Emotional Inattention

Lukas Bolte[*] Collin Raymond[*]

August 5, 2022

PRELIMINARY DRAFT

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

Abstract

A decision-maker chooses an attention allocation across consumption problems, states, or time periods. In addition to being instrumentally valuable, attention determines how payoffs are aggregated. Optimal attention to a problem (or state or period) is increasing in its payoff and the instrumental value of attention. The attention-reweighted environment is interpreted as (as-if) belief distortions and (as-if) time preferences—whose forms are determined by the solution of the attention allocation problem—thus generating distinct behavioral phenomena within a single framework. We apply our model to information acquisition and portfolio choice and discuss implications for policymaking.

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 acknowledge financial support from the Leonard W. Ely and Shirley R. Ely Graduate Student Fellowship.

1 Introduction

Attention has two fundamental features. First, attention is instrumentally valuable: It allows individuals to acquire and process information, or to simply take an action. Take an investor, for instance, who wants to execute a financial trade. This action requires some cognitive resources, in other words, they must devote attention. The instrumental value of attention and the consequences of attention as a scarce resource are well recognized and studied (see, e.g., Sims (2003) for the canonical paper and Mackowiak et al., 2022 for a recent review).

Second, attention determines how the various experiences an individual makes are aggregated into a single measure of satisfaction—their overall payoff. For example, the lack of attention to something negative may allow an individual to lessen its impact, and, conversely, excessive attention to something positive may magnify its impact, with both improving the overall payoff. Indeed, by executing the trade, the investor may inevitably devote attention to their finances and increase the impact of how they experience their financial situation on their overall payoff. This second fundamental feature of attention is relatively understudied. Yet, it is plausibly a key determinant of behavior. Indeed, if the investor's portfolio has been performing poorly, they may prefer to not execute the desired trade after all.

In this paper, we introduce a model of attention allocation, where attention has these two fundamental features—it is instrumentally valuable and determines how experiences are aggregated into an overall payoff—and study the implications. Attention is allocated trading off its instrumental value with the consequences for how experiences are aggregated, potentially leading to avoidance of low-payoff situations (such as the investor who does not execute the trade). In addition to such straightforward avoidance behavior (and, conversely, excessive attention to high-payoff situations), our model generates various behavioral phenomena. Attention as an aggregator of experiences essentially reweights the decision environment which, for example, can look like distorted beliefs (when attention is allocated across states) or lead to endogenous time discounting (when attention is allocated across time). In addition, our model predicts default effects, generates intrinsic preferences over information, and provides guidance for policy-making.

A central starting point of our model is that attention is (at least in some part) voluntarily directed by the individual (in the context of visual cognition, neuroscientists refer to this premise as "top down" attention). This assumption is standard in, for example, the rational inattention literature (Sims, 2003)Mackowiak et al., 2022). Here, as we show, it is arguably as plausible, and delivers a host of implications, some with extensive empirical support. Furthermore, individuals are aware of their ability to freely allocate attention: Falk and Zimmermann (2016) find that information about an electrical shock is viewed differently depending on whether participants in their experiment have a distracting task available.

Attention is, of course, also determined involuntarily, when it is captured by salient features of the environment, and, even if attention is seemingly voluntary, individuals, unlike in our model, may devote it to situations that predictably make them worse off (e.g., worrying about bad outcomes). Whatever model of attention allocation one wishes to employ, given a resulting attention allocation, our framework still allows us to study the attention-reweighted decision environment and the ensuing behavioral phenomena. For instance, a present focus or distorted subjective probabilities are the result of 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.

Another crucial premise of the model is that attention reweights the decision environment with increased weights on aspects of it receiving high attention. Conceptually, we find this modeling approach plausible. Indeed, anticipatory utility or remembrance utility from noncontemporaneous consumption can only occur if it is in fact 'anticipated' or 'remembered,' in other words, if the individual devotes attention to it. Thus, attention to such noncontemporaneous consumption should affect the weight the payoff from consumption takes. Our model with attention reweighting the decision environment can thus be understood equivalently as one where attention leads to "attention utility" (e.g., anticipatory or remembrance utility).[1] A more subtle assumption of the model is then that attention devoted to, say, improve a payoff from some consumption, generates (unavoidable) attention utility from that payoff. In other words, it is not possible to devote attention without generating attention utility (neuroscientists may refer to this premise as "bottom up" attention).

We begin by considering a decision-maker (henceforth, DM—they) who chooses an action and an attention allocation to maximize the weighted sum of consumption payoffs (the aforementioned "experiences" or "situations"). Attention is a measure over a set of consumption problems that determines: 1) which actions are available, and 2) the weight the consumption payoff of each problem takes. The action taken determines the value of each consumption payoff. Equivalently, the DM's objective can be interpreted as the some of (unweighted) consumption payoffs and "attention

[1] This of the dual role of attention as being instrumentally valuable and creating attention utility is shared by Schelling (1988): "[...] we have at least two distinct roles for our minds at play, that of the information processing and reasoning machine by which we choose what to consume out of the array of things that our resources can be exchanged for, and that of the pleasure machine or consuming organ, the generator of direct consumer satisfaction."

utility," taking the form of the product of attention devoted to a problem and its payoff.

The optimal (action, attention)-pair is governed by some relatively straightforward comparative statics: The DM weights the instrumental value of attention against its role in aggregating different payoffs, and suitable formulation of variations in each, lead attention to be geared more towards one or the other.

In particular, the payoff levels of the consumption problems matter (and not how much attention can increase them): Ceteris paribus, the DM devotes more attention to consumption problems with higher payoff levels. The DM may thus avoid a low-payoff problem, even though attending to them would increase its payoff, while devoting excessive attention, beyond the point where it is instrumentally valuable, to others. There is extensive empirical support for this prediction: People log in to their investment portfolio less frequent when the market is down (Karlsson et al., 2009; Gherzi et al., 2014; Sicherman et al., 2015; Quispe-Torreblanca et al., 2020); check their bank accounts less frequent when the balance is low Olafsson and Pagel (2017); avoid getting tested if at-risk for a severe disease (Shouldson and Young, 2011; Oster et al., 2013; Ganguly and Tasoff, 2017); and fail to take their medicine (Becker and Mainman, 1975; Sherbourne et al., 1992; Lindberg and Wellisch, 2001; DiMattero et al., 2007). We note that in some of these settings, information is unlikely to play a major role: For instance, Quispe-Torreblanca et al. (2020) find that individuals log in to their investment portfolio if its value is already known more so when it is up. Thus, information avoidance due to adjustments of reference points (Karlsson et al., 2009) or nonlinearities in anticipatory utility (Caplin and Leahy, 2001) cannot explain them.

In addition to concurrent consumption problems, we consider two more dimensions of attention allocation—uncertain states and time—beginning with the former. Still, the optimal (action, attention)-pair is governed by the instrumental value of attention and its role in aggregating payoffs, now in different states. (In fact, we provide close analogs of the comparative statics results for attention across consumption problems.) The ensuing weight of a state can be interpreted as the subjective probability of that state occurring; importantly, this is not to say that the DM chooses a belief (directly); indeed, they are always aware of the objective probability of a state occurring.

Similar to before, the payoff level of a state matters for optimal attention with the DM devoting more attention (and consequently assigning the higher subjective probability) the higher the payoff level of a state, ceteribs paribus. When attention is non-instrumental (i.e., the available actions are independent of the attention allocation), the DM devotes all to high-payoff states, leading them to appear 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 documents in various settings, e.g., individuals playing lotto (Garrett and Sobel, 1999; Forrest et al., 2002).

But the DM's optimism is not universal; it can be mitigated or overturned when the attention allocation is predominantly guided by its instrumental value. We next consider environments with instrumental value of attention and show what kind of probability weighting (subjective probability of a state as a function of its objective probability) ensue in different environments. For instance, an inverted-S-shaped probability weighting occurs in environments where some 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 mechanism with predictions on how environments map into the type of probability weighting

The final dimension of attention allocation we consider is time. Again, the optimal (action, attention)-pair is governed by the instrumental value of attention and its role in aggregating payoffs, now across time. (The comparative statics (varying each of attention's role) are somewhat more complex since the DM takes an action and allocation attention in each period.) The ensuing weight on the different time periods are interpreted as (endogenous) discount factors.

For instance, the DM discounts the future (they are "present-focused") if the payoff in the present is particularly high or attention is of particular instrumental value. Additionally, the DM may optimally generate payoffs that vary across time (in otherwise completely symmetric environments) as 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).

Having considered three canonical dimensions in turn, we next consider them jointly in order to explore their interactions. In particular, 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 note three implications. First, the DM takes actions that create varied future payoffs, since those lead to the largest benefit from distorting the environment; second, when choosing a default, an action that only matter if unchanged in the future, the DM focuses on low-payoff realizations of the future, since those are the ones for which the default binds; and third, the DM's subjective probabilities overweight high-payoff realizations of the future leading to optimistic actions.

This concludes our general model; we next turn to applying our insights to two settings: information acquisition and portfolio choice. A consumption problem can nest information acquisition:

E.g., attention could lead to information that the DM uses to guess some underlying state or simply reveal a random payoff—all captured by reduced-form formulation. The basic comparative static with regards to payoff levels then implies that the DM acquires information about questions that generally involve a high payoff, or those for which their prior puts much weight on high payoff states (Möbius et al. (2022) provides some related laboratory evidence). Our model also predicts a preference for early, as opposed to late, resolution of uncertainty—early information allows the DM to condition their future attention, a "hidden action," on the realized information—which is broadly consistent with laboratory evidence (Masatlioglu et al., 2017; Nielsen, 2020); and a preference over the shape of information, since information that can change future attention (relative to no information) is particularly valuable.

The mechanisms in our model leading to the aforementioned preferences over information are distinct from those in models of anticipatory utility (Kreps and Porteus, 1978; Caplin and Leahy, 2001). There, when there is no instrumental value of attention, preferences over information depend on whether the objective is convex or concave in the expected payoff (the anticipatory utility). Briefly: Here, the attention-weighted objective can be interpreted as resulting from attention utility, which, in turn, can be thought of as anticipatory utility that is only received when the DM devotes attention; however, without varying attention, this anticipatory utility enter the DM's objective linearly.

In our second application, 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 (to our knowledge) 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 mechanically generating a disposition effect Shefrin and Statman (1985); Odean (1998); Barberis and Xiong (2009, 2012). Third, the DM is excessively risk-seeking (in a precise sense) in some, and, forth, risk-averse in other situations. More generally, they demand an "attention premium" from assets, i.e., they prefer assets that do not require continual attention. Fifth, the DM has a high discount factor because they invest in financial assets, and not vice versa.

We next study the implications of our model for policy-making (broadly construed)—highlighting the differences to perhaps more standard models. (The policy-maker can be the DM themselves or a second party, such as the government.) When allocating resources, the policy-maker should take the DM's attention allocation and ensuing attention-weighted environment into account. For instance, the value of increasing a payoff of a particular consumption problem is also attention-weighted, and so most effective for problems receiving a lot of attention.

We also show that incentivizing the DM to take certain actions operates very differently depending on whether the policy-maker employs rewards (uses a "carrot") or penalties (a "stick"). While, in a certain sense, equivalently effective in many standard models, here, the DM may shy away a problem if the penalty results in a low expected payoff. Negative commitment devices, those that penalize for deviation from the action committed to, may be ineffective (and those providing rewards may be too expensive).

Lastly, we consider how the policy-maker would construct the set of consumption problems from a meta set of smaller ones: I.e., how should they build/perceive the environment—as one grand problem (thus nesting more standard models) or as distinct consumption problems whose payoffs are attention-weighted? We (partially) characterize such optimal bracketing. Intuitively, the brackets two consumption problems together when they otherwise would devote more attention to the lower-payoff one. We note that this formulation microfounds the set of consumption problems, a premise of the model.

Section 5 discusses how our model differs from several other classes of models. We consider in turn models of rational inattention, anticipatory emotions, news utility, time preferences, subjective beliefs, and other models of attention.

2 Model

A decision-maker (henceforth, DM—they) chooses an (action, attention)-pair to maximize some objective. The choice of attention has two implications: First, it determines the available actions. For example, executing a trade (an action) may only be possible if the DM devotes attention to their portfolio (we will be precise about what this means shortly). Second, it affects the weights of the different terms in the DM's objective. For example, attention to their portfolio increases the weight of an associated payoff in the objective.

Formally, attention is a measure over additive payoff terms in the DM's objective. These terms can correspond to payoffs from different consumption dimensions, uncertain states, or time periods (all required to be additive in utility space). Attention to a term—a consumption dimension (we will refer to them as consumption problems), state, or time—increases its weight in the DM's objective.

The weighting of payoff terms essentially reweights the DM's environment: they behave like a standard DM (a formal definition follows shortly), but one who has distorted beliefs or discounts time. Since attention is endogenous—the DM optimally trades off its instrumental value (enabling actions) and its consequence for how the payoff terms are aggregated—so are the resulting behavioral phenomena, e.g., belief distortions and time discounting.

To keep the model tractable, we begin by considering each dimension of attention allocation—across consumption problems, uncertain states, and time periods—one at a time (Sections 2.1–2.3). In the following section (Section 2.4), we combine some of the dimensions to, among other things, study interactions; similarly, (in Section 3) we apply our findings to study information acquisition (Section 3.1) and portfolio choice (Section 3.2), also combining previous findings within a single environment. We then consider implications for policy-making (Section 4), before relating our model and findings to the existing literature (Section 5).

2.1 Attention allocation across consumption problems

We begin with attention allocation across consumption problems. This simple environment allows us to formally express two fundamental features of attention: it determines 1) which actions are available to the DM, and 2) how the payoffs from the consumption problems are aggregated in the DM's objective. After deriving our model's implications for optimal attention—e.g., avoidance behavior or excessive attention—we discuss its relationship to existing empirical evidence.

The DM faces a finite number of consumption problems $\mathcal{C}$ with generic consumption problem c. Consumption problem c is associated with a consumption payoff denoted by V_c. We refer to problem c with payoff V_c if there is no risk of ambiguity. The DM chooses an (action, attention)-pair denoted by (x, α). The action determines the payoffs from each problem: given x, the payoff from problem c is $V_c(x)$, where V_c is continuous in x. Attention has two implications, but first we define it as a measure on the set of consumption problems with total measure of 1 (a normalization), i.e., $\alpha = (\alpha_c)_{c\in\mathcal{C}}$, where α_c denotes the attention devoted to problem c; with $\alpha_c \geq 0$ and $\sum_{c\in\mathcal{C}} \alpha_c = 1$. It will be useful to let $V_{-c} := (V_{c'})_{c'\in\mathcal{C}\setminus\{c\}}$.

Attention has two implications. First: the actions available depend on the attention allocation, i.e., given α, x must be chosen from $X(\alpha)$, where X is compact- and non-empty-valued and upper hemicontinuous. Additionally, we assume a form of monotonicity of X: let X be defined on the set of measures on $\mathcal{C}$ with total measure of at most 1, then $X(\mu') \supseteq X(\mu)$ for any such measures μ, μ if $\mu' \geq \mu$ element-wise. Thus, while the DM always devotes a full unit of attention, whether a particular action is available depends on whether the attention allocation is satisfies some minimal attention.

(We note that this formulation nests, e.g., attention leading to information acquisition. Attention α_c to problem c may enable an action x that represents acquiring information increasing the payoff V_c (perhaps an expected payoff over states). However, at this stage, we abstract away from such details and instead take a reduced-form approach. In Section XYZ we provide examples of our formulation nesting models with information acquisition, recall of memories, and attention reducing "trembles.")

