Emotional Inattention

Lukas Bolte[*] Collin Raymond[*]

August 31, 2022

PRELIMINARY DRAFT: DO NOT CIRCULATE

https://lukasbolte.github.io/papers/emotionalInattention.pdf

Abstract

A decision-maker allocates attention across additively separable dimensions (e.g., consumption problems, states, or time periods). In addition to being instrumentally valuable, attention generates attention utility, and so the decision-maker maximizes an attention-weighted objective function. Optimal attention to a dimension is increasing in its payoff and the instrumental value of attention. The attention-weighted objective generates behavioral phenomena such as belief distortions and time preferences, as well as predictions on when and in which form they occur. We apply our model to information acquisition and portfolio choice and discuss implications for policy interventions designed to increase overall utility or improve decisions.

JEL Classification Codes: D81, D83

Keywords: Attention, attention utility, 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. Bolte gratefully acknowledges financial support from the Leonard W. Ely and Shirley R. Ely Graduate Student Fellowship.

1 Introduction

Attention—"a state in which [non-infinite] cognitive resources are focused on certain aspects of the environment rather than on others" (VandenBos, ed, 2015)—has (at least) two fundamental features. First, it is instrumentally valuable: it allows individuals to better match their actions to the environment through, e.g., information acquisition and processing or changing a default. This feature is well recognized and studied by both psychologists (Desimone et al., 1995) and economists (Sims, 2003). Second, attention governs how individuals generate and regulate emotions (see Dixon et al., 2017 and Gross, 1998 for reviews). Indeed, anticipatory utility or remembrance utility from non-contemporaneous consumption, which has been discussed as a key component of generating emotional utility (as in, e.g., Caplin and Leahy (2001) and Loewenstein (1987)) can only occur if the consumption is, in fact, 'anticipated' or 'remembered;' in other words, if the individual devotes attention to it. Despite widespread interest in this second feature among psychologists and cognitive scientists, it has received less study by economists.[1]

This paper combines these two insights about the role of attention and makes three main contributions: (i) we formalize a model of attention allocation—that considers both of the aforementioned features of attention jointly; (ii) with it, we provide a unifying explanation for a host of existing behavioral phenomena; and (iii) we make novel (and testable) predictions about new ones. Two key insights drive many of our results. First, because attention to a situation increases the emotional valence of that situation, individuals will ignore low-payoff ones even though attention to them would be instrumentally valuable and, conversely, devote excess attention to high-payoff situations. Second, because attention to a situation increases its weight in the emotional component of utility, attention, in the words of Todd et al. (2012), "is proactive in shaping perceptual experience." It essentially reweights the decision environment leading to a variety of behavioral phenomena. In

[1] That is not to say it has been completely ignored. For example, Schelling (1988) highlights the role of the "mind" as a "pleasure machine or consuming organ, the generator of direct consumer satisfaction," in addition to the role of the "information processing and reasoning machine." Schelling also implicitly suggests that these roles should be considered jointly and writes: "Marvelous it is that the mind does all these things. Awkward it is that it seems to be the same mind from which we expect both the richest sensations and the most austere analyses." Our following analysis of the interaction of these roles hopes to alleviate this "awkwardness."

a world with uncertainty, the this second effect manifests as probability weighting, while a dynamic setting it generates as-if non-exponential discounting.

A key tenet of our model is that attention is (at least in some part) voluntarily directed by the individual, a premise often referred to as top-down attention. Such assumption is standard in much of economics such as in the rational inattention literature, and the empirical literature has emphasized that at least some attention capacity is directed (e.g., Corbetta and Shulman (2002); Buschman and Miller (2007)). This is not to say that we think that involuntarily allocated attention (often referred to as bottom-up attention) is not important. One can imagine our model as determining the use of the residual stock of attention after involuntarily attentional allocations have been made. We return, in Section 7, to discussing how voluntary and involuntary attention may interact.

In order to fix ideas, we begin by sketching out a simplified version of our model. A decision-maker (henceforth, DM—they) faces an environment that has a number of dimensions, indexed by i, each associated with a payoff V_i. The DM devotes a unit mass of attention across these dimensions, with attention to dimension i denoted by α_i, and the vector of attention as α. The DM also takes an action x from a set $X(\alpha)$, which depends on the attention allocation, that determines the payoff associated with each dimension i: $V_i(x)$. The DM experiences two kinds of utility, both of which are additively separable across dimensions. First, they gain a material payoff from each dimension due to the act of consumption which equals the payoff associated with a dimension: $V_i(x)$. Second, they gain emotional utility from each dimension in proportion to the payoff from that dimension, and the amount of attention devoted to it: $\alpha_i V_i(x)$. In spirit this approach is similar to models of anticipatory utility (e.g., Loewenstein (1987); Caplin and Leahy (2001)), where the flow utility of emotions is a function of beliefs about future payoffs. Our innovation is to make this flow utility of emotions from a dimension a function of the attention paid to it. For instance, by paying less attention to a low-payoff dimension, and more to a high-payoff dimension, the DM will be able to improve their emotional state. We refer to this second component as attention utility. Total utility is simply the sum of these two components over all dimensions, where attention utility receives relative weight λ:

$$\underbrace{\sum_i V_i(x)}_{\text{material payoff}} + \lambda \underbrace{\sum_i \alpha_i V_i(x)}_{\text{attention utility}} . \tag{1}$$

Thus, attention is a measure on a set of dimensions that determines: 1) which actions are available and 2) the weights of the different dimensions (notice that the total weight on dimension i is $1 + \lambda\alpha_i$). Although the resulting objective is simple, it will allow us to capture a rich set of implications in a tractable way.[2]

In Section 2 we introduce our model, a generalization of Equation (1), and characterize the optimal attention allocation and action. We begin by highlighting some key comparative statics with respect to the payoffs. In particular, the payoff levels of the dimensions matter: ceteris paribus, the DM devotes more attention to dimensions with higher payoff levels. The DM may thus avoid a low-payoff dimension, even though attending to it would increase its payoff while devoting excessive attention beyond the point where it is instrumentally valuable to others.

We then study how the weight on attention utility (λ) affects optimal attention and actions. For $\lambda = 0$, our model collapses to what we call the "standard model," one where the DM chooses attention to maximizes material payoffs. As λ increases, the DM monotonically departs from the material-payoff maximizing solution and so takes worse actions when seen through the lens of the standard model. We also show how accounting for attention utility leads to more extreme attention allocations. Essentially, attention to a dimension generates attention utility, and more so if the payoff associated with that dimension is high, which may be the case when there is a lot of attention devoted it to begin with.

Having presented results characterizing the optimal attention allocation in a general setting, we next discuss their implications in several standard economic environments. We focus on understanding what occurs when the DM must make a single choice about how to allocate attention (i.e. a static decision problem). In Section 2.1 we suppose that each

[2]One could also allow for non-additivity across dimensions, the sub-utility functions, or for non-multiplicative interactions in attention utility. We suspect that most of our results would remain qualitatively similar.

dimension is a separate consumption problem—a reduced-form decision problem where attention may be used to acquire information or take an action. There is extensive empirical evidence consistent with the implications of our general characterization of the optimal attention and actions in this type of environment. For instance, the aforementioned attentional patterns of avoidance behavior and excessive attention depending on payoff levels has been documented both in financial decision-making (starting with Karlsson et al. (2009)) and medical decision-making (e.g., Becker and Mainman (1975) and Oster et al. (2013)) and our model provides a rationalization.[3]

In Section 2.2 we suppose that the different dimensions correspond to different states of the world, each with an associated probability of realizing. The ensuing attention-dependent weight of a state can be interpreted as if it is the subjective probability of that state occurring (although our DM understands the true probabilities perfectly). When attention is non-instrumental (i.e., available actions are independent of the attention allocation), our general comparative statics results imply the DM devotes all to high-payoff states. As a result, the DM appears optimistic (relative to a standard DM), in particular, more risk-seeking and with a preference for positively-skewed payoffs. Such optimism is ubiquitous (Sharot, 2011), and a preference for positively-skewed payoffs has been documented in various settings, e.g., individuals playing lotto (Garrett and Sobel, 1999; Forrest et al., 2002).

But the DM's optimism is not universal and can be mitigated or even overturned into pessimism when attention is predominantly guided by its instrumental value. Indeed, we next show how the kind of probability weighting (subjective probability of a state as a function of its objective probability) depends on the details of the environments. For instance, an inverted-S-shaped probability weighting occurs in environments where a minimal amount of attention to each state is necessary to ensure a good payoff. We thus add to the extensive literature on probability weighting by offering attention as a novel mechanism and predictions on how environments map into the shape of the weighting function.

In Section 3 we extend our results to dynamic settings by allowing the DM to make a choice about contemporaneous attention in each period. Thus DM's behavior is thus

[3] We note that in some of these settings information, a key feature in existing models used to explain such avoidance behavior, is unlikely to play a major role. For instance, Quispe-Torreblanca et al. (2020) finds that individuals log in to their investment portfolio more frequently after a known increase in its value than after a known decrease.

the equilibrium of an intrapersonal game played between selves, which we solve for using backwards induction (and assuming sophistication). In Section 3.1 we begin with a simple assumption that different dimensions represent different time periods. Now attention leads to endogenous weights on time periods that can be interpreted as (endogenous) time preferences. The DM takes action and devotes attention in each period. Hence, our general comparative statics characterize the DM's best response to their future selves' actions and attention allocations. For instance, the DM may discount the future (i.e., they are "present-focused") if the payoff in the present is particularly high or attention to it is of particular instrumental value. But we also study the solution to the DM's full problem, i.e., the set of subgame-perfect equilibria of the game played between the selves across different periods, in particular cases. Here the DM may optimally take actions to create payoffs that vary across time (and avoid consumption-smoothing) to devote attention to high-payoff periods while "ignoring" others. Preference for such "memorable consumption" has been documented (Gilboa et al., 2016; Hai et al., 2020).

In Section 3.2 we explore the implications of a (random) future intra-period attention allocation problem (as when the DM faces multiple concurrent consumption problems) for actions today. We formalize how the DM trades off two considerations. First, when the DM chooses over actions today that can be freely adjusted in the future, but if not adjusted, act as a default, they tend to overweight future bad states. Second, when actions today irreversibly affect future payoffs, the DM acts in a way that overweights future good states. We demonstrate how the tradeoff between overweighting good and bad states varies with the parameters in the model, allowing us to discuss when pessimism or optimism dominate.

In Section 3.3 we study situations in which there is no instrumental value of information, in order to highlight the novel implications of attention utility. We show three implications: the DM acquires more information when the expected payoff is high; attention as a hidden action implies a preference for early information acquisition (broadly consistent with laboratory evidence in (Masatlioglu et al., 2017; Nielsen, 2020)); and the DM exhibits strict preferences over the skeweness of information.

In Section 4, we turn to applying our model to a concrete setting where the DM takes the role of an investor who repeatedly makes a portfolio decision. Our general insights lead to a host of results in this context. First, we find a new mechanism behind the positive relationship between wealth and participation in financial markets (Mankiw and Zeldes, 1991; Poterba and Samwick, 2003; Calvet et al., 2007; Briggs et al., 2021): Participation requires continual attention to one's wealth which low-wealth individuals may want to avoid. Second, this (in)attention on the extensive margin extends to the intensive margin: If the portfolio is performing poorly, the DM may ignore it, which mechanically generates a disposition effect Shefrin and Statman (1985); Odean (1998); Barberis and Xiong (2009, 2012). Third, we can provide conditions under which the DM acts excessively risk-seeking or risk-averse. In particular we show they demand an "attention premium" from assets, i.e., they prefer assets that do not require continual attention, which can help explain premium on equities. Fourth, the DM has a high discount factor because they invest in financial assets as devote attention to their eventual payout, and not vice versa.

In Section 5 we study the implications of our model for interventions of various kinds. These interventions may be effected by either an external policymaker or the DM themselves, in order to improve either the material payoff or total utility of the DM. For example, if an intervention can shift payoffs between consumption problems, total utility will be increased if payoffs are moved towards (away from) problems which already attract a lot of (little) attention. If the goal is to incentive the DM to take specific actions, we show that the effect of an intervention can vary depending on whether it uses rewards (a "carrot") or penalties (a "stick"). While equivalently effective in many standard models, here, the DM may shy away a problem if the penalty reduces the expected payoff. Hence, negative commitment devices, those that penalize for deviation from the action committed to, may in some cases be not only be ineffective, but in fact counterproductive. Lastly, we consider the optimal way to mentally bundle different dimensions. In the rest of the paper we assume that the set of dimensions is exogenously given. However, in reality, whether the DM considers two dimensions as distinct or as one larger dimension (with payoff $V_i + V_{i'}$) may be amendable. For example, the DM may want to ensure that whenever they think about flying a long distance (a low-payoff problem) they also think about the vacation at the end of the flight (a high-payoff problem), thus increasing their attention utility. Given a set of small primitive dimensions, we characterizing the optimal bracketing, providing a microfoundation for one of the key premises of our model—the set of dimensions.

Section 6 discusses how our model relates to several other classes of models. We consider, in turn, models of rational inattention, Bayesian and non-Bayesian models with anticipatory utility, and other models of attention.

Section 7 concludes with a discussion of some of the limitations of our approach and avenues for future research.

2 Model

We consider a decision-maker (henceforth, DM—they) and begin by introducing a general objective to formally express the two fundamental features of attention: attention determines which actions are available to the DM, and 2) generates attention utility. We first characterize the optimal attention allocation in an abstract environment and explore the role of attention utility. We then turn to exploring our models implications in two simple static environments: concurrent consumption problems and uncertain states of the world, and time (Sections 2.1–2.2). This allows us to focus on understanding the implications of our model when there is a single choice of how to allocate attention. In Sections 3.1, 3.3 (Section 3.2) we turn to understanding out model's implications in situations where the agent must make multiple decisions about how to allocate attention, one in each time period. Agents treat the multi-period decision problem as an interpersonal game, forcing us to extend our analysis in this section.

The DM faces a finite number of dimensions $\mathcal{D}$ with generic dimension i.[4] A dimension can correspond to a time period, some consumption problem, an uncertain state, or a combination of these. Each dimension i is associated with a payoff V_i. The DM chooses an (action, attention)-pair denoted by (x, α). The action determines the payoffs from each dimension: given x, the payoff from dimension i is $V_i(x)$, where V_i is continuous in x. Attention is a measure on the set of dimensions with total measure of 1 (a normalization), i.e., $\alpha = (\alpha_i)_{i \in \mathcal{D}}$, where α_i denotes the attention devoted to dimension i with $\alpha_i \geq 0$ and

[4] We take here the dimensions as given. In practice, the boundaries between dimensions may not always be obvious. While we acknowledge this underspecification, we provide some guidance as to how our model can be applied in practice (Section 7) and study a meta optimization problem in which the DM chooses how to define a consumption problem (Section 5.3).

$\sum_{i \in \mathcal{D}} \alpha_i = 1$.[5] As a notational convention used throughout, for any variable that is indexed by $i \in \mathcal{D}$, say b_i, we let $b_{-i} \coloneqq (b_{i'})_{i' \in \mathcal{D} \setminus \{i\}}$.

Attention has two implications. First, the available actions depend on the attention allocation, i.e., given α, action x must be chosen from $X(\alpha)$, where X is compact- and non-empty-valued and upper hemicontinuous. We also assume a form of monotonicity of X: let X be defined on the set of measures on $\mathcal{D}$ with total measure of at most 1; then $X(\mu) \subseteq X(\mu')$ for any such measures μ, μ' if $\mu' \geq \mu$ element-wise. In other words, whether a particular action x is available depends on whether the attention allocation satisfies $\alpha \geq \mu_x$, for some μ_x. This reduced-form formulation of the instrumental value of attention nests a number of canonical decision-problems (as dimension i giving rise to $V_i(x)$), most notably information acquisition where attention α_i to dimension i allows for information x used to increase the payoff V_i. In Appendix A.1 we show formally how our approach nests decision-problems with information acquisition, attention reducing "trembles," and recall of memories (Examples 1–3).

Second, in line with the psychology literature on attention and emotion regulation, the DM values attention utility, a payoff in addition to the material payoff from eventual consumption that is jointly determined by the amount of attention devoted to a dimension and its payoff. The relative importance of attention utility to material payoff is given by parameter λ. In some settings, attention utility can be interpreted as anticipatory utility (Loewenstein, 1987; Caplin and Leahy, 2001), but one that is only generated when the DM devotes attention to (future) consumption.

We view the first consequence of attention—its instrumental value—as relatively standard, and for $\lambda = 0$, it is the only consequence of attention and so the DM maximizes their material payoff. We thus refer to the case when $\lambda = 0$ as the "standard model" and the corresponding DM as the "standard DM." In the general case, the DM's objective is the

[5] Alternatively, one can impose an upper bound on the measure of attention (with no lower bound). Such a model is nested in ours by adding a trivial dimension with payoff 0 to $\mathcal{D}$.

weighted sum of material payoff and attention utility:

$$\underbrace{\sum_{i\in\mathcal{D}}\omega_i V_i(x)}_{\text{material payoff}} + \lambda \underbrace{\sum_{i\in\mathcal{D}}(\alpha_i+\psi_i)V_i(x)}_{\text{attention utility}}, \tag{2}$$

where ω_i and ψ_i are nonnegative parameters that generalize (1) in two ways. First, we allow material payoffs to be weighted differentially across dimensions (via ω_i's). Although in many settings it may be natural to weight different dimensions the same in standard material payoff, in other settings different dimensions may generate differentially weighted material payoff: for example, when the dimensions are different states, the weights are the ex-ante probabilities of each state, or when the dimensions are different time periods, the weight may capture some exogenous discounting of future material payoffs.

Second, we allow the amount of attention utility from dimension i to not only depend on the amount of attention to i and its associated payoff, but also on the dimension itself. This allows us to capture the fact that although attention regulates emotional reactions, it may not be able to fully control it, and so each dimension generate a baseline level of emotional reaction independent of the attention paid to the dimension. [6] This approach will allow us to capture the DM's concerns for (fixed) future attention utility in their objective.

These generalizations enable us to directly employ (2), and the general comparative statics we derive shortly, to characterize the optimal allocation of attention in simple static environments such as attention across consumption problems (Section 2.1) and states (Section 2.2). In dynamic environments, where the DM chooses an (action, attention)-pair in each period (Sections 3.1- 3.3), Equation 2 allows us to characterize the optimal attentional choice and action in any given time period, conditional on the attentional choices and actions in all other time periods. Of course, actual behavior will then be given by the fixed point solution to the set of such best response functions across all time periods, as we explore.

Our formulation is also applicable in situations where overall material payoffs are gen-

[6]Thus, we can nest the case where attention utility is independent of the amount of directed attention α_i (e.g., as is true in anticipatory utility models like Loewenstein (1987); Caplin and Leahy (2001)): let λ go to 0, and ψ_i go to infinity, keeping their product constant.

erated via aggregating across different types of dimensions. For example, our model could address situations where each i is a tuple of time period, consumption problem, and state of the world. In this case there would be a set of time periods, indexed by t; within each time period, individuals would face a set of consumption problems, indexed by c; and each consumption problem may have payoffs that depend on the realized state of the world, s. Thus, it could be that $i = (t, c, s)$.

Note that attention determines the total weight of dimension i (and its payoff) as $\omega_i + \lambda(\alpha_i + \psi_i)$. Attention thus reweights the environment and it is this reweighting of the environment that leads to (as if) belief distortions and time preferences.

To state how the (action, attention)-pairs are optimally determined, it is useful to introduce a parameterization of the environment as well as to define a particular restriction. First, the parameterization: For each dimension i, we fix some $\tilde{V}_i$ and define i's payoff as $V_i = \beta_i \tilde{V}_i + \gamma_i$, for scalars $\beta_i \geq 0$ and γ_i. Intuitively, increasing γ_i shifts the payoff level of dimension i, and increasing β_i increases the payoff difference induced by different actions.

Second, the restriction: We say that the environment is separable if action x is a vector $x = (x_i)_{i \in \mathcal{D}}$, payoff $V_i(x_i, x_{-i})$ is independent of x_{-i} for all i and x_i, and $X(\alpha) = \Pi_{i \in \mathcal{D}} X_i(\alpha_i)$. In words, the DM takes separate actions for each dimension and whether a dimension-specific action is available depends on the amount of attention devoted to the dimension. Note that maximizing (2) with respect to an (action, attention)-pair is then equivalent to maximizing $\sum_{i \in \mathcal{D}} (\omega_i + \lambda(\alpha_i + \psi_i)) \hat{V}_i(\alpha_i)$, where $\hat{V}_i(\alpha_i) := \max_{x_i \in X_i(\alpha_i)} V_i(x_i, \cdot)$; note that $\tilde{V}$ is increasing due to the monotonicity of X.

An increase in the payoff level of a particular dimension i does not affect which (action, attention)-pair maximizes the material payoff and so does not affect the standard (i.e. $\lambda = 0$) DM's solution. However, the attention utility from dimension i increases in proportion to the attention devoted it and so the DM devotes more attention. If the environment is separable, this increase in attention, in turn, allows for a better action. An increase in the payoff difference from different actions increases the importance of taking an action suitable for dimension i. It may also move the payoff up or down (e.g., V_i increases everywhere if $\tilde{V}_i$ is nonnegative), inducing the DM to change their attention just as above. In the proposition below, we offset such level change and the DM always chooses an action better suited for i.

If the environment is separable, the "more suitable" action can only be available if the DM increases their attention.

For the rest of this section, when stating the results, we maintain the assumption, for the sake of simplicity, that the optimal solution is unique. However, we state and prove a general version of the propositions that does not assume unique solutions in Appendix C. All other results in the rest of the paper are also proved there.

