Emotional Inattention

Lukas Bolte* Collin Raymond*

October 30, 2022

https://www.lukasbolte.com/papers/emotionalInattention.pdf

Abstract

We propose a framework where a decision-maker allocates attention across payoff-dimensions which can be different dimensions of consumption, realizations of an unknown state, or time periods. Attention has two features: (1) it is instrumentally valuable by allowing the decision-maker to take actions, and (2) it leads to an emotional response, which is proportional to the attention devoted to a dimension and the associated payoff. The framework provides a unifying explanation for a number of behavioral phenomena. We discuss implications for policy interventions designed to increase overall utility or improve decisions.

JEL Classification Codes: D81, D83

Keywords: Attention, attention utility, information, adaptive preferences

*Lukas Bolte; Department of Economics, Stanford University, Stanford, CA; lbolte@stanford.edu. Collin Raymond; SC Johnson Graduate School of Management, Cornell University, Ithaca, NY; collinbraymond@gmail.com. We thank Douglas Bernheim, Gabriel Carroll, Matthew Jackson, and Muriel Niederle, and seminar participants from Stanford University, BEAM, BRIC, M-BEES/M-BEPS, and SABE for helpful comments. Bolte gratefully acknowledges financial support from the Leonard W. Ely and Shirley R. Ely Graduate Student Fellowship.

1 Introduction

Over the last decades, economists and psychologists have documented many important deviations from the predictions of the standard economic model. For instance, rather than optimizing in the good times and the bad, individuals have been shown to ignore their investment portfolio when the market is down ("ostrich effect," Karlsson et al. (2009)); rather than assessing the likelihood of outcomes objectively, they often down-weigh the possibility of a bad outcome ("optimism," Brunnermeier and Parker (2005); Sharot (2011)); and rather than smoothing consumption, they choose to have lavish weddings and luxurious vacations ("memorable consumption," Hai et al. (2020)). A variety of theories have been developed to explain each of these and many other behavioral anomalies. However, it is often the case that each anomaly is given its own model and distinct source of deviation from the classic predictions. While useful for their respective applications, the multiplicity of theories poses modeling challenges in new environments where there is little guidance as to which theory is most applicable. Indeed, Fudenberg (2006) suggests: "For [behavioral economics] to advance further, it should devote more attention to the foundations of its models, and develop unified explanations for a wider range of phenomena."

This project develops a unifying theory of various behavioral phenomena, including the three previously mentioned. We do so by positing the role of *attention* as a single mechanism underlying them. Individuals choose how to allocate their attention, selecting a subset of things (e.g., items, information, and ideas) to attend to. They do so, taking into consideration two important consequences of attention. First, it improves decision-making and leads to higher material payoffs. For instance, it allows individuals to reason, process information, and (consciously) take actions. This instrumental role of attention is well recognized and studied by both psychologists (since at least James (1890); see Desimone et al. (1995) for a review) and economists (e.g., Sims (2003) and the ensuing rational inattention literature).

Attention, however, has a second consequence, one that is typically absent from economic models: it generates and regulates emotions (see Dixon et al. (2017) and Gross (1998) for reviews).[1] For instance, attending to a news article about the terrors

[1]That is not to say it has been completely ignored by economists. For example, Schelling (1988) highlights the role of the "mind" as a "pleasure machine or consuming organ, the generator of direct consumer satisfaction," in addition to the role of the "information processing and reasoning machine." Schelling also implicitly suggests that these roles should be considered jointly and writes:

occurring in the world may lead to a negative visceral reaction.

It is this second consequence of attention—the emotional responses it induces—that, once added to the standard model, can explain the previously mentioned behavioral anomalies. The ostrich effect arises because, although ignoring one's investment portfolio may lead to a worse investment performance, it avoids the negative emotional response from thinking about one's poor finances. Optimism occurs because, although inattention to bad possible outcomes may lead to suboptimal decisions, it also leads to more positive anticipatory emotions. And "memorable consumption," e.g., a lavish wedding or luxurious vacation, can be optimal because, although consumption will be low during other times, attention to the memories created (and anticipation of the event) leads to positive emotions long before and after the event takes place.

