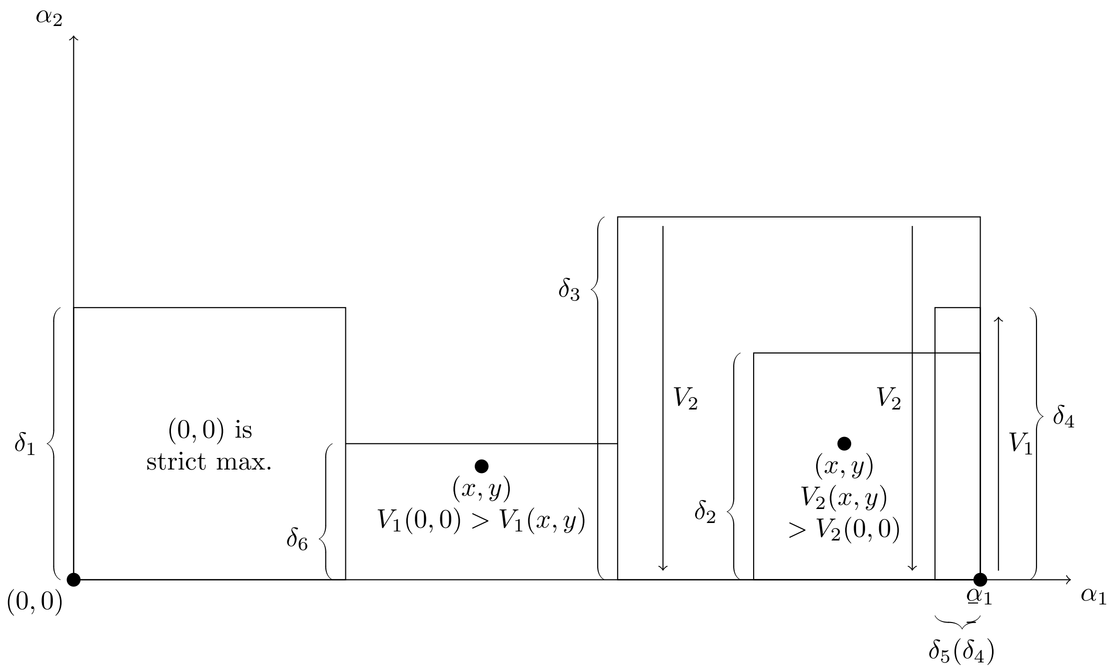

Figure 4: Constructing the bounds. Arrows indicate preferences.

We now use these results to show the existence of five new variables which each defines a region.

1. As $(0, 0)$ is a strict local maximum, there exists $\delta_1 > 0$ so that period-1 self strictly prefers $(0, 0)$ to any point in $[0, \delta_1]^2$.

2. By Claim 3, there exists $\delta_2 > 0$ so that period-2 self strictly prefers any point in $[\underline{\alpha}_1 - \delta_2, \underline{\alpha}_1] \times [0, \delta_2]$ to $(0, 0)$.

3. By V^D being continuously differentiable and condition iii), there exists $\delta_3 > 0$ so that for any α_1 in $[\underline{\alpha}_1 - \delta_3, \underline{\alpha}_1]$, period-2 self strictly most prefers to devote no attention.

4. By V^D being continuously differentiable and condition ii), there exists $\delta_4 > 0$ so that for any C in $[0, \delta_4]$, there exists $\delta_5(\delta_4) > 0$, so that $\max_{\alpha_1 \in [0, \underline{\alpha}_1 - \delta_5(\delta_4)], \alpha_2 \in [0, C]} V_1(\alpha_1, \alpha_2) > V(\underline{\alpha}_1, 0)$.

5. By definition of $\underline{\alpha}_1$, period-1 self strictly prefers $(0, 0)$ to any point in $[\delta_1, \delta_3] \times \{0\}$. By continuity of V^D, there exists $\delta_6 > 0$ so that period-1 self strictly prefers $(0, 0)$ to any point in $[\delta_1, \delta_3] \times [0, \delta_6]$.

Figure 4 illustrates the choice of the δs and the regions they define. Let $\delta \equiv \min\{\delta_1, \delta_2, \delta_3, \delta_4, \delta_5(\delta_4), \delta_6\}$, and let the bounds on attention be given by $(C_1, C_2) \equiv (\underline{\alpha}_1, \delta)$.

First, we show that the solution without commitment is $(0,0)$. Consider any attention level α_1. If $\alpha_1 \leq \underline{\alpha}_1 - \delta_3$, then period-1 self strictly prefers $(0,0)$. Otherwise, period-2 self chooses no attention, and as $\alpha_1 < \underline{\alpha}_1$ and by definition of $\underline{\alpha}_1$, period-1 self, again, strictly prefers $(0,0)$.