Second: attention determines the weights on the consumption payoffs in the DM's objective. In particular, the weight of problem c (and its payoff) is given by $1 + \lambda\alpha_c$, where $\lambda \geq 0$ governs the extent to which the payoffs are attention-weighted. One interpretation is that the DM always values the "material payoff" from actual consumption (with weight 1); however, additionally, they also value "attention utility," an additional payoff from problem c proportional to the amount of attention devoted and its consumption payoff. λ then governs the relative importance of the "material payoff" and "attention utility." In this interpretation, attention utility can be interpreted as anticipatory utility, but on that is only generated when the DM devotes attention to the (future) consumption problem. We consider the first consequence of attention—its instrumental value—as relatively standard and for $\lambda = 0$, the it is the only consequence of attention. We thus refer to the case when $\lambda = 0$ as the "standard model" and the corresponding DM as the "standard DM."

The DM's objective is then the attention-weighted sum of consumption payoffs (or equivalently, the sum of "material payoffs" and "attention utility")

$$\sum_{c \in \mathcal{C}} (1 + \lambda\alpha_c) V_c(x); \tag{1}$$

$$\text{or equivalently:} \quad \underbrace{\sum_{c \in \mathcal{C}} V_c(x)}_{\text{material payoff}} + \lambda \underbrace{\sum_{c \in \mathcal{C}} \alpha_c V_c(x)}_{\text{attention utility}}.$$

To state how the (action, attention)-pairs are optimally determined, we find it useful to introduce parameterization of the set of consumption problems as well as define a particular restriction. First, the parameterization: For each problem c, we fix $\tilde{V}_c$ and define its payoff as $V_c = \beta_c \tilde{V}_c + \gamma_c$, for any scalars $\beta_c \geq 0$ and γ_c. Intuitively, increasing γ_c shifts the payoff level of problem c, and increasing β_c increases the payoff difference induced by different actions.

Next, we say that an environment (consisting of $(V_c)_{c\in\mathcal{C}}$ and X) is separable if action x is a vector $x = (x_c)_{c\in\mathcal{C}}$ and, letting $x_{-c} := (x_{c'})_{c'\in\mathcal{C}\setminus\{c\}}$, $V_c(x_c, x_{-c})$ is independent of x_{-c} for all c and x_c, and $X(\alpha) = \Pi_{c\in\mathcal{C}} X_c(\alpha_c)$. Intuitively, the DM takes separate actions for each problem and whether such problem-specific action is available depends on the amount of attention devoted to the problem. (Note that maximizing (1) with respect to an (action, attention)-pair is then equivalent to maximizing $\sum_{c\in\mathcal{C}}(1+\lambda\alpha_c)\hat{V}_c(\alpha_c)$, where $\hat{V}_c(\alpha_c) := \max_{x_c\in X_c(\alpha_c)} V_c(x_c,\cdot)$ is increasing in α_c.)

An increase in the payoff level of a particular problem c has no effect on the instrumental value of attention, but the DM wants to increase its weight in their objective (to capture this increase) and hence devotes additional 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 problem c. It may, additionally, move the payoff up or down, also inducing the DM to change their attention. In the lemma below, we offset such level change. If the environment is separable, the "more suitable" action can only be available if the DM increases their attention.

Lemma 1A. *Consider a particular consumption problem $c \in \mathcal{C}$ with $\tilde{V}_c$, fix V_{-c}, and let $\Gamma(\gamma_c, \beta_c)$ denote the set of optimal (action, attention)-pairs.*

- *If $\lambda > 0$: If $\gamma_c^{'} > \gamma_c$ then $\min_{(x,\alpha)\in\Gamma(\gamma_c',\beta_c)} \alpha \geq \max_{(x,\alpha)\in\Gamma(\gamma_c,\beta_c)} \alpha$. If, in addition, the environment is separable, then $\min_{(x,\alpha)\in\Gamma(\gamma_c',\beta_c)} V_c(x) - \max_{(x,\alpha)\in\Gamma(\gamma_c,\beta_c)} V_c(x) \geq \gamma_c' - \gamma_c$.*
- *If for β_c and γ_c, $\max_{(x,\alpha)\in\Gamma(\gamma_c,\beta_c)} V_c(x) = \min_{(x,\alpha)\in\Gamma(\gamma_c,\beta_c)} V_c(x)$, then for any $\beta_c^{'} > \beta_c$ and $\gamma_c^{'} = \gamma_c - (\beta_c^{'} - \beta_c)\tilde{V}_c(x)$, where $(x,\alpha) \in \Gamma(\gamma_c, \beta_c)$, we have $\min_{(x,\alpha)\in\Gamma(\gamma_c',\beta_c')} V_c(x) \geq \max_{(x,\alpha)\in\Gamma(\gamma_c,\beta_c)} V_c(x)$. If, in addition, the environment is separable, then $\min_{(x,\alpha)\in\Gamma(\beta_c',\gamma_c')} \alpha_c \geq \max_{(x,\alpha)\in\Gamma(\beta_c,\gamma_c)} \alpha_c$.*

(The proof of this lemma, as are all other formal statements, is relegated to Appendix G.)

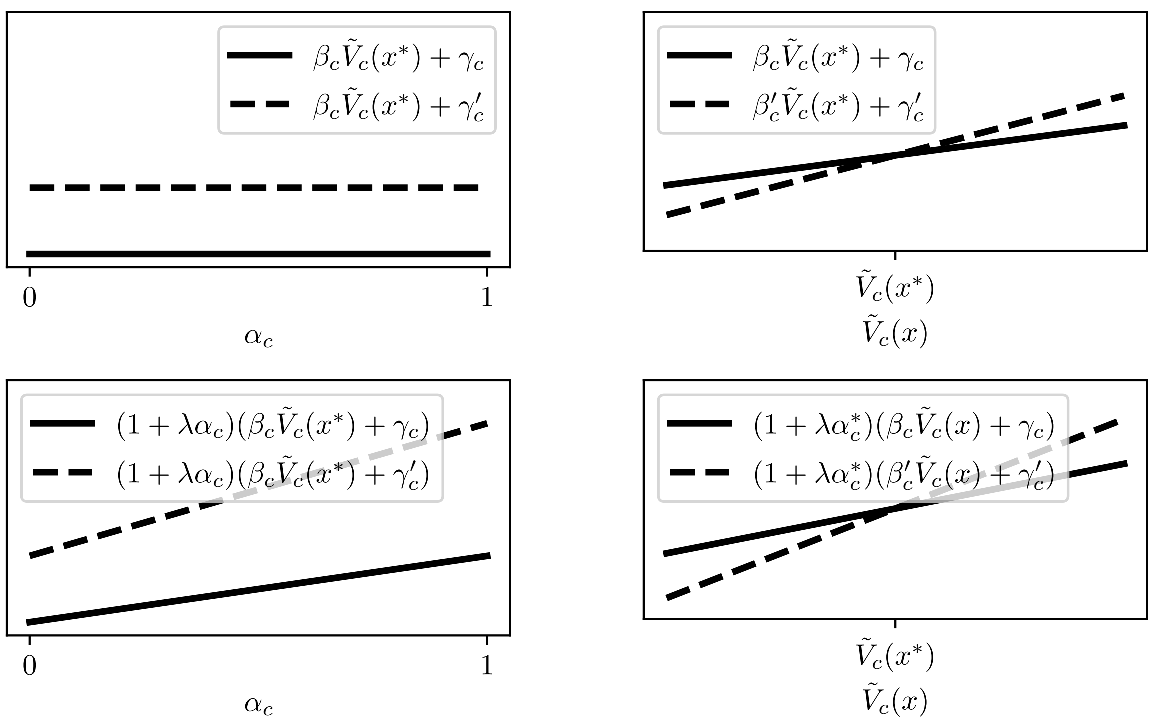
Thus, increasing the payoff level leads to an increase in attention (the first part of the lemma); increasing payoff differences from different actions leads to an improved action (the second part). And, if the environment is separable, increased attention and improved action go hand in hand.

This lemma has analogs when we consider attention allocation across uncertain states and time periods (Sections 2.2 and 2.3).

Figure 1 illustrates Lemma 1A. Panel (a), showing the consumption payoff (top figure) and attention-weighted payoff (bottom figure) as a functions of attention, depicts an increase in γ_c

to γ_c'. Throughout, the optimal action x^* is held fixed (and hence the consumption payoff is independent of α_c). This increase simply shifts the payoff up (top figure). However, the attention-weighted increase is larger for higher α_c (bottom figure). Thus, the DM increases their attention in response (first part of Lemma 1A).

Panel (b), showing the consumption payoff (top figure) and attention-weighted payoff (bottom figure) as functions of $\tilde{V}_c(x)$, depicts an increase in β_c to β_c' with an offsetting change in γ_c to γ_c' as described in the second case of Lemma 1A. Throughout, the optimal attention α_c^* 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 $\tilde{V}_c(x)$ than before. When the payoff is attention-weighted (bottom figure), the DM still benefits from increasing $\tilde{V}_c(x)$ than before and thus increases $\tilde{V}_c(x)$ in response to the change (second part of Lemma 1A).



(a) (b)

Figure 1: Panel (a) corresponds to an increase γ_c to γ_c'. We hold the optimal action x^* fixed. The consumption payoff is shifted up (top figure), independently of α_c. However, increasing α_c now increases the attention-weighted payoff now by more than before (bottom figure). Panel (b) corresponds to an increase in β_c to β_c' with an offsetting change of γ_c to γ_c'. We hold the optimal attention α_c^* fixed. The consumption payoff pivots around $\tilde{V}_c(x)$ (top figure). Already here, the DM benefits more from increasing their action, i.e., $\tilde{V}_c(x)$, than before. When considering the attention-weighted payoff (holding optimal attention fixed), does not change this conclusion.

Having considered how the consumption problems determine the optimal (action, attention)-pair, we turn to the role of λ; recall that λ governs the extent to which attention reweights the payoffs in the DM's objective (or equivalently, the weight on attention utility). A standard DM maximizes the (unweighted) sum of consumption payoffs (they maximize the "material payoffs" as they do not value "attention utility"); one can think of the standard DM as fully utilizing the instrumental value of attention. A DM with $\lambda > 0$, instead, may not maximize the (unweighted) sum of consumption payoffs in order to devote attention to high-payoff problems.

Lemma 2A. *Let $\Gamma(\lambda)$ denote the set of optimal (action, attention)-pairs given λ. If $\lambda' > \lambda$, then $\max_{(x,\alpha)\in\Gamma(\lambda')} \sum_{c\in\mathcal{C}} V_c(a) \leq \min_{(x,\alpha)\in\Gamma(\lambda)} \sum_{c\in\mathcal{C}} V_c(a)$.*

The lemma states that the DM monotonically departs from the (unweighted) payoff-maximizing standard DM, who fully utilizes the instrumental value of attention, as λ increases. In other words, through a standard lens, that is looking at the non-distorted environment, the DM's action becomes worse as λ increases; however, the DM, of course, is still maximizing their objective—the attention-weighted sum of payoffs. As $\lambda \to \infty$, the DM only values payoffs of problems they devote attention to, i.e., they fully disregard problems they do not devote attention to.

These first two results are intuitive and to be expected given the goal of the model to formally express the two consequences of attention. We next note some additional implications.

Recall our previous parameterization of the payoff from problem c as $V_c = \beta_c \tilde{V}_c + \gamma_c$, for fixed $\tilde{V}_c$ and scalars $\beta_c \geq 0$ and γ_c. The DM's objective (1) is (evidently) linear in $(\gamma_c)_{c\in\mathcal{C}}$; thus, the DM's value (i.e., (1) for optimal (action, attention)-pairs) is convex in $(\gamma_c)_{c\in\mathcal{C}}$.

Lemma 3A. *The DM's value is convex in $(\gamma_c)_{c\in\mathcal{C}}$.*

An implication of this lemma is that the DM has a preference for "extreme" payoffs: they prefer to have more varied payoffs (holding the average payoff fixed). We illustrate this point in a simple example: there are two consumption problems, $\mathcal{C} = \{c, c'\}$; the DM chooses $x = (x_c, x_{c'}, x_\gamma)$ from set $X(\alpha) = \tilde{X}(\alpha_c) \times \tilde{X}(\alpha_{c'}) \times [0,1]$; given x, the payoffs are $V_c(x) = f(x_c) + x_\gamma$ and $V_{c'}(x) = f(x_{c'}) + (1 - x_\gamma)$, for some function f. Note that the problems are symmetric. However, the DM strictly prefers to choose x^* with $x^*_\gamma \in \{0,1\}$ (with full attention to problem c if $x^*_\gamma = 1$, and c' if$x^*_\gamma = 0$). Intuitively, the DM chooses an (action, attention)-pair to make the ex-ante symmetric problem ex-post asymmetric, which, by devoting attention to the high-payoff problem, increases their objective.

A second, perhaps more subtle implication, of Lemma 3A is that when consumption 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 for a standard DM.

Lemma 4A. *Suppose the environment is separable. Let* $F(\lambda)$ *denote the DM's objective* (1) *given* λ. *For any* $\lambda' > \lambda$, *if* $F(\lambda)$ *is convex in* α, *then so is* $F(\lambda')$.

In a separable environment, increasing α_c increases the payoff from problem c (in addition to the weight problem c takes in the DM's objective). Consequently, the added weight as α_c increases increases the DM's objective by more for higher values of α_c potentially making the objective convex in attention. This effect is driven by λ; hence, it may be that a standard DM's objective is not convex, whereas it is for a DM with $\lambda > 0$.

In Section 3, we explore how the convexity of the objective discussed here relates to a demand for (non-instrumental) information (Section 3.1) and a preference for assets with varied return (Section 3.2).

Empirical evidence

We now discuss how the simple predictions of our model outlined above relate to existing empirical evidence. Our focus is on the on the first prediction of Lemma 1A—attention varies in the level of the payoff. Evidence of (attentional) avoidance of low-payoff situations, potentially further lowering the payoff, as well as excessive attention to high-payoff situations has been extensively document 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; Karlsson et al. (2009); Sicherman et al. (2015). Although Gherzi et al. (2014) find that increase monitoring following market downturns). This behavior, in particular the tendency to avoid one's portfolio in bad market conditions, is often referred to as an "ostrich effect"—individuals bury their figurative heads in the send like the proverbial ostrich.[2] 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 then suggests that doing so co-varies with the payoff associated with the portfolio; reasonably, this payoff may be eventual

[2]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 actually do not display this behavior. Instead, they put their heads into their nests (which are built on the ground) in order to check temperatures and rotate eggs.

consumption. A down market (for most investors) implies low future consumption and by decreasing attention to their portfolio, investors can minimize the effect of the associated payoff decrease.[3]

As eluded to, 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 in modeling consumption problems nests problems of information acquisition (see Section XYZ for an example), and so our predictions are in line with such 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 can be revealed; similarly, 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) look at individuals' attention to their financial accounts and find 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 pay-dates and amounts, as well as overdrawn status, and so such avoidance information avoidance may be implausible.

Similar behaviors have been documented in other domains. For instance, researcher 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 than 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,

[3]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 adding the attention-weighted sum of payoffs relative to a reference point.

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 ("attention") between two games ("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 Lemma 1A—once controlling for payoff levels, increasing the payoff differences different action induce on a particular problem leads the DM's to choose an action better suited for that problem—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.

We do not know of evidence relating the extent to which attention determines the weight of payoffs, λ, relates to the extent to which they utilize attention's instrumental value (Lemma 2A). More generally, we are not aware of studies getting at the distribution of λ.

We are also unaware of evidence relating to Lemmas 3A and 4A—each discussing convexities in the DM's objective. However, it seems plausible that individuals tend to focus on particular problems and specialize in them. Typical explanations include, perhaps, skill acquisition from specialization. We offer an alternative reason: focusing on one problem simultaneously allows individuals to achieve a high payoff from that problem and increase its weight in their objective.