We operationalize the two consequences of attention as follows. The decision-maker (henceforth, DM—they) devotes attention across a number of dimensions. These dimensions can correspond to different consumption dimensions but also to realizations of an unknown state or different time periods. Each is associated with what we call a "material utility" or just "payoff" for brevity which corresponds to the standard (non-emotional) utility from actual consumption, e.g., the actual enjoyment experienced while on vacation. The attention allocation determines which actions are available to the DM and the action taken (which may be multi-dimensional) affects the material utility from each dimension. This formulation captures the instrumental role of attention in a reduced form. It nests models of rational inattention (Sims, 2003) but also cases where executing an action requires attention. For example, the DM may not be able to take the action 'reoptimize investment portfolio' without devoting attention to their finances, and taking this action improves the material utility.

The DM values the sum or expectation (depending on the context) of the material utilities and what we call "attention utility." Attention utility from a dimension is proportional to both the dimension's payoff and the amount of attention devoted to it and captures attention's emotional role. For example, when the DM devotes attention to an upcoming vacation, they receive some additional utility, say, because of feelings

"Marvelous it is that the mind does all these things. Awkward it is that it seems to be the same mind from which we expect both the richest sensations and the most austere analyses." Our analysis of the interaction of these roles seeks to alleviate this "awkwardness."

of anticipation. In spirit, this approach is similar to models of anticipatory utility (e.g., Loewenstein (1987); Caplin and Leahy (2001)), which assume that agents derive flow utility as a function of beliefs about future payoffs. Our innovation is to make this flow utility a function of the attention paid to the future payoffs. Notice that the total weight a dimension (and its material utility) takes in the DM's objective is determined by the attention allocation.

After formally introducing this framework in Section 2.1, we derive general properties of the optimal attention allocation in Section 2.2. These properties apply regardless of whether attention is allocated across the different dimensions of consumption, realizations of an unknown state, or time periods. Our modeling approach is standard with a single twist: the DM maximizes a combination of their material utility (standard) and their attention utility (twist). Hence, the DM responds to many changes in the environment in a way similar to a standard agent without attention utility: increasing the instrumental value of attention for a dimension increases the attention devoted to it. However, unlike in the standard model, a key determinant of attention is the levels of material utility across dimensions: ceteris paribus, the DM devotes more attention to dimensions with higher material utility. The DM may thus ignore a low-payoff dimension, even though attending to it would increase their material utility, while they may devote excessive attention beyond the point where it is instrumentally valuable to dimensions with higher material utilities.

In the context of attention across consumption dimensions (Section 2.3), the aforementioned attentional patterns lead to the well-documented ostrich phenomenon mentioned in the introductory paragraphs: individuals tend to be inattentive to (and possibly avoid information about) consumption dimensions with low payoffs; e.g., they ignore their investment portfolio when the market is down (Karlsson et al., 2009). Such behavior has also been noted in medical decision-making (e.g., Becker and Mainman (1975) and Oster et al. (2013)), as well as in the lab (Avoyan and Schotter, 2020). We show that this evidence can be organized through attention's emotional role.

Section 2.2 also shows that attention utility introduces a preference for varied payoffs across dimensions. This is because the DM can exploit such variation and increase their attention utility by focusing on high-payoff dimensions. For instance, temporally varying payoff streams, such as a single luxurious vacation and low consumption during other times, can be exploited by devoting attention to the high payoffs before and after they occur. Since the payoffs themselves are driven by attention, the DM's objective is more likely to be convex once attention's emotional role is incorporated. Allocating attention in the face of a convex objective naturally generates "sparse" attention allocations by our agent, as in Gabaix (2014).