Proposition 1. *Consider dimension $i \in \mathcal{D}$. Fix V_{-i} and consider changing parameters (γ_i, β_i) to (γ_i', β_i'). Denote the optimal (action, attention)-pairs for each parameter set as (x, α) and (x', α'), respectively.*

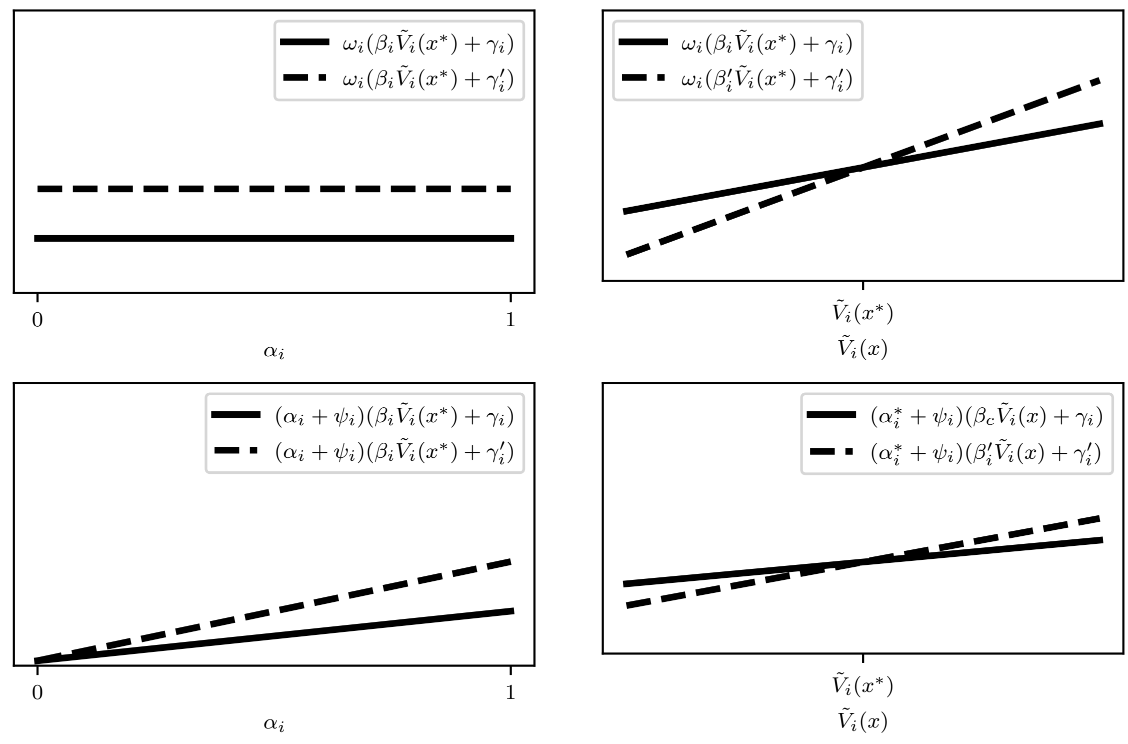
- *If $\gamma_i' \geq \gamma_i$ and $\beta_i = \beta_i'$, then $\alpha_i' \geq \alpha_i$. If, in addition, the environment is separable, then $V_i(x) - V_i(x') \geq \gamma_i' - \gamma_i$.*
- *If $\beta_i' \geq \beta_i$ and $\gamma_i' = \gamma_i - (\beta_i' - \beta_i)\tilde{V}_i(x)$, then $V_i(x') \geq V_i(x)$. If, in addition, the environment is separable, then $\alpha_i' \geq \alpha_i$.*

Thus, increasing the payoff level leads to an increase in attention (the first part of the proposition); increasing payoff differences from different actions leads to an action better suited for the dimension (the second part). And, if the environment is separable, increased attention and better-suited actions go hand in hand.

Figure 1 illustrates Proposition 1. Panel (a) considers an increase of γ_i to γ_i' depicting the material payoff (top figure) and attention utility (bottom figure) of dimension i as functions of attention. The optimal action x^* is held fixed (hence the consumption payoff is independent of α_i). This increase simply shifts the material payoff up (top figure). However, the increase in attention utility is larger for higher α_i (bottom figure). Thus, the DM increases their attention in response (first part of Proposition 1).

Panel (b) considers an increase in β_i to β_i' with an offsetting change in γ_i to γ_i' as described in the second case of Proposition 1 and shows the material payoff (top figure) and attention utility (bottom figure) of dimension i as functions of $\tilde{V}_i(x)$. Throughout, the optimal attention α_i^* is held fixed. This change then pivots the payoff around its initial optimal value (top figure). Already here, the DM benefits relatively more from increasing

$\tilde{V}_i(x)$ than before. The same pivoting occurs for attention utility (bottom figure). Thus, the DM increases $\tilde{V}_i(x)$ in response to the change (second part of Proposition 1).



(a) (b)

Figure 1: Panel (a) corresponds to an increase of γ_i to γ_i'. We hold the optimal action x^* fixed. The payoff V_i is shifted up and hence so is the material payoff (top figure), independently of α_i. However, increasing α_i now increases the attention utility now by more than before (bottom figure). Panel (b) corresponds to an increase of β_i to β_i' with an offsetting change of γ_i to γ_i'. We hold the optimal attention α_i^* fixed. The material payoff pivots around $\tilde{V}_i(x)$ (top figure) as does attention utility.

We next turn to understanding how the optimal (action, attention)-pair depends on the novelty of our model and vary the relative weight on attention utility λ. We show three results. First, a standard DM (one with $\lambda = 0$) maximizes the material payoff only; one can think of the standard DM as fully utilizing the instrumental value of attention. A DM with $\lambda > 0$, instead, may not maximize material payoff in favor of attention utility. We show that the material payoff decreases monotonically as the weight on attention utility increases. Second, the DM's objective is evidently linear in payoff levels γ_i and hence, the DM's value, i.e., (2) for optimal (action, attention)-pairs, is convex. Third, when the environment is separable and $\lambda > 0$, then increasing attention to dimension i increases both the payoff V_i but also the weight on dimension i; hence, accounting for attention utility makes the DM's objective more convex. The proposition below provides the formal statements.

Proposition 2. *Consider a change of parameter λ to λ' and let x and x' denote the optimal action, respectively. If $\lambda' > \lambda$, then*

- $\sum_{i\in\mathcal{D}} \omega_i V_i(x) \geq \sum_{i\in\mathcal{D}} \omega_i V_i(x')$*;*
- *the DM's value is convex in* $(\gamma_i)_{i\in\mathcal{D}}$*;*
- *if the environment is separable and the objective given λ is convex in α it is the case that it is also convex in α given λ'.*

The first part of the proposition states that the DM monotonically departs from the standard DM, who fully utilizes the instrumental value of attention, as λ increases. In other words, through a standard lens, that is looking at material payoff, the DM's action becomes worse as λ increases.

The second part implies that the DM has a preference for "extreme" payoffs: they prefer to have more varied payoffs (holding the average payoff fixed). This is because not only do material payoffs in dimension i increase with γ_i, but the DM can re-allocate attention to take further advantage of this increase (and conversely mitigating a decrease of $\gamma_{i'}$). One way of interpreting this result is that the DM has a preference for specialization—it is better to be outstanding in one area and relatively poor in many others, rather than doing mediocre in all.

The third part states that when payoffs are increasing in attention, as is the case in a separable environment, then the DM's objective may be convex in attention, even though it may not be for a standard DM. Accounting for attention utility may thus lead to more extreme or "sparse" attention allocations (Gabaix, 2014) relative to those of a standard DM. In Section **??**, we explore how the convexity of the objective discussed here relates to a demand for (non-instrumental) information (Section 3.3) and a preference for assets with varied return (Section 4).

Lastly, we note the effects of ω_i, the weight on payoff V_i in the material payoff, and ψ_i, the involuntary or future attention devoted to dimension i. Note that both ω_i and ψ_i play a similar role to β_i in Proposition 1. Thus, the following proposition follows straightforwardly and a formal proof is omitted.

Proposition 3. *Consider dimension $i \in \mathcal{D}$. Fix V_{-i} and consider changing parameters (ω_i, ψ_i) to (ω_i', ψ_i'), with $(\omega_i', \psi_i') \geq (\omega_i, \psi_i)$ elementwise. Denote the optimal (action, attention)-pairs for each parameter set as (x, α) and (x', α'), respectively. Then $V_i(x') \geq V_i(x)$. If, in addition, the environment is separable, then $\alpha_i' \geq \alpha_i$.*

We explore the implication of Proposition 3 and the other general comparative statics in more specific contexts next.

2.1 Attention across consumption problems

The first type of dimension we consider is concurrent consumption problems. There is only a single period and the DM cannot allocate attention across the random realizations; instead, nontrivial attention allocation occurs across a set of consumption problems $c \in \mathcal{C} = \mathcal{D}$. (We alter notation from i to c for convenience; nothing of substance has changed.) Each problem c is a reduced-form of some underlying optimization problem that only depends on action x whose availability depends on the attention allocation. As eluded to previously, this approach nests, e.g., information acquisition, attention-reducing "trembles," and memory recall (Examples 1–3 in Appendix A). For more concrete examples, problems may be 'arranging a retirement home for a relative', 'vacation', 'personal health', 'financial situation', etc., all of which are associated with some payoff that may be improved by some action that may require attention.

All consumption problems are weighted equally in the material payoff, i.e., $\omega_c = 1$, and there is no exogenous emotional utility flow that is outside the DM's control, i.e., $\psi_c = 0$. The DM's objective is thus to choose (x, α) with $x \in X(\alpha)$ to maximize

$$\sum_{c \in \mathcal{C}} V_c(x) + \lambda \sum_{c \in \mathcal{C}} \alpha_c V_c(x).$$

We first discuss how the payoffs across consumption problems determine the optimal (action, attention)-pair, i.e., Proposition 1. We begin by discussing how the first part of the proposition, which says that attention varies with the payoff level, whenever $\lambda > 0$, corresponds to findings in empirical research. Both evidence of (attentional) avoidance of low-payoff situations, as well as excessive attention to high-payoff situations, has been extensively documented in economics, health, psychology, and related fields.

For instance, retail investors' propensity to check their portfolios generally comoves with the market (both with market levels and changes; see Karlsson et al. (2009); Sicherman et al. (2015), although Gherzi et al. (2014) finds increased monitoring following market downturns). This behavior, in particular the tendency to avoid one's portfolio in bad market conditions, is often referred to as the "ostrich effect"—individuals bury their figurative heads in the sand like an ostrich.[7] Such behavior is consistent with our model. Accessing one's portfolio to (for example) acquire information about the performance of one's portfolio after receiving news about the aggregate market requires attention to the portfolio. Our model suggests that doing so co-varies with the payoff associated with the portfolio; reasonably, this payoff may be eventual consumption. A down market (for most investors) implies low future consumption, and by decreasing attention to their portfolio, investors can improve their attention utility.[8]

Information avoidance may be a contributing factor to such ostrich effect behavior. In fact, Karlsson et al. (2009) define the ostrich effect in their context as "avoiding exposing oneself to information that one fears will cause psychological discomfort." Our reduced form approach to modeling consumption problems nests problems of information acquisition (see Example 1 in Appendix A), and so our predictions are in line with such an explanation. However, our model predicts that, even when no additional information can be acquired, attention will still be devoted to high-payoff problems. And indeed, there is evidence suggesting that information avoidance may not fully explain the ostrich effect: Sicherman et al. (2015) find a 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

[7] It appears that the term was coined in Galai and Sade (2006), where it describes individuals avoiding risky financial situations by pretending they do not exist. Although the idea of ostriches burying their heads in the sand to avoid predators dates back to Roman times, they do not display this behavior. Instead, they put their heads into their nests (which are built on the ground) to check temperatures and rotate eggs.

[8] While the propensity to check one's portfolio comoves with the market in both levels and changes, individuals may be avoiding payoffs that are low relative to some reference point (and not absolutely). Our model can be enriched to capture such behavior by supposing that attention utility is proportional to consumption payoffs relative to some reference point.

can be revealed—and Quispe-Torreblanca et al. (2020) find that individuals devote excessive attention to positive information that is already known. In a related setting, Olafsson and Pagel (2017) study individuals' attention to their financial accounts and finds increased attention after they are paid and decreased attention when the account balance becomes low, in particular, when it turns negative. Arguably individuals often know about their payment dates and amounts and their overdrawn status, so information avoidance may be an implausible motive.

Similar behavior has been documented in other domains. For instance, researchers have noted low rates of testing for serious medical illnesses (Huntington's disease (Shouldson and Young, 2011; Oster et al., 2013); sexually transmitted diseases (Ganguly and Tasoff, 2017)). Our model predicts that an individual at risk of such a disease may have a low (expected) payoff related to the consumption problem 'health' and hence avoids any actions, such as taking a test, that require attention to it. Indeed, Ganguly and Tasoff (2017) document that the demand for medical testing for sexually transmitted diseases is decreasing as the expected health outcome worsens.

Our model also predicts that, to the extent that taking a non-default action requires attention, individuals will avoid altering defaults in low-payoff problems. The health literature has found that individuals often fail to follow medical recommendations, both with respect to information-generating activities (e.g., self-screening) but also with non-information-generating activities, such as taking medicine (Becker and Mainman, 1975; Sherbourne et al., 1992; Lindberg and Wellisch, 2001; DiMattero et al., 2007). For instance, DiMattero et al. (2007) find that, among individuals experiencing serious medical conditions, individuals with worse health status tend to adhere less to medical regimes.

Avoyan and Schotter (2020) provide evidence in a stylized laboratory environment, where experimental participants choose to allocate time (i.e., attention) between two games (i.e., consumption problems). In line with our model, they find that "the game with the largest maximum payoff attracts more attention, as does the game with the greatest minimum payoff" and further that "games that have zero payoffs attract less attention than identical games in which all payoffs are positive."

The second part of Proposition 1—once controlling for payoff levels, increasing the payoff differences from different actions on a particular dimension leads the DM to choose an action better suited for that dimension—is already present for the standard DM and indeed, simply put, much of economics builds on this premise, and so we refrain from discussing related evidence.

The aforementioned evidence suggests an important role of attention utility in determining behavior and material payoffs (and hence $\lambda > 0$). Our model suggests that the presence of attention utility leads to worse material payoffs (first part of Proposition 2). And indeed, low compliance may negate the benefits from medical therapies or limited medical testing in the context of Hunting's disease leads to worse decision about "childbearing, retirement, education, participation in clinical research" (Oster et al., 2013).

While we consider the remaining comparative statics of Proposition 2, pertaining to the convexity of the DM's value in payoff levels and objective in attention, intuitive, we do not know of empirical work speaking to them.

Proposition 2 allows for a continuous variation on λ, i.e., it goes beyond simply comparing a DM with $\lambda > 0$ to the standard DM. Indeed, in principle, one may expect λ to vary by individual or circumstances. To what extend λ is a personal trait and situation-dependent is, to our knowledge, an open question and we welcome quantitative studies estimating λ.

2.2 Attention across states

We next consider attention allocation across uncertain states and the implications of the two fundamental features of attention in this context. That is, there is a single consumption problem (and time period) but the DM can devote attention across a set of realizations of an uncertain state $s \in \mathcal{S} = \mathcal{D}$. This formulation is applicable to consumption problems as discussed in Section 2.1, but where the DM is given additional flexibility in their attention allocation now to payoff-relevant states instead of unconditional consumption problem and its expected payoff.

A state s is weighted by $\omega_s = p_s$ in the DM's material payoff, where p_s denotes the objective probability of state s occurring. We maintain that $\psi_s = 0$, i.e., there is no exogenous emotional utility flow outside the DM's control. The DM's objective is then to choose (x, α) with $x \in X(\alpha)$ to maximize

$$\sum_{s \in \mathcal{S}} p_s V_s(x) + \lambda \sum_{s \in \mathcal{S}} \alpha_s V_s(x),$$

i.e., the expected material payoff plus attention utility.

It is useful to divide the DM's objective by $1+\lambda$ (an affine transformation) and gather the terms in front of V_s to define $q_s := \frac{p_s + \lambda \alpha_s}{1+\lambda}$. Note that $q_s \in [0, 1]$ for all s and $\sum_{s \in \mathcal{S}} q_s = 1$, i.e., q_s describes a probability distribution. The DM, conditional on their attention allocation, behaves like a subjective expected payoff maximizer, but one who assigns probability q_s to state s. Note that as attention to state s increases, so does the subjective probability q_s assigned to that state. Thus, our general comparative static results for α_s (Propositions 1–3) have direct implications for q_s.

For example, Proposition 1 implies that individuals devote more attention to high-payoff states and thus (at least in a separable environment) will take actions suited for those states relative to those with a low payoff. Simply put, in the context of an individuals devising a "plan" for different contingencies, they will know what to do with a financial windfall (as they have contemplated such contingency) but not which expenses to cut when they are laid off (as this scenario has been ignored). While perhaps intuitive, we do not know of empirical studies testing this prediction.

Proposition 2 indicates that these plans lead to a lower material payoff. Again, while we do not know of studies testing this comparative static, we highlight day dreaming as an activity where individuals devote attention to a low probability state, although doing so clearly provides low expected material utility.[9]

Proposition 3 immediately implies a standard intuition: increasing the objective probability of a state p_s leads to an action better suited for that state.[10] Such a comparative static, in addition to the second part of Proposition 1, is standard in economics and we

[9] Anecdotally, individuals engage in "Zillow surfing," a form of escapism where they browse through home buying sites and imagine themselves in different houses, possibly much more expensive ones compared to their current accommodation.

[10] The one nuance in applying Proposition 3 is that there is a constraint on the set of probabilities: increasing the probability of one state means reducing the probability of another. In order for the result to hold it must be the case that the probability shift to s comes from a "trivial" state—one where attention has no material benefit.

refrain from discussing it further.

We now turn to understanding novel results when the dimensions are states—the implications of our model for q_s. We begin by supposing that there is no instrumental value of attention, that is, the available actions $X(\alpha)$ do not depend on α and consider the DM's preference over lotteries. Thus, let X be the set of available lotteries. (With minor abuse of notation) given a set of states $S' \subseteq \mathcal{S}$, lottery $x \in X$ has a monetary payoff $x_{S'}$ in all states $s' \in S'$. The DM is equipped with a Bernoulli utility u and hence, the payoff in state s given lottery x is $V_s(x) = u(x_s)$. For the second and third cases of the following proposition, we consider binary lotteries, those where any state either pays a low payoff $L(x)$ or a high payoff $H(L)$ (with $L(X) < H(x)$). It will be useful to let $X(\mu, L)$ denote the set of binary lotteries with mean μ and low payoff L.

Proposition 4.

- *Let DM(λ) refer to the DM given λ. DM(λ) is more risk-averse than DM(λ') for any $\lambda' > \lambda$.*[11]
- *Suppose u is unbounded and $\lambda > 0$. For any μ, L and $x \in X(\mu, L)$, there exists a lottery $\hat{x} \in X(\mu, L)$ so that if a lottery $x' \in X(\mu, L)$ has high payoff $H(x') > H(\hat{x})$ then the DM's prefers x' to x.*
- *For any μ, L and $x, x' \in X(\mu, L)$ with $H(x) > H(x')$, the DM prefers x to x' if λ is large enough.*

Proposition 4 first states that the DM has an additional preference for risk. Intuitively, given a lottery x, the DM devotes attention to the high-payoff states—the "upside" of the lottery—resulting in those states receiving a higher subjective probability q_s, i.e., the upside is as-if more likely. The second and third cases of the proposition state that the DM has a preference for positively skewed lotteries (where the skew of a lottery is defined as its third standardized moment; fixing a low outcome and a mean for a set of binary lotteries, comparing the skewness of two lotteries is equivalent to comparing their high payoffs). Intuitively, positive skew always increases attention utility since the DM devotes

[11] Let δ_y be the lottery with monetary payoff y in each state. Given two preference relations on the set of lotteries, $\succeq$ and $\succeq'$, $\succeq$ is more risk averse than $\succeq'$ if $x \succeq \delta_y \implies x \succeq' \delta_y$ for all lotteries x and payoff y.

their attention exclusively to high-payoff states. Then, if the high payoff is large enough (second case) or the DM puts enough weight on attention utility (third case), the DM has a preference for such an increase in the skew of a lottery.

An implication of Proposition 4 is that individuals appear more optimistic than objective probabilities warrant. Such optimism has been documented in a wide range of circumstances. Sharot (2011) summarizes: "we underrate our chances of getting divorced, being in a car accident, or suffering from cancer. We also expect to live longer than objective measures warrant, overestimate our success in the job market, and believe that our children will be especially talented." In our model, the DM devotes little attention to low-payoff states, thus acting as if they "underrate" them, and, conversely, overweights the high-payoff ones. There is laboratory evidence consistent with optimism, e.g., Mayraz (2011).[12] There, participants guess the realization of a random variable and are rewarded for accuracy, and some participants for high and others for low realizations. Our model predicts that participants whose payoff is high for high realizations devote attention to those realizations, which then have a large weight in their objective, leading them to guess a high realization. Indeed, this is what Mayraz (2011) finds.

There is also extensive evidence for preferences for positively skewed lotteries in the context of portfolio choice (Blume and Friend, 1975), betting on horses (Golec and Tamarkin, 1998; Jullien and Salanié, 2000; Snowberg and Wolfers, 2010), individuals playing lotto (Garrett and Sobel, 1999; Forrest et al., 2002), as well as in various laboratory settings (Ebert and Wiesen, 2011; Grossman and Eckel, 2015; Ebert, 2015; Åstebro et al., 2015; Dertwinkel-Kalt and Köster, 2020). Furthermore, consistent with our model, Jullien and Salanié (2000); Snowberg and Wolfers (2010) suggest that the preference for skewness is driven by subjective probabilities, as in our model, rather than the Bernoulli utility u.

Proposition 4 relies on the absence of instrumental value of attention; we next consider how the presence of it, and the details of how attention is valuable, can affect the attention allocation and thus the ensuing subjective probabilities. Our results focus on describing the mapping of objective probabilities p_s to subjective probabilities q_s—i.e., the probability weighting function—which we denote $q_s(p_s)$. For simplicity, we focus on the situation with

[12] See also Mijović-Prelec and Prelec (2010); Engelmann et al. (2019); Orhun et al. (2021) for related evidence in both monetary and non-monetary domains.

only two states, $\mathcal{S} = \{s, s'\}$ and consider separable environment.

First, we consider probability weighting when there is no instrumental value (i.e., a counterpart to Proposition 4). Here, the DM devotes full attention to the state with the higher payoff and behaves as if they overweight this state. Second, we suppose that each state requires some minimum amount of attention, or otherwise, the payoff in that state is extremely low, and demonstrate this leads the DM to overweight low-probability states, i.e., the probability weighting takes the form of an inverse-S-shape. Third, instead of decreasing returns to attention as in the second case, we suppose that the returns to attention are increasing, i.e., the payoff in a state is convex in the amount of attention devoted to that state; then, the DM may not devote any attention to a low-probability state while devoting full attention if that state is relatively likely, i.e., the probability weighting is S-shaped.