2.2 Attention allocation across states

We next consider attention allocation across uncertain states, and the implications of the two fundamental features of attention in this context. As before, attention both allows the DM to take actions that affect the consumption payoffs across states, and also determines how these payoffs are aggregated in the DM's objective. The attention-weighted environment, in this context of an uncertain state, implies that DM acts as if their belief is distorted, thus generating as-if probability weighting. But first, we introduce the environment and derive results analogous to those in the previous section with regards to the determinants of the DM's optimal (action, attention)-pair.

The DM faces an uncertain state with a finite number of realizations denoted by $\mathcal{S}$ with generic state s. State s occurs with (known) probability p_s and is associated with a consumption payoff denoted by V_s. Similar to before, the DM chooses an (action, attention)-pair denoted by (x, α).

The action determines the payoff in each state: given x, the payoff in state s is $V_s(x)$, where V_s is continuous in x. Attention is a measure on the set of states with total measure 1, i.e., $\alpha = (\alpha_s)_{s\in\mathcal{S}}$, where α_s denotes the attention devoted to state s; with $\alpha_s \geq 0$ and $\sum_{s\in\mathcal{S}} = 1$. We also let $V_{-s} \coloneqq (V_{s'})_{s'\in\mathcal{S}\setminus\{s\}}$.

Attention (still) has two implications. First: which actions are available depend on the attention allocation, i.e., given α, x must be chosen from $X(\alpha)$, where X is compact- and non-empty-valued and upper hemicontinuous. We maintain the monotonicity assumption on X (now with respect to measures on $\mathcal{S}$, but otherwise, identically defined). For example, an action could be choosing a lottery with state-contingent payoffs, or action x is a vector of state-contingent plans and taking the (sub-)action 'think about state s,' the DM begins to solve the corresponding optimization problem which increases their consumption payoff V_s in that state.

Second: attention determines the weights on the consumption payoff in each state in the DM's objective. The weight of state s (and the payoff in that state) is given by $p_1 + \lambda\alpha_s$, where $\lambda \geq 0$, as before, governs the extent to which attention weights the payoff terms. One can still interpret the objective as the DM always valuing the "expected material payoff" from actual consumption, and additionally value "attention utility" given by the sum across states of attention devoted to that state times the payoff in that state.[4]

Thus, the DM's objective is analogous to that in (1), mutatis mutandis. However, here, we renormalize the objective and divide by $1 + \lambda$. Such a increasing transformation does not change the optimal choice of (action, attention)-pair but normalizes the weights on the different states so that they sum to 1. Thus, the DM's objective is then the attention-weighted sum of the consumption payoff in different states (or equivalently, the sum of "expected material payoff" and "attention uility")

$$\sum_{s\in\mathcal{C}} \underbrace{\frac{p_s + \lambda\alpha_s}{1+\lambda}}_{=:q_s} V_s(x); \tag{2}$$

$$\text{or equivalently:} \qquad \underbrace{\sum_{s\in\mathcal{S}} p_s V_s(x)}_{\text{expected material payoff}} + \lambda \underbrace{\sum_{s\in\mathcal{S}} \alpha_s V_s(x)}_{\text{attention utility}} .$$

[4] Alternatively, we could have specified the weight as $p_s(1 + \lambda\alpha)$. The ensuing results would be similar. We opted for our preferred specification as it lends itself to the interpretation of attention α_s as the time spent (or intensity) mentally simulating state s.

Importantly, the weights on the payoff in different states (as determined by attention) can be interpreted as probabilities; thus the DM, conditional on their attention allocation, behaves like a subjective expected payoff maximizer, but one who assigns probability q_s to state s (where q_s is defined in (2)).

We begin our analysis by noting that the simple comparative statics determining the optimal allocation of attention (and the optimal action) across consumption problems (Section 2.1) reassuringly hold in this setting as well, mutatis mutandis. But first, we introduce a parameterization of the set of states. For each state s, we fix $\tilde{V}_s$ and define the payoff in that state as $V_s = \beta_s \tilde{V}_s + \gamma_s$ for any scalars $\beta_s \geq 0$ and γ_s. Similar to before, increasing γ_s shifts the payoff level in state s, and increasing β_s increases the payoff difference in state s induced by different actions. We also say an environment (consisting of $(V_s)_{s\in\mathcal{S}}$ and X) is separable if action x is a vector $x = (x_s)_{s\in\mathcal{S}}$ and, letting $x_{-s} := (x_{s'})_{s'\in\mathcal{S}\setminus\{s\}}$, $V_s(x_s, x_{-s})$ is independent of x_{-s} for all s and x_s, and $X(\alpha) = \Pi_{s\in\mathcal{S}} X_s(\alpha_s)$. For such environment, for each s, we define $\hat{V}_s(\alpha_s) := \max_{x_s\in X_s(\alpha_s)} V_s(x_s, \dots)$.

The following lemma provides analogous results to those in Lemma 1A: Increasing the payoff level in a state increases attention to that state, and increasing the payoff difference in a state from different actions increases the importance of taking an action suitable for that state. Furthermore, when the environment is separable, an increase in attention allows for a better action, so that both comparative static results, both increase attention to and improve the action for a state.

Additionally, the following lemma also considers a new comparative static: changes in the probability of state s, p_s. An increase in p_s also increases the (expected) payoff difference from different actions, and hence also leads the DM to choose an action more suitable for that state (and to devote more attention if the environment is separable). A complication in conducting the last comparative static is that increasing p_s must decrease $p_{s'}$ for some s', which may complicate effects. We address this by assuming that there exists a state $\bar{s}$ with $V_{\bar{s}}(x)$ independent of x, and increasing p_s to $p_{s'}$ leaves all $s' \notin \{s, \bar{s}\}$ constant and decreases $p_{\bar{s}}$ by an equivalent amount.

Lemma 1B. *Consider a particular state $s \in \mathcal{S}$ with $\tilde{V}_s$, fix V_{-s}, and let $\Gamma(\gamma_s, \beta_s, p_s)$ denote the set of optimal (action, attention)-pairs.*

- *If $\gamma_s^{'} > \gamma_s$, then $\min_{(x,\alpha)\in\Gamma(\gamma_s', \beta_s, p_s)} \alpha \geq \max_{(x,\alpha)\in\Gamma(\gamma_s,\beta_s,p_s)} \alpha$. if, in addition, the environment is separable, then $\min_{(x,\alpha)\in\Gamma(\gamma_s',\beta_s,p_s)} V_s(x) \geq \max_{(x,\alpha)\in\Gamma(\gamma_s,\beta_s,p_s)} V_s(x) \geq \gamma_s^{'} - \gamma_s$.*

- *Suppose $p_s > 0$. If for β_s and γ_s, $\max_{(x,\alpha)\in\Gamma(\gamma,s,\beta,c,p_s)} V_s(x) = \min_{(x,\alpha)\in\Gamma(\gamma,s,\beta,c,p_s)} V_s(x)$, then for any $\beta_s' > \beta_s$ and $\gamma_s' = \gamma_s - (\beta_s' - \beta_s)\tilde{V}_s(x)$, where $(x, \alpha) \in \Gamma(\gamma_s, \beta_s, p_s)$, we have*

$\min_{(x,\alpha)\in\Gamma(\gamma'_s,\beta'_s,p_s)} V_s(x) \geq \max_{(x,\alpha)\in\Gamma(\gamma_s,\beta_s,p_s)} V_s(x)$. *If, in addition, the environment is separable, then* $\min_{(x,\alpha)\in\Gamma(\gamma'_s,\beta'_s,p_s)} \alpha_s \geq \max_{(x,\alpha)\in\Gamma(\gamma_s,\beta_s,p_s)} \alpha_s$.

- *If* $p'_s > p_s$, *then* $\min_{(x,\alpha)\in\Gamma(\gamma'_s,\beta_s,p_s)} V_s(x) \geq \max_{(x,\alpha)\in\Gamma(\gamma_s,\beta_s,p_s)} V_s(x)$. *If, in addition, the environment is separable, then* $\min_{(x,\alpha)\in\Gamma(\gamma'_s,\beta_s,p_s)} \alpha_s \geq \max_{(x,\alpha)\in\Gamma(\gamma_s,\beta_s,p_s)} \alpha_s$.

Note that as attention to a state s increases, so does the subjective probability q_s assigned to that state. Thus, increasing the payoff level of a state, or increasing the payoff differences in that state from different actions (and the environment is separable), leads the DM to behave as if the state is relatively more likely.

Parameter λ continues to govern how close the DM is to a standard DM who, in this context, maximizes the expected consumption payoff. As λ increases, the expected consumption payoff decreases, that is, through a standard lens, the DM's action becomes worse.

Lemma 2B. *Let* $\Gamma(\lambda)$ *denote the set of optimal (action, attention)-pairs given* λ. *If* $\lambda' > \lambda$, *then* $\max_{(x,a)\in\Gamma(\lambda')} \sum_{s\in\mathcal{S}} V_s(x) \leq \min_{(x,a)\in\Gamma(\lambda)} \sum_{s\in\mathcal{S}} V_s(x)$.

Lastly, Lemmas 3A and 4A also have their straightforward analogs here.

Let the payoff in state s be parameterized as $V_s = \beta_s \tilde{V}_s + \gamma_s$, for fixed $\tilde{V}_s$ and scalars $\beta_s \geq 0$ and γ_s, and let the DM's value be (2) for optimal (action, attention)-pairs.

Lemma 3B. *The DM's value is convex in* $(\gamma_s)_{s\in\mathcal{S}}$.

Lemma 4B. *Suppose the environment is separable. Let* $F(\lambda)$ *denote the DM's objective* (2) *given* λ. *For any* $\lambda' > \lambda$, *if* $F(\lambda)$ *is convex in* α, *then so is* $F(\lambda')$.

Lemmas 1B–4B establish some basic comparative static that govern the optimal (action, attention)-pair and thus the subjective probabilities; here, we continue this path and consider particular environments and the type of subjective probabilities they give rise to (via optimal attention). We begin 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 consumption payoff in state s given lottery x is $V_s(x) = u(x_s)$. For the second and third part of the ensuing proposition, we consider binary lotteries, those where any state either pays a low payoff $L(x)$ or 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 1.

- *Let DM(λ) refer to the DM given λ. DM(λ) is more risk-averse than DM(λ') for any $\lambda' > \lambda$.*[5]
- *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 1 first states that the DM has an additional preference for risk. Intuitively, given a lottery x, the DM devotes attention the 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. (This generic optimism relies on attention playing no instrumental role.) 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, using the interpretation of (2) as the weighted sum of expected material payoff and attention utility, positive skew always increases the latter, sine the DM devotes 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 increase in the skew of a lottery.

Proposition 1 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. We consider the mapping of objective probabilities p_s to subjective probabilities q_s—i.e., probability weighting—varying the details of how attention is instrumentally valuable. Given the dependence of q_s on p_s (both directly and through attention), we write $q_s(p_s)$; as the attention allocation may not be unique, q_s is set-valued in general and if there is a unique solution, we take $q_s(p_s)$ to be the scalar associated with that solution. For simplicity, we focus on the case with only two states, $\mathcal{S} = \{s, s'\}$ and consider a separable environment.

First, we consider probability weighting when there is no instrumental value (i.e., a counterpart to Proposition 1). Here, the DM devotes full attention to the state with the higher payoff and

[5] 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.

behaves 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 low, and how this minimum attention to even low-probability states result in the DM to overweight such 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 2. *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*

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

  (If $q(p)$ is a set, then the above comparisons apply to each element of $q(p)$.)

- *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*

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

(We note that in the second case of the proposition, although we obtain 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.)

Figure 2 illustrates the different forms of probability weighting (and attention allocations) occurring in the three cases discussed in Proposition 2. A panel corresponds to an environment with the top subfigure showing the optimal attention 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 corresponding 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 and we now provide an example. Choosing $\hat{V}(\alpha) = -\frac{1}{\alpha}$ as the functional form for the payoff as a function of attention turns out to be tractable. Maximizing $\sum_{s\in\mathcal{S}}(p_s + \lambda\alpha_s)\hat{V}(\alpha_s)$ gives $\alpha_s = \frac{p_s-\sqrt{p_s(1-p_s)}}{2p_s-1}$. Thus, optimal attention is inverse-S-shapes as is shown in the top subfigure of Panel (b). When state s does not occur for sure $p_s = 0$, then the DM devotes full attention to state s'. When p_s increases, the DM devotes some attention to s as otherwise their expected payoff, $\sum_{s\in\mathcal{S}} p_s\hat{V}(\alpha_s)$, is $-\infty$. In fact, the instrumental value of attention for low levels of attention is so large, that the DM chooses $\alpha_s > p_s$. Conversely, by symmetry, we have $\alpha_s < p_s$ for large p_s strictly less than 1. As the subjective probability q_s is linear in the objective probability p_s and the α_s, the probability weighting function inherits the inverse-S shape of attention.

Before discussing related empirical evidence, we briefly contrast our model, and in particular the probability weighting it produces, to that of existing models. Consider a model with rank-dependent probability weighting, e.g., cumulative prospect theory (Tverseky and Kahneman, 1992). With two states, the probability assigned to a state (q_s) depends 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, $V_s - V_{s'}$, and not just the ranking, as well as the instrumental value of attention (e.g., in a separable environment, on $\frac{\partial}{\partial a}\hat{V}_s(a)$). Thus, an increase in the payoff or instrumental value of attention in a state, increases the subjective probability of that state (Lemma 1B), producing predictions beyond those made by cumulative prospect theory.

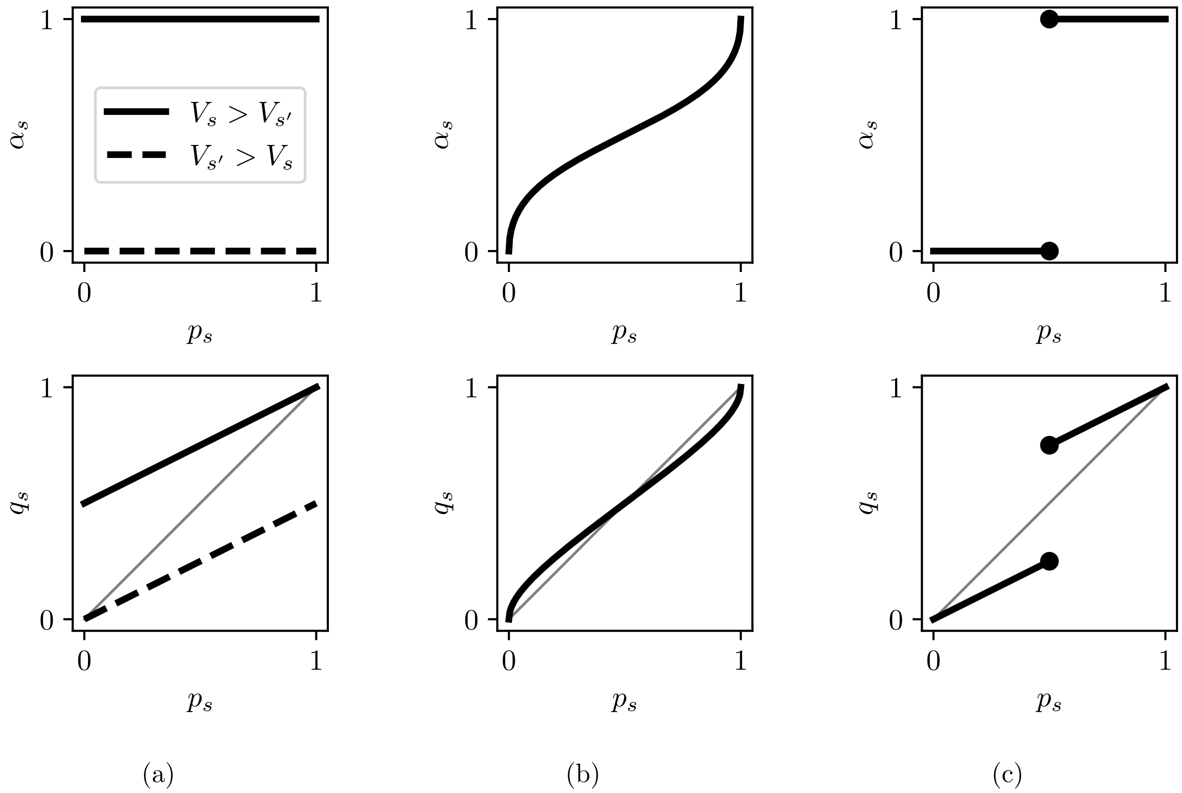


Figure 2: Panels (a), (b) and (c) correspond to the three environments discussed in Proposition 2—no instrumental value, minimum attention requirement, increasing returns to attention—respectively. The weight on attention utility is always $\lambda = 1$. The functional forms in Panels (a) and (c) do 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 $\hat{V}(\alpha) = -\frac{1}{\alpha}$.