Next, note that period-1 self strictly prefers the outcome with commitment as she strictly prefers (C_1, C_2) to $(0,0)$. Also note that the outcome with commitment, (α_1, α_2), is in $[\underline{\alpha}_1 - \delta_3, \underline{\alpha}_1] \times [0, \delta_3]$, so that period-2 self also strictly prefers it to $(0,0)$. $\square$

Proof of Proposition 6. We formalize V^D *is not too concave in* α_2 and V^D *has not too decreasing differences in* (α_1, α_2), both given V^D. The former is

$$\frac{2\underline{\tau}_2}{\underline{\tau}_2\alpha_2 + 1} > -\frac{\frac{\partial^2}{\partial\alpha_2^2}V^D(\alpha_1,\alpha_2)}{\frac{\partial}{\partial\alpha_2}V^D(\alpha_1,\alpha_2)}, \quad \text{for all } (\alpha_1,\alpha_2), \tag{8}$$

and the latter is

$$\frac{\underline{\tau}_2}{\underline{\tau}_2\alpha_2 + 1} \geq -\frac{\frac{\partial^2}{\partial\alpha_1\partial\alpha_2}V^D(\alpha_1,\alpha_2)}{\frac{\partial}{\partial\alpha_1}V^D(\alpha_1,\alpha_2)}, \quad \text{for all } (\alpha_1,\alpha_2), \tag{9}$$

$$\frac{1}{\alpha_1} > -\frac{\frac{\partial^2}{\partial\alpha_1\partial\alpha_2}V^D(\alpha_1,\alpha_2)}{\frac{\partial}{\partial\alpha_1}V^D(\alpha_1,\alpha_2)}, \quad \text{for all } (\alpha_1,\alpha_2) \tag{10}$$

with $\underline{\tau}_2$ defined in Claim 1.

Given τ_1 large enough, we first show that we can find $\underline{\alpha}_1, \underline{\alpha}_2$ and τ_2 so that two conditions hold:

i) τ_1, τ_2 are large enough so that by Claim 1, $(0,0)$ is a strict local maximum,

ii) period-2 self is indifferent between $(\underline{\alpha}_1, \underline{\alpha}_2)$ and $(\underline{\alpha}_1, 0)$.

Let $\bar{\alpha}_2$ solve $\underline{U} = V^D(\underline{\alpha}_1(\bar{\tau}_1), \bar{\alpha}_2)$, if such solution exists, and equal ∞ otherwise.

Then for all $\underline{\alpha}_2$, with $0 < \underline{\alpha}_2 < \bar{\alpha}_2$, there exists $\tau_1 > \underline{\tau}_1$, such that

$$\tau_2 \equiv \frac{V^D(\underline{\alpha}_1(\tau_1), \underline{\alpha}_2) - V^D(\underline{\alpha}_1(\tau_1), 0)}{\underline{\alpha}_2[\underline{U} - V^D(\underline{\alpha}_1(\tau_1), \underline{\alpha}_2)]} > \underline{\tau}_2;$$

to see this, note that the numerator is strictly positive and bounded away from 0 for all τ_1, and the denominator is strictly positive for $\tau_1 = \underline{\tau}_1$, continuous in τ_1, and strictly negative for large enough τ_1 (using part iv) of Claim 2).

Note that given τ_2, period-2 self is indifferent between $(\underline{\alpha}_1, \underline{\alpha}_2)$ and $(\underline{\alpha}_1, 0)$; in other words, we have found the parameters we set out to find.

The fact that period-2 self is indifferent between $(\underline{\alpha}_1, \underline{\alpha}_2)$ and $(\underline{\alpha}_1, 0)$ implies that period-1 self strictly prefers the former to the latter. Next, consider period-2 self's attentional best response, given different constraints. period-2 self's payoff is strictly convex in α_2 as the left-hand side of (8) is increasing in $\underline{\tau}_2$ and $\tau_2 > \underline{\tau}_2$ and by Lemma 2. Hence, it suffices to find the attention required to have the same payoff (for period-2 self) as with no attention (if such attention level exists) and check whether it falls within the constraint. For $\alpha_1 = \underline{\alpha}_1$, this level of attention is given by $\alpha_2 = \underline{\alpha}_2$, as per our construction.

As the left-hand side of (9) is increasing in $\underline{\tau}_2$ and $\tau_2 > \underline{\tau}_2$, and using some simple algebra, (9) implies that

$$\frac{\partial^2}{\partial \alpha_1 \partial \alpha_2} V_2(\alpha_1, \alpha_2) \geq 0,$$

everywhere. It follows that the attention required to have the same payoff as with no attention is weakly decreasing in α_1.

Finally, let $(C_1, C_2) = (\underline{\alpha}_1, \underline{\alpha}_2)$. Consider any $\delta > 0$ and suppose the attention bounds are $(C_1 - \delta, C_2 - \delta)$. As per the previous observation for any attention level α_1, period-2 self chooses to devote no attention. By construction of $\underline{\alpha}_1$, period-1 self strictly prefers $(0, 0)$ to any other outcome she can implement. Hence, the unique outcome without commitment is $(0, 0)$. Suppose the attention bounds are $(C_1 + \delta, C_2 + \delta)$. Then period-1 self can implement attention levels $(\alpha_1, C_2 + \delta)$ for $\alpha_1 \in [C, C + \delta)$. Together, (9) and (10) imply that $\frac{\partial^2}{\partial \alpha_1 \partial \alpha_2} V_1(\alpha_1, \alpha_2) > 0$ so that period-1 self implements attention levels $(C_1 + \delta, C_2 + \delta)$. This level of attention is preferred to $(0, 0)$ by period-1 self with the difference bounded away from 0 as $\delta \to 0$, as period-1 self strictly prefers (C_1, C_2) to $(0, 0)$ which she can always implement. Lastly, this level of attention is also preferred by period-2 self to $(0, 0)$, with the difference bounded away from 0 as $\delta \to 0$. This is because, by construction, period-2 self prefers to devote attention to not devoting attention which she strictly prefers to $(0, 0)$ as period-1 self devotes attention. $\square$