Proposition 5. *Suppose there are two states, $\mathcal{S} = \{s, s'\}$ and the environment is separable.*

- *If $\hat{V}_s$ and $\hat{V}_s$ are constant and $\hat{V}_s > \hat{V}_{s'}$, then:*

$$q_s(p_s) = \frac{p_s + \lambda}{1 + \lambda}, \quad \textit{and} \quad q_{s'}(p_{s'}) = \frac{p_{s'}}{1 + \lambda}$$

- *Suppose $\hat{V}_s = \hat{V}_{s'} = \hat{V}$, $\hat{V}$ is continuously differentiable, $\lim_{a \to 0} a \frac{\partial}{\partial a} \hat{V}(a) = \infty$, and $\frac{\partial}{\partial a} \hat{V}(1) < \infty$. Then, $q_s = q_{s'} = q$ and there exist some $\bar{p}$ with $0 < \bar{p} < 1/2$, such that*[13]

$$q(p) \begin{cases} = 0 & \textit{if } p = 0 \\ > p & \textit{if } 0 < p < \bar{p} \\ < p & \textit{if } 1 - \bar{p} < p < 1 \\ = 1 & \textit{if } p = 1. \end{cases}$$

[13] Although this result generates two classic features of inverse-S-shaped probability weighting (underweighting of high probabilities and overweighting of low probabilities), the probability weighting need not be concave and then convex (as is often assumed). Intuitively, the instrumental value of attention needs to be small for high values of attention, i.e., $\hat{V}(1) - \hat{V}(1/2)$ small, to guarantee the inverse-S shape probability weighting everywhere.

- *Suppose $\hat{V}_s = \hat{V}_{s'} = \hat{V}$ and that $\hat{V}$ is convex and not constant. Then, $q_s = q_{s'} = q$ and*

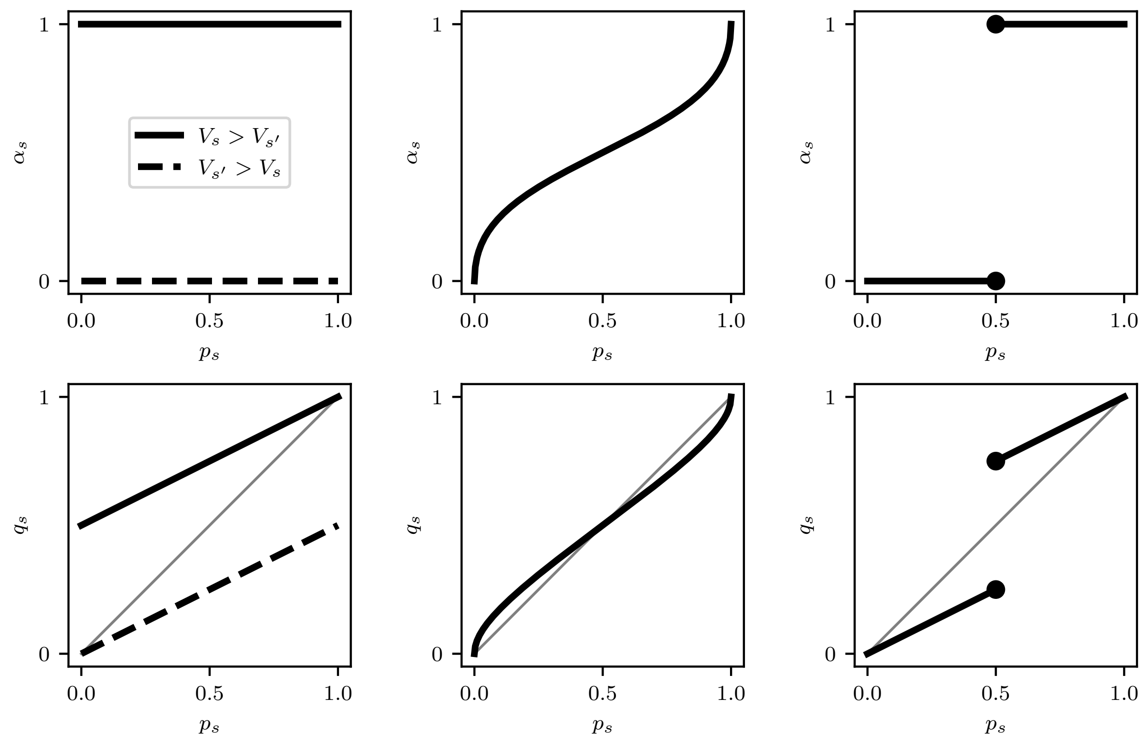
$$q(p) = \begin{cases} \frac{p}{1+\lambda} & \text{if } p < \frac{1}{2} \\ \{\frac{\frac{1}{2}}{1+\lambda}, \frac{\frac{1}{2}+\lambda}{1+\lambda}\} & \text{if } p = \frac{1}{2} \\ \frac{p+\lambda}{1+\lambda} & \text{if } p > \frac{1}{2}. \end{cases}$$

Figure 2 illustrates the different forms of probability weighting (and attention allocations) occurring in the three cases discussed in Proposition 5. A panel corresponds to an environment with the top subfigure showing the optimal attention, α_s, as a function of the probability with which state s occurs, p_s, and the bottom subfigure showing the resulting probability weighting, $q_s(p_s)$. Panels (a) and (c), which correspond to environments with no instrumental value of attention and increasing returns to attention, respectively, are straightforward; in fact, the allocation of attention and ensuing probability weighting are independent of the specifics of the environment beyond these assumptions. The probability weighting does, however, depend on λ, which governs the degree of reweighting of the environment. Throughout, we chose $\lambda = 1$. For Panel (b), which visualizes the second case where each state requires some minimum amount of attention, the details of $\hat{V}$ matter; we use $\hat{V}(a) = -\frac{1}{a}$ as tractable functional form.[14]

Probability weighting has been extensively studied since first discussed in Kahneman (1979), and there is now a voluminous literature analyzing and empirically estimating prospect-theory models (see Wakker (2010); Barberis (2013) for two surveys). We know of few models that link the shape of the weighting function to environmental details like the level of payoffs across states. The classic finding is that individuals' probability weighting follows an inverse-S shape (Wu and Gonzalez, 1996)—which our model can generate under the conditions described in Proposition 5. Thus, it provides a mechanism giving rise to the inverse-S shape. However, in contrast to much of the literature, our model predicts other forms of probability weighting, e.g., an S-shaped probability weighting, if the details of the environment change. To our knowledge, these kinds of predictions are yet to be tested.[15]

[14] One can easily show that the optimal attention is $\alpha_s = (p_s - \sqrt{p_s(1-p_s)})/(2p_s - 1)$.

[15] Our model has implications for probability weighting that distinguishes it from other models. For example, cumulative prospect theory (Tverseky and Kahneman, 1992) or rank dependent utility (Quiggin,



(a) (b) (c)

Figure 2: Panels (a), (b) and (c) correspond to the three environments discussed in Proposition 5—no instrumental value, minimum attention requirement, increasing returns to attention—respectively. The weight on attention utility is always $\lambda = 1$. The functional forms in Panels (a) and (c) do not depend on the specifics of the environment and are given in the proposition. For the environment depicted in Panel (b), the consumption payoff as a function of attention is given by $\hat{V}(\alpha) = -\frac{1}{\alpha}$.

3 Extension: Multi-period Attentional Allocation

The previous section explored the implications of our model an environment where attention is allocated once. However, in many important situations a DM may engage in repeated choices about attention — a choice each time period. Because agents cannot commit their future selves to a particular attentional allocation, the solution to the DM's problem is the equilibrium of an intrapersonal game played between each period's self.

Thus, we must extend our analysis in the previous section. As discussed, one can treat

1982) predicts that the probability assigned to a state (q_s) depends only on the ranking of the states ($V_s > V_{s'}$ or $V_s < V_{s'}$) and the objective probabilities of each state occurring (p_s). In contrast, in our model, q_s additionally depends on the difference in payoffs, $V_s - V_{s'}$, as well as the instrumental value of attention.

the results in Proposition 1-3 as defining the best response of the DM in a single period, given the choices of the DM in all other periods.

Of course, these interpersonal games requires us to make an assumption about what any given time period's self believes about the strategies of other players. As a benchmark, we make the assumption that our DM is fully sophisticated about their future choices. Thus, our results our driven solely by the role of attention in generating utility, rather than any kind of misprediction. Bronchetti et al. (2020) provides evidence of some albeit not full sophistication, while Falk and Zimmermann (2016)) provide evidence that individuals seem to alter their choices in anticipation of being able to redirect attention in the future.

3.1 Attention across time periods

We next consider a third and final environment: attention across time. The DM now faces a sequence of time periods $\mathcal{T} = \{1, \ldots, T\} = \mathcal{D}$, with generic period t. Each period is associated with a single consumption problem and there is no uncertainty (we consider the implications of a dynamic model features multiple consumption problems or states in Section 3.2). For simplicity, we assume that there is no exogenous discounting, i.e., $\omega_t = 1$ for all t. However, the attention-dependent weights on the different periods (the dimensions) can be interpreted as endogenous time preferences.

In each period t, the DM chooses an (action, attention)-pair denoted by (x_t, α_t), i.e., there are multiple "selves." As as consequence, we need to be careful in applying Propositions 1–3 to this setting. The actions jointly determine the payoff in each period: given $x := (x_t)_{t=1}^T$, the payoff in period t is $V_t(x)$ with natural assumptions on future actions' impact on past payoffs. Attention is a measure on the set of dimensions, now time periods, with total measure 1 i.e., $\alpha_t = (\alpha_{t\to t'})_{t'\in\mathcal{T}}$, where $\alpha_{t\to t'}$ denotes the attention (in period t) devoted to period t'; with $\alpha_{t\to t'} \geq 0$ and $\sum_{t'\in\mathcal{T}} \alpha_{t\to t'} = 1$; we also let $\alpha = (\alpha_t)_{t=1}^T$. We assume that the available actions in period t only depend on attention in period t, i.e., x_t must be in $X_t(\alpha_t)$ which is compact- and non-empty-valued and upper hemicontinuous.

As before, in each period, the DM receives a material payoff and attention utility. We assume that the DM maximizes the sum of such payoffs across periods and first consider the best response function in any period t holding fixed x_{-t}, α_{-t}: The DM chooses (x_t, α_t)

with $x_t \in X_t(\alpha_t)$ to maximize

$$\sum_{t'=t}^{T} \left(\underbrace{V_t(x_t, x_{-t})}_{\text{material payoff in } t'} + \lambda \underbrace{\sum_{t''=1}^{T} \alpha_{t' \to t''} V_{t''}(x_t, x_{-t})}_{\text{attention utility in } t'} \right). \tag{3}$$

Notice that (3) can be written as (2) for $\psi_t = 0$ and $\psi_{t'} := \sum_{t''=1}^{T} \alpha_{t' \to t''}$ for $t' \neq t$ (and $\omega_{t'} = 1$ for all t').

Note that in period t the weight on period t' is given by $1 + \lambda(\alpha_{t \to t'} + \psi_{t'})$. These weights across periods t' can be interpreted as discounting: Fixing (x, α) the DM behaves like a standard DM, but one who discount period t' (relative to period t) by $\delta_{t \to t'} := \frac{1+\lambda(\alpha_{t \to t'} + \psi_{t'})}{1+\lambda \alpha_{t \to t}}$. Thus, e.g., as attention to the present period increases, the DM discounts future periods by more. In other words, the time discounting—whether the DM is present or future-focused—is thus endogenous and depends on circumstances.

Propositions 1–3 then provide simple comparative statics with predictions on time preferences. In particular, Proposition 1 suggests that the DM may weight a period more, if the payoff level or the instrumental value of attention to that period increases. Proposition 3 implies that some (exogenous) flow emotional utility derived from period t' leads the DM to increase the payoff in period t'.

The best response function formulation, i.e., holding (action, attention)-pairs in other periods fixed, provides potentially useful intuition. However, the DM may be sophisticated and understand how their current (action, attention)-pair affects future behavior (Bronchetti et al., 2020). We make this assumption and solve the DM's problem by backward induction (and so solve for the subgame perfect equilibria of the intrapersonal game). We demonstrate that our model naturally leads to apparent non-exponential discounting (a well-documented fact in economics, biology, and related social sciences, surveyed by Frederick et al. (2002)), which is often, although not exclusively, in the direction of excessive over-weighting of present consumption (i.e., $\delta_{t \to t+1}$ decreasing in t; see Thaler (1981) for an early example of this phenomenon).

To this end, let $\mathcal{H}_t := (x_{t'}, \alpha_{t'})_{t'=1}^{t-1}$ denote the (action, attention)-pairs the DM chose up to (and excluding) period t. Let $\Gamma_t(\mathcal{H}_t)$ denote the set of credible (x, α) when the DM has chosen $\mathcal{H}_t$ so far and now chooses (x_t, α_t), where credibility requires that the DM in each future period chooses their corresponding (action, attention)-pair optimally. Specifically, for $t < T$, $\Gamma_t(\mathcal{H}_t)$ is recursively defined as argmax of (3) over (x, α), with $(x, \alpha) \in \Gamma_{t+1}(\mathcal{H}_t, (x_t, \alpha_t))$ and $x \in X(\alpha)$; and $\Gamma_T(\mathcal{H}_T)$ as the argmax of (3) over (x, α), with $(x, \alpha) \in \{\mathcal{H}_T, (x_T, \alpha_T)\}$ and $x \in X(\alpha)$.

Our comparative static results (Propositions 1–3) suitably adjusted cease to hold due to coordination motives in the DM's problem. Example 4 in Appendix A.2 provides an example where increasing a future payoff leads to less attention to that period; Example 5 shows that varying λ can affect the material payoff non-monotonically.

However, under reasonable restrictions our model generates clean results linking as-if discounting and λ. We consider an environment where the payoff in a period can be written as the sum of attention devoted to that period; formally: suppose that x_t takes the form $x_t = (x_{t \to t'})_{t'=1}^T$, let $X(\alpha) = \{x : x_{t \to t'} \leq \alpha_{t \to t'} \forall t, t'\}$, and suppose $V_t(x) = V(\sum_{t'=1}^t x_{t' \to t})$ for some increasing function V. We assume the symmetry across periods (except that attention to past periods cannot retrospectively increase the payoff) to not bias the model, at least directly, in favor of a particular attention allocation.

Patterns of attention vary depending on whether attention across periods—i.e., between $\alpha_{t \to t''}$ and $\alpha_{t' \to t''}$—are complements or substitutes, and this environment is sufficiently rich to accommodate both. Complementarities may arise both because an increase in $\alpha_{t \to t''}$ makes the value of attending to t'' in all other periods higher, as well as because V is convex. The former mechanism is novel to our approach, and so to focus on it we rule out the latter, assuming V is concave. The following result demonstrates that whether or not attention across periods are substitutes or complements depends on λ.

To simplify the statement of the following proposition, in addition to the assumption of concave V, we also assume that the V is satiated at exactly K, i.e., $V(K) = V(K')$ for all $K' \geq K$ and $V(K) > V(K')$ for all $K' < K$, and suppose K is a divisor of T.[16]

Proposition 6. *There exist $\underline{\lambda} > 0$ and $\bar{\lambda} < \infty$, such that*

[16]If V was not strictly concave, then $\underline{\lambda}$ in the ensuing proposition could be 0. If K is not a divisor of T, then when $\lambda > \bar{\lambda}$ (again, defined in the proposition) the last payoff in period T would be less than those in other high-payoff periods.

- *if* $\lambda < \underline{\lambda}$*, in each period* t*,* $\alpha_{t \to t} = 1$*; and*

- *if* $\lambda > \bar{\lambda}$*, in each period* t*,* $\alpha_{t \to K(t)} = 1$ *where* $K(t) \equiv \lceil \frac{t}{K} \rceil K$*.*

The proposition shows that two patterns of attention and payoffs emerge, and λ governs which one. Moreover, because attention drives as-if discounting, λ determines the pattern of discounting.

When λ is small, i.e., the DM is close to a standard DM, the DM maximizes the material payoff: Here, due to the concavity of V, this is achieved by devoting full attention to the present period t (in each period). This implies that the individual acts as if they are non-exponential discounters where the present is over-weighted: in period t, the DM discounts each period $t' > t$ by $\frac{1}{1+\lambda}$. The DM thus falls in the class of quasi-hyperbolic discounters (Laibson, 1997). In light of our model, the widely used term "present focus," rather than "present bias," to describe this behavior stylized fact is indeed the more aptly chosen term to describe decreasing discount factors.[17]

When λ is large, the additional complementarity of attention from different periods to a particular period outweighs the concavity of V. Consequently, the DM allocates attention to generate periods with a particularly high payoff which are subsequently exploited for attention utility. In between those periods, the payoff is low; however, as the DM does not devote attention to these periods, including when they are in one of those periods, their attention utility is unaffected. Gilboa et al. (2016) and Hai et al. (2020) point out that individuals often construct such payoff sequences; that is, they do not fully smooth payoffs but rather have periods of "memorable consumption"—weddings, vacations, and celebrations. For instance, Hai et al. (2020) notes that the average expenditure on weddings is about USD 20,000 and that the average annual household income of a newly married couple is USD 55,000. Our prediction is consistent with such payoff patterns. In these cases, the DM is present-focused (in terms of the endogenous discount factors) in periods with such "memorable consumption" and future-focused (with respect to those periods) otherwise. The existence of future focus has found support, albeit more limited than present focus (e.g. Loewenstein and Sicherman (1991); Åstebro et al. (2015); Rubinstein (2003)).

[17] In fact, the recent survey article on intertemporal choice by Ericson and Laibson (2019) has present focus as the second keyword, and present bias as the third.

We do not know of systematic empirical evidence testing a key prediction of our model—that the structure of discounting should change with the weight on attention utility and so the economic environment. The closest is that our agents should discount periods with high payoffs less, because they attract more attention. Magnitude dependent discounting is a well known empirical regularity (e.g., Green et al. (1997) is an early paper), although it has not been directly linked to attention.

3.2 Defaults

We now consider an enriched setting, which allows us to understand how consumption problems, states and time may interact in generating novel behavior. Our results are two-fold. First, we consider how the DM chooses "pure defaults" where a period-2 payoff from a consumption problem depends on the action taken in period 1, but only if the DM does not devote attention to the consumption problem in period 2. We show the DM anticipates their inattention to the consumption problem when its payoff is low and chooses a default optimal for those realizations, i.e., one that is "pessimistic." We then suppose that the Period 1 action is an impure default: it affects the future payoff even when the DM does devote attention. In this case, the DM may choose an "optimistic" action, one that performs well if the future payoff is high since that is the state the DM devotes attention to today. Our model can thus, depending on the effect of period 1 actions on future payoffs, endogenously generate either as-if pessimism or as-if optimism.

Formally, there are two periods: period 1 and period 2. The setup in period 2 is that of Section 2 with two consumption problems, $\mathcal{D} = \mathcal{C} = \{c_0, c_1\}$, where the payoff from problem c_0 is parameterized by some state s and the DM's action in the first period, and c_1 is trivial, in that its payoff V_{c_1} is independent of any action.

In period 1, the DM chooses a default action that they can revise in period 2. Formally with state s realized; if $\alpha_{2\to c_0} < \eta$, then $V_{c_0}(x_1, x_2|s)$ (with $x_2 \in X_2(\alpha_2)$) is independent of x_2, and if $\alpha_{2\to c_0} \geq \eta$, then $\max_{x_2 \in X_2(\alpha_2)} V_{c_0}(x_1, x_2|s)$ is independent of x_1 for all x_1. In period 2, the DM knows the realized state s and chooses (x_2, α_2) with $x_2 \in X_2(\alpha_2)$ to maximize the material payoff and their attention utility, i.e.,

$$V_{c_0}(x_1, x_2|s) + V_{c_1} + \lambda(\alpha_{2\to c_0} V_{c_0}(x_1, x_2|s) + \alpha_{2\to c_1} V_{c_1}). \tag{4}$$

We denote the value of the above expression by $U_2(x_1, s)$ and the corresponding action and attention by $x_2(x_1, s)$ and $\alpha_2(x_1, s)$, respectively (we suppose that the solution is unique for ease of notation).

In period 1, the DM values their period-2 sum of material payoff and attention utility, i.e., U_2, plus their attention utility in period 1. We allow the DM in period 1 to devote attention to (c_0, s) for different states s and c_1 and so their objective is to chose (x_1, α_1) with $x_1 \in X_1(\alpha_1)$ to maximize

$$\lambda(\sum_{s\in\mathcal{S}} \alpha_{1\to(c_0,s)} V_{c_0}(x_1, x_2(x_1, s)|s) + \alpha_{1\to c_1} V_{c_1}) + \sum_{s\in\mathcal{S}} p_s U_2(x_1, s). \tag{5}$$

The following proposition characterizes the DM's optimal default.

Proposition 7. *Let* $B(x_1) \coloneqq \{s : \alpha_{2\to c_0}(x_1, s) < \eta\}$.

- *If* $X_1(\alpha_1)$ *is independent of* α_1*, then the optimal* x_1 *satisfies*

$$x_1 = \underset{x_1\in X_1}{\arg\max} \sum_{s\in B(x_1)} p_s V_{c_0}(x_1, x_2(x_1, s)|s).$$

- *For each* x_1*, if* λ *is large enough, then*

$$B(x_1) = \{s : \max_{(x_2,\alpha_2), x_2\in X_2(\alpha_2)} V_{c_0}(x_1, x_2) < V_{c_1}\}.$$

The first part of Proposition 7 states that the DM chooses a default focusing solely on realization of s for which the default is binding, i.e., $s \in B(x_1)$. For the remaining realizations, the DM devotes sufficient attention to c_0, rendering the default immaterial. On first sight, this observation may be considered similar to that of how a default is chosen in models of, e.g., rational inattention: If changing the default is costly (say, in terms of cognitive or physical resources), the DM may not always do so; at an optimum, the DM only conditions the default on cases on which it binds.