Empirical evidence

We next discuss some empirical evidence, and how it compares to our model's predictions. Additionally, we note predictions that, while perhaps intuitive, are not (yet) tested. One such prediction is that individuals devote more attention to high-payoff states (Lemma 1B), 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 a 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).

When attention plays no instrumental role (or it is small), or model predicts optimism with the DM acting as if the high-payoff states are more likely than they are (first part of Proposition 1). (An implication is that they appear more risk-seeking than the curvature on their Bernoulli utility

u would indicate.) Optimism, defined in this way, 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 would 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 such low-payoff states, thus acting as if they "underrate" them (see Orhun et al. (2021) for a recent example of optimism with respect to health risks)

There is laboratory evidence of optimism, e.g., Mayraz (2011). There, participants guess the realization of a random variable and are rewarded for accuracy and, additionally, 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, in turn, then have a large weight in their objective, leading the participants to guess a high realization. Indeed, this is what Mayraz (2011) finds.

Our model predicts a preference for positively skewed payoffs (second and third part of Proposition 1). Evidence for such preference has been documented 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). (In all these settings, our model predicts that individuals value the possibility of a (very) high payoff, as it allows them to devote attention to the associated state.) Furthermore, consistent with our model, Jullien and Salanié (2000); Snowberg and Wolfers (2010) suggest that that the preference for skewness is driven by subjective probabilities, as in our model, rather than the Bernoulli utility u.

Our model makes predictions about probability weighting (first discussed in Kahneman (1979)) and how it depends on the details of the environment (Proposition 2). Empirically (and also theoretically), there is now a voluminous literature analyzing and empirically estimating prospect-theory models (see Wakker (2010); Barberis (2013) for two surveys). The classic finding is that individuals probability weighting follows an inverse S (Wu and Gonzalez, 1996). Our model generates such probability weighting, if payoff increase from devoting a small amount of attention, rather than none, is large (second case of Proposition 2); thus, it provides a mechanism giving rise the inverse-S shape. Our model predicts other forms of probability weighting, e.g., an S-shaped probability weighting if the instrumental value of attention is increasing (third case of Proposition 2); to our knowledge, those predictions are yet to be tested.

2.3 Attention allocation across time periods

We next consider a third, and final, dimension—attention allocation across time—and the implications of the two fundamental features of attention in this context. The DM now faces a sequence of time periods $\mathcal{T} = \{1, \ldots, T\}$, with generic period t. To isolate attention allocation across time, we suppose that each period is associated with a single consumption problem and that there is no uncertainty (we consider implications of a dynamic model for intra-period attention allocation in Section 2.4). As when attention is allocated across consumption problems (Section 2.1) and states (Section 2.2), attention both allows the DM to take actions that affect consumption payoff in each period, and determines how these payoffs are aggregated in the DM's objective. The attention-weighted environment, in this context of consumption across time, implies that the marginal value of consumption in a period depends on how much attention is allocated to that period. In other words, it can be interpreted as endogenous discounting. Once more, we first introduce the environment and derive some results similar to those in previous section highlighting the determinants of the DM's optimal (action, attention)-pair.

Period t is associated with a single consumption problem whose payoff is denoted by V_t. We also refer to period t with payoff V_t. In each period t, the DM the chooses an (action, attention)-pair denoted by (x_t, α_t), i.e., there are multiple "selves" and we do not assume the DM can commit their future selves. The actions jointly determine the payoff in each period: given $x \coloneqq (x_t)_{t=1}^T$, the payoff in period t is $V_t(x)$ with natural assumptions on future actions' impact on past payoffs.[6] Attention is, once more, a measure on the set of time period 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 $\sum_{t' \in \mathcal{T}} \alpha_{t \to t'} = 1$.

Attention (still) has two implications. First: the available actions depend on the attention allocation, i.e., letting $\alpha \coloneqq (\alpha_t)_{t=1}^T$, given α, x must be chosen from $X(\alpha) \coloneqq \Pi_{t=1}^T X_t(\alpha_t)$, where X_t is finite- and non-empty-valued and upper hemicontinuous. We also maintain the monotonicity assumption on X_t (now with respect to measures on $\mathcal{T}$, but otherwise, identically defined). As before, the possible actions could vary depending on the application. For example, the DM could decide how much to consume in the current and future periods. In this case, increased, or better chosen, consumption in the current period may require attention to that period. Similarly, saving

[6] Unlike for V_c and V_s, we do not assume that V_t is continuous since we make a stronger topological assumption on available actions rending continuity not needed to guarantee existence of a solution, which is more demanding given the multiple "selves."

for future consumption, or planning for future actions, may require attention to future periods. Another example is thinking about the different optimization problems they have to solve in the different time periods.

Second: attention determines the weights on the consumption payoff in each period in the DM's objective, but we have to be careful in applying our approach from previous sections. We assume that the weight of period t' in the DM's objective in period t is given by $1 + \lambda \sum_{t'' \geq t}^{T} \alpha_{t'' \to t'}$, where $\lambda \geq 0$, as before, governs the extent to which attention weights the payoff terms. In words, the weight on period t' increases in the attention the DM devotes currently or in the future to period t'. We chose this formulation as it naturally lends itself to the interpretation of the DM valuing the "material payoff" from actual consumption in period t' as well as the "attention utility" each "future self" receives from period t'. In period t, the DM's objective is then the sum over payoffs from current or future periods (we do not add exogenous discount to emphasize the role of attention in generating endogenous discounting)

$$\sum_{t'=t}^{T} (1 + \lambda \sum_{t''=t}^{T} \alpha_{t'' \to t'}) V_{t'}(x); \tag{3}$$

$$\text{or equivalently:} \quad \sum_{t'=t}^{T} \left(\underbrace{V_t(x)}_{\text{material payoff in } t} + \underbrace{\lambda \sum_{t''=1}^{T} \alpha_{t' \to t''} V_{t''}(x)}_{\text{attention utility in } t} \right).$$

We do not assume that the DM can commit to a particular (action, attention)-pair in future periods; instead, the DM anticipates their future behavior, i.e., the DM's overall problem is solved using backward induction. 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 $x \in X(\alpha)$.

The weights on payoffs across periods can be interpreted as discounting. Let (x, α) be a solution. Then, the DM behaves like a standard DM, but one who discount period t' (relative to period t) by $\delta_{t \to t'} := \frac{1+\lambda \sum_{t''=t}^{T} \alpha_{t'' \to t'}}{1+\lambda \alpha_{t \to t}}$. (We cannot rewrite (3) and divide through by $1 + \lambda \alpha_{t \to t}$, as this term is endogenous.)

Once more, we again being our analysis by considering variants of the simple comparative statics encountered previously that determine the optimal allocation of attention and action. We introduce a familiar parameterization: For each period t, we fix $\tilde{V}_t$ and define the payoff in that period as $V_t = \beta_t \tilde{V}_t + \gamma_t$ for any scalars $\beta_t \geq 0$ and γ_t. Similar to before, increasing γ_t shifts the payoff level in period t, and increasing β_t increases the payoff difference in period t induced by different actions. We also say an environment (consisting of $(V_t)_{t=1}^T$ and X) is separable if, letting $\alpha_{t\to -t} := (\alpha_{t\to t'})_{t'=1}^T, t' \neq t$ and $x_{-t} := (x_{t'})_{t'=1}^T, t' \neq t$, $V_t(x_t, x_{-t})$ and $X_t(\alpha_{t\to t}, \alpha_{t\to -t})$ are independent of x_{-t} for all t and x_t and $\alpha_{t\to -t}$ for all t and $\alpha_{t\to t}$, respectively. For such environment, for each t, we define $\hat{V}_t(\alpha_{t\to t}) := \max_{x_t \in X_t(\alpha_t)} V_t(x_t, \cdot)$.

The following lemma provides a result similar to those in Lemmas 1A and 1B, however with some restriction due to the lack of commitment the DM faces.

Lemma 1C. *Consider period* 1 *with* $\tilde{V}_1$, *fix* V_{-1} *and let* $\Gamma_t(\gamma_t, \beta_t)$ *denote the set of optimal profile of (action, attention)-pairs. Suppose it is never optimal to devote attention to a past period, i.e.,* $\alpha_{t\to t'} = 0$ *if* $t' < t$.

- *If* $\gamma_1' > \gamma_1$, *then* $\min_{(x,\alpha)\in\Gamma_1(\gamma_1',\beta_1')} \alpha_{1\to 1} \geq \max_{(x,\alpha)\in\Gamma_1(\gamma_1,\beta_1)} \alpha_{1\to 1}$. *If, in addition, the environment is separable, then* $\min_{(x,\alpha)\in\Gamma_1(\gamma_1',\beta_1)} V_1(x) - \max_{(x,\alpha)\in\Gamma_1(\gamma_1,\beta_1)} V_1(x) \geq \gamma_1' - \gamma_1$.
- *If for* β_1 *and* γ_1, $\max_{(x,\alpha)\in\Gamma_1(\gamma_1,\beta_1)} V_1(x) = \min_{(x,\alpha)\in\Gamma_1(\gamma_1,\beta_1)} V_1(x)$, *then for any* $\beta_1' > \beta_1$ *and* $\gamma_1' = \gamma_1 - (\beta_1' - \beta_1)\tilde{V}_1(x)$, *where* $(x,\alpha) \in \Gamma_1(\gamma_1,\beta_1)$, *we have* $\min_{(x,\alpha)\in\Gamma_1(\gamma_1',\beta_1')} V_1(x) \geq \max_{(x,\alpha)\in\Gamma_1(\gamma_1,\beta_1} V_1(x)$. *If, in addition, the environment is separable, then* $\min_{(x,\alpha)\in\Gamma_1(\gamma_1',\beta_1')} \alpha_{1\to 1} \geq \max_{(x,\alpha)\in\Gamma_1(\gamma_1,\beta_1} \alpha$.

Note that as attention to the present period increases, e.g., as its payoff level increases or the payoff differences from different actions increase (and the environment is separable), the DM must discount some future periods by more. The time discounting—whether the DM is present or future-focused—is thus endogenous and depends on circumstances.

Although this is the analogue of the "simple comparative statics" noted in the previous settings of attention allocation across consumption problems (Lemmas 1A and 2A) and across states (Lemma 1B and 2B) it is more restrictive as the DM takes action at multiple points in time. Similar comparative statics in future periods may thus, in general, lead to nonmonotonicities due to nonmonotonicities in the DM's coordination problem. Example 1 in Appendix F provide examples where increasing the a future payoff leads to less attention to that period. By forcing the DM to not devote attention to past periods, future selves are unaffected by changes to the consumption payoff in the first period, unless the DM changes their behavior, and thus the unambiguous implications follow.

Similarly, in previous settings, we could state a straightforward comparative with respect to λ. No such result is possible in this dynamic setting: varying λ can affect the sum of consumption payoffs non-monotonically as Example 2 in Appendix F demonstrates.

Lastly, Lemmas 3A and 3B as well as Lemmas 4A and 4B also have counterparts here.

Let the payoff in time t be parameterized as $V_t = \beta_t \tilde{V}_t + \gamma_t$, for fixed $\tilde{V}_t$ and scalars $\beta_t \geq 0$ and γ_t, and let the DM's value be (3) for $(x, \alpha) \in \Gamma_1$.

Lemma 3C. *Suppose it is never optimal to devote attention to a past period, i.e., $\alpha_{t \to t'} = 0$ if $t' < t$. Then the DM's value is convex in γ_1.*

Lemma 4C. *Suppose the environment is separable. Let $F(\lambda)$ denote the DM's objective* (3) *given λ. For any $\lambda' > \lambda$ and $t \in \mathcal{T}$, if $F(\lambda)$ is convex in α_t, then so is $F(\lambda')$.*

(We briefly comment on the restrictions: The convexity result regarding payoff levels is restricted to the present period; the DM's value may not be convex in payoff levels more generally as a result of nonmonotonicities arising due to the lack commitment. The convexity result regarding attention is restricted to a single period; this restriction is necessary as even when the environment is separable, more attention devoted to a period does not necessarily increase its payoff, and so the "attention utility" for convex combinations of attention may be larger than the convex combination of attention utilities, while the same does not hold for the associated "material payoff.")

Since general statements about the impact of λ are difficult to make, we next consider a particular environment to relate λ to the pattern of attention (and hence discounting) and observable payoffs (actions). Essentially, 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 (with the exception that attention to past periods cannot retrospectively increase the payoff) as to not bias the model, at least directly, in favor of a particular attention allocation.

The DM's overall payoff in period t is then given by

$$\sum_{t'=t}^{T} (1 + \lambda \sum_{t''=t}^{T} \alpha_{t'' \to t'}) V_{t'} (\sum_{t''=1}^{t} x_{t'' \to t'}). \tag{4}$$

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. Such complementarities may occur for two reasons. One is due to the novelty in our model as attention distortion the environment: an increase in $\alpha_{t\to t''}$ increases the consumption payoff in period t'', given by $V_{t'}$, which makes increasing $\alpha_{t'\to t''}$, and thus the weight on period t'', more beneficial. Note that effect increases in the weight on attention utility λ; indeed, it is absent for the standard DM.

A second reason may simply due to the shape of V (and thus present for the standard DM already), e.g., if V is convex. In order to focus on the novelty of our model, we focus on the case where V is strictly concave. As we show below, this implies that depending on the weight on attention utility λ, attention across periods may either be substitutes or complements.

To simplify the statement of the following proposition, we 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. We do not consider these assumption of substantive importance and briefly discuss their roles of these assumptions after the proposition.

Proposition 3. *There exist $\underline{\lambda} > 0$ and $\bar{\lambda} < \infty$, such that*

- *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. When λ is small, i.e., the DM is close to a standard DM, the DM maximizes the sum of consumption payoffs: here, due to the concavity of V , this is achieved by devoting full attention to the present period t (in each period).

When λ is large, the additional complementary across 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—a period where the DM "parties." 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, these low payoffs only have a small weight.

As promised, the roles of the assumption stated just before the propositions are not essential: If V is strictly increasing everywhere, for large enough weights on attention utility, the DM would devote attention always to period T and have one "big party." If K is not a divisor of T, then the last "party" would be "smaller" than the previous ones. If V was not strictly concave, then there may be "parties" even if λ is very small as doing so does not come at a cost to the sum of consumption payoffs.