However, the choice of default in our model is distinct, in particular, the default is chosen asymmetrically as the second part of the proposition states. It identifies set $B(x_1)$ for large λ as those realizations s for which the max V_{c_0} given s is strictly less than the payoff from the trivial problem. In other words, the DM devotes attention to c_0 if and only if it has a high payoff.

There are many situations where asymmetric defaults may play a role. For example, individuals may plan, and then change, consumption depending on the realization of income. Emotionally inattention agents may only adjust their plan after good shocks to income, and in anticipation, choose a low initial (planned) consumption level. Similarly, emotionally inattentive agents may not want to revisit previously negotiated contracts if bad events occur, leading to asymmetric adjustment processes, and again a default contract which is tailored to bad contingencies. In Section 4 we work out the implications in detail in a particular setting: portfolio choice, and show the implications of asymmetric adjustments. More generally, our model suggests individuals fall back on heuristics and decision processes that can be implemented automatically ("System 1") when the situation at hand has a low payoff, whereas they engage with the situation and use System 2 when its payoff is high.

Proposition 7 states that for λ large, the DM chooses a "pessimistic" default. This default maximizes the material payoff regarding problem c_0, since it is always non-binding if the DM were to devote attention to c_0. But for large λ, the DM is mostly concerned about their attention utility. We next show that Proposition 7 relies on the default being a "pure default," one that has no implications beyond its role as a default action. We maintain the setting as described above with one modification: the payoff of problem c_0 is now given by $V_{c_0}(x_1, x_2|s) = \tilde{V}_{c_0}(x_1, x_2|s) + \beta F(x_1)$, where $\tilde{V}_{c_0}$ has the "default property" from above, and $\beta \geq 0$. Thus, the previous setting is nested with $\beta = 0$.

Proposition 8. *Suppose $\beta > 0$, that the solution is unique, and that $X_1(\alpha_1)$ is independent of α_1 and finite. Let $(s^*, x_1^*) = \arg\max_{s \in \mathcal{S}, x_1 \in X_1} V_{c_0}(x_1, x_2(x_1, s)|s)$. If $V_{c_0}(x_1^*, x_2(x_1^*, s^*)|s^*) > V_{c_1}$ and λ is large enough, then $x_1^* = \arg\max_{x_1 \in X} F(x_1)$*

In the setting of Proposition 8, the DM chooses an action that maximizes F, i.e., x_1 is chosen independent of $\tilde{V}_{c_0}$ (and p_s's), thus, contrasting Proposition 7. There, if λ is large, the DM chooses an action optimal for realized s's for which the DM does not devote attention in the second period. But, as λ is large, it is those realizations that do not matter (much) in the DM's objective (they only value what they devote attention to). Thus, whenever x_1 does affect the payoffs, the DM chooses it, so the payoff of the realization the DM devotes attention to increases. In a sense, the DM then behaves "optimistically" and now fully ignores the consequence of x_1 for when the default binds, and it determines the payoff. Using the language from Section 2.2, the DM's subjective probability is only large for states the DM devotes attention to in the first period (s^*) or the second period ($\mathcal{S} \setminus B(x_1^*)$); in fact, for λ large, the subjective probability of all other states is essentially 0, and so they do not matter for the DM's optimal action.

3.3 Information acquisition

We now turn to a distinct problem that involves multiple choices regarding attention: the dynamics of information acquisition. We consider when emotinoally inattentive agents demand (or avoid) information about a consumption problem, and what form does optimal information take. We consider a setting in which there is no instrumental value of information (for a standard DM), e.g., an uncertain payoff independent of the DM's action. Although the standard DM is indifferent to information in this setting, our DM may nevertheless have strict preferences over when to acquire information and the structure of the information.

The setting is as follows. There are two periods and two consumption problems in the second period: One that is non-trivial and denoted by c and one trivial problem with payoff $\bar{V}$. Problem c is as follows: is an eventual payment (in utility space) that is either high V_H or low V_L with $V_L < \bar{V} < V_H$. At the beginning of the first period, the DM believes that the eventual payment is high with probability p_1; at the beginning of the second period, the DM's belief evolves to p_2. Formally, c is a consumption problem in the second period. Thus, given p_2, its payoff is $V_c = p_2 V_H + (1 - p_2) v_L$. (Alternatively, we could include a third period; nothing substantive would change.)

For simplicity, we restrict the DM in the first period to allocate attention across future realizations of c proportional to their likelihood (i.e., they devote attention to the expected payoff; we denote this by $\alpha_{1\to E[c]}$) or to the trivial problem. In the second period, the DM devotes attention to the now realized problem c or the trivial problem.

The action taken in the first period x_1 encodes information acquisition that determines the distribution of p_2. Formally, given attention $\alpha_{1\to E[c]}$ devoted to the (expected) c, the DM can acquire distribution over posteriors x_1 from $X_1(\alpha_1) = \{x_1 \in \Delta([0,1]) : \mathrm{Var}(x_1) \leq \beta\alpha_{1\to E[c]}\}$, where β governs how easy information acquisition is.

Of course, in such a setting, a standard DM is indifferent between all attention allocations—in other words, they do not value (or disvalue) information. In contrast, if $\lambda > 0$, the DM has value in conditioning their attention in the second period on V_c (i.e., on their posterior p_2—increasing its weight when it is high and decreasing it otherwise—creating a preference for information.

Define $\bar{p} := \frac{v - v_L}{v_H - v_L}$ (i.e., $\bar{p}V_H + (1-\bar{p})V_L = \bar{V}$).

Proposition 9.

1. *There exists a $\tilde{p} \leq \bar{p}$ such that the following attention allocation is optimal.*
   In the first period, *in the second period,*

$$\alpha_{1\to E[c]} \begin{cases} = 0 & \text{if } p_1 < \tilde{p} \\ > 0 & \text{if } \tilde{p} \leq p_1 < \bar{p} \\ = 1 & \text{if } \bar{p} \geq p_1; \end{cases} \qquad \alpha_{2\to c} = \begin{cases} 0 & \text{if } p_2 < \bar{p} \\ 1 & \text{if } \bar{p} \geq p_1. \end{cases}$$

2. *For any $p_1 > \tilde{p}$ there exists $\bar{\beta}$ such that for all $\beta < \bar{\beta}$, if also $p_1 \in (\tilde{p}, \bar{p})$, then x_1 positively skewed, and if and $p_1 \in (\bar{p}, 1)$, then x_1 is negatively skewed.*

3. *$\bar{p}$ and $\tilde{p}$ are decreasing as V_L, V_H increase or $\bar{V}$ decreases; holding $\bar{p}$ and $V_L + (1 - p_1)V_H - \bar{V}$ fixed, $\tilde{p}$ is decreasing as $V_H - V_L$ increases; $\bar{p}$ and $\tilde{p}$ are independent of λ.*

The DM's information acquisition in the second period, i.e., $\alpha_{2\to c}$, is an instantiation of Proposition 1—that the DM avoids or pays excess attention to a problem depending on the level of its payoff. In the first period, the DM has an additional reason to acquire information (and hence $\tilde{p} < \bar{p}$): doing so allows them to condition their future attention on the revealed information. Note that the gain from acquiring information is due to the fact that the future period's payoff is convex in the payoff level of the non-trivial problem—an immediate implication of Proposition 2.

If the agent only acquires a "small amount" of information (that is, the variance of the distribution of posteriors must be small), then, say, a symmetric signal may not provide sufficient information to change the DM's second-period attention, a "hidden action" (relative to not receiving any information). Thus, the DM acquires information that, with a small probability, leads to a posterior that does alter their attention, i.e., a skewed information structure. In contrast, if the agent can obtain "a lot" of information (say, the information is fully revealing), then the skew of posterior depends on the prior: if $p_1 < \frac{1}{2}$, the skew is positive if $p_1 > \frac{1}{2}$, it is negative. Lastly, the comparative statics (the third part of the proposition) are natural; we only highlight that the DM acquires more information when the spread of the payoffs $V_H - V_L$ increases, i.e., when there is more to learn.

Thus, the DM acquires information early (in the first period) for two reasons. First, just as in the second period, the DM devotes attention to high-payoff problems in general, so acquiring information (and devoting attention) is optimal if their prior is high enough. Second, early information allows the DM to condition their future attention allocation on the realized information (and hence $\bar{p} < \bar{p}$. This latter force implies that if the DM has to acquire the information in either period, they prefer to do so early. Experimental evidence has been consistent with this prediction (Masatlioglu et al., 2017; Nielsen, 2020).

An implication of the behavior characterized in Proposition 9 is that the DM is better informed about good states; if the initial news is good, they continue acquiring information and may become more certain of the good state, whereas they stop learning and remain pessimistic, but uncertain, about the state when the initial news is bad. For example, Möbius et al. (2022) provides experimental evidence that participants' willingness-to-pay to learn about their performance in an IQ test increases if they received initial good, instead of bad, news.

4 Portfolio choice

We now turn to applying our model to an important, and widely studied, economic environment: portfolio choice where the DM takes the role of an investor. This exercise allows us to demonstrate many of our model's implications—as derived in Section 2—in a single environment. Moreover, it shows that our model can generate results that accord with intuitions in particular applications and not just the abstract environments considered previously. Of course, as previously mentioned, we can perform similar exercises in other canonical economic environments, such as consumption-savings problems or contracting.

We consider a 2-period portfolio choice model. We begin by describing the DM's actions and only then introduce attention. At the beginning of the first period, the DM may allocate wealth w across a risky and a safe asset, where x_1 denotes the amount invested in the risky asset. We impose that $x_1 \in [0, w]$ (i.e., preclude borrowing). After the DM makes their initial portfolio choice, the first period's return of the risky asset, denoted by r_1, and future market conditions, denoted by N, which determines the distribution of future returns, are realized. We assume that $(r_1, N) \sim G$, where G has discrete support, and $p_{(r_1,N)}$ denotes the probability of (r_1, N).

At the beginning of the second period, the DM may readjust their portfolio; in particular, without readjustment, the amount invested in the risky asset is $x_1(1 + r_1)$ (and $w - x_1$ is invested in the safe asset), with readjustment the DM chooses any amount $x_2 \in [0, w+x_1r_1]$. Then, the second period's return of the risky asset is realized according to $r_2 \sim F(N)$, and the DM consumes their final wealth given by $w + x_1r_1 + x_2r_2$, giving a payoff in utils according to Bernoulli utility function u, which is continuously differentiable.

In our framework, this portfolio choice problem is a consumption problem in the second period whose realization, a function of r_1 and N, we denote by $\rho = (r_1, N)$ (for ρortfolio choice) and with payoff $V_\rho(x_1, x_2) \coloneqq E_{r_2 \sim F(N)}[u(w + x_1r_1 + x_2r_2)]$. In addition to choosing their portfolio and consuming its proceeds, the DM also faces a trivial problem that yields an action-independent payoff $\bar{V}$ in the second period.

In each period, the DM allocates attention. We begin specifying the DM's behavior in the second period. Here, the DM allocates attention across the portfolio problem and the trivial problem with the respective levels of attention denoted by $\alpha_{2\to\rho}$ and $\alpha_{2\to t}$. In addition to increasing the weight a problem has in the DM's objective, attention is instrumentally valuable: readjusting the portfolio requires $\alpha_{2\to\rho} \geq \eta_2$ for some η_2. If the DM does not readjust their portfolio (formally, they take some action $\underset{\sim}{x}_2$), the consumption payoff is given by $V_\rho(x_1, \underset{\sim}{x}_2) \coloneqq E_{r_2 \sim F(N)}[u(w - x_1 + x_1(1+r_1)(1+r_2)]$; if they do, then they choose $x_2 \in [0, w + x_1 r_1]$ to maximize $V_\rho(x_1, x_2)$. Since V_ρ (optimally) only takes these two values (corresponding to when the DM does not readjust or readjusts their portfolio), it is without loss to only consider $\alpha_{2\to\rho} \in \{0, \eta_2, 1\}$ with payoffs in the second period given by

$$\begin{aligned}
&(\alpha_{2\to\rho} = 0) \quad V_\rho(x_1, \underset{\sim}{x}_2) + (1+\lambda)\bar{V}, \\
&(\alpha_{2\to\rho} = \eta) \quad (1+\lambda\eta_2) \max_{x_2 \in [0, w + x_1 r_1]} V_\rho(x_1, x_2) + (1 + \lambda(1-\eta_2))\bar{V}, \\
&(\alpha_{2\to\rho} = 1) \quad (1+\lambda) \max_{x_2 \in [0, w + x_1 r_1]} V_\rho(x_1, x_2) + \bar{V},
\end{aligned} \tag{6}$$

respectively.

Taking the max over these gives the payoff of ρ given optimal second-period behavior (we assume throughout the solution is unique to simplify notation). We denote second-period attention and action by $\alpha_{2\to\rho}(\rho, x_1), \alpha_{2\to t}(\rho, x_1)$ and action by $x_2(\rho, x_1)$, respectively, with ties broken in favor of high attention to ρ since that will be preferred by the DM in the first period.

In the first period, the DM allocates attention across all possible consumption problems—i.e., each possible ρ (the different "states") as well as the (deterministic) trivial problem—with levels of attention denoted by $\alpha_{1\to\rho}$ and $\alpha_{1\to t}$, respectively. Attention (again) increases the weight that a realization of ρ takes. Moreover, it is instrumentally valuable: Making an initial portfolio choice requires attention η_1 to the expected portfolio choice problem, i.e., $\alpha_{1\to\rho} \geq p_\rho \eta_1$ for some η_1. In such cases, the DM chooses $x_1 \in [0, w]$ to invest in the risky asset. Otherwise, the DM does not participate in the portfolio choice problem (including the second period) and optimally devotes attention to the trivial problem. In this case their overall payoff in the first period is $u(w) + (1 + \lambda 2)\bar{V}$. Otherwise, the DM's objective is

$$\sum_\rho (p_\rho + \lambda(\alpha_{1\to\rho} + p_\rho \alpha_{2\to\rho}(\rho, x_1))) V_\rho(x_2(\rho, x_1)) + (1 + \lambda(\alpha_{1\to t} + \sum_\rho p_\rho \alpha_{2\to t}(\rho, x_1)))\bar{V}. \tag{7}$$

We first note a simple comparative static with respect to the DM's participation in the portfolio choice problem: Their participation increases the lower the payoff from the trivial problem. A standard DM, of course, always participates.

Result 1. *The DM's value when participating in the portfolio choice problem, i.e.,* (7) *for optimal* x_1, α_1*, is increasing in* $\bar{V}$ *by less than the DM's value of not participating* $(u(w) + (1 + \lambda 2)\bar{V})$*.*

One may think of a DM with a high $\bar{V}$ as one with relatively low wealth. In this case, the proposition suggests that low-wealth individuals abstain from investing their wealth because doing so requires attention to their low eventual consumption. Expressing this intuition through varying $\bar{V}$ instead of the wealth w directly allows us to abstract away from other factors influencing the investment decision, such as changing preferences over risk as wealth varies. This finding is consistent with empirical findings about the relationship between wealth and participation found in Mankiw and Zeldes (1991); Poterba and Samwick (2003); Calvet et al. (2007); Briggs et al. (2021) (and distinct from the typical assumption of an exogenous cost of participation in the market).

In the context of individuals investing their wealth, researchers have noted an ostrich effect: differential attention to one's portfolio depending on market conditions (see our discussion at the end of Section 2.1). Result 1 can be understood as a similar ostrich effect but on the extensive margin of investing: The DM in our model may abstain completely from investing.

For the remainder of the section, we assume that the DM participates in the portfolio choice problem, i.e., they make an initial portfolio allocation. We also assume that the solution is unique to ease notation. Our following result states a version of the aforementioned (more standard) ostrich effect.

Result 2. *Fix* $x_1 \in [0, w]$*. Suppose that the solution is unique. Let* $B(\bar{V}, \lambda) := \{\rho : \alpha_{2\to\rho}(\rho, x_1) < \eta_2\}$*. Pick any* $\bar{V}, \bar{V}', \lambda, \lambda'$ *with* $\bar{V}' > \bar{V}$ *and* $\lambda' > \lambda$*.*

- *If* $\lambda > 0$*,* $B(\bar{V}, \lambda) \supseteq B(\bar{V}', \lambda)$*, with* $\lim_{\bar{V}\to-\infty} B(\bar{V}, \lambda) = \text{support}(F)$ *and* $\lim_{\bar{V}\to+\infty} B(\bar{V}, \lambda) = \emptyset$*.*

- $B(\bar{V}, \lambda) \subseteq B(\bar{V}, \lambda')$, *with* $\lim_{\lambda \to 0} B(\bar{V}, \lambda) = \emptyset$ *and* $\lim_{\lambda \to +\infty} B(\bar{V}, \lambda) = \{\rho : \max_{x_2 \in [0, w + x_1 r_1]} V_\rho(x_1, x_2) < \bar{V}\}$.

This result reflects the basic comparative static results (for each realized ρ) given in Proposition 1; we, again, refer to the discussion at the end of Section 2.1 for related empirical support in the current context.

This result is also linked to the disposition effect—"the disposition to sell winners too early and ride losers too long" (Shefrin and Statman (1985); see also Odean (1998)). This effect is often explained with reference-depend preferences and concave utility over gains and convex utility over losses (Barberis and Xiong, 2009). Here, we provide a distinct (partial) explanation: The DM does not sell (nor buy) the risky asset when the market conditions are poor (and the payoff is low)—i.e., when the risky asset is a "loser." poorly. They do execute trades when the market conditions are good (and the payoff is high)—i.e., when the risky asset is a "winner." Mechanically, this then induces a type of disposition effect, in particular with respect tod the volume of trade depending on market conditions.

Conceptually, our model is related to, but distinct from, models where the DM generates "realization utility" from selling an asset (Barberis and Xiong, 2009, 2012). Recall that our DM's objective can be interpreted as the sum of (unweighted) consumption payoffs plus attention utility. Hence, utility is "generated" by attention instead of realizing assets, but attention also allows the DM to realize (sell) assets.

The DM thus requires a type of attention premium to both participate in the portfolio choice problem (Result 1) and to continuously reoptimize their portfolio (Result 2). This premium reflects a "cost of attention" from increasing the weight on low pay-off consumption problems. In contrast, while other models also feature such attention premium, say because of computational cognitive costs or physical (time) costs associated with attending to one's portfolio, the resulting attention is typically symmetric, whereas here, it is asymmetric: Result 2 states that the DM devotes attention only to high-payoff realizations of ρ (in the sense as stated in the result).

Although we focus on a setting with a single risky asset, one can extend our results to where there are multiple risky assets, and the DM decides how to allocate attention across them (with implications for feasible trades). Formally, this may be modeled by letting each asset constitute a distinct consumption problem. We suspect that such a model leads to a within-portfolio ostrich effect—differential attention across assets depending on their (individual) performance—very much in the spirit of Proposition 1.

Having noted two types of inattention—non-participation, and non-reoptimization—we next consider the DM's optimal portfolio choice—their chosen mix of assets—for when they participate. Multiple of our previous findings may apply: Roughly, Proposition 4 suggests that attention available to roam freely, $1 - \eta_1$, is devoted to high-payoff ρ's increasing its subjective probability, which, in turn, leads to an added preference for the risky asset; Proposition 2 suggests that the DM seeks to face a varied future payoff, with the same effect; however, Proposition 7 suggests, instead, that the DM may choose a portfolio that performs well for those ρ's for which the DM does not reoptimize their portfolio.

In the following result, the last effect, which pushes towards the safe asset, is muted by not allowing the DM to reoptimize.

Result 3. *Suppose $x_2(x_1, \rho) = \underline{x}_2$ for all x_1, ρ; $F(N)$ is deterministic; and the solution is unique. Then x_1 is increasing λ and $1 - \eta_1$.*

The comparative static with respect to λ expresses the intuition above, but what about $1-\eta_1$? $1-\eta_1$ is the amount of attention the DM allocates freely (after having devoted η_1 to the expected portfolio choice problem). The DM devotes this attention to states where the return of the risky asset is highest; thus, as in Proposition 4, their subjective probability of such high returns increases, and more so the higher $1 - \eta_1$.

We next perform the reverse exercise, highlighting when the DM may invest more in the safe asset compared to the standard DM; assume that G constant on $r_1 = 0$, and $\eta_1 = 0$.

Then, a preference for the safe asset comes from the DM's anticipation that they may not reoptimize their portfolio in poor market conditions N; hence, the DM may want to set a pessimistic "default." (Choosing $\eta_1 = 0$ leads to time consistency and is essentially a technical trick to state the following result.)

Result 4. *Suppose G is deterministic on $r_1 = 0$ and $\eta_1 = 0$. Suppose that the solution is unique. Let $B \coloneqq \{\rho : \alpha_{2\to\rho}(\rho, x_1) < \eta_2\}$.*

- $x_1 = \arg\max_{x_1\in[0,w]} \sum_{\rho\in B} p_\rho V_\rho(x_1, x_2(\rho, x_1))$.

Note that we can easily define set B for large λ by Result 2.

To summarize, the DM is "optimistic," if they cannot reoptimize their portfolio in the second period—i.e., when there is no instrumental value of attention. And the DM is "pessimistic" if their action in the first period does not affect the high-payoff realization of ρ (it is "non-binding;" see Section 3.2), but can insure against future inattention.

Thus, the DM can be, in a sense, excessively risk averse, and we would be remiss not to relate our findings to risk premia and the equity premium puzzle (Mehra and Prescott, 1985). As just discussed in Result 4, the DM may choose a portfolio that performs well in poor market conditions—when they do not devote attention—which may lead to excessive risk aversion. More generally, the DM prefers assets that do not require attention (here, the risky asset as it may need to be readjusted). In principle, assets that require attention may not be the same as those that are risky. Indeed, a risky but illiquid asset may not require attention, whereas a safe asset, for which the DM needs to perform some administrative (but payoff-irrelevant) work, may require attention. Thus, the risk premium occurring in our model may be more aptly described as an "attention premium." Hence, the mechanism through which our model may lead to excessive risk aversion is different from other (related) approaches incorporating non-standard decision-making: For instance, Caplin and Leahy (2001) use anticipatory utility, while Benartzi and Thaler (1995) and Barberis and Huang (2006) use reference points, while Sarver (2018) uses both.

The final result we present in the context of portfolio choice concerns time preference. A typical view is that time preference—determined elsewhere—affects the portfolio choice. Here, the reverse holds, in a sense: the portfolio choice determines the time preferences.

So far, there is no consumption problem in the first period; hence, the DM only devotes attention to the future. We thus introduce an arbitrary (parameterized) consumption problem in the first period $V_{c_1} = \tilde{V}_{c_1}(x_1) + \gamma_{c_1}$ (action x_1 now does not only denote the DM's choice of portfolio but more generally affects their payoff in consumption problem x_1; one can think of it as a tuple).

For the following result, we do not necessarily assume that the DM always participates in the portfolio choice problem; the statement of the result includes both cases.

Result 5. *$\alpha_{1\to c_1}$ is increasing in γ_{c_1}.*

In the presence of multiple consumption problems in a single period, such as is the case in the second period, discount factors are consumption problem specific. Result 5 thus implies that the discount factors between the first-period consumption problem and (at least some) second-period consumption problems increase as the DM becomes more wealthy. Standard formulations of the lifetime income hypotheses predict that differences in discount rates cause differences in wealth (Epper et al., 2018). Here, the story is reversed: A difference in wealth causes inattention to the present and leads to high discount factors.

5 Implications for (self-imposed) policies

In this section, we discuss our model's implications for policymaking broadly construed and ask how a policymaker—a second party, say, a government or the DM themselves—should intervene in the environment. Incorporating our model into policymaking is necessary to fully understand the behavioral changes a policy induces—which may be different from those if the DM was standard—and its implications for the DM's overall payoff. We focus on the case where the DM allocates attention across consumption problems, but the other two dimensions (states and periods) are similar. We consider three broad classes of policies: optimal resource allocation, incentivization of actions, and optimal construing of consumption problems.

5.1 Optimal resource allocation

We first consider transfers, both in payoff space and in "input space," as we explain shortly.

Suppose first that the policymaker can increase one (or multiple) of the payoffs of the consumption problems by some total amount. The DM need not need to devote attention to receive the transfer; formally, consumption problem c with consumption payoff $\tilde{V}_c$ has consumption payoff $\tilde{V}_c + \gamma_c$ after transfer $\gamma_c \geq 0$ to it, and the resource constraint faced by the policymaker is $\sum_{c \in \mathcal{C}} \gamma_c \leq \gamma$. What is the optimal way—in terms of overall payoff—to allocate this equi-utility transfer?

To a standard DM, the choice of $(\gamma_c)_{c \in \mathcal{C}}$ would not matter (as long as the resource constraint binds) as each increase receives the same weight; here, however, an increase of the payoff from problem c is weighted by $1 + \lambda\alpha_c$. It follows that the policymaker should transfer the utility amount to the consumption problem which receives the most attention. Intuitively, increasing the payoff associated with a situation that an individual ignores does little to that individual's overall payoff.

Suppose next that the policymaker allocates an "input." Formally, let the environment be separable and, additionally, let the payoff of problem c given α_c and input r_c be $\hat{V}_c(\alpha_c + r_c)$, i.e., the input is a perfect substitute of attention. For example, α_c and r_c could represent amounts of information, where the DM acquires the former and the latter provided by the policymaker (Note that this information is processed without devoting attention). We ask: What is the optimal allocation of inputs $(r_c)_{c\in\mathcal{C}}$, subject to $\sum_{c\in\mathcal{C}} r_c \leq r$?

For ease of exposition, suppose that $\hat{V}_c$ is continuously differentiable and consider the marginal benefit of increasing r_c (given an input allocation). By the Envelope theorem, the DM's value from increasing r_c increases by

$$\frac{\partial}{\partial r_c}\hat{V}_c(\alpha_c + r_c)(1 + \lambda\alpha_c) \tag{8}$$

(take (2), substitute $\hat{V}_c$ for V_c, and differentiate). Furthermore, if optimal attention is interior, the first-order condition for α_c is given by $F(\alpha_c, r_c) := \frac{\partial}{\partial\alpha_c}\hat{V}_c(\alpha_c + r_c)(1 + \lambda\alpha_c) + \lambda\hat{V}_c(\alpha_c + r_c) - \mu = 0$, where μ is the Lagrange multiplier on the constraint for the sum of attention. Using the fact that $\frac{\partial}{\partial r_c}\hat{V}_c(\alpha_c + r_c) = \frac{\partial}{\partial\alpha_c}\hat{V}_c(\alpha_c + r_c)$ and substituting it into (8) gives $-\lambda\hat{V}_c(\alpha_c + r_c) + \mu$.

The policymaker should then give the marginal unit of input to the problem with the lowest payoff (if attention is interior) or (possibly) to the problem that already receives full attention. Intuitively, at an optimum, the benefit from increasing attention equals its cost, which decreases in the payoff; the policymaker does not bear this cost, and thus the benefit of the input is largest for the problem with the highest cost, i.e., that with the lowest payoff.

We note that increasing r_c may not increase the payoff of c due to an endogenous reduction in α_c.[18]

[18] Indeed, let $\alpha_c(r_c)$ denote the optimal level of attention as input r_c varies; implicitly differentiating the DM's first-order condition gives $\frac{\partial}{\partial r_c}\alpha_c(r_c) = -\frac{\frac{\partial}{\partial r_c}F(\alpha_c, r_c)}{\frac{\partial}{\partial\alpha_c}F(\alpha_c, r_c)} \leq -1$. Thus, attention to consumption problem c decreases by more than the input increases so that $r_c + \alpha_c$, and hence $\hat{V}_c$, decreases.

When the policymaker is a government, these results may guide how the government's resources are best allocated, which tasks should be left to the individual, and which tasks are better completed by the government.

5.2 Providing incentives to induce better actions

We next consider two ways of inducing the DM to take better actions—increasing the rewards for "success" and increasing the penalty for "failure." As we show, for a standard DM, their effects are similar; but when $\lambda > 0$, they may have very different consequences.

Formally, consider a separable environment. Attention α_c to problem c, which may be interpreted as effort, leads to "success" with probability $p(\alpha_c)$, where p is increasing, continuously differentiable and bounded away from 0 and 1, and "failure" otherwise. We thus have $\hat{V}_c(\alpha_c) = p(\alpha_c)V_H + (1 - p(\alpha_c))V_L$, with $V_H > V_L$.

The standard DM's optimal attention is unchanged when V_H and V_L are shifted by the same amount. Also, as expected, the standard DM increases α_c in response to an increase in $V_H - V_L$. Thus, the standard DM responds to both "carrots" (an increase in V_H) and "sticks" (a decrease in V_L).

In contrast, when $\lambda > 0$, the DM increases α_c in response to a shift of V_H, V_L (Proposition 1). They also respond positively to carrots: increasing V_H increases α_c. However, increasing the stick can, in fact, decrease attention, i.e., worsen the action.

Proposition 10. *Consider the environment as introduced prior to this proposition and suppose the optimal α_c is unique.*

1. *Increasing V_H, V_L by the same amount increases α_c.*

2. *Increasing V_H increases α_c.*

3. *Decreasing V_L decreases α_c if $p(\alpha_c)+\alpha_c\frac{\partial}{\partial\alpha_c}p(\alpha_c) < 1$ everywhere and λ is large enough.*

In the third part of Proposition 10, attention is not very effective in increasing p ($\frac{\partial}{\partial\alpha_c}p(\alpha_c)$ is low), and success is never guaranteed ($p(\alpha)$ is also low). In these circumstances, the stick may induce the DM to shy away from problem c instead of increasing their attention to it so that they can decrease the weight of the associated payoff. An implication is that the DM may not demand commitment contracts that involve penalties (while those with rewards may be too expensive).

5.3 Optimal bracketing of consumption problems

Lastly, consider how the environment is optimally construed; that is, when should consumption problems be perceived as distinct, and when should they be thought of jointly? For instance, the DM may be able to learn to associate one problem with another, either through some purely cognitive process or with the help of, say, physical cues that the policymaker installs. Such bracketing of consumption problems (or compartmentalization)is a form of intentional use of associations: Attending to one problem forces the DM to ponder about another and vice versa. We note that such optimal bracketing serves as a microfoundation for the set of consumption problems in Section 2.1, which may be understood as the optimally bracketed set of smaller consumption problems.

The setup is that of Section 2.1, where, in addition to choosing (x, α) with $x \in X(\alpha)$, the DM also chooses a bracketing $B \in \mathcal{P}(\mathcal{C})$, a partition of the consumption problems. Let $B(c)$ be defined by $c \in B(c) \in B$. Whenever the DM devotes attention to c and there is $c' \neq c$ in $B(c)$, then both c and c' "come to mind." As multiple payoffs come to mind, the DM's attention is diluted uniformly among them. Thus, given (x, α) and B, the DM overall payoff is

$$\sum_{c \in \mathcal{C}} (1 + \lambda \alpha_c) \bar{V}_{B(c)}(x), \tag{9}$$

where $\bar{V}_C(x) := \frac{\sum_{c \in C} V_c(x)}{|C|}$ for $C \subseteq \mathcal{C}$; also let $\bar{\alpha}_C(x) := \frac{\sum_{c \in C} \alpha_c}{|C|}$ for $C \subseteq \mathcal{C}$.

Note that the model in Section 2.1 is recovered when B consists of singleton sets and that a DM who uses one bracket, i.e., B is a singleton, is equivalent to the standard DM.

Let $\bar{V}_C(x) := \frac{\sum_{c \in C} V_c(x)}{|C|}$ for $C \subseteq \mathcal{C}$.

Proposition 11. *Consider any (x, α) and B optimal given (x, α). Then $\bar{V}_C(X) > \bar{V}_{C'}(x)$ implies $\bar{\alpha}_C \geq \bar{\alpha}_{C'}$ for all $C, C' \in B$.*

Intuitively, when the DM devotes much attention to a low-payoff problem, they would like to associate it with other (high-payoff) problems. If, instead, they devote a lot of attention to a high-payoff problem, they do not want to dilute their attention utility from it by associating it with another lower-payoff problem. Consequently, the DM distorts the environment if doing so helps them, which is precisely when the distortion is in the direction of high-payoff consumption problems; they consider the problems simultaneously and behave like a standard DM when their distortion is in the direction of low-payoff consumption problems.

6 Relation to existing models

In this section, we compare our model to other related approaches.

Rational inattention: In models of rational inattention (e.g., Sims (2003) and Mackowiak et al., 2022), attention serves an instrumental role as in ours. Additionally, attention is (pecuniary or non-pecuniary) costly; while we do not model these costs explicitly, they can be captured in the functional form of the payoffs V_i. The key difference is thus that in our model, unlike in models of rational inattention, the consumption payoff terms in the DM's objective are weighted by attention—i.e., the second feature of attention: its role as an aggregator of experiences.

These features imply that our model rationalizes behaviors that are at odds with rational inattention. For example, 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, information avoidance (inattention) varies with the level of future payoffs (Karlsson et al. (2009); Sicherman et al. (2015) in the context of investors; and Ganguly and Tasoff (2017) in the context of health) with no (obvious) corresponding change in rational costs or benefits. Even more basic, there is no reason in models of rational inattention to devote attention to already known information (Quispe-Torreblanca et al., 2020). These examples suggest that there is a "cost" (or benefit) of attention missing from the consideration; our model, and in it attention's explicit role in reweighting the environment, provides this cost (see also the discussion in Section 2.1).

Belief-based utility with Bayesian agents: There is by now an extensive literature in economics modeling agents who directly gain anticipatory utility from their (rational) beliefs (see Loewenstein (1987); Loewenstein and Elster (1992) for early contributions, and recent efforts of Caplin and Leahy (2001); Kőszegi (2010); Dillenberger and Raymond (2020)), or gain utility from changes in beliefs, or news utility (e.g., Kőszegi and Rabin (2009)). Broadly speaking both classes of models assume that some present utility may be generated via beliefs, or changes in beliefs, about future payoffs.

There are some similarities between models of anticipatory utility and our approach: In our model, the DM values material payoffs and attention utility. Attention utility, when stemming from a future problem, can be thought of as anticipatory utility. However, unlike in the aforementioned models, the DM only "receives" this anticipatory utility if they devote attention to its underlying payoff; not otherwise. The same applies to models where the DM receives gain utility from changes in their belief (e.g., Kőszegi and Rabin (2009)). There, the DM "receives" the gain utility regardless of their attention.

Just as models of rational inattention, models where attention is allocated to induce changes in anticipatory utility or gain utility (via information acquisition) rely on the presence of uncertainty. Such models thus also fail to make predictions in situations where information is unlikely to play a major role, such as in much of the evidence presented in Section 2.1).

Belief based utility with chosen beliefs: Our model, in Section 2.2, also relates to those where subjective beliefs are optimally chosen to (for example) increase anticipatory utility as in Bénabou and Tirole (2002); Brunnermeier and Parker (2005); Bracha and Brown (2012); Caplin and Leahy (2019) (for a recent summary of the larger literature see Bénabou and Tirole (2016)). While our model is, of course, conceptually very different (beliefs do not feature in Sections 2.1 and 3.1, and in Section 2.2, the DM chooses an attention allocation that leads to weights we interpret a subjective belief), there are some similarities. Optimal attention and optimal beliefs are both determined by a tradeoff of "optimism" (here, devoting attention to high-payoff states) and the instrumental value of attention. Our model is not, however, observationally equivalent to models of chosen beliefs. For instance, the DM may, in fact, overweight a low-payoff state if states require some minimum amount of attention to ensure a good expected payoff (see the second case of Proposition 5 for an example).

Temporal discounting: There is a huge theoretical literature devoted to temporal discounting (see Frederick et al. (2002) for an overview). In our model, when attention is allocated across time, endogenous weights on periods appear, and the DM develops a preference for the timing of consumption. Our formulation is somewhat related to the ideas in Loewenstein (1987). There, as in our model, the DM may, e.g., negatively discount a high future payoff since it creates (high) anticipatory utility until it is realized.

However, as for other models with anticipatory utility (see above), the weight of a future payoff in today's objective is fixed, in particular, independent of whether the DM devotes attention to it or not. It thus cannot capture our basic comparative static that discounting varies with the instrumental value of attention or the payoff level. One additional implication of this difference is that a non-smooth consumption path is generally not beneficial to a DM in Loewenstein (1987) whereas it is valued by ours since they ignore low-payoff periods and devote excessive attention to high-payoff ones.

Other Models of Attention:

A couple of other papers directly model the two fundamental features of attention in ways similar to ours. The model of Tasoff and Madarasz (2009) is closest to ours. A DM faces a decision problem with multiple dimensions (analogous to our consumption problems) and receives anticipatory utility from each as a function of its payoff and the attention devoted to it. Attention to a dimension increases when its payoff changes because the DM chooses an action different from a default or receives payoff-relevant information. Similar to Proposition 1, the DM is more likely to take a non-default action or acquire information if the payoff is high. Such formulation is, in some sense, nested in ours: Let $x_d \in X(\alpha)$ for all α be a default action that is always available, and let acquiring information be encoded as some action x (providing payoff-relevant information for an underlying (not modeled) consumption problem) and suppose x is only available for some attention allocations. A difference between the two formulations is that we allow for attention allocation with no instrumental consequence. More broadly, in our model, attention is chosen to enable non-default actions and information acquisition, whereas in theirs, the order is reversed.

Their subsequent focus is on how information provision (as requested by the DM or forced by an advertiser) can increase consumption, even when the DM learns their marginal payoff is less than what they expected (this follows from the increase in attention and hence the importance in the DM's objective; this intuition can be expressed in our framework as we show in Example 6 in Appendix A.3). Instead, our focus is on attention allocation across uncertain states and time and the ensuing behavioral phenomena due to the attention-weighted environments. Our applications (information acquisition and portfolio choice) are also very different from their main one (a monopolist manipulating information provided to—and thus the attention of—consumers), as are the implications we draw for policymaking.

Another related model is that in Karlsson et al. (2009): The DM gains utility not from anticipatory emotions but rather as gain-loss utility from changes in expected future payoffs. Devoting attention to some initial news and acquiring further information increases the relative impact of gain-loss utility and speeds up the reference point adjustment. Under some conditions, the DM acquires additional information in response to positive initial news and not otherwise.

Our model is similar in that attention also increases the impact (or weight) of a payoff. We abstract away, however, from attention's effect on reference points and instead explicitly include actions whose availability depends on the attention allocation. We also construct our model more general, allowing us to consider different dimensions of attention allocation with different insights.

7 Conclusion

This paper has presented a model of attention allocation. Attention has two fundamental features: It helps the DM make better decisions, and it determines how payoffs are aggregated. We study our model in a variety of economic environments focusing on two key lessons. First, the DM may ignore a low-payoff situation (even if doing so is instrumentally harmful) to decrease its weight in their objective (and conversely devote excessive attention to high-payoff ones). Second, due attention reweighting the objective function, our model can lead to a variety of behavioral phenomena, where the exact form reflects the underlying economic environments.

Our model, of course, has limits in terms of what it can explain. There are situations where individuals choose to engage with negative emotion-generating activities with low instrumental value. For instance, the premise of our model seems at odds with pessimists who constantly focus on the negative aspects of any situation and overweight those, or the fact that many people doom-scroll and look at Twitter feeds that induce negative feelings. Our framework still allows us to study the attention-weighted decision environment and the ensuing behavioral phenomena, regardless of what model of attention allocation (e.g., negativity bias or salience) produces them. For instance, a present focus or distorted subjective probabilities result from excessive attention to the present or a particular state—regardless of whether that attention is directed as in our model or simply because the present or the state are salient.

Of course, our approach, which assumes that the entire stock of attention is under the control of the DM (i.e., the top down approach), is likely not entirely true. Involuntarily allocated attention, as is highlighted by recent models on attribute-based choice, e.g., Bordalo et al. (2013); Kőszegi and Szeidl (2013); Bushong et al. (2021), can also play important role. That said, as Desimone et al. (1995) points out, the "attentional system, however, would be of little use if it were entirely dominated by bottom-up biases." Thus, we believe our model is a useful first step in understanding how an individual may utilize the remaining stock of attention after bottom-up allocations have been made.

Like many other models of attention, our model also suffers from a recursion problem (see Lipman (1991) for an discussion infinite regress issues in economic models). We suppose that the DM fully understands all parameters of the model, and is able to conduct the optimization procedure of allocating attention and taking an action without reweighting the payoffs. Although such an approach is tractable, it does beg the question of how the implications of the model might be changed if even the act of optimization itself—during which the DM arguably devotes attention to different payoffs—reweights the payoffs. One can embed higher-level learning by the DM about the parameters of the model and consider a multi-period model with consumption payoffs only in the final period; but we do not provide formal results.

Our model also requires carefully specifying the environment: a key component is a way of partitioning the environment into sets of consumption problems, states, and time periods. In many real economic environments natural partitions exist. However, in many situations it may not be as obvious what the correct sets are. Although Section 5.3 provides some guidance given a finest possible partition, there are also likely situations the environment is determined differently.

Thankfully, the novel primitives of our model, the set of dimensions and λ, can be identified from the data. The details would vary by the environment, but here we provide the intuition for a situation where the dimensions are states. We first can identify whether two states are considered jointly (they are in the same "bracket," Section 5.3) by reducing the payoff of one problem and increasing the other the same amount and seeing whether the (action, attention)-pair changes. If we can find some shift such that it does, then the two states are not part of the same bracket. Then choices over lotteries allow us to identify the degree of overweighting of the high payoff state(s) and thus λ.

Our paper also focused on the DM's problem. In Section B we consider what happens in strategic interactions where many agents gain attention utility. In a setup similar to that of Brunnermeier and Parker (2005) Section III, ex-ante identical agents are placed in an endowment economy and in equilibrium choose to hold idiosyncratic risk. Such risk-taking is optimal since it allows agents to increase their attention utility, and possibly through agents taking opposing gambles so that which states are the high-payoff states differs by agent. Similarly, ex-ante identical agents would "trade consumption payoffs" to create payoffs that vary across problems and agent groups. Thus, in strategic settings, agents may sort into ex-post different groups and naturally some sort of polarization occurs.

There are other strategic implication: agents may fail to readjust strategies in dynamic games, or firms may change the framing of their product in order to be more appealing to emotionally inattentive consumers. To the extent that firms may be able to exogenously shift the attention of consumers, and even change what are the relevant dimensions, opens up a new way to view framing effects in product markets.

References

Åstebro, Thomas, José Mata, and Luís Santos-Pinto, "Skewness seeking: risk loving, optimism or overweighting of small probabilities?," *Theory and Decision*, 2015, *78* (2), 189–208.

Avoyan, Ala and Andrew Schotter, "Attention in games: An experimental study," *European Economic Review*, 2020, *124*, 103410.

Barberis, Nicholas and Wei Xiong, "What drives the disposition effect? An analysis of a long-standing preference-based explanation," *the Journal of Finance*, 2009, *64* (2), 751–784.

_ **and** _ , "Realization utility," *Journal of Financial Economics*, 2012, *104* (2), 251–271.

Barberis, Nicholas C, "Thirty years of prospect theory in economics: A review and assessment," *Journal of Economic Perspectives*, 2013, *27* (1), 173–96.

_ **and Ming Huang**, "The loss aversion/narrow framing approach to the equity premium puzzle," 2006.

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

Bénabou, Roland and Jean Tirole, "Self-confidence and personal motivation," *The quarterly journal of economics*, 2002, *117* (3), 871–915.

_ **and** _ , "Mindful economics: The production, consumption, and value of beliefs," *Journal of Economic Perspectives*, 2016, *30* (3), 141–64.

Benartzi, Shlomo and Richard H Thaler, "Myopic loss aversion and the equity premium puzzle," *The quarterly journal of Economics*, 1995, *110* (1), 73–92.

Blume, Marshall E and Irwin Friend, "The asset structure of individual portfolios and some implications for utility functions," *The Journal of Finance*, 1975, *30* (2), 585–603.

Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer, "Salience and consumer choice," *Journal of Political Economy*, 2013, *121* (5), 803–843.

Bracha, Anat and Donald J Brown, "Affective decision making: A theory of optimism bias," *Games and Economic Behavior*, 2012, *75* (1), 67–80.

Briggs, Joseph, David Cesarini, Erik Lindqvist, and Robert Östling, "Windfall gains and stock market participation," *Journal of Financial Economics*, 2021, *139* (1), 57–83.