Because attention is directly linked to discounting, we observe different patterns of discounting depending on λ. When λ is small (and attention is devoted only to the present), we observe a form of non-exponential discounting: in period t, the DM discounts each period $t' > t$ by $\frac{1}{1+\lambda}$. This non-exponential discounting implies that the DM is time-inconsistent. In particular, they are indifferent between shifting the consumption payoff between two future periods, but strictly prefer earlier consumption as one of those periods becomes the present. The DM thus falls in the class of quasi-hyperbolic discounters (Laibson, 1997).

In contrast, when λ is large, the DM assigns weight 1 to any period that is not a "party period," and weight $1 + \lambda$ on the "party period," effectively the DM discounts all periods other than the "party period"—they are future-focused. In the "party period," the DM is present-focused and discounts future periods by $\frac{1}{1+\lambda}$.

Empirical evidence

Lemmas 1C, 3C and 4C predict that time discounting (as defined by $\delta_{t\to t'}$) follows basic comparative statics: E.g., the DM may weight a period more, if the payoff level or the the instrumental value of attention to that period increase. Wile discounting (i.e., $\delta_{t\to t'} \neq 1$), including non-exponential (where exponential discounting is $\delta_{t\to t'} = \delta^{t'-t}$) is well documented (see Frederick et al. (2002) for a review), we are not aware of any empirical evidence testing these predictions.

However, we formalize the perspective of discounting being driven by attention which 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). In light of our model, "present focus" instead of "present bias" to describe this stylized fact is indeed the more aptly chosen term to describe decreasing discount factors. Similarly, our model suggests a "future focus" and that the same individual may vary between the modes depending on circumstances.

Proposition 3 offers a result when the extent to which attention reweights payoffs, λ, is varied. When the DM is close to standard, they behave like a quasi-hyperbolic discounter (i.e., $\delta_{t\to t'} = \beta\delta^{t'-t}$ for $t' > t$, with $\beta = \frac{1}{1+\lambda}$ and $\delta = 1$; Laibson (1997)).[7] Thus, our model suggests that such present

[7]Many applications of quasi-hyperbolic discounting assume that $\delta = 1$ (e.g. O'Donoghue and Rabin (1999).

focus may occur when attention to the present maximizes the (unweighted) payoff stream and attention does little reweighting; in this case, the present focus is increasing in λ.

When λ is large, the DM instead deviates from the (unweighted)-payoff-maximizing allocation of attention and instead creates "parties," periods with a particularly high payoff, which subsequently receive a large weight. 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) note that the average expenditure on weddings is about USD 20,000, and that the average annual household income of newly married couple is USD 55,000. Our prediction is consistent with such payoff patterns. Other studies have, such as Frederick (1999), Rubinstein (2000), Loewenstein and Sicherman (1991) and Loewenstein and Prelec (1993), have found increasing discount rates or a preference for delaying a better outcomes.

Of course, our model also makes a rich set of predictions regarding how discount rates should change as the environment changes. We do not know of many clear tests of the predictions that can emerge from our model other than one. Our model makes the prediction that larger payoffs should be discounted less, because they attract more attention. This has been found repeatedly (e.g., Green et al. (1997) is an early paper).

2.4 Combining all three dimensions of attention allocation

So far, we considered attention allocation across consumption problems, states, and time periods—one at a time; here, we include them in a single setting allowing us to derive their joint implications. In particular, we focus on the implications of a future intra-period attention allocation problem (taken from Section 2.1) on the optimal (action, attention)-pair today.

For a first (informal) result, recall that the DM's objective in the future period is convex in payoff levels by Lemmas 3A and 3B (and also Lemma 3C). Thus, the DM today chooses an action that induces a mean-preserving spread of the future payoff levels over no action. It is this logic that creates a preferences for information—even if it is non-instrumental for the standard DM. We discuss such informational preferences as an application in Section 3.1. Additionally, the DM likes to induce a possibly high future payoff, so that they can devote attention today to that "state" (the same intuition drives the preference for skewness as expressed in Proposition 1).

Although the formal results show the equivalence to this form, it is an artifact of the fact we have no pure time discounting—with endogenous discount factors we would get a richer structure of present biased preferences.

We next consider actions whose effect on future payoffs goes beyond simply inducing a mean-preserving spread. We first suppose that a future consumption problem's payoff depends on today's action, but only if the DM does not devote attention to the consumption problem in the future; in other words, we consider default actions. The future consumption problem the DM faces is random—e.g., it may vary in payoff levels and which action increases its payoff most. The DM, anticipating their inattention to the consumption problem when its payoff is low, chooses a default optimal for those realizations, i.e., one that is "pessimistic." We then suppose that the action affects the future payoff even when the DM does devote attention (i.e., the action is not a (pure) default). 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.

There are two periods, period 1 ("today") and period 2 ("the future"). To easily define a default in our context, suppose that there are only two consumption problems in the second period with one being trivial—its payoff, denoted by $\bar{V}$, is independent of any action taken—and one non-trivial one which we denote by c. Problem c is random with finitely many realizations $\mathcal{C}$ where a particular one, c, has probability p_c. Let $\alpha_{1\to c}$ denote the attention devoted in the first period to the yet to be realized problem c (with the remainder, $\alpha_{1\to t} = 1 - \sum_{c\in\mathcal{C}} \alpha_{1\to c}$, devoted to the trivial problem; the vector of attention in the first period is denoted α_1), and let $\alpha_{2\to c}$ denote devoted in the second period to the now realized problem c (with the remainder, $\alpha_{2\to t} = 1 - \alpha_{2\to c}$ devoted to the trivial problem). Let α_1, α_2 denote the full vectors of attention in the first and second period, respectively.

The DM chooses a default action in the first period $x_1 \in X_1(\alpha_1)$, that she can revise in the second period, $x_2 \in X_2(\alpha_2)$; formally, this means that if $\alpha_{2\to c} < \eta$, then $V_c(x_1, x_2)$ (with $x_2 \in X_2(\alpha_2)$) is independent of x_2, and if $\alpha_{2\to c} \geq \eta$, then $\max_{x_2\in X_2(\alpha_2)} V_c(x_1, x_2)$ is independent of x_1 for all x_1. In the former case we drop x_2 and in the latter x_1 to highlight that their values do not matter. Furthermore, for ease of notation, we assume that the solution is unique. In the second period, the DM's attention, denoted by $\alpha_{2\to c}(x_1, c), \alpha_{2\to t}(x_1, c)$, scalars jointly with $x_2 \in X_2(\alpha_2)$ solving

$$(1 + \lambda\alpha_{2\to c})V_c(x_1, x_2) + (1 + \lambda\alpha_{2\to t}))\bar{V}; \tag{5}$$

denote $x_2(x_1, c)$ the corresponding solution.

But our interest is in the choice of x_1, the default action. The DM solves

$$\sum_{c\in\mathcal{C}}(p_c + \lambda(\alpha_{1\to c} + p_c\alpha_{2\to c}(x_1)))V_c(x) + (1 + \lambda(\alpha_{1\to t} + \sum_{c\in\mathcal{C}} p_c\alpha_{2\to t}(x_1, c)))\bar{V},$$

subject to $x_1 \in X_1(\alpha_1)$ and $x_2(x_1, c), \alpha_{2\to c}(x_1, c)$ and $\alpha_{2\to t}(x_1, c)$ solving (5). The following proposition characterizes the DM's optimal default.

Proposition 4. *Suppose that the solution is unique. Let $B(x_1) \coloneqq \{c : \alpha_{2\to c}(x_1, c) < \eta\}$.*

- *If $X_1(\alpha_1)$ is independent of α_1. The optimal x_1 satisfies*

$$x_1 = \operatorname*{argmax}_{x_1 \in X_1(\cdot)} \sum_{c \in B(x_s)} p_c V_c(x_1, \cdot).$$

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

$$B(x_1) = \{c : \max_{\alpha_2, x_2 \in X_2(\alpha_2)} V_c(\cdot, x_2) < \bar{V}\}.$$

The first part of Proposition 4 states that the DM chooses a default focusing solely on realization of c for which the default is binding, i.e., $c \in B(x_1)$. For the remaining realization, the DM devotes sufficient attention to c 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 of c for which the max V_c is strictly less than the payoff from the trivial problem. In other words, the DM devotes attention to c if and only if it has a high payoff.

Notice that Proposition 4 does not require the default to be chosen optimally; it thus applies to endogenous defaults, i.e., situations where the DM chooses the default, as well as exogenous ones, where a second party, say a policy-maker or a company, chooses the it. Indeed, defaults are relevant in a variety of situations. For instance, consider an individual who chooses how much to consume before learning their income. After observing their income, they can revise their planned consumption, but doing so requires attention to it. Such individual would possibly react asymmetrically to these income shocks and, in anticipation, choose a low initial (planned) consumption level. More generally, individuals often employ only partially specified plans, those that do not take all future contingencies into account—"heuristics." Our model suggests that these heuristics may be devised pessimistically and that individuals fall back on heuristics (system 1) when the situation (the "consumption problem") at hand has a low payoff, whereas they engage with the situation and use system 2 when its payoff is high. Another example is that of a firm with an incompletely specified organizational design (incomplete contracts). Our DM represents the firm's leadership, or perhaps the owners. When the firm faces hardship, say, because the economy is sliding into a recession, the consumption payoff associated with the firm may decrease, and the firm's leadership may devote attention elsewhere. As a result, the design of the organization—business plan, contracts, and structures—is not adjusted to reflect the new reality. A final example is portfolio choice; here, today's portfolio is tomorrow's default, if it were not for market changes. We discuss this application in detail in Section 3.2.

Lastly, we consider actions that also affect future payoffs even if the DM devotes attention to the consumption problem in the future. We maintain the setting as described above with one modification: the payoff of problem c is now given by $V_c(x) = \tilde{V}_c(x) + \beta F(x_1)$, where $\tilde{V}_c$ has the "default property" from above, and $\beta \geq 0$. Thus, the previous setting is nested with $\beta = 0$.

Proposition 5. *Suppose $\beta > 0$, that the solution is unique, and that $X_1(\alpha_1)$ is independent of α_1 and finite. Let $(c^*, x_1^*) \in \operatorname{argmax}_{c\in\mathcal{C},x_1\in X_1(\cdot)} V_c(x_1, x_2(x_1, c))$. If $V_{c^*}(x_1^*, x_2(x_1^*, c^*)) > \bar{V}$ and λ is large enough, then $x_1^* \in \operatorname{argmax}_{x_1\in X(\cdot)} F(x_1)$*

In the setting of Proposition 5, the DM chooses an action that maximizes F, i.e., x_1 is chosen independent of $\tilde{V}_c$'s (and p_c's), thus, contrasting Proposition 4. There, if λ is large, the DM chose an action optimal for realized c's for which the DM does not devote attention in the second period. But, as λ is large, it is those realization that do not matter (much) in the DM's objective (they only value what they devote 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 (c^*) or the second period ($\mathcal{C} \setminus B(x_1^*)$); 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 Applications

We now turn to relating our model, and many of the results derived in the previous section to two well known, concrete economic environments: information acquisition and portfolio choice.

3.1 Information acquisition

Our first application concerns the DM's preferences over information acquisition: when do they demand information about a consumption problem? And if so they do, 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 that is independent of the action the DM takes. Although classical economic agents will be indifferent to information in this setting, DM may nevertheless have strict preferences over when to acquire information, and the structure of the information. We consider this environment because it allows us to highlight the novel implications of our model by shutting down any behavior on the part of classic economic motivations. Intuitively, our DM's future attention allocation is a "hidden action" and they may benefit from conditioning this action on the information acquired—if the information is good, they can devote a lot of attention, and devote attention elsewhere otherwise. Anticipating future attentional allocations, agents in prior periods have preferences over information in order to maximize the future option value of attention.

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 also model 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 6.

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 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 instatiation of Lemma 1A—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 Lemma 3A.

If the agent only acquires a "small amount" of information (that is, the variance of distribution of posteriors) 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 if their prior is high enough, acquiring information (and devoting attention) is optimal. 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, then 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 6 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) provide experimental evidence that participants' willingness-to-pay to learn about their performance in an IQ test increases if they received good, instead of bad, news before.

3.2 Portfolio choice

We apply our model to a problem of 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 $x_2 \in [0, w + x_1 r_1]$. 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 $w + x_1 r_1 + x_2 r_2$ 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 $V_\rho(x_1, x_2) := E_{r_2\sim F(N)}[u(w + x_1r_1 + x_2r_2)]$. 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 $\underline{x}_2$), the consumption payoff is given by $V_\rho(x_1, \underline{x}_2) := E_{r_2\sim F(N)}[u(w - x_1 + x_1(1+r_1)(1+r_2)]$; if they do, then they choose $x_2 \in [0, w + x_1r_1]$ 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

$$\begin{aligned}
&(\alpha_{2\to\rho} = 0) \quad V_\rho(x_1, \underline{x}_2) + (1 + \lambda)\bar{V}, \\
&(\alpha_{2\to\rho} = \eta) \quad (1 + \lambda\eta_2) \max_{x_2\in[0,w+x_1r_1]} V_\rho(x_1, x_2) + (1 + \lambda(1 - \eta_2))\bar{V}, \\
&(\alpha_{2\to\rho} = 1) \quad (1 + \lambda) \max_{x_2\in[0,w+x_1r_1]} V_\rho(x_1, x_2) + \bar{V},
\end{aligned} \tag{6}$$

respectively.

Taking the max over these gives the payoff of ρ given optimal second-period behavior (we assume throughout that 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 factor influencing the investment decision such as changing preferences over risk as wealth varies. Such a 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, researcher 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, in order to ease notation. Our next 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* $B(\bar{V}, \lambda) \coloneqq \{\rho : \alpha_{2\to\rho}(\rho, x_1) < \eta_2\}$*. 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 Lemma 1A; we, again, refer to the discussion in 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 rid 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 to 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 (in the sense stated in the result) realizations of ρ.

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 implication for feasible trades). Formally, this may be modeled by letting each asset constitute a distinct consumption problem. We suspect that such model leads to a within-portfolio ostrich effect—differential attention across assets depending on their (individual) performance—very much in the spirit of Lemma 1A.

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 1 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; Lemma 3A suggests that the DM seeks to face a varied future payoff, with the same effect; however, Proposition 4 suggests, instead, that the DM may choose a portfolio that performs well for those ρ for which the DM does not 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 aforementioned intuition, but what about $1-\eta_1$? $1-\eta_1$ is the amount of attention that the DM allocates freely (after having devoted amount η_1 to the expected portfolio choice problem). The DM devotes this attention to state where the return of the risky asset are highest; thus, as in Proposition 1, 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 largely 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 $B := \{\rho : \alpha_{2\to\rho}(\rho, x_1) < \eta_2\}$.*

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

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 2.4), 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 to 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 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 causes 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.

4 Implications for (self-imposed) policies

In this section, we discuss our model's implications for policy-making broadly construed and ask how a policy-maker—which may be a second party, say, a government, or the DM themselves—should intervene in the environment. Broadly speaking, incorporating our model into policy-making is necessary to fully understand the behavioral changes a policy induces—which may be different to 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 time periods) are similar. We consider three broad classes of policies: optimal resource allocation, incentivization of actions, and optimal construing of consumption problems.

4.1 Optimal resource allocation

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

Suppose first, that the policy-maker can increase one (or multiple) of the payoffs of the consumption problems by some total amount of payoff. 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 policy-maker 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 policy-maker 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 policy-maker 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 an example, α_c and r_c could both represent amounts of information, where the former is acquired by the DM, and the latter provided by the policy-maker (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 easy 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 of 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 (1), 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 policy-maker 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 policy-maker does not bear this cost and thus the benefit of the input is largest for the problem with the largest 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.[8]

When the policy-maker is a government, these results may guide how the governments resources are best allocated; which tasks should be left to the individual and which tasks are better completed by the government.