Bronchetti, Erin T, Judd B Kessler, Ellen B Magenheim, Dmitry Taubinsky, and Eric Zwick, "Is Attention Produced Optimally? Theory and Evidence from Experiments with Bandwidth Enhancements," Working Paper 27443, National Bureau of Economic Research June 2020.

Brunnermeier, Markus K. and Jonathan A. Parker, "Optimal Expectations," *American Economic Review*, September 2005, *95* (4), 1092–1118.

Buschman, Timothy J and Earl K Miller, "Top-down versus bottom-up control of attention in the prefrontal and posterior parietal cortices," *science*, 2007, *315* (5820), 1860–1862.

Bushong, Benjamin, Matthew Rabin, and Joshua Schwartzstein, "A model of relative thinking," *The Review of Economic Studies*, 2021, *88* (1), 162–191.

Calvet, Laurent E, John Y Campbell, and Paolo Sodini, "Down or out: Assessing the welfare costs of household investment mistakes," *Journal of Political Economy*, 2007, *115* (5), 707–747.

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

— **and John V Leahy**, "Wishful thinking," Technical Report, National Bureau of Economic Research 2019.

Corbetta, Maurizio and Gordon L Shulman, "Control of goal-directed and stimulus-driven attention in the brain," *Nature reviews neuroscience*, 2002, *3* (3), 201–215.

Dertwinkel-Kalt, Markus and Mats Köster, "Salience and skewness preferences," *Journal of the European Economic Association*, 2020, *18* (5), 2057–2107.

Desimone, Robert, John Duncan et al., "Neural mechanisms of selective visual attention," *Annual review of neuroscience*, 1995, *18* (1), 193–222.

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.

Dixon, Matthew L, Ravi Thiruchselvam, Rebecca Todd, and Kalina Christoff, "Emotion and the prefrontal cortex: An integrative review.," *Psychological bulletin*, 2017, *143* (10), 1033.

Ebert, Sebastian, "On skewed risks in economic models and experiments," *Journal of Economic Behavior & Organization*, 2015, *112*, 85–97.

_ **and Daniel Wiesen**, "Testing for prudence and skewness seeking," *Management Science*, 2011, *57* (7), 1334–1349.

Engelmann, Jan, Maël Lebreton, Peter Schwardmann, Joel van der Weele, and Li-Ang Chang, "Anticipatory anxiety and wishful thinking," Technical Report 2019.

Epper, Thomas, Ernst Fehr, Helga Fehr-Duda, Claus Thustrup Kreiner, David Dreyer Lassen, Søren Leth Petersen, and Gregers Nytoft Rasmussen, "Time Discounting, Savings Behavior and Wealth Inequality," Technical Report, Working paper 2018.

Ericson, Keith Marzilli and David Laibson, "Intertemporal choice," in "Handbook of Behavioral Economics: Applications and Foundations 1," Vol. 2, Elsevier, 2019, pp. 1–67.

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.

Forrest, David, Robert Simmons, and Neil Chesters, "Buying a dream: Alternative models of demand for lotto," *Economic Inquiry*, 2002, *40* (3), 485–496.

Frederick, Shane, George Loewenstein, and Ted O'donoghue, "Time discounting and time preference: A critical review," *Journal of economic literature*, 2002, *40* (2), 351–401.

Gabaix, Xavier, " A Sparsity-Based Model of Bounded Rationality *," *The Quarterly Journal of Economics*, 09 2014, *129* (4), 1661–1710.

Galai, Dan and Orly Sade, "The "ostrich effect" and the relationship between the liquidity and the yields of financial assets," *The Journal of Business*, 2006, *79* (5), 2741–2759.

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.

Garrett, Thomas A and Russell S Sobel, "Gamblers favor skewness, not risk: Further evidence from United States' lottery games," *Economics Letters*, 1999, *63* (1), 85–90.

Gherzi, Svetlana, Daniel Egan, Neil Stewart, Emily Haisley, and Peter Ayton, "The meerkat effect: Personality and market returns affect investors' portfolio monitoring behaviour," *Journal of Economic Behavior & Organization*, 2014, *107*, 512–526. Empirical Behavioral Finance.

Gilboa, Itzhak, Andrew Postlewaite, and Larry Samuelson, "Memorable consumption," *Journal of Economic Theory*, 2016, *165*, 414–455.

Golec, Joseph and Maurry Tamarkin, "Bettors love skewness, not risk, at the horse track," *Journal of political economy*, 1998, *106* (1), 205–225.

Green, Leonard, Joel Myerson, and Edward McFadden, "Rate of temporal discounting decreases with amount of reward," *Memory & cognition*, 1997, *25* (5), 715–723.

Gross, James J, "The emerging field of emotion regulation: An integrative review," *Review of general psychology*, 1998, *2* (3), 271–299.

Grossman, Philip J and Catherine C Eckel, "Loving the long shot: Risk taking with skewed lotteries," *Journal of Risk and Uncertainty*, 2015, *51* (3), 195–217.

Hai, Rong, Dirk Krueger, and Andrew Postlewaite, "On the welfare cost of consumption fluctuations in the presence of memorable goods," *Quantitative economics*, 2020, *11* (4), 1177–1214.

Jullien, Bruno and Bernard Salanié, "Estimating preferences under risk: The case of racetrack bettors," *Journal of Political Economy*, 2000, *108* (3), 503–530.

Kahana, Michael Jacob, *Fundamental Principles of Optical Lithography*, Foundations of Human Memory, 2012.

Kahneman, Daniel, "Prospect theory: An analysis of decisions under risk," *Econometrica*, 1979, *47*, 278.

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

Kőszegi, Botond, "Utility from anticipation and personal equilibrium," *Economic Theory*, 2010, *44* (3), 415–444.

_ **and Adam Szeidl**, "A model of focusing in economic choice," *The Quarterly journal of economics*, 2013, *128* (1), 53–104.

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

Laibson, David, "Golden Eggs and Hyperbolic Discounting*," *The Quarterly Journal of Economics*, 05 1997, *112* (2), 443–478.

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.

Lipman, Barton L, "How to decide how to decide how to...: Modeling limited rationality," *Econometrica: Journal of the Econometric Society*, 1991, pp. 1105–1125.

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

_ **and Jon Elster**, "Utility from memory and anticipation," *Choice over time*, 1992, pp. 213–234.

_ **and Nachum Sicherman**, "Do workers prefer increasing wage profiles?," *Journal of Labor Economics*, 1991, *9* (1), 67–84.

Mankiw, N Gregory and Stephen P Zeldes, "The consumption of stockholders and nonstockholders," *Journal of financial Economics*, 1991, *29* (1), 97–112.

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

Matějka, Filip and Alisdair McKay, "Rational Inattention to Discrete Choices: A New Foundation for the Multinomial Logit Model," *American Economic Review*, January 2015, *105* (1), 272–98.

Mayraz, Guy, "Wishful Thinking," Working Paper, SSRN October 2011.

Mehra, Rajnish and Edward C Prescott, "The equity premium: A puzzle," *Journal of monetary Economics*, 1985, *15* (2), 145–161.

Mijović-Prelec, Danica and Drazen Prelec, "Self-deception as self-signalling: a model and experimental evidence," *Philosophical Transactions of the Royal Society B: Biological Sciences*, 2010, *365* (1538), 227–240.

Möbius, Markus M, Muriel Niederle, Paul Niehaus, and Tanya S Rosenblat, "Managing self-confidence: Theory and experimental evidence," *Management Science*, 2022.

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

Odean, Terrance, "Are investors reluctant to realize their losses?," *The Journal of finance*, 1998, *53* (5), 1775–1798.

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.

Orhun, A Yesim, Alain Cohn, and Collin Raymond, "Motivated optimism and workplace risk," *Available at SSRN*, 2021.

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.

Poterba, James M and Andrew A Samwick, "Taxation and household portfolio composition: US evidence from the 1980s and 1990s," *Journal of Public Economics*, 2003, *87* (1), 5–38.

Quiggin, John, "A theory of anticipated utility," *Journal of economic behavior & organization*, 1982, *3* (4), 323–343.

Quispe-Torreblanca, Edika, John Gathergood, George Loewenstein, and Neil Stewart, "Attention Utility: Evidence from Individual Investors," Technical Report, Center for Economic Studies and ifo Institute (CESifo) 2020.

Rubinstein, Ariel, ""Economics and psychology"? The case of hyperbolic discounting," *International Economic Review*, 2003, *44* (4), 1207–1216.

Sarver, Todd, "Dynamic Mixture-Averse Preferences," *Econometrica*, 2018, *86* (4), 1347–1382.

Schelling, T. C., "THE MIND AS A CONSUMING ORGAN," in David E. Bell, Howard Raiffa, and Amos Tversky, eds., *Decision Making: Descriptive, Normative, and Prescriptive Interactions*, Cambridge University Press, 1988, pp. 343—357.

Sharot, Tali, "The optimism bias," *Current biology*, 2011, *21* (23), R941–R945.

Shefrin, Hersh and Meir Statman, "The disposition to sell winners too early and ride losers too long: Theory and evidence," *The Journal of finance*, 1985, *40* (3), 777–790.

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.

Snowberg, Erik and Justin Wolfers, "Explaining the favorite–long shot bias: Is it risk-love or misperceptions?," *Journal of Political Economy*, 2010, *118* (4), 723–746.

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

Thaler, Richard, "Some empirical evidence on dynamic inconsistency," *Economics letters*, 1981, *8* (3), 201–207.

Todd, Rebecca M, William A Cunningham, Adam K Anderson, and Evan Thompson, "Affect-biased attention as emotion regulation," *Trends in cognitive sciences*, 2012, *16* (7), 365–372.

Tverseky, Amos and Daniel Kahneman, "Advances in Prospect Theory: Cumulative Representation of Uncertainty," *Journal of Risk and Uncertainty*, 1992, *5* (4), 297–323.

VandenBos, Gary R., ed., *APA Dictionary of Psychology*, second ed., American Psychological Association, 2015.

Wakker, Peter P, *Prospect theory: For risk and ambiguity*, Cambridge university press, 2010.

Wu, George and Richard Gonzalez, "Curvature of the probability weighting function," *Management science*, 1996, *42* (12), 1676–1690.

A Addition examples

A.1 Examples of canonical problems

In Examples 1–3, we consider a separable environment (defined in Section 2) when attention is allocated across consumption problems (Section 2.1, a special case of the general abstract model in Section 2) and a particular problem $c \in \mathcal{C}$. "Actions" and "payoffs" shall refer to those in the now explicitly modeled problem c and those in the reduced-form setup of Section 2.

Example 1. *Problem c is the reduced form of a canonical choice problem with imperfect information and information acquisition (using the framework of Matějka and McKay (2015)).*

The DM chooses an action i from set $A = \{1, \ldots, N\}$. The state of nature is a vector $v \in \mathbb{R}^N$ where v_i is the payoff of action $i \in A$. When the DM's belief is $B \in \Delta(\mathbb{R}^N)$, they receive payoff $v(B) := \max_{i \in A} E_B[v_i]$. The DM is initially endowed with some belief $G \in \Delta(\mathbb{R}^N)$. They can receive signals $s \in \mathbb{R}^N$ on the state: They choose $F(s, v) \in \mathcal{F}(\alpha_c) \subseteq \Delta(\mathbb{R}^{2N})$, where $\mathcal{F}$ is compact-valued, increasing and non-empty, and for all α_c and $F \in \mathcal{F}(\alpha_c)$, F the law of iterated expectations holds, $\int_s F(ds, v) = G(v)$ for all $v \in \mathbb{R}^N$

The DM's payoff from consumption problem c given α_c is then

$$\hat{V}_c(\alpha_c) := \max_{F \in \mathcal{F}(\alpha_c)} \int_v \int_s v(F(\cdot|s))F(ds|v)G(dv).$$

For an example of a particular $\mathcal{F}$, suppose that the information structure is fully flexible subject to a capacity constraint; i.e., let $\bar{\mathcal{F}} := \{F \in \Delta(\mathbb{R}^{2N}) : \int_s F(ds, v) = G(v)$ for all $v \in \mathbb{R}^N\}$ (set of posterior distribution satisfying law of iterated expectations) and

$$\mathcal{F}(\alpha_c) = \{F \in \bar{\mathcal{F}} : \kappa(H(G) - E_{\mathbf{s}}[H(F(\cdot|\mathbf{s}))) \leq \alpha_c]\},$$

for some $\kappa \geq 0$ and where $H(B)$ denotes the entropy of belief B.[19]

Example 2. *Problem c is the reduced form of a canonical choice problem with trembles.*

The DM chooses an element i from set $A = \{1, \ldots, N\}$. The vector $v \in \mathbb{R}^N$ where v_i is the payoff of element $i \in A$ is known. The DM's choice is random, they "tremble": They choose $B \in \mathbb{F}(\alpha_c) \subseteq \Delta(A)$, where $\mathcal{F}$ is compact-valued, increasing and non-empty. The

[19] When the distribution of states is discrete, $H(B) = -\sum_k p_k \log(p_k)$, where p_k is the probability of state k; and for distribution that has a probability density function f, entropy is $-\int_v f(v)\log(f(v))dv$.

DM's payof from consumption problem c given α_c *is then*

$$\hat{V}_c(\alpha_c) \coloneqq E_B[v_i].$$

For an example of a particular $\mathcal{F}$*, consider*

$$\mathcal{F}(\alpha_c) = \{B \in \Delta(A) : \kappa(H(\mathcal{U}) - H(B)) \leq \alpha_c\},$$

for some $\kappa \geq 0$*, where* $H(B)$ *denotes the entropy of belief* B *(see footnote 19 for the definition) and* $\mathcal{U}$ *the uniform distribution on* A*; i.e., if the DM devotes no attention, they will make each choice with equal probability.*

Example 3. *The setup is as in Example 1; we interpret a particular* $\mathcal{F}$ *as corresponding to the DM accessing information from their memory as we describe next.*

We follow memory recall models as discussed in Kahana (2012). Endow the DM with memory $M \in \mathbb{R}^{KN}$ *which is a set of* $|M|$ *signal realization from some* $F_1(s, v) \in \Delta(\mathbb{R}^{2N})$ *with* $\int_s F_1(ds, v) = G(v)$ *for all* $v \in \mathbb{R}^N$*.* F_1 *corresponds to the distribution of individual memories (a signal) the DM has made. Given* α_c*, the DM can make up to* $\lfloor \alpha_c \frac{1}{\kappa} \rfloor$ *uniform draws with replacement from* M*. With* K *draws, probability of* J *distinct draws is* $P(J|K) \coloneqq \binom{|M|}{J} \left(\frac{J}{|M|}\right)^K$*. Define* $F_J(s_1, \ldots, s_J, v) \coloneqq \Pi_{j=1,\ldots,J} F_1(s_j|v)G(v)$ *as joint distribution of* J *distinct memories and the state.*

Finally, let $\mathcal{F}$ *be*

$$\mathcal{F}(\alpha_c) = \{\sum_{J=1}^{M} P(J|K)F_J : K \in \mathbb{N}, K \leq \lfloor \alpha_c \frac{1}{\kappa} \rfloor\}.$$

As the DM devotes more attention to c, they make more draws from their memory; a form of information acquisition.

A.2 Examples for Section 3.1

Example 4. *There are three time periods,* $T = 3$*. The consumption payoffs in periods 1 and 2 are constant and equal and denoted by* $\bar{V}$*. The consumption payoff in period 3 is either high* $\bar{V}_3$ *or low* $\underline{V}_3$ *depending on the action the DM chooses in period 1 and 2. In each* *period $t \in \{1,2\}$, the available actions are*

$$X_t(\alpha_t) = \begin{cases} \{\underline{x}\} & \text{if } < \eta_t \\ \{\underline{x}, x^* *\} & \text{if } \alpha_{t\to 3} \geq \eta_t, \end{cases}$$

in particular, taking the action x^* requires attention devoted to period 3. The payoff in period 3 is high if the DM takes action x^* in at least one period, otherwise, it is low. We also force $\alpha_{2\to 3} \geq \underline{\alpha}_2$, with $0 < \underline{\alpha}_2 < \eta_2$ (formally, this is modeled by assuming any payoff is negative infinity if the DM's attention differs).

Suppose the payoff in period 3 is lower than that in periods 1 and 2, i.e., $\underline{V}_3 < \bar{V}_3 < \bar{V}$.We construct an example where the DM in period 1 prefers action x^* to be taken in period 2 over it being taken in period 1 over it never being taken. Initially, however, the DM in period 2 would not take x^* including if the DM in period 1 did not take it and so the DM takes x^* (and devotes attention to period 3) in period 1. As the payoff in period 3 increases, this changes: the DM in period 2 now takes x^* and so the DM in period 1 does not, and hence reduces their attention to period 3. Let us derive the conditions.

In period 3, the DM devotes all their attention to $\bar{V}$ (from either other period) and takes a degenerate action. If the DM took action x^* in period 1, then in period 2, they choose $\alpha_{2\to 2} = 1 - \underline{\alpha}_2$ and $\alpha_{2\to 3} = \underline{\alpha}_2$. Otherwise, they take action x^* (and $\alpha_{2\to 2} = 1 - \eta_2$ and $\alpha_{2\to 3} = \eta_2$) over $\underline{x}$ (and $\alpha_{2\to 2} = 1 - \underline{\alpha}_2$ and $\alpha_{2\to 3} = \underline{\alpha}_2$) if

$$(1 + \lambda(1 - \eta_2))\bar{V} + (1 + \lambda\eta_2)\bar{V}_3 \geq (1 + \lambda(1 - \underline{\alpha}_2))\bar{V} + (1 + \lambda\underline{\alpha}_2)\underline{V}_3. \tag{10}$$

In period 1, the DM prefers to take action $\underline{x}$ (and $\alpha_{1\to 1} = 1$) and the DM in period 2 taking action x^* (with aforementioned attention) over taking action x^* (and $\alpha_{1\to 1} = 1 - \eta_1$ and $\alpha_{1\to 3} = \eta_1$) and the DM in period 2 taking $\underline{x}$ (with aforementioned attention) if

$$(1+\lambda(1+(1-\eta_2))\bar{V} + (1+\lambda\eta_2)\bar{V}_3 \geq (1+\lambda((1-\eta_1)+1)\bar{V} + (1+\lambda\eta_1)\bar{V}_3 \iff \eta_1 \geq \eta_2. \tag{11}$$

Still in period 1, the DM prefers taking action x^* (with aforementioned attention and action in period 2) over always taking action $\underline{x}$ (with no attention to period 3 in period 1 and *minimal in period 2) if*

$$(1 + \lambda(1 - \eta_1))\bar{V} + (1 + \lambda(\eta_1 + \underset{\sim}{\alpha}_2))\bar{V}_3 \geq (1 + \lambda)\bar{V} + (1 + \lambda\underset{\sim}{\alpha}_2)\underline{V}_3. \tag{12}$$

Since $\underline{V}_3 < \bar{V}_3 < \bar{V}$*, there exists* $\lambda > 0$ *such that* (12) *holds with equality. For such* λ*, since* $\underset{\sim}{\alpha}_2 > 0$*, there exists* $\eta_2 < \eta_1$ *(i.e.,* (11) *holds) so that* (10) *does not hold. Furthermore,* λ *can be slightly decreased so that* (12) *now holds strictly and* (10) *still does not hold. Thus, for these parameter values, the unique solution is for the DM to take action* x^* *in period 1 and hence devote* $\eta_1 > 0$ *to period 3.*

Now, increase both $\underline{V}_3$ *and* $\bar{V}_3$ *by* γ*. If* γ *is large enough (but still* $\bar{V}_3 + \gamma < \bar{V}$*), then* (10) *holds (and* (11) *and* (12) *remain to hold), so that the unique solution is for the DM to take action* x^* *in period 2 only, i.e., the DM reduces their attention to period 3 in period 1.*

A non-monotonicity of the attention devoted to period 3 as a function of β_3 *(as in the parameterization used for the comparative statics) can be constructed similarly, but is omitted.*

Example 5. *Take the setting of Example 4. The construction proceeds almost identically.*

Since $\underline{V}_3 < \bar{V}_3 < \bar{V}$*, there exists* $\lambda > 0$ *such that* (12) *holds with equality. For such* λ*, since* $\underset{\sim}{\alpha}_2 > 0$*, there exists* $\eta_2 < \eta_1$ *(i.e.,* (11) *holds) so that* (10) *does not hold. Furthermore,* λ *can be slightly decreased so that* (12) *now holds strictly and* (10) *still does not hold. Thus, for these parameter values, the unique solution is for the DM to take action* x^* *in period 1 and hence devote* $\eta_1 > 0$ *to period 3.*

Now decrease λ *to something still strictly positive but so that* (10) *holds. As before, the DM now takes action* x^* *in period 2. Of course, the unweighted consumption payoff is unchanged. However, all comparisons in our constructions are strict; thus, assuming that taking action* x^* *in period 2 only leads to a payoff of* $\bar{V}_3 - \epsilon$ *in period 3 does not change the construction for* $\epsilon > 0$ *small enough. In this case, decreasing* λ *leads to a decrease in the unweighted consumption payoff.*

A.3 Example of bad news about the quality of a product increasing consumption

Example 6. *This example builds on the ideas of Tasoff and Madarasz (2009).*

Consider the setup of Section 2.1 and suppose $\mathcal{C} = \{c, m\}$*. Consumption problem* c *corresponds to the DM purchasing a quantity of a consumption good at unit price* 1*. Their valuation of quantity* k *is* $\theta u(k)$*, where* u *is strictly concave and continuously differentiable, and* $\theta \in \{\theta_L, \theta_H\}$ *with* $P(\theta = \theta_H) = p \in (0, 1)$*. The DM has wealth* 1 *available and whatever amount they do not consume,* $1 - k$*, leads to payoff* $1 - k$ *as part of problem* m *(the "money" problem).*

We assume that $\lim_{k \to 0} \frac{\partial}{\partial k} u(k) = \infty$ *and* $\frac{\partial}{\partial k} u(1) = 0$ *so that the DM always chooses an interior* k*.*

The DM can learn the value of θ *by choosing* $\alpha_c = 1$ *(formally, such attention allows for some action* x *that corresponds to learning the value of* θ*). Otherwise, they decide on* k *before knowing* θ *and receive the expected payoff from consumption. The DM will optimally either choose* $\alpha_c = 1$ *or* $\alpha_c = 0$*.*

Suppose the DM learns the value of θ*. Then they choose* c *to satisfy*

$$(1 + \lambda)\theta u'(c) = 1.$$

If they do not learn θ*, the DM chooses* c *to satisfy*

$$E[\theta] u'(c) = 1 + \lambda.$$

(The values of $V_c(x), V_m(x)$ *are the expected payoffs with the just derived optimal level of consumption.)*

Thus, if $1 + \lambda > \frac{E[\theta]}{\theta_L}$*, the DM consumes more of the good if they receive the information and learn it is of low value compared to when they do not receive the information.*

B Attention utility in a strategic environment

In this section we extend the environment of Section 2.2 to allow for strategic interaction. We use the a setup similar to that of Brunnermeier and Parker (2005) Section III, suitably adjusted to our model. That is, there is a unit mass of agents with the same continuously differentiable Bernoulli utility function u situated in an exchange economy with no aggregate risk. Each agent is initially endowed with one unit of a safe and can purchase a risky asset that is in zero net supply. The price of the risky asset is P and determined in equilibrium. The risky asset's net return is random and denoted by x^r with payoff x^r_s in state s, where $x^r_s \neq x^r_{s'}$ for all $s \neq s'$.