4.2 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 $\hat{V}_c(\alpha_c) = p(x_c)V_H + (1 - p(x_c))V_L$, 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).

[8]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.

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

Proposition 7. *Consider the environment as introduced prior to this proposition and suppose 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 7, 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).

4.3 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 policy-maker installs. Such bracketing of consumption problems is a form intentional use of bottom-up attention: 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. The DM applies the average distortion to problems in the same bracket, i.e., let $B(c) \in B$ with $c \in B(c)$, the DM's objective is

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

where $\bar{\alpha}_C := \frac{\sum_{c \in C} \bar{\alpha}_c}{|C|}$ for $C \subseteq \mathcal{C}$. We can equivalently interpret the above through our interpretation of the model as the DM maximizing material payoff and attention utility.[9]

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 8. *If (x, α) and B are optimal, 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 a lot of attention to a low-payoff problem, then 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 the distortion in the direction of this problem (their "attention utility") 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 problem simultaneously and behave like a standard DM, when their distortion is in the direction of low-payoff consumption problems.

5 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_c. 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

[9]Here, 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 $\sum_{c \in \mathcal{C}} (1 + \lambda \alpha_c) \bar{V}_{B(c)}(x)$, where $\bar{V}_C(x) := \frac{\sum_{c \in C} V_c(x)}{|C|}$ for $C \subseteq \mathcal{C}$.

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 explicit feature of of attention in reweighting the environment, takes this role (see also the discussion in Section 2.1).

Bayesian agents with anticipatory utility: 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)). There are some similarities: In our model, the attention-weighted objective can be interpreted as the result of the DM valuing "material payoffs" (the unweighted sum of consumption payoffs) and "attention utility" taking the form of attention to a problem times its payoff (see (1)). Attention utility, when stemming from a future problem, can be thought of as anticipatory utility. However, and unlike in the aforementioned models, the DM only "receives" this anticipatory utility if they devote attention to its underlying payoff; not otherwise.

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

(The point that models of information avoidance that rely on beliefs struggle to rationalize much of the presented evidence also applies to models of news utility (Kőszegi and Rabin, 2009).)

Chosen beliefs (with anticipatory utility): 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 2.3, and in Section 2.2, the DM chooses an attention allocation that lead 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 2 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 periods, endogenous weights on time periods appear and the DM develops 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 our 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 valued by ours since they ignore low-payoff periods and devote excessive attention to high-payoff ones.

Other Models of Attention:

There are a couple other papers that directly model the two fundamental features of attention in ways similar to us.

The model of Tasoff and Madarasz (2009) is closest to ours. A DM faces a decision problem with multiple dimensions (analogous to our different 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 Lemma 1A, 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 in the two formulations is that we allow for attention allocation with no instrumental consequence. More broadly, in our model, attention is chosen as to enable non-default actions and information acquisition, whereas in theirs, the order is reversed.)

Their subsequent focus is on how information provided (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 importance in the DM's objective). Instead, our focus is on attention allocation across uncertain states and time, and the ensuing behavioral phenomena due to the attention-reweighted environments. Our applications (information acquisition and portfolio choice) are also very different to their main one (a monopolist manipulating information provided to—and thus the attention of—consumers), as are the implications we draw for policy-making.

However, there are also important differences. First, they model the acquisition of information as causing attention. In other words, while in our model the choice of attention can determine the information structure, in theirs the choice of information structure determines (in expectation) the attention. Thus, for example, they do study the situation where attention can be directed even when no information has come to light a key driving force for some of our results. This distinction also motivates very different applications than us (their focus is on a monopolist manipulating information provided to—and thus the attention of—consumers).

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 also speeds up the adjustment of the reference point. 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 also construct our model more general allowing us to consider different dimensions of attention allocation with different insights.

(We note that Golman and Loewenstein (2018); Golman et al. (2021) provide models that incorporate attention, but model its provision as automatic, rather than as a choice variable.)

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.

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.

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

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.

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

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.

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.

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.

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.

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.

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

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

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

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.

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.

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

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

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.

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.

Loewenstein, George F and Dražen Prelec, “Preferences for sequences of outcomes.,” *Psychological review*, 1993, *100* (1), 91.

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.

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.

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.

O’Donoghue, Ted and Matthew Rabin, “Doing it now or later,” *American economic review*, 1999, *89* (1), 103–124.

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.

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.

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.

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

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 Identification

B Microfoundations

C Attention as processing signals

D Non-multiplicative

E Non-linear AU

F Additional examples

Example 1. *There are three time periods $T = 3$. The consumption payoffs in periods 1 and 2 are constant and equal and denoted by v, respectively. 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*

$$A_t(\alpha_t) = \begin{cases} \{a_{1,d}\} & \text{if } < \eta_t \\ \{a_{1,d}, a^*\} & \text{if } \alpha_{t\to 3} \geq \eta_t, \end{cases}$$

in particular, taking the action a^* requires attention devoted to period 3. The consumption payoff in period 3 is high if the DM takes action a^* in at least one period, otherwise, it is low.

Suppose the consumption payoffs in period 3 is lower than that in periods 1 and 2 $v > \bar{v}_3 > \underline{v}_3$. We solve for the equilibrium via backward induction. In period 3, the DM can only devote attention to the current period and take the default action. There are two histories to consider when determining the behavior of the DM in period 2. If the DM took action a^* in period 1, i.e., $h_2 = (a^*)$, then, since the consumption payoff in period 2 is larger than that in period 3 $v_2 > \bar{v}_3$, the DM takes action $a_{2,d}$ and devotes full attention to period 2 $\alpha_{2\to 2}$. If, instead, the DM took action $a_{1,d}$ in period 1, i.e., $h_2 = (a_{1,d})$, then, the DM takes action a^* and (optimally) devotes the *minimum amount of attention to period 3 if*

$$\underbrace{(1+\lambda(1-\eta_2))v+(1+\lambda(\eta_2+1))\bar{v}_3}_{\text{payoff if } \alpha_{2\to 3}=\eta_2,\, a_2=a^*} \geq \underbrace{(1+\lambda\cdot 1)v+(1+\lambda(0+1))\underline{v}_3}_{\text{payoff if } \alpha_{2\to 3}=0,\, a_2=a_{2,d}} \iff (1+\lambda)(\bar{v}_3-\underline{v}_3) \geq \lambda\eta_2(v-\bar{v}_3). \tag{10}$$

If (10) *holds, in period 1, the DM chooses action $a_1 = a_{1,d}$ and devotes attention to the current period, knowing that they will take action $a_2 = a^*$ in period 2. If* (10) *fails, a derivation similar to the above gives that the DM in period 1 chooses action $a_1 = a^*$ if*

$$(1+\lambda)(\bar{v}_3-\underline{v}_3) \geq \lambda\eta_1(v-\bar{v}_3). \tag{11}$$

Suppose that the attention requirement to take action a^* is weaker in period 1 than in period 2, i.e., $\eta_1 < \eta_2$. Then, condition (11) *is a weaker condition than condition* (10) *implying that there exist parameter values such that the DM chooses action $a_1 = a^*$ in the first period. Indeed, varying the consumption payoff in period 3 by increasing both $\bar{v}_3$ and $\underline{v}_3$ by γ results in three regions: For low values of γ (below the value for which* (11) *holds with equality), in no period does the DM choose action a^*; for intermediate values of γ (up to the value for which* (10) *holds with equality), the DM chooses action $a_1 = a^*$ in period 1; for high values of γ (the remaining values), the DM chooses action $a_1 = a_{1,d}$ in period 1 and $a_2 = a^*$ in period 2.*

Note that the attention allocation of the DM in period 1 is non-monotone in γ. In particular, as γ increases from an intermediate value, for which the DM in period 1 devotes attention $\alpha_{1\to 3}$ to period 3, to a high value, for which they do not devote any attention to period 3, the attention devoted to period 3 decreases.

Similarly, one can vary the instrumental value of attention and multiply the consumption payoff in period 3 by $\beta \geq 0$. A non-monotonicity of the attention devoted to period 3 as a function of β can be constructed similarly.

Example 2. *Take the setting of Example 1. Recall that condition* (11) *is a weaker condition than condition* (10)*. In particular, by varying the weight on attention utility λ, again, can again result in three regions; however, we further restrict parameters so that $\bar{v}_3 - \underline{v}_3 < \eta_2(v - \bar{v}_3)$. Then, for low values of λ (below the value for which* (10) *holds with equality), the DM chooses action $a_1 = a_{1,d}$ in period 1 and $a_2 = a^*$ in period 2; for intermediate values of λ (up to the value for which* (11) *holds with equality is possible since $\bar{v}_3 - \underline{v}_3 < \eta_2(v - \bar{v}_3)$), the DM chooses action $a_1 = a^*$ in period 1; for high values of λ (the remaining values), in no periods does the DM choose action a^*.*

In the specified environment, the sum of consumption payoffs is monotonically decreases in the weight on attention utility. However, suppose that the consumption payoff in period 3 is slightly larger when the DM takes action a^* in period 1 as opposed to period 2, i.e., replace $\bar{v}_3$ in condition (11) *by $\bar{v}_3 + \epsilon$. Take any weights on attention utility λ and λ' with $\lambda' > \lambda$ such that: (1) given λ, both* (10) *and* (11) *hold strictly; and 2) given λ',* (10) *does not hold and* (11) *holds, both strictly. For $\epsilon > 0$ small enough, whether the conditions hold or not in the respective cases is unchanged. Additionally, consumption payoff is now increasing in the weight on attention utility.*

G Proofs

Proof of Lemma 1A. Take any γ_c', γ_c with $\gamma_c' > \gamma_c$ and β_c. Let (x, α) and (α', a') denote a solution given γ_c and γ_c', respectively. Optimality of (x, α) and (α', a') implies

$$\underbrace{\sum_{c' \in \mathcal{C} \setminus \{c\}} (1 + \lambda \alpha_{c'}) V_{c'}(x)}_{:= \kappa_0} + (1 + \lambda \alpha_c)(\beta_c \tilde{V}_c(x) + \gamma_c) \geq \underbrace{\sum_{c' \in \mathcal{C} \setminus \{c\}} (1 + \lambda \alpha'_{c'}) V_{c'}(a')}_{:= \kappa_1} + (1 + \lambda \alpha'_c)(\beta_c \tilde{V}_c(a') + \gamma_c)$$

$$\underbrace{\sum_{c' \in \mathcal{C} \setminus \{c\}} (1 + \lambda \alpha'_{c'}) V_{c'}(a')}_{= \kappa_1} + (1 + \lambda \alpha'_c)(\beta_c \tilde{V}_c(a') + \gamma'_c) \geq \underbrace{\sum_{c' \in \mathcal{C} \setminus \{c\}} (1 + \lambda \alpha_{c'}) V_{c'}(x)}_{= \kappa_0} + (1 + \lambda \alpha_c)(\beta_c \tilde{V}_c(x) + \gamma'_c).$$

Combining the above, gives

$$\begin{aligned} &- \Big((1 + \lambda \alpha_c)(\beta_c \tilde{V}_c(x) + \gamma'_c) - (1 + \lambda \alpha'_c)(\beta_c \tilde{V}_c(a') + \gamma'_c) \Big) \\ \geq &\kappa_0 - \kappa_1 \\ \geq &- \Big((1 + \lambda \alpha_c)(\beta_c \tilde{V}_c(x) + \gamma_c) - (1 + \lambda \alpha'_c)(\beta_c \tilde{V}_c(a') + \gamma_c) \Big) . \end{aligned}$$

The outer inequality implies

$$-\lambda(\alpha_c - \alpha'_c)(\gamma'_c - \gamma_c) \geq 0,$$

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

If the environment is separable, then $\tilde{V}_c$ is increasing in the amount of attention α_c devoted to problem c, and the result follows.

Take any $\beta_c, \beta'_c \geq 0$ with $\beta'_c > \beta_c$ and γ_c and suppose that $\max_{(x,\alpha) \in \Gamma^*(\gamma_c, \beta_c)} V_c(x) = \min_{(a,\alpha) \in \Gamma(\gamma_c, \beta_c)} V_c(x)$. Let $\gamma'_c = \gamma_c - (\beta'_c - \beta_c) \tilde{V}_c(x)$, where $(x, \alpha) \in \Gamma(\gamma_c, \beta_c)$. Let (x, α) and (α', a') denote a solution given

(β_c, γ_c) and (β'_c, γ'_c), respectively. Optimality of (x, α) and (α', a') implies

$$\underbrace{\sum_{c' \in \mathcal{C}\setminus\{c\}} (1+\lambda\alpha_{c'})V_{c'}(x)}_{:=\kappa_2} + (1+\lambda\alpha_c)(\beta_c\tilde{V}_c(x)+\gamma_c) \geq \underbrace{\sum_{c' \in \mathcal{C}\setminus\{c\}} (1+\lambda\alpha'_{c'})V_{c'}(a')}_{:=\kappa_3} + (1+\lambda\alpha'_c)(\beta_c\tilde{V}_c(a')+\gamma_c)$$

$$\underbrace{\sum_{c' \in \mathcal{C}\setminus\{c\}} (1+\lambda\alpha'_{c'})V_{c'}(a')}_{=\kappa_3} + (1+\lambda\alpha'_c)(\beta'_c\tilde{V}_c(a')+\gamma'_c) \geq \underbrace{\sum_{c' \in \mathcal{C}\setminus\{c\}} (1+\lambda\alpha_{c'})V_{c'}(x)}_{=\kappa_2} + (1+\lambda\alpha_c)(\beta'_c\tilde{V}_c(x)+\gamma'_c).$$

Combining the above and substituting for γ'_c gives

$$\begin{aligned}
&-\Big((1+\lambda\alpha_c)(\beta_c\tilde{V}_c(x)+\gamma_c) - (1+\lambda\alpha'_c)(\beta'_c\tilde{V}_c(a') - (\beta'_c-\beta_c)\tilde{V}_c(x))\Big)\\
\geq&\kappa_2 - \kappa_3\\
\geq&-\Big((1+\lambda\alpha_c)(\beta_c\tilde{V}_c(x)+\gamma_c) - (1+\lambda\alpha'_c)(\beta_c\tilde{V}_c(a')+\gamma_c)\Big).
\end{aligned}$$

The outer inequality implies

$$-(1+\lambda\alpha'_c)(\tilde{V}_c(x)-\tilde{V}_c(a'))(\beta'_c-\beta_c) \geq 0,$$

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

If the environment is separable, then $\tilde{V}_c$ is increasing in the amount of attention α_c devoted to problem c, and the result follows. □

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

$$\begin{aligned}
\sum_{c\in\mathcal{C}} V_c(x) + \lambda\sum_{c\in\mathcal{C}}\alpha_c V_c(x) &\geq \sum_{c\in\mathcal{C}} V_c(a') + \lambda\sum_{c\in\mathcal{C}}\alpha'_c V_c(a'), \quad \text{and}\\
\sum_{c\in\mathcal{C}} V_c(a') + \lambda'\sum_{c\in\mathcal{C}}\alpha'_c V_c(a') &\geq \sum_{c\in\mathcal{C}} V_c(a') + \lambda'\sum_{c\in\mathcal{C}}\alpha'_c V_c(a').
\end{aligned}$$

Combining the above, gives

$$-\lambda'\left(\sum_{c\in\mathcal{C}}\alpha_c V_c(x) - \sum_{c\in\mathcal{C}}\alpha'_c V_c(a')\right) \geq \sum_{c\in\mathcal{C}} V_c(x) - \sum_{c\in\mathcal{C}} V_c(a') \geq -\lambda\left(\sum_{c\in\mathcal{C}}\alpha_c V_c(x) - \sum_{c\in\mathcal{C}}\alpha'_c V_c(a')\right).$$

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 claim follows. $\square$

Proof Lemma 3A. Take two sets of payoff levels, (γ_c) and (γ'_c), and scalar $\chi \in [0,1]$. Then

$$\begin{aligned}
&\max_{\alpha, x \in X(\alpha)} \sum_{c \in \mathcal{C}} (1+\lambda\alpha_c)(\beta_c \tilde{V}_c(x) + \chi\gamma_c + (1-\chi)\gamma'_c) \\
=& \max_{\alpha, x \in X(\alpha)} \left(\chi \sum_{c \in \mathcal{C}} (1+\lambda\alpha_c)(\beta_c \tilde{V}_c(x) + \gamma_c) + (1-\chi) \sum_{c \in \mathcal{C}} (1+\lambda\alpha_c)(\beta_c \tilde{V}_c(x) + \gamma'_c) \right) \\
&\geq \chi \max_{\alpha, x \in X(\alpha)} \sum_{c \in \mathcal{C}} (1+\lambda\alpha_c)(\beta_c \tilde{V}_c(x) + \gamma_c) + (1-\chi) \max_{\alpha, x \in X(\alpha)} \sum_{c \in \mathcal{C}} (1+\lambda\alpha_c)(\beta_c \tilde{V}_c(x) + \gamma'_c).
\end{aligned}$$

$\square$

Proof of Lemma 4A. First, note that (1) convex in α is equivalent to $(1+\lambda\alpha_c)\hat{V}_c(\alpha_c)$ convex in α_c for all $c \in \mathcal{C}$, where $\hat{V}_c(\alpha_c) := \max_{x_c \in X_c(\alpha_c)} V_c(x_c)$. Note that as $X_c(\alpha_c)$ is increasing in α_c, $\hat{V}_c$ is increasing. Thus, it suffices to show that if $(1+\lambda\alpha_c)\hat{V}_c(\alpha_c)$ convex in α_c, then so is $(1+\lambda'\alpha_c)\hat{V}_c(\alpha_c)$ for $\lambda' > \lambda$; to simplify notation, consider an increasing function f and suppose $(1+\lambda a)f(a)$ is convex in a.

Take any $\chi \in [0,1]$ and a, a' with $a' < a$. Suppose that $\chi f(a) + (1-\chi)f(a') \geq f(\chi a + (1-\chi)a')$. Then $\chi a f(a) + (1-\chi)a' f(a') \geq (\chi a + (1-\chi)a')f(\chi a + (1-\chi)a')$. Formally,

$$\begin{aligned}
&\chi a f(a) + (1-\chi)a' f(a') \\
=& a(\chi f(x) + (1-\chi)af(a')) + (a'-a)(1-\chi)f(a') \\
\geq& a f(\chi a + (1-\chi)a') + (a'-a)(1-\chi)f(a') \\
=& \chi a f(\chi a + (1-\chi)a') + (1-\chi)(af(\chi a + (1-\chi)a') + (a'-a)f(a')) \\
\geq& \chi a f(\chi a + (1-\chi)a') + (1-\chi)(af(\chi a + (1-\chi)a') + (a'-a)f(\chi a + (1-\chi)a')) \\
=& \chi a f(\chi a + (1-\chi)a') + (1-\chi)(a' f(\chi a + (1-\chi)a')),
\end{aligned}$$

where the first inequality follows by assumption, and the second as f is increasing. Thus, since $(1+\lambda a)f(a)$ is a linear combination of $f(a)$ and $af(a)$, if $\chi(1+\lambda a)f(a) + (1-\chi)(1+\lambda a')f(a') \geq (1+\lambda(\chi a + (1-\chi)a')f(\chi a + (1-\chi)a')$, then it must be that that $\chi a f(a) + (1-\chi)a' f(a') \geq \chi a f(\chi a + (1-\chi)a') + (1-\chi)(a' f(\chi a + (1-\chi)a'))$. But then increasing λ to λ' only loosens the initial inequality and the result follows. $\square$

Proof of Lemma 1B. Take any γ'_s, γ_s with $\gamma'_s > \gamma_s$ and β_s, p_s. Let (x, α) and (α', a') denote a solution given γ_s and γ'_s, respectively. Optimality of (x, α) and (α', a) implies

$$\underbrace{\sum_{s' \in \mathcal{S} \setminus \{s\}} (p_{s'} + \lambda \alpha_{s'}) V_{s'}(x)}_{:=\kappa_0} + (p_s + \lambda \alpha_s)(\beta_s \tilde{V}_s(x) + \gamma_s) \geq \underbrace{\sum_{s' \in \mathcal{S} \setminus \{s\}} (p_{s'} + \lambda \alpha'_{s'}) V_{s'}(a')}_{:=\kappa_1} + (p_s + \lambda \alpha'_s)(\beta_s \tilde{V}_s(a') + \gamma_s)$$

$$\underbrace{\sum_{s' \in \mathcal{S} \setminus \{s\}} (p_{s'} + \lambda \alpha'_{s'}) V_{s'}(a')}_{=\kappa_1} + (p_s + \lambda \alpha'_s)(\beta_s \tilde{V}_s(a') + \gamma'_s) \geq \underbrace{\sum_{s' \in \mathcal{S} \setminus \{s\}} (p_{s'} + \lambda \alpha_{s'}) V_{s'}(x)}_{=\kappa_0} + (p_s + \lambda \alpha_s)(\beta_s \tilde{V}_s(x) + \gamma'_s).$$

Combining the above, gives

$$\begin{aligned} & - \Big((p_s + \lambda \alpha_s)(\beta_s \tilde{V}_s(x) + \gamma'_s) - (p_s + \lambda \alpha'_s)(\beta_s \tilde{V}_s(a') + \gamma'_s) \Big) \\ \geq & \kappa_0 - \kappa_1 \\ \geq & - \Big((p_s + \lambda \alpha_s)(\beta_s \tilde{V}_s(x) + \gamma_s) - (p_s + \lambda \alpha'_s)(\beta_s \tilde{V}_s(a') + \gamma_s) \Big). \end{aligned}$$

The outer inequality implies

$$-\lambda(\alpha_s - \alpha'_s)(\gamma'_s - \gamma_s) \geq 0,$$

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

If the environment is separable, then $\tilde{V}_s$ is increasing in the amount of attention α_s devoted to state s, and the result follows.

Suppose $p_s > 0$ and take any $\beta_s, \beta'_s \geq 0$ with $\beta'_s > \beta_s$ and γ_s and suppose that $\max_{a \in a^*(\gamma_s, \beta_s, p_s)} V_s(x) = \min_{a \in a^*(\gamma_s, \beta_s, p_s)} V_s(x)$. Let $\gamma'_s = \gamma_s - (\beta'_s - \beta_s)\tilde{V}_s(x)$, where $a \in \Gamma_1(\gamma_s, \beta_s, p_s)$. Let (x, α) and (α', a') denote a solution given (β_s, γ_s, p_s) and $(\beta'_s, \gamma'_s, p'_s)$, respectively. Optimality of (x, α) and (α', a) implies

$$\underbrace{\sum_{s' \in \mathcal{S} \setminus \{s\}} (p_{s'} + \lambda \alpha_{s'}) V_{s'}(x)}_{:=\kappa_2} + (p_s + \lambda \alpha_s)(\beta_s \tilde{V}_s(x) + \gamma_s) \geq \underbrace{\sum_{s' \in \mathcal{S} \setminus \{s\}} (p_{s'} + \lambda \alpha'_{s'}) V_{s'}(a')}_{:=\kappa_3} + (p_s + \lambda \alpha'_s)(\beta_s \tilde{V}_s(a') + \gamma_s)$$

$$\underbrace{\sum_{s' \in \mathcal{S} \setminus \{s\}} (p_{s'} + \lambda \alpha'_{s'}) V_{s'}(a')}_{=\kappa_3} + (p_s + \lambda \alpha'_s)(\beta'_s \tilde{V}_s(a') + \gamma'_s) \geq \underbrace{\sum_{s' \in \mathcal{S} \setminus \{s\}} (p_{s'} + \lambda \alpha_{s'}) V_{s'}(x)}_{=\kappa_2} + (p_s + \lambda \alpha_s)(\beta'_s \tilde{V}_s(x) + \gamma'_s).$$

Combining the above and substituting for γ_s' gives

$$\begin{aligned}
&-\Big((p_s+\lambda\alpha_s)(\beta_s\tilde{V}_s(x)+\gamma_s)-(p_s+\lambda\alpha_s')(\beta_s'\tilde{V}_s(a')-(\beta_s'-\beta_s)\tilde{V}_s(x))\Big)\\
\geq&\kappa_2-\kappa_3\\
\geq&-\Big((p_s+\lambda\alpha_s)(\beta_s\tilde{V}_s(x)+\gamma_s)-(p_s+\lambda\alpha_s')(\beta_s\tilde{V}_s(a')+\gamma_s)\Big).
\end{aligned}$$

The outer inequality implies

$$-(p_s+\lambda\alpha_s')(\tilde{V}_s(x)-\tilde{V}_s(a'))(\beta_s'-\beta_s)\geq 0,$$

and thus, as $p_s>0$, it must be that $\tilde{V}_s(a')\geq\tilde{V}_s(x)$.

If the environment is separable, then $\tilde{V}_s$ is increasing in the amount of attention α_s devoted to state s, and the result follows.

Take any p_s',p_s with $p_s'>p_s$ and $\beta_s,\geq 0,\gamma_s$. Let (x,α) and (α',a') denote a solution given p_s and p_s', respectively. Optimality of (x,α) and (α',a') implies

$$\begin{aligned}
&\underbrace{\sum_{s'\in\mathcal{S}\setminus\{s,s_0\}}(p_{s'}+\lambda\alpha_{s'})V_{s'}(x)}_{:=\kappa_2}+(p_{s_0}+\lambda\alpha_{s_0})V_{s_0}(x)+(p_s+\lambda\alpha_s)V_s(x)\\
\geq&\underbrace{\sum_{s'\in\mathcal{S}\setminus\{s,s_0\}}(p_{s'}+\lambda\alpha_{s'}')V_{s'}(a')}_{:=\kappa_3}+(p_{s_0}+\lambda\alpha_{s_0}')V_{s_0}(a')+(p_s+\lambda\alpha_s')V_s(a'),\quad\text{and}\\
&\underbrace{\sum_{s'\in\mathcal{S}\setminus\{s,s_0\}}(p_{s'}+\lambda\alpha_{s'}')V_{s'}(a')}_{=\kappa_3}+(p_{s_0}'+\lambda\alpha_{s_0}')V_{s_0}(a')+(p_s'+\lambda\alpha_s')V_s(a')\\
\geq&\underbrace{\sum_{s'\in\mathcal{S}\setminus\{s,s_0\}}(p_{s'}+\lambda\alpha_{s'})V_{s'}(x)}_{=\kappa_2}+(p_{s_0}'+\lambda\alpha_{s_0})V_{s_0}(x)+(p_s'+\lambda\alpha_s)V_s(x).
\end{aligned}$$

Combining the above gives

$$\begin{aligned}
&-\big((p_{s_0}'+\lambda\alpha_{s_0})V_{s_0}(x)+(p_s'+\lambda\alpha_s)V_s(x)-\big((p_{s_0}'+\lambda\alpha_{s_0}')V_{s_0}(a')+(p_s'+\lambda\alpha_s')V_s(a')\big)\big)\\
\geq&\kappa_2-\kappa_3\\
\geq&-\big((p_{s_0}+\lambda\alpha_{s_0})V_{s_0}(x)+(p_s+\lambda\alpha_s)V_s(x)-\big((p_{s_0}+\lambda\alpha_{s_0}')V_{s_0}(a')+(p_s+\lambda\alpha_s')V_s(a')\big)\big).
\end{aligned}$$

Since $V_{s_0}(x) = V_{s_0}(a')$, the outer inequality implies

$$-(p'_s - p_s)(V_s(x) - V_s(a') \geq 0,$$

and thus it must be that $V_s(a') \geq V_s(x)$.

If the environment is separable, then V_s is increasing in the amount of attention α_s devoted to state s, and the result follows. □

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

$$\begin{aligned} \sum_{s\in\mathcal{S}} p_s V_s(x) + \lambda \sum_{s\in\mathcal{S}} \alpha_s V_s(x) &\geq \sum_{s\in\mathcal{S}} p_s V_s(a') + \lambda \sum_{s\in\mathcal{S}} \alpha'_s V_s(a'), \quad \text{and} \\ \sum_{s\in\mathcal{S}} p_s V_s(a') + \lambda' \sum_{s\in\mathcal{S}} \alpha'_s V_s(a') &\geq \sum_{s\in\mathcal{S}} p_s V_s(a') + \lambda' \sum_{s\in\mathcal{S}} \alpha'_s V_s(a'). \end{aligned}$$

Combining the above, gives

$$-\lambda' \left(\sum_{s\in\mathcal{S}} \alpha_s V_s(x) - \sum_{s\in\mathcal{S}} \alpha'_s V_s(a') \right) \geq \sum_{s\in\mathcal{S}} p_s V_s(x) - \sum_{s\in\mathcal{S}} p_s V_s(a') \geq -\lambda \left(\sum_{s\in\mathcal{S}} \alpha_s V_s(x) - \sum_{s\in\mathcal{S}} \alpha'_s V_s(a') \right).$$

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 claim follows. □

Proof of Lemma 3B. Take two sets of payoff levels, (γ_s) and (γ'_s), and scalar $\chi \in [0, 1]$. Then

$$\begin{aligned} &\max_{x,\alpha\in X(\alpha)} \sum_{s\in\mathcal{S}} (p_1 + \lambda\alpha_s)(\beta_s \tilde{V}_s(x) + \chi\gamma_s + (1-\chi)\gamma'_s) \\ &= \max_{x,\alpha\in X(\alpha)} \left(\chi \sum_{s\in\mathcal{S}} (p_s + \lambda\alpha_s)(\beta_s \tilde{V}_s(x) + \gamma_s) + (1-\chi) \sum_{s\in\mathcal{S}} (p_s + \lambda\alpha_s)(\beta_s \tilde{V}_s(x) + \gamma'_s) \right) \\ &\geq \chi \max_{x,\alpha\in X(\alpha)} \sum_{s\in\mathcal{S}} (p_s + \lambda\alpha_s)(\beta_s \tilde{V}_s(x) + \gamma_s) + (1-\chi) \max_{x,\alpha\in X(\alpha)} \sum_{s\in\mathcal{S}} (p_s + \lambda\alpha_s)(\beta_s \tilde{V}_s(x) + \gamma'_s). \end{aligned}$$

□

Proof of Lemma 4B. First, note that (2) convex in α is equivalent to $(p_s + \lambda\alpha_s)\hat{V}_s(\alpha_s)$ convex in α_s for all $s \in \mathcal{S}$, where $\hat{V}_s(\alpha_s) := \max_{x_s \in X_s(\alpha_s)} V_s(a_s)$. Note that as $X_s(\alpha_s)$ is increasing in α_s, $\hat{V}_s$ is increasing. Thus, it suffices to show that if $(p_s + \lambda\alpha_s)\hat{V}_s(\alpha_s)$ convex in α_s, then so is

$(p_s + \lambda' \alpha_s)\hat{V}_s(\alpha_s)$ for $\lambda' > \lambda$; to simplify notation, consider an increasing function f and $p \in [0,1]$ and suppose $(p + \lambda a)f(a)$ is convex in a.