Agent i acquiring an amount ξ^i of the risky assets leads to monetary payoff of $c^i_s = 1 - \xi^i + \xi^i \frac{1+x^r_s}{P}$ in state s. Thus, each agent i takes the price of the risky asset P as given and chooses a lottery x^i from the set $X(P) = \{\delta_1 - \xi^i + \xi^i \frac{1+x^r}{P} : c^i_s \geq 0, \forall s \in \mathcal{S}\}$. An equilibrium is a price P and choice of lottery x^i and attention allocation α^i for each agent i (with amount ξ^i purchased of the risky asset) such that each agent maximizes their objective

$$\sum_{s \in \mathcal{S}} p_s u(c_s) \tag{13}$$

and $\int_i \xi^i = 0$.

Proposition 12. *An equilibrium exists. If $\lambda > 0$, for $|\mathcal{S}| \geq 2$, agents have heterogeneous subjective beliefs q_s such that there exists a subset $\mathcal{I}$ some agents hold the risky asset and some agents short the risky asset.*

Proof of Proposition 12. If $|\mathcal{S}| = 1$, then it must be that $P = 1 + x^r_s$, and each agent maximizes their objective, e.g., with $\xi^i = 0$.

Suppose $|\mathcal{S}| > 1$. Let $\bar{s} = \arg\max_{s \in \mathcal{S}} x^r_s$ and $\underline{s} = \arg\min_{s \in \mathcal{S}} x^r_s$. First note that $\xi^i \neq 0$ as thus it must be that $\alpha^i_{\bar{s}} = 1$ or $\alpha^i_{\underline{s}} = 1$. To see this, suppose $\xi^i = 0$. Given $\xi^i = 0$, both $\alpha^i_{\bar{s}} = 1$ and $\alpha^i_{\underline{s}} = 1$ are optimal, but $\xi^i = 0$ cannot be optimal for both $\alpha^i_{\bar{s}} = 1$ and $\alpha^i_{\underline{s}} = 1$, a contradiction. Conditional on $\alpha^i_{\bar{s}} = 1$ ($\alpha^i_{\underline{s}} = 1$), the payoff (13) is continuously decreasing (increasing) in P. Furthermore, for large enough P, the payoff given $\alpha^i_{\bar{s}} = 1$ is less than that given $\alpha^i_{\underline{s}} = 1$ with optimal ξ^i. Thus, there exists a unique P^* such that each agent i is indifferent between $\alpha^i_{\bar{s}} = 1$ and $\alpha^i_{\underline{s}} = 1$. Furthermore, at P^*, it must be that $\alpha^i_{\bar{s}} = 1$ implies be that $\xi^i > 0$ and $\alpha^i_{\underline{s}} = 1$ implies be that $\xi^i < 0$ (otherwise, the associated payoffs cannot be the same). But then $\int_i \xi^i = 0$ is easily achieved by assigning the appropriate masses of agents to either $\alpha^i_{\bar{s}} = 1$ or $\alpha^i_{\underline{s}} = 1$ with the corresponding optimal ξ^i. The heterogeneity in attention implies the heterogeneity in subjective beliefs. $\square$

C Proofs

C.1 Proposition 1

We first state a version of Proposition 1 that does not rely on uniqueness of the solutions which we subsequently prove.

Proposition 1*. *Take consumption problem $i \in \mathcal{D}$. Fix V_{-i}, and let $\Gamma(\gamma_i, \beta_i)$ denote the set of optimal (action, attention)-pairs.*

- *If $\lambda > 0$: If $\gamma_i^{'} > \gamma_i$ then $\min_{(x,\alpha)\in\Gamma(\gamma_i^{'},\beta_i)} \alpha_i \geq \max_{(x,\alpha)\in\Gamma(\gamma_i,\beta_i)} \alpha_i$. If, in addition, the environment is separable, then $\min_{(x,\alpha)\in\Gamma(\gamma_i^{'},\beta_i)} V_i(x) - \max_{(x,\alpha)\in\Gamma(\gamma_i,\beta_i)} V_i(x) \geq \gamma_i^{'} - \gamma_i$.*
- *If for β_i and γ_i, $\max_{(x,\alpha)\in\Gamma(\gamma_i,\beta_i)} V_i(x) = \min_{(x,\alpha)\in\Gamma(\gamma_i,\beta_i)} V_i(x)$, then for any $\beta_i^{'} > \beta_i$ and $\gamma_i^{'} = \gamma_i - (\beta_i^{'} - \beta_i)\tilde{V}_i(x)$, where $(x,\alpha) \in \Gamma(\gamma_i,\beta_i)$, we have $\min_{(x,\alpha)\in\Gamma(\gamma_i^{'},\beta_i^{'})} V_i(x) \geq \max_{(x,\alpha)\in\Gamma(\gamma_i,\beta_i)} V_i(x)$. If, in addition, the environment is separable, then $\min_{(x,\alpha)\in\Gamma(\beta_i^{'},\gamma_i^{'})} \alpha_i \geq \max_{(x,\alpha)\in\Gamma(\beta_i,\gamma_i)} \alpha_i$.*

It is immediate that Proposition 1* implies Proposition 1*.

Proof of Proposition 1.* Take any $\gamma_i^{'}, \gamma_i$ with $\gamma_i^{'} > \gamma_i$ and β_i. Let (x,α) and $(\alpha^{'}, a^{'})$ denote a solution given γ_i and γ_i', respectively. Optimality of (x, α) and (α', a') implies

$$\begin{aligned}
&\underbrace{\sum_{i' \in \mathcal{D} \setminus \{i\}} (\omega_{i'} + \lambda(\alpha_{i'} + \psi_{i'}))V_{i'}(x)}_{:=\kappa_0} + (\omega_i + \lambda(\alpha_i + \psi_i))(\beta_i \tilde{V}_i(x) + \gamma_i) \\
\geq &\underbrace{\sum_{i' \in \mathcal{D} \setminus \{i\}} (\omega_{i'} + \lambda(\alpha'_{i'} + \psi_{i'}))V_{i'}(x')}_{:=\kappa_1} + (\omega_i + \lambda(\alpha'_i + \psi_i))(\beta_i \tilde{V}_i(x') + \gamma_i) \quad \text{and} \\
&\underbrace{\sum_{i' \in \mathcal{D} \setminus \{i\}} (\omega_{i'} + \lambda(\alpha'_{i'} + \psi_{i'}))V_{i'}(x')}_{=\kappa_1} + (\omega_i + \lambda(\alpha'_i + \psi_i))(\beta_i \tilde{V}_i(x') + \gamma'_i) \\
\geq &\underbrace{\sum_{i' \in \mathcal{D} \setminus \{i\}} (\omega_{i'} + \lambda(\alpha_{\prime} + \psi_{i'}))V_{i'}(x)}_{=\kappa_0} + (\omega_i + \lambda(\alpha_i + \psi_i))(\beta_i \tilde{V}_i(x) + \gamma'_i).
\end{aligned}$$

Combining the above, gives

$$\begin{aligned}
&-\left((\omega_i + \lambda(\alpha_i + \psi_i))(\beta_i \tilde{V}_i(x) + \gamma'_i) - (\omega_i + \lambda(\alpha'_i + \psi_i))(\beta_i \tilde{V}_i(x') + \gamma'_i)\right) \\
\geq &\kappa_0 - \kappa_1 \\
\geq &-\left((\omega_i + \lambda(\alpha_i + \psi_i))(\beta_i \tilde{V}_i(x) + \gamma_i) - (\omega_i + \lambda(\alpha'_i + \psi_i))(\beta_i \tilde{V}_i(x') + \gamma_i)\right).
\end{aligned}$$

The outer inequality implies

$$-\lambda(\alpha_i - \alpha'_i)(\gamma'_i - \gamma_i) \geq 0,$$

and thus, it must be that $\alpha'_i \geq \alpha_i$ as $\lambda > 0$.

If the environment is separable, then $\tilde{V}_i$ is increasing in the amount of attention α_i devoted to dimension i, and the result follows.

Take any $\beta_i, \beta'_i \geq 0$ with $\beta'_i > \beta_i$ and γ_i and suppose that $\max_{(x,\alpha) \in \Gamma^*(\gamma_i, \beta_i)} V_i(x) = \min_{(a,\alpha) \in \Gamma(\gamma_i, \beta_i)} V_i(x)$. Let $\gamma_i^{'} = \gamma_i - (\beta_i^{'} - \beta_i)\tilde{V}_i(x)$, where $(x, \alpha) \in \Gamma(\gamma_i, \beta_i)$. Let (x, α) and (x', α') denote a solution given (β_i, γ_i) and (β'_i, γ'_i), respectively. Optimality of (x, α) and

(x', α') implies

$$\underbrace{\sum_{i' \in \mathcal{D} \setminus \{i\}} (\omega_{i'} + \lambda(\alpha_{i'} + \psi_{i'}))V_{i'}(x)}_{:=\kappa_2} + (\omega_i + \lambda(\alpha_i + \psi_i))(\beta_i \tilde{V}_i(x) + \gamma_i)$$
$$\geq \underbrace{\sum_{i' \in \mathcal{D} \setminus \{i\}} (\omega_{i'} + \lambda(\alpha'_{i'} + \psi_{i'}))V_{i'}(x')}_{:=\kappa_3} + (\omega_i + \lambda(\alpha'_i + \psi_i))(\beta_i \tilde{V}_i(x') + \gamma_i)$$
$$\underbrace{\sum_{i' \in \mathcal{D} \setminus \{i\}} (\omega_{i'} + \lambda(\alpha'_{i'} + \psi_{i'}))V_{i'}(x')}_{=\kappa_3} + (\omega_i + \lambda(\alpha'_i + \psi_i))(\beta'_i \tilde{V}_i(x') + \gamma'_i) \quad \text{and}$$
$$\geq \underbrace{\sum_{i' \in \mathcal{D} \setminus \{i\}} (\omega_{i'} + \lambda(\alpha_{i'} + \psi_{i'}))V_{i'}(x)}_{=\kappa_2} + (\omega_i + \lambda(\alpha_i + \psi_i))(\beta'_i \tilde{V}_i(x) + \gamma'_i).$$

Combining the above and substituting for γ'_i gives

$$\begin{aligned}
&-\Big((\omega_i + \lambda(\alpha_i + \psi_i))(\beta_i \tilde{V}_i(x) + \gamma_i) - (\omega_i + \lambda(\alpha'_i + \psi_i))(\beta'_i \tilde{V}_i(x') - (\beta'_i - \beta_i)\tilde{V}_i(x))\Big) \\
\geq& \kappa_2 - \kappa_3 \\
\geq& -\Big((\omega_i + \lambda(\alpha_i + \psi_i))(\beta_i \tilde{V}_i(x) + \gamma_i) - (\omega_i + \lambda(\alpha'_i + \psi_i))(\beta_i \tilde{V}_i(x') + \gamma_i)\Big).
\end{aligned}$$

The outer inequality implies

$$-(\omega_i + \lambda(\alpha'_i + \psi_i))(\tilde{V}_i(x) - \tilde{V}_i(x'))(\beta'_i - \beta_i) \geq 0,$$

and thus, it must be that $\tilde{V}_i(x') \geq \tilde{V}_i(x)$.

If the environment is separable, then $\tilde{V}_i$ is increasing in the amount of attention α_i devoted to dimension i, and the result follows. $\square$

C.2 Proof of Proposition 2

Proof of Proposition 2. Take any $\lambda', \lambda \geq 0$ with $\lambda' > \lambda$. Let (x, α) and (x', α') denote a solution given λ and λ', respectively. Optimality of (x, α) and (x', α') implies

$$\sum_{i\in\mathcal{D}} \omega_i V_i(x) + \lambda \sum_{i\in\mathcal{D}} (\alpha_i + \psi_i) V_i(x) \geq \sum_{i\in\mathcal{D}} \omega_i V_i(x') + \lambda \sum_{i\in\mathcal{D}} (\alpha'_i + \psi_i) V_i(x'), \quad \text{and}$$
$$\sum_{i\in\mathcal{D}} \omega_i V_i(x') + \lambda' \sum_{i\in\mathcal{D}} (\alpha'_i + \psi_i) V_i(x') \geq \sum_{i\in\mathcal{D}} \omega_i V_i(x) + \lambda' \sum_{i\in\mathcal{D}} (\alpha_i + \psi_i) V_i(x).$$

Combining the above, gives

$$\begin{aligned} -\lambda' \left(\sum_{i\in\mathcal{D}} (\alpha_i + \psi_i) V_i(x) - \sum_{i\in\mathcal{D}} (\alpha'_i + \psi_i) V_i(x') \right) &\geq \sum_{i\in\mathcal{D}} \omega_i V_i(x) - \sum_{i\in\mathcal{D}} \omega_i V_i(x') \\ &\geq -\lambda \left(\sum_{i\in\mathcal{D}} (\alpha_i + \psi_i) V_i(x) - \sum_{i\in\mathcal{D}} (\alpha'_i + \psi_i) V_i(x') \right). \end{aligned}$$

If the expression in the middle is strictly negative, so must be the right one; but then it is strictly larger than the left one as $\lambda' > \lambda$. Thus, the first claim follows.

Now consider two sets of payoff levels, $(\gamma_i)_{i\in\mathcal{D}}$ and $(\gamma'_i)_{i\in\mathcal{D}}$, and scalar $\chi \in [0, 1]$. Then

$$\begin{aligned} &\max_{\alpha, x\in X(\alpha)} \sum_{i\in\mathcal{D}} (\omega_i + \lambda(\alpha_i + \psi_i))(\beta_i \tilde{V}_i(x) + \chi\gamma_i + (1-\chi)\gamma'_i) \\ =& \max_{\alpha, x\in X(\alpha)} \left(\chi \sum_{i\in\mathcal{D}} (\omega_i + \lambda(\alpha_i + \psi_i))(\beta_i \tilde{V}_i(x) + \gamma_i) + (1-\chi) \sum_{i\in\mathcal{D}} (\omega_i + \lambda(\alpha_i + \psi_i))(\beta_i \tilde{V}_i(x) + \gamma'_i) \right) \\ \geq& \chi \max_{\alpha, x\in X(\alpha)} \sum_{i\in\mathcal{D}} (\omega_i + \lambda(\alpha_i + \psi_i))(\beta_i \tilde{V}_i(x) + \gamma_i) + (1-\chi) \max_{\alpha, x\in X(\alpha)} \sum_{i\in\mathcal{D}} (\omega_i + \lambda(\alpha_i + \psi_i))(\beta_i \tilde{V}_i(x) + \gamma'_i), \end{aligned}$$

and so the second claim follows.

Now suppose the environment is separable and that the objective given λ is convex in α. First, since the environment is separable, the additive separability of (2) in i implies that (2) convex in α is equivalent to $(\omega_i + \lambda(\alpha_i + \psi_i))\hat{V}_i(\alpha_i)$ convex in α_i for all $i \in \mathcal{D}$ (where $\hat{V}_i(\alpha_i) := \max_{x_i \in X_i(\alpha_i)} V_i(x_i)$). Note that as $X_i(\alpha_i)$ is increasing in α_i, $\hat{V}_i$ is increasing. Thus, it suffices to show that if $(\omega_i + \lambda(\alpha_i + \psi_i))\hat{V}_i(\alpha_i)$ is convex in α_i, then so is $(\omega_i + \lambda'(\alpha_i + \psi_i))\hat{V}_i(\alpha_i)$.

Observe that when $\hat{V}_i(\alpha_i)$ is convex in α_i, then so is $\alpha_i \hat{V}_i(\alpha_i)$. Formally, take any $\chi[0, 1]$

and α_i, α_i' with $\alpha_i < \alpha_i'$. Then

$$
\begin{aligned}
&\chi\alpha_i\hat{V}_i(\alpha_i) + (1-\chi)\alpha_i'\hat{V}_i(\alpha_i') \\
=&a(\chi\hat{V}_i(\alpha_i) + (1-\chi)\hat{V}_i(\alpha_i')) + (\alpha_i' - \alpha_i)(1-\chi)\hat{V}_i(\alpha_i') \\
\geq&\alpha_i\hat{V}_i(\chi\alpha_i + (1-\chi)\alpha_i') + (\alpha_i' - \alpha_i)(1-\chi)\hat{V}_i(\alpha_i') \\
=&\chi\alpha_i\hat{V}_i(\chi\alpha_i + (1-\chi)\alpha_i') + (1-\chi)(\alpha_i\hat{V}_i(\chi\alpha_i + (1-\chi)\alpha_i') + (\alpha_i' - \alpha_i)\hat{V}_i(\alpha_i')) \\
\geq&\chi\alpha_i\hat{V}_i(\chi\alpha_i + (1-\chi)\alpha_i') + (1-\chi)(\alpha_i\hat{V}_i(\chi\alpha_i + (1-\chi)\alpha_i') + (\alpha_i' - \alpha_i)\hat{V}_i(\chi\alpha_i + (1-\chi)\alpha_i')) \\
=&\chi\alpha_i\hat{V}_i(\chi\alpha_i + (1-\chi)\alpha_i') + (1-\chi)(\alpha_i'\hat{V}_i(\chi\alpha_i + (1-\chi)\alpha_i')),
\end{aligned}
$$

where the first inequality follows by assumption, and the second as $\hat{V}_i$ is increasing. Thus, since $(\omega_i + \lambda(\alpha_i + \psi_i)\hat{V}_i(\alpha_i)$ is a linear combination of $\hat{V}_i(\alpha_i)$ and $\alpha_i\hat{V}_i(\alpha_i)$ with the relative weight on the latter increasing in λ, the third claim follows. □

C.3 Proof of Proposition 4

Proof of Proposition 4. Take any λ, λ' with $\lambda' > \lambda$, lottery a and x, and suppose that the DM(λ) prefers x to δ_y for arbitrary payoff y, i.e.,

$$\frac{p_{\bar{S}} + \lambda}{1+\lambda}u(x_{\bar{S}}) + \sum_{s\in\mathcal{S}\setminus\bar{S}} \frac{p_s}{1+\lambda}u(x_s) \geq u(y),$$

where the DM optimally devotes full attention to the states with the highest payoff, $\bar{S} \coloneqq \arg\max_{s\in\mathcal{S}} u(x_s)$. We rewrite the above as the expected material payoff plus attention utility (divided by $1+\lambda$), i.e.,

$$\frac{1}{1+\lambda}\sum_{s\in\mathcal{S}} p_s u(x_s) + \frac{\lambda}{1+\lambda}u(x_{\bar{S}}).$$

As $u(x_{\bar{S}}) \geq \sum_{s\in\mathcal{S}} u(x_s)$, the above is increasing in λ and so DM(λ') also prefers x to δ_y.

Take any μ, L and $x \in X(\mu, L)$. Consider lottery x'; we can bound the DM's payoff from x as

$$\frac{p_{\bar{S}} + \lambda}{1+\lambda}u(H(x')) + \frac{p_{\underline{S}}}{1+\lambda}u(L(x')) \geq \frac{\lambda}{1+\lambda}u(H(x')) + \frac{1}{1+\lambda}u(L(a')).$$

Since u is unbounded, the above goes to infinity as $H(x')$ goes to infinity. Thus, there exists some lottery $\hat{x} \in X(\mu, L)$ such that for all x' with $H(x') > H(\hat{x})$, the DM prefers x' to x.

Take any μ, L and $x, x' \in X(\mu, L)$ with $H(x) > H(x')$. The DM's payoff from x is

$$\frac{p_{\bar{S}} + \lambda}{1+\lambda} u(H(x)) + \frac{p_{\underline{S}}}{1+\lambda} u(L(x)),$$

and similarly for lottery x'. The above converges to $u(H(x))$ as λ goes to infinity. As $u(H(x)) > u(H(x'))$, the final result follows. $\square$

C.4 Proof of Proposition 5

Proof of Proposition 5. When $\hat{V}_s, \hat{V}'_s$ are constant, the DM chooses α to maximize

$$\frac{p_s + \lambda}{1+\lambda} \hat{V}_s + \frac{p_{s'} + \lambda}{1+\lambda} \hat{V}_{s'};$$

which is strictly increasing in α_s (using $\alpha_s + \alpha_{s'} = 1$) since $\hat{V}_s > \hat{V}_{s'}$, and hence $\alpha_s = 1$ and $\alpha_{s'} = 1$, and the claim follows.