Take any $\chi \in [0,1]$ and a, a' with $a' < a$. Suppose that $\chi f(a) + (1-\chi)f(a') \geq f(\chi a + (1-\chi)a')$. Then $\chi a f(a) + (1-\chi)a' f(a') \geq (\chi a + (1-\chi)a')f(\chi a + (1-\chi)a')$. Formally,

$$\begin{aligned}
&\chi a f(a) + (1-\chi)a' f(a') \\
=&a(\chi f(x) + (1-\chi)af(a')) + (a'-a)(1-\chi)f(a') \\
\geq&af(\chi a + (1-\chi)a') + (a'-a)(1-\chi)f(a') \\
=&\chi a f(\chi a + (1-\chi)a') + (1-\chi)(af(\chi a + (1-\chi)a') + (a'-a)f(a')) \\
\geq&\chi a f(\chi a + (1-\chi)a') + (1-\chi)(af(\chi a + (1-\chi)a') + (a'-a)f(\chi a + (1-\chi)a')) \\
=&\chi a f(\chi a + (1-\chi)a') + (1-\chi)(a' f(\chi a + (1-\chi)a')),
\end{aligned}$$

where the first inequality follows by assumption, and the second as f is increasing. Thus, since $(p+\lambda a)f(a)$ is a linear combination of $f(a)$ and $af(a)$, if $\chi(p+\lambda a)f(a) + (1-\chi)(p+\lambda a')f(a') \geq (p + \lambda(\chi a + (1-\chi)a')f(\chi a + (1-\chi)a')$, then it must be that that $\chi a f(a) + (1-\chi)a' f(a') \geq \chi a f(\chi a + (1-\chi)a') + (1-\chi)(a' f(\chi a + (1-\chi)a'))$. But then increasing λ to λ' only loosens the initial inequality and the result follows. $\square$

Proof of Proposition 1. 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, $\bar{S} \coloneqq \operatorname{argmax}_{s \in \mathcal{S}} u(x_s)$. 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. $\square$

Proof of Proposition 2. 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. $\square$

Proof of Lemma 1C. Take any γ_1', γ_1 with $\gamma_1' > \gamma_1$ and β_1. Let (α_1, a_1) and (α_1', a_1') denote a solution (for the DM's choice in the first period) given γ_1 and γ_1', respectively. Let

$$\kappa_0 \coloneqq \sum_{t'=2}^{T}(1+\lambda\sum_{t''=2}^{t'}\alpha_{t''\to t'})V_{t'}(h_1, a_1, (a_\tau(h_\tau))_{t'\geq\tau>1}) \quad \text{and}$$
$$\kappa_1 \coloneqq \sum_{t'=2}^{T}(1+\lambda\sum_{t''=2}^{t'}\alpha'_{t''\to t'})V_{t'}(h_1, a_1', (a_\tau(h_\tau'))_{t'\geq\tau>1}).$$

Optimality of (α_1, a_1) and (α_1', a_1') implies

$$(1+\lambda\alpha_{1\to1})(\beta_1\tilde{V}_1(h_1,a_1)+\gamma_1)+\kappa_0 \geq (1+\lambda\alpha'_{1\to1})(\beta_1\tilde{V}_1(h_1,a_1')+\gamma_1)+\kappa_1, \quad \text{and}$$
$$(1+\lambda\alpha'_{1\to1})(\beta_1\tilde{V}_1(h_1,a_1')+\gamma_1')+\kappa_0 \geq (1+\lambda\alpha_{1\to1})(\beta_1\tilde{V}_1(h_1,a_1)+\gamma_1')+\kappa_1.$$

Combining the above gives

$$\begin{aligned}
&-\Big((1+\lambda\alpha_{1\to1})(\beta_1\tilde{V}_1(h_1,a_1)+\gamma_1') - (1+\lambda\alpha'_{1\to1})(\beta_1\tilde{V}_1(h_1,a_1')+\gamma_1')\Big)\\
\geq&\kappa_0-\kappa_1\\
\geq&-\Big((1+\lambda\alpha_{1\to1})(\beta_1\tilde{V}_1(h_1,a_1)+\gamma_1) - (1+\lambda\alpha'_{1\to1})(\beta_1\tilde{V}_1(h_1,a_1')+\gamma_1)\Big)
\end{aligned}$$

The outer inequality implies

$$-(\alpha_{1\to1}-\alpha'_{1\to1})(\gamma_1-\gamma_1)$$

and thus, it must be that $\alpha_{1\to1} \geq \alpha_{1\to1}$ as $\lambda > 0$.

If the environment is separable, then $\tilde{V}_1$ is increasing in the amount of attention $\alpha_{1\to1}$ devoted to period 1, and the result follows.

Take any $\beta_1, \beta_1' \geq 0$ with $\beta_1' > \beta_1$ and γ_1 and suppose that $\max_{(\alpha_1,a_1)\in\Gamma^*(\gamma_c,\beta_c)} V_1(x) = \min_{(a_1,\alpha_1)\in\Gamma_1(\gamma_1,\beta_1)} V_c(a_1)$. Let $\gamma_1^{'} = \gamma_1 - (\beta_1' - \beta_1)\tilde{V}_1(x)$, where $(\alpha_1, a_1) \in \Gamma(\gamma_1, \beta_1)$. Let (α_1, a_1) and (α_1', a_1') denote a solution given (β_1, γ_1) and (β_1', γ_1'), respectively. Optimality of (α_1, a_1) and (α_1', a_1') implies

$$(1+\lambda\alpha_{1\to1})(\beta_1\tilde{V}_1(h_1,a_1)+\gamma_1)+\kappa_0 \geq (1+\lambda\alpha_{1\to1}')(\beta_1\tilde{V}_1(h_1,a_1')+\gamma_1)+\kappa_1, \quad \text{and}$$
$$(1+\lambda\alpha_{1\to1}')(\beta_1'\tilde{V}_1(h_1,a_1')+\gamma_1')+\kappa_0 \geq (1+\lambda\alpha_{1\to1})(\beta_1'\tilde{V}_1(h_1,a_1)+\gamma_1')+\kappa_1.$$

Combining the above and substituting for γ_1' gives

$$\begin{aligned}
&-\Big((1+\lambda\alpha_{1\to1})(\beta_1\tilde{V}_1(h_1,a_1)+\gamma_1)-(1+\lambda\alpha_{1\to1}')(\beta_1'\tilde{V}_1(h_1,a_1')+\gamma_1')\Big)\\
\geq&\kappa_0-\kappa_1\\
\geq&-\Big((1+\lambda\alpha_{1\to1})(\beta_1\tilde{V}_1(h_1,a_1)+\gamma_1)-(1+\lambda\alpha_{1\to1}')(\beta_1\tilde{V}_1(h_1,a_1')+\gamma_1)\Big)
\end{aligned}$$

The outer inequality implies

$$-(1+\lambda\alpha_{1\to1}')(\tilde{V}_1(a_1)-\tilde{V}_1(a_1'))(\beta_1'-\beta_1) \geq 0,$$

and thus, it must be that $\tilde{V}_1(a_1') \geq \tilde{V}_1(a_1)$.

If the environment is separable, then $\tilde{V}_1$ is increasing in the amount of attention $\alpha_{1\to1}$ devoted to period 1, and the result follows. □

Proof of Lemma 3C. Take two payoff levels (γ_1) and (γ_1'), and scalar $\chi \in [0,1]$. Then

$$\begin{aligned}
&\max_{\alpha, x \in X(\alpha), (x,\alpha) \in \Gamma_2((x_1,\alpha_1))} (1+\lambda\alpha_{1\to 1})(\beta_1 \tilde{V}_t(x) + \chi\gamma_1 + (1-\chi)\gamma_1') + \sum_{t'>t}^{T}(1+\lambda \sum_{t''=1}^{T} \alpha_{t''\to t'})V_{t'}(x) \\
= &\max_{\alpha, x \in X(\alpha), (x,\alpha) \in \Gamma_2((x_1,\alpha_1))} \Bigg(\chi \Bigg((1+\lambda\alpha_{1\to 1})(\beta_1 \tilde{V}_t(x) + \gamma_1) + \sum_{t'>t}^{T}(1+\lambda \sum_{t''=1}^{T} \alpha_{t''\to t'})V_{t'}(x)\Bigg) \\
&+ (1-\chi)\Bigg((1+\lambda\alpha_{1\to 1})(\beta_1 \tilde{V}_t(x) + \gamma_1') + \sum_{t'>t}^{T}(1+\lambda \sum_{t''=1}^{T} \alpha_{t''\to t'})V_{t'}(x)\Bigg)\Bigg) \\
&\geq \chi \max_{\alpha, x \in X(\alpha), (x,\alpha) \in \Gamma_2((x_1,\alpha_1))} \Bigg((1+\lambda\alpha_{1\to 1})(\beta_1 \tilde{V}_t(x) + \gamma_1) + \sum_{t'>t}^{T}(1+\lambda \sum_{t''=1}^{T} \alpha_{t''\to t'})V_{t'}(x)\Bigg) \\
&+ (1-\chi) \max_{\alpha, x \in X(\alpha), (x,\alpha) \in \Gamma_2((x_1,\alpha_1))} \Bigg((1+\lambda\alpha_{1\to 1})(\beta_1 \tilde{V}_t(x) + \gamma_1') + \sum_{t'>t}^{T}(1+\lambda \sum_{t''=1}^{T} \alpha_{t''\to t'})V_{t'}(x)\Bigg).
\end{aligned}$$

□

Proof of Lemma 4C. First, note that $F(\lambda)$ is linear in $\alpha_{1\to -1}$. Thus, we need to show that it is convex in $\alpha_{1\to 1}$. Note that as $X_1(\alpha_1)$ is increasing in $\alpha_{1\to 1}$, $\hat{V}_1$ is increasing. Thus, it suffices to show that if $(1+\lambda\alpha_{1\to t})\hat{V}_t(\alpha_{1\to 1})$ convex in $1 \to 1$, then so is $(1+\lambda'\alpha_{1\to 1})\hat{V}_1(1\to 1)$ for $\lambda' > \lambda$; to simplify notation, consider an increasing function f and suppose $(1+\lambda a)f(a)$ is convex in a.

Take any $\chi \in [0,1]$ and a, a' with $a' < a$. Suppose that $\chi f(a) + (1-\chi)f(a') \geq f(\chi a + (1-\chi)a')$. Then $\chi a f(a) + (1-\chi)a' f(a') \geq (\chi a + (1-\chi)a')f(\chi a + (1-\chi)a')$. Formally,

$$\begin{aligned}
&\chi a f(a) + (1-\chi)a' f(a') \\
=&a(\chi f(x) + (1-\chi)a f(a')) + (a'-a)(1-\chi)f(a') \\
\geq&a f(\chi a + (1-\chi)a') + (a'-a)(1-\chi)f(a') \\
=&\chi a f(\chi a + (1-\chi)a') + (1-\chi)(a f(\chi a + (1-\chi)a') + (a'-a)f(a')) \\
\geq&\chi a f(\chi a + (1-\chi)a') + (1-\chi)(a f(\chi a + (1-\chi)a') + (a'-a)f(\chi a + (1-\chi)a')) \\
=&\chi a f(\chi a + (1-\chi)a') + (1-\chi)(a' f(\chi a + (1-\chi)a')),
\end{aligned}$$

where the first inequality follows by assumption, and the second as f is increasing. Thus, since $(1+\lambda a)f(a)$ is a linear combination of $f(a)$ and $af(a)$, if $\chi(1+\lambda a)f(a) + (1-\chi)(1+\lambda a')f(a') \geq (1+\lambda(\chi a + (1-\chi)a')f(\chi a + (1-\chi)a')$, then it must be that that $\chi a f(a) + (1-\chi)a' f(a') \geq \chi a f(\chi a + (1-\chi)a') + (1-\chi)(a' f(\chi a + (1-\chi)a'))$. But then increasing λ to λ' only loosens the initial inequality and the result follows. $\square$

Proof of Proposition 3. 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 (4) 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. $\square$

Proof of Proposition 4. Suppose $X_1(\alpha_1)$ is independent of α_1. Let $V^* := \max_{c\in\mathcal{C}} V_c(x_1, x_2(x_1, c))$. If $\bar{V} \geq V^*$, then the DM devotes all their attention in the first period to the trivial problem; if $\bar{V} < V^*$, then the DM devotes attention in the second period when the realized problem is the corresponding argmax. In either case, the DM is time consistent, and the first part of the proposition follows.

For the second part, fix x_1 and consider a realized c. Clearly, if $\max_{x_2\in X_2(\alpha_2)} V_c(\cdot, x_2) \geq \bar{V}$, solving (5) gives $\alpha_{2\to c} = 1$ (for any λ). If $\max_{x_2\in X_2(\alpha_2)} V_c(\cdot, x_2) < \bar{V}$, then (5) for $\alpha_{2\to c} = 0$ (and some finite $V_c(x_1, \cdot)$) is larger than (5) for any $\alpha_{2\to c} \geq \eta$ for $\max_{x_2\in X_2(\alpha_2)} V_c(\cdot, x_2)$ when λ is large enough. Since $\mathcal{C}$ is finite, taking the max λ implies the result. $\square$

Proof of Proposition 5. Consider $x_1 \notin \mathrm{argmax}_{x_1\in X_1(\cdot)} F(x_1)$. For lambda large enough, by Proposition 4, and since $\max_{\alpha_2, x_2\in X_2(\alpha_2)} \tilde{V}_c(\cdot, x_2) + \beta F(x_1) < \max_{\alpha_2, x_2\in X_2(\alpha_2)} \tilde{V}_c(\cdot, x_2) + \beta F(x_1^*)$, we have

$B(x_1) \subseteq B(x_1^*) \neq$ (where non-emptiness follows from $V_{c^*}(x_1^*, x_2(x_1^*, c^*)) > \bar{V}$) and hence, again for λ large, the DM's second-period payoff (i.e., (5)) 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 (5) 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^*. □

Proof of Proposition 6. 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 Lemma 3A.

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. $\square$

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 $\max_{x_2 \in [0, w + x_1 r_1]} V_\rho(x_1, x_2)$ and $V_\rho(x_1, \underline{x}_2)$ are finite, it must be that 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 $\max_{x_2 \in [0, w + x_1 r_1]} V_\rho(x_1, x_2) \geq V_\rho(x_1, \underline{x}_2)$, 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 $\tilde{r}\coloneqq r_1+r_2+r_1r_2$ (with r_2 deterministic given N).

In the second period, the DM's attention satisfies $\alpha_{2\to\rho}(\rho,x_1)=\operatorname{argmax}\{V_\rho(x_1,\bar{x}_2),\bar{V}\}$. Let $\rho^*=\operatorname{argmax}_\rho \tilde{r}$. 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

$$\lambda\left(\underbrace{(1-\eta_1)\overbrace{\max\{V_{\rho^*}(x_1,\underline{x}_2),\bar{V}\}}^{\coloneqq C}+\eta_1\overbrace{\sum_\rho p_\rho V_\rho(x_1,\underline{x}_2)}^{\coloneqq D}+\overbrace{\sum_\rho p_\rho\max\{V_\rho(x_1,\underline{x}_2),\bar{V}\}}^{\coloneqq E}}_{\coloneqq A}\right) + \underbrace{\sum_\rho p_\rho V_\rho(x_1,\underline{x}_2)}_{\coloneqq B}. \tag{12}$$

Also note that the DM is time consistent, in particular, the DM's attention in the second period maximizes (12) (given ρ,x_1).

We begin with the comparative static in λ. Suppose $x_1>0$ (otherwise, it cannot decrease when λ increases). Then (12) 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 (12), 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 $1-\eta_1$. $\square$

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 Lemma 1A, $\alpha_{1\to c_1}$ increases in γ_{c_1}.

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

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. □

Proof of Proposition 7. The first claim follows from Lemma 1A, 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. □

Proof of Proposition 8. 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 somplifications, 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 $\bar{\alpha}_C \geq \bar{\alpha}_{C'}$. □