Suppose that $\hat{V}_s = \hat{V}_{s'} = V$, with $\hat{V}$ continuously differentiable, $\lim_{a \to 0} a \frac{\partial}{\partial a} \hat{V}(a) = \infty$, and $\frac{\partial}{\partial a} \hat{V}(1) < \infty$. $q_s = q_{s'}$ since the labels, s, s', can be exchanged in the DM's objective. For $p_s = 0$, since $\hat{V}$ is increasing and not constant (by the limit condition), the DM optimally devotes full attention to state s'. Hence, $q_s(p) = 0$. The DM's overall payoff is given by

$$\frac{p_s + \lambda \alpha_s}{1+\lambda} \hat{V}(\alpha_s) + \frac{1 - p_s + \lambda(1-\alpha_s)}{1+\lambda} \hat{V}(1-\alpha_s).$$

Differentiating the above gives

$$\frac{(p_s + \lambda \alpha_s) \frac{\partial}{\partial \alpha} \hat{V}(\alpha_s) - ((1-p_s) + \lambda(1-\alpha_s)) \frac{\partial}{\partial \alpha} \hat{V}(1-\alpha_s)}{1+\lambda} + \frac{\lambda(\hat{V}(\alpha_s) - \hat{V}(1-\alpha_s))}{1+\lambda}.$$

The above is decreasing in p_s. Furthermore, for $p_s = 0$ and as $\alpha_s \to 0$, the above tends to infinity. Thus, there exists a set $(0, \bar{\alpha}_s)$, with $\bar{\alpha} > 0$, such that the above is strictly positive for all $\alpha_s \in (0, \bar{\alpha}_s)$ for any p_s. For any $p_s > 0$ and $\alpha_s = 0$, the above is strictly positive as $\frac{\partial}{\partial \alpha} \hat{V}(\alpha_s) = \infty$. Thus, for any $p_s > 0$, the DM chooses $\alpha_s > 0$. Furthermore, they choose

$\alpha_s \geq \bar{\alpha}_s$. Thus, for $0 < p_s < \bar{\alpha}_s$, we have $\alpha_s > p_s$ and thus $q_s(p_s) > p_s$. (If $q_s(p_s)$ is a set, then the comparison applies to each element of $q_s(p_s)$.) The remaining comparisons follow from the symmetry of q_s.

Lastly, suppose that $\hat{V}_s = \hat{V}_{s'} = \hat{V}$ and that $\hat{V}$ is convex and not constant. Since $\hat{V}$ is convex, the DM's overall payoff is convex ($\hat{V}(\alpha_s)$ and α_s are increasing and nonnegative convex functions, and so $\alpha_s\hat{V}(\alpha_s)$ is convex, and adding convex functions also preserves convexity). Given p_s, the DM's payoff from $\alpha_s = 1$ and $\alpha_s = 0$ is

$$\frac{p_s + \lambda}{1+\lambda}\hat{V}(1) + \frac{1-p_s}{1+\lambda}\hat{V}(0), \quad \text{and}$$
$$\frac{1-p_s+\lambda}{1+\lambda}\hat{V}(1) + \frac{p_s}{1+\lambda}\hat{V}(0),$$

respectively. The former is strictly greater than the later if $p_s > 1/2$, strictly less if $p_s < 1/2$, and equal for $p_s = 1/2$. Thus, the probability weighting for $q(p) \neq 1/2$ follows. For $p_s = 1/2$, note that either of the above is larger than, e.g., $\hat{V}(1/2)$, the DM's payoff if they devote equal attention, since $\hat{V}(1) > \hat{V}(0)$. Hence, full or no attention is uniquely optimal, completing the proof of the final claim. $\square$

C.5 Proof of Proposition 6

Proof of Proposition 6. Notice that when $\lambda = 0$, the DM, in each period t, maximizes the unweighted sum of consumption payoffs. Since V is strictly concave, by Jensen's inequality, this sum is uniquely maximized when $\sum_{t''=1}^{t} x_{t''\to t'} = 1$ for all t'. If in each previous period, the DM only devoted attention to that period, then for $t' = t$, this sum equals $\alpha_{t\to t}$; hence, the unique attention allocation achieving this optimum is $\alpha_{t\to t}$ for all periods t. λ changes the overall payoff continuously; hence, for λ small enough, the above still maximizes the DM's overall payoff in each period. Furthermore, this attention allocation is implementable in equilibrium. Hence, the first claim follows.

Normalizing (3) by $1+\lambda$, when $\lambda = \infty$, the DM's overall payoff in each period t is given by

$$\sum_{t'=t}^{T}\sum_{t''=t}^{T} \alpha_{t''\to t'} V_{t'}(\sum_{t''=1}^{t} x_{t''\to t'}).$$

This expression is maximized when the DM, in each period t'', devotes attention to a period t', with $\sum_{t''=1}^{t} x_{t''\to t'} \geq K$. The unweighted consumption payoff given one of these optimal attention allocations for $\lambda = \infty$ is maximized when this inequality holds with equality; the unique such attention allocation is the one mentioned in the proposition statement. Thus, as λ changes the overall payoff continuously, increasing the weight on the unweighted consumption payoffs, the claim follows. $\square$

C.6 Proof of Proposition 7

Proof of Proposition 7. Suppose $X_1(\alpha_1)$ is independent of α_1. We show that the DM is time-consistent which implies the first claim as a simple optimality condition. Let $V^* :=$ $\max\{\max_{s\in\mathcal{S}} V_{c_0}(x_1, x_2(x_1,s)|s), V_{c_1}\}$. If $V_{c_1} \geq V^*$, then the DM devotes all their attention in period-1 to c_1; if $V_{c_1} < V^*$, then the DM devotes attention to c_0 and s^*, where s^* is the corresponding argmax. In either case, the DM's action in the second period always maximizes the DM's objective in the first (they are time-consistent) and the first part of the proposition follows.

For the second part, fix x_1 and consider a realized s. Clearly, if $\max_{x_2\in X_2(\alpha_2)} V_{c_0}(x_1, x_2|s) \geq$ V_{c_1}, solving (4) gives $\alpha_{2\to c_0} = 1$ (for any λ). If $\max_{x_2\in X_2(\alpha_2)} V_{c_0}(x_1, x_2|s) < V_{c_1}$, then (4) for $\alpha_{2\to c_0} = 0$ (and some finite $V_{c_0}(x_1, x_2|s)$) is larger than (4) for any $\alpha_{2\to c_0} \geq \eta$ for $\max_{x_2\in X_2(\alpha_2)} V_{c_0}(x_1, x_2|s)$ when λ is large enough. Since $\mathcal{C}$ is finite, taking the max λ implies the result. $\square$

C.7 Proof of Proposition 8

Proof of Proposition 8. Consider $x_1 \notin \arg\max_{x_1\in X_1} F(x_1)$. For λ large enough, by Proposition 7, and since $\max_{\alpha_2, x_2\in X_2(\alpha_2)} \tilde{V}_{c_0}(x_1, x_2|s) + \beta F(x_1) < \max_{\alpha_2, x_2\in X_2(\alpha_2)} \tilde{V}_{c_0}(x_1^*, x_2|s) +$ $\beta F(x_1^*)$, we have $B(x_1) \subseteq B(x_1^*) \neq$ (where non-emptiness follows from $V_{c_0}(x_1^*, x_2(x_1^*, s^*)) >$ V_{c_1}) and hence, again for λ large, the DM's second-period payoff (i.e., (4)) is strictly larger with x_1^* than with x_1.

For the first period, the DM's objective can be written as the sum of (4) and their attention utility in the first period. But the latter is also increasing when choosing x_1^* instead of x_1, for large enough λ, as the DM then gets V^*. $\square$

C.8 Proof of Proposition 9

Proof of Proposition 9. Let $V_{c:p} := pV_H + (1-p)V_L$. We begin with the second period. The DM chooses $\alpha_{2\to c}$ to maximize

$$(1 + \lambda\alpha_{2\to c})V_{c:p_2} + (1 + \lambda(1 - \alpha_{2\to c})\bar{V};$$

and the claim about optimal $\alpha_{2\to c}$ follows.

Given this optimal attention allocation, in the first period, using the fact that $E_{p_2\sim x_1}[V_{c:p_2}]] = V_{c:p_1}$, and subtracting $V_{c:p_1}$ and $\bar{V}$ from the DM's objective, the DM's objective in the first period is given by

$$\lambda\alpha_{1\to E[c]}V_{c:p_1} + \lambda(1 - \alpha_{1\to E[c]})\bar{V} + E_{p_2\sim x_1}\left[\max\{\lambda V_{c:p_2} - \bar{V}, 0\}\right].$$

If $p_1 \geq \bar{p}$, then $V_{c:p_1} \geq \bar{V}$, and so devoting full attention (weakly) maximizes the DM's payoff considering the first period only; and strictly so if $p_1 > \bar{p}$. Furthermore, the second part of the DM's overall objective is also (weakly) increasing in $\alpha_{1\to E[c]}$ (for optimally chosen x_1), by Proposition 2.

Next, note that the DM does not devote attention to c in either period if $p_1 = 0$. Thus, there exists some $\tilde{p} \leq \bar{p}$ such that it is optimal for the DM to devote attention to c when $p_1 = \tilde{p}$, but not for any $p_1 < \tilde{p}$.

Take any $p_1 < p_1' \leq \bar{p}$, what remains to show is that if it is optimal for the DM to devote some attention when their prior is p_1, then it is optimal to devote some attention when their prior is p_1'. Note that it is without loss to assume the DM acquires a binary signal. (Formally, any posterior distribution can be replaced with a binary distribution with values $E[p|p_2 \geq \bar{p}]$ and $E[p|p_2 < \bar{p}]$, with probability $P(p_2 \geq \bar{p})$ and $P(p_2 < \bar{p})$, respectively. This distribution of posteriors has a lower variance than the original distribution and gives the same overall payoff.) Let $p_H > p_1$ and $p_L < p_1$ be the posteriors that result from the acquired information given prior p_1, and let $P(p_2 = p_L)$ be the probability of the low posterior. Consider $p_H' = p_H$ and $p_L' = \frac{p_1' - p_1}{P(p_2 = p_L)}$, occurring with the same probabilities as p_H and p_L, respectively. This new posterior distribution has mean p_1', its variance is less than that given p_1, and the second-period payoff it induces is unchanged. Since its variance is less than the variance of p_H and p_L and since $V_{c:p_1} < V_{c:p_1'}$, the cost, i.e., the reduction in attention utility in the first period decreases. Hence, $\alpha_{1\to E[c]} > 0$ is optimal given p_1'.

Next, we prove the skewness result. Suppose $p_1 \in (\tilde{p}, \bar{p})$. If the DM acquires information, it is uniquely optimal to acquire a binary signal leading to posterior p_H or p_L. The variance of such posteriors is given by $P(p_2 = p_L)P(p_2 = p_H)(p_H - p_L)^2$.

For optimal x_1, it must be that $p_H > \bar{p}$. Thus, it must be that $P(p_2 = p_L)P(p_2 = p_H) \leq \frac{\beta}{(\bar{p}-p_1)^2}$. If the right-hand side is small enough, it must be that either the high posterior or the low posterior are likely. But the p_H cannot be more likely than p_L, since it is bounded away from p_1 (as $p_1 < \bar{p} < p_H$) and so it must be that the low posterior is more likely; i.e., $P(p_2 = p_L) > 1/2$ and so the distribution of posteriors is positively skewed.

Now, suppose that $p_1 > \bar{p}$. Then, $\alpha_{1\to E[c]} = 1$. Now, it must be that $p_L < \bar{p}$ (and $p_H > \bar{p}$) for x_1 optimal. Hence, $P(p_2 = p_L)P(p_2 = p_H) \leq \frac{\beta}{(p_1-\bar{p})^2}$. For κ small enough, it must again be that either the high or low posterior are likely. Now, the low posterior cannot be likely since it is bounded away from p_1 (as $p_L < \bar{p} < p_1$) and so it must be that the high posterior is likely; i.e., $P(p_2 = p_L) > 1/2$ and so the distribution of posteriors is negatively skewed.

For the third part of the proposition, first note that the comparative static results regarding $\bar{p}$ following straight from its definition.

For the comparative statics with respect to $\tilde{p}$, take any p_1 and suppose that it is optimal to acquire a positive amount of information; we show that for any increase in v_L, v_H, decrease in v or $v_H - v_L$ holding $\bar{p}$ and $V_L + (1 - p_1)V_H - \bar{V}$ fixed, it is still optimal for the DM to acquire some information. In fact, we show that 1) the increase in the DM's second-period payoff from acquiring the same information, i.e., devoting attention $\alpha_{1\to E[c]}$ and choosing $x_1 \in X_1(\alpha_1)$, weakly increases, and 2) the "cost" in terms of attention utility in the first period decreases.

Indeed, increasing V_H or V_L clearly implies 1). To see 2), note that $E_{p_2 \sim x_1}\left[\max\{V_{c:p_2} - \bar{V}, 0\}\right]$ increases as it the the inner term increases for each p_2. If the second-period payoff (the inner term) for $p_2 = p_1$ also increases, then it must be that p_1 is now at least $\bar{p}$, in which case the DM devotes full attention. Hence, 2) follows as well.

Since a decrease in $\bar{V}$ is equivalent to an increase of V_H and V_L by an equal amount, this comparative static also follows. By construction, when $V_H - V_L$ increases while $\bar{p}$ and $V_L+(1-p_1)V_H-\bar{V}$ are fixed, the DM's "cost" of acquiring information, effect on first-period "attention utility," is unchanged. Furthermore, $V_{c:p_2}$ as a function of p_2, when $v_H - v_L$ is increased, still intersecting $\bar{V}$ at $p_2 = \bar{p}$. Hence, $E_{p_2 \sim x_1}\left[\max\{V_{c:p_2} - \bar{V}, 0\}\right] - \max\{V_{c:p_1} - \bar{V}, 0\}$ increases.

Lastly, the fact that $\bar{p}$ is independent of λ follows from its definition; the fact that $\tilde{p}$ is also independent follows from the fact that the DM's objective is a multiple of λ. $\square$

C.9 Proofs of Results 1–5

Proof of Result 1. Take two payoffs from the trivial problem $\bar{V}, \bar{V}'$ with $\bar{V}' > \bar{V}$, and let (x_1, α_1) and (x_1', α_1') maximize (7) (that is, the DM's objective conditional on participating in the portfolio choice problem). We compare the change in (7) to the change in the payoff from not participating as the consumption payoff from the trivial problem decreases from $\bar{V}'$ to $\bar{V}$.

The latter is simply $(1+\lambda 2)(\bar{V}' - \bar{V})$. Consider (7) given $\bar{V}$ evaluated at (x_1', α_1'). Note that (7) can be written as

$$\underbrace{\lambda\left(\sum_\rho \alpha_{1\to\rho} V_\rho(x_1, x_2(\rho, x_1)) + \alpha_{1\to t}\bar{V}\right)}_{:=A_1}$$
$$+\underbrace{\sum_\rho p_\rho \left((1+\lambda\alpha_{2\to\rho}(\rho, x_1))V_\rho(x_1, x_2(\rho, x_1)) + (1+\lambda\alpha_{2\to t}(\rho, x_1))\bar{V}\right)}_{:=V_2},$$

i.e., using our equivalent interpretation of the objective as unweighted consumption payoffs plus attention utility, the A_1 corresponds to the attention utility in the first period, and V_2 to the expected consumption payoff and attention utility in the second period. A decrease from $\bar{V}'$ to $\bar{V}$ (holding α_1' constant) decreases V_2 by at most $(1+\lambda)(\bar{V}' - \bar{V})$, since for every realization (of ρ), the DM's payoff is given by the max of (6) and each of those terms decreases by at most that amount (and so also their max). Next, notice that $\alpha_{2\to\rho}$ is increasing (still for each realization ρ); that is as the decrease in the terms of (6) is decreasing in $\alpha_{2\to\rho}$. Hence, V_ρ is increasing; and so A_1.

Thus, holding the action x_1' and attention α_1' fixed, the decrease in (7) is strictly less than $(1+2\lambda)(\bar{V}'-\bar{V})$; furthermore, this decrease only becomes smaller when the DM chooses their attention allocation and action optimally given $\bar{V}$, and so the result follows. $\square$

Proof of Result 2. For the first bullet point, first, for each ρ, the DM's payoff in the second period (6) has increasing differences in $\alpha_{2\to\rho}$ and $-\bar{V}$ and so the claim about set inclusion follows. For the extreme values of $B(\bar{V},\lambda)$, again for each ρ, since $\lambda > 0$ and $\max_{x_2\in[0,w+x_1r_1]} V_\rho(x_1,x_2)$ and $V_\rho(x_1,\underline{x}_2)$ are finite, it must be that $\alpha_{2\to\rho} = 0$ is uniquely optimal for $\bar{V}$ large enough and $\alpha_{2\to\rho} = 1$ for $\bar{V}$ low enough; since there are finitely many realizations of ρ, the result follows.

For the second bullet point, first note that, for each ρ, when $\max_{x_2\in[0,w+x_1r_1]} V_\rho(x_1,x_2) \geq V_\rho(x_1,\underline{x}_2)$, then, regardless of λ, $\alpha_{2\to\rho} = 1$. Since this inequality does not involve λ, we can ignore such ρ. If the inequality does not hold, (6) has increasing differences in $\alpha_{2\to\rho}$ and $-\lambda$ and so the claim about set inclusion follows. For the extreme values of $B(\bar{V},\lambda)$, $B(\bar{V},0) = \emptyset$ and B finite-valued, implies the first, and, again for each ρ, if $\max_{x_2\in[0,w+x_1r_1]} V_\rho(x_1,x_2) < V_\rho(x_1,\underline{x}_2)$, then it must be that $\alpha_{2\to\rho} = 0$ is uniquely optimal for λ large enough; since there are finitely many realizations of ρ, the result follows. $\square$

Proof of Result 3. Note that $V_\rho(x_1,\underline{x}_2) = u(w + x_1\tilde{r})$, where $\tilde{r} := r_1 + r_2 + r_1r_2$ (with r_2 deterministic given N).

In the second period, the DM's attention satisfies $\alpha_{2\to\rho}(\rho,x_1) = \arg\max\{V_\rho(x_1,\bar{x}_2),\bar{V}\}$. Let $\rho^* = \arg\max_\rho \tilde{r}$. In the first period, the DM devotes η_1 attention to the expected portfolio choice problem, and the remainder to the max of $\bar{V}$ or $V_{\rho^*}(x_1,\underline{x}_2)$.

Thus, the DM's objective (7) is

$$\lambda\left(\underbrace{(1-\eta_1)\overbrace{\max\{V_{\rho^*}(x_1,\underline{x}_2),\bar{V}\}}^{:=C}+\eta_1\overbrace{\sum_\rho p_\rho V_\rho(x_1,\underline{x}_2)}^{:=D}+\overbrace{\sum_\rho p_\rho \max\{V_\rho(x_1,\underline{x}_2),\bar{V}\}}^{:=E}}_{:=A}\right) \tag{14}$$
$$+\underbrace{\sum_\rho p_\rho V_\rho(x_1,\underline{x}_2)}_{:=B}.$$

Also note that the DM is time consistent, in particular, the DM's attention in the second period maximizes (14) (given ρ, x_1).

We begin with the comparative static in λ. Suppose $x_1 > 0$ (otherwise, it cannot decrease when λ increases). Then (14) differentiated with respect to x_1 must be nonnegative. If $\frac{\partial}{\partial x_1} B \geq 0$, then $\frac{\partial}{\partial x_1} A$. (This holds as the max operator selects on high $\tilde{r}$, and "removing" one with positive derivative implies that none for which the derivative is negative is kept.) Thus, $\frac{\partial}{\partial x_1} A \geq 0$ at the optimum. Then, e.g., by the Implicit Function Theorem, x_1 is increasing in λ.

For $1 - \eta_1$, similarly to before, if $\frac{\partial}{\partial x_1} D \geq 0$, then so is the derivative of (14), and strictly so, as $\frac{\partial}{\partial x_1} C > 0$ (otherwise, the DM would not invest to begin with), and $x_1 = 1$ in a neighborhood of $1 - \eta_1$. Thus, $\frac{\partial}{\partial x_1} D < 0$. Then, e.g., by the Implicit Function Theorem, x_1 is increasing in $1 - \eta_1$. □

Proof of Result 4. Since $\eta_1 = 0$, in the first period, the DM devotes all attention to the portfolio choice problem with the highest payoff or the trivial problem. In either case, the DM in the second period maximizes the corresponding payoff; hence, the DM is time consistent. In this case, the result follows from optimality. □

Proof of Result 5. If the DM does not participate in the portfolio choice problem, then they optimally do not devote attention to their (trivial) portfolio choice problem. Furthermore, the DM is time consistent and their first-period action has no effect on the payoff in the second period. Thus, the DM chooses α_1 and $x_1 \in X(\alpha_1)$ to maximize $(1+\lambda\alpha_{1\to c_1})V_{c_1}(x_1) + (1 + \lambda\alpha_{1\to t})\bar{V}$. By Proposition 1, $\alpha_{1\to c_1}$ increases in γ_{c_1}.

When the DM participates, an increase in γ_{c_1} also $\alpha_{1\to c_1}$, again by Proposition 1.

Furthermore, by arguments similar to those in the proof of Result 1, the DM's decision to participate in the portfolio choice problem is decreasing in γ_{c_1}, and $\alpha_{1\to c_1}$ could only increase when the DM changes from non-participation to participation if $\alpha_{1\to t}$ decreases; but that contradicts optimality when the DM does not participate. □

C.10 Proof of Proposition 10

Proof of Proposition 10. The first claim follows from Proposition 1, the second from, e.g., Topkis since the DM's objective has increasing differences in α_c and V_H. For the third, note that the cross-partial derivative of the DM's objective with respect to α_c and V_L is given by

$$\lambda(1 - p(\alpha_c)) - (1 + \lambda\alpha_c)\frac{\partial}{\partial\alpha_c}p(\alpha_c).$$

If $p(\alpha_c) + \alpha_c \frac{\partial}{\partial\alpha_c}p(\alpha_c) < 1$ everywhere, then the above becomes positive for large enough λ (e.g., take $\lambda > \frac{\max_{\alpha_c} \frac{\partial}{\partial\alpha_c}p(\alpha_c)}{\min_{\alpha_c}(1-p(\alpha_c)+\alpha_c\frac{\partial}{\partial\alpha_c}(\alpha_c)}$), and the claim follows from Topkis. $\square$

C.11 Proof of Proposition 11

Proof of Proposition 11. Take any $C, C' \in B$ and consider $B' := (B \cup \{C \cup C'\}) \setminus \{C, C'\}$. Evaluate (9) at (x, α) and B' and subtract its value given (x, α) and B; after some simplifications, we have

$$-\frac{|C||C'|\lambda}{|C| + |C'|}(\bar{\alpha}_C - \bar{\alpha}_{C'})(\bar{V}_C(x) - \bar{V}_{C'}(x)).$$

Optimality the implies that the above is non-positive, i.e., if $\bar{V}_C(X) > \bar{V}_{C'}(x)$, then $\bar{\alpha}_C \geq \bar{\alpha}_{C'}$. $\square$