In Section 2.4, we let different dimensions correspond to different states of the world. The DM values the expected material utility, and attention utility from a state is still proportional to that state's payoff and the attention devoted to it. The attention-dependent weight of a state leads to an as-if subjective probability of that state occurring, even though our DM understands the true probabilities perfectly: Fixing the attention allocation, our DM chooses the same action as a standard DM—who only maximizes material utility—but who has distorted beliefs. When attention is non-instrumental, the DM devotes all their attention to high-payoff states while ignoring the ones with low payoffs, which leads to optimism (Brunnermeier and Parker, 2005; Sharot, 2011). But the DM's optimism is not universal and can be mitigated or even turned into pessimism when attention is instrumentally valuable and it is the low-payoff states that receive most attention. Our model also rationalizes individuals' preferences for positively-skewed payoffs. For instance, the DM buys a lotto ticket to devote attention to the state where they win the jackpot, e.g., via day-dreaming.[2]

Attention's emotional role can also lead to more general forms probability weighting, as in (Kahneman and Tversky, 1979), where the DM distorts objective probabilities when calculating expected payoffs. A typical finding is that low probabilities are over- and high ones under-weighted. In our model, such a weighting function occurs when the DM devotes a disproportionate amount of attention to unlikely states relative to the probability with which they occur. For instance, this occurs when the DM is forced to devote at least some minimum amount of attention to each state or when the material gains of doing so are large. We thus add to the extensive literature on probability weighting by offering attention as a novel mechanism that predicts how the economic environment affects the shape of the weighting function.

In Section 3, we extend our model to dynamic settings in which the DM makes a choice about contemporaneous attention in multiple periods. Doing so allows us to explore behavioral phenomena with dynamic aspects. Since the DM makes decisions

[2]For empirical evidence documenting such preferences, see, e.g., Blume and Friend (1975) in the context of portfolio choice, or Garrett and Sobel (1999); Forrest et al. (2002) in the context of individuals playing lotto.

at various points in time, the DM's behavior is the equilibrium of an intrapersonal game played between different selves. In Section 3.1, the dimensions across which the DM allocates attention represent time periods, i.e., there is only inter-period (and no intra-period) attention allocation. Our model leads to endogenous weights on time periods, i.e., preferences over the timing of consumption. For instance, the DM may as-if discount future periods (i.e., be present-focused) if the payoff in the present is particularly high or attention to it is of high instrumental value. Conversely, high attention to the future manifests as seemingly negative discounting. Our model rationalizes so-called memorable consumption periods, like the lavish wedding or luxurious vacation previously mentioned (Gilboa et al., 2016; Hai et al., 2020). When attention utility is important enough, the consumption path is non-smooth in equilibrium, allowing the DM to devote attention to high-payoff periods long before or after they happen.

Attention's emotional role predicts default effects, which we explore next in Section 3.2. Defaults are ubiquitous: many individuals have a default monthly contribution to their savings, today's investment portfolio is the default for tomorrow's modulo investment returns, and heuristics are defaults in unknown situations. By definition, the default action is implemented without devoting attention, while deviating from it may necessitate some. Because of this attention requirement, the DM may not change the default when the associated payoff is low, as in the static model, leading to asymmetric default effects. In the portfolio investment example, the default portfolio binds when the market is down and the payoff is low. Our model thus suggests that defaults matter most for the eventual choice when the associated payoff is low. For instance, defaults for end-of-life medical treatments should matter, whereas they should matter less for planning a vacation. When the DM chooses the default themselves, anticipating their future self's inattention, they may opt for a default that performs well when it binds, i.e., in low-payoff situations.

Dynamically extending the model also allows us to study preferences over multi-period information acquisition (Section 3.3). Even without any instrumental value of information, we show that the DM has strict preferences over the resolution of uncertainty: they 1) acquire more information when the expected payoff is high, 2) have a preference for early information acquisition to condition their future attention allocation, and 3) prefer positively-skewed (negatively-skewed) information if their prior is low (high), again, to condition their future attention allocation. These predictions, broadly speaking, rationalize laboratory evidence (e.g., CHEW and HO (1994); see Nielsen (2020) for a recent contribution).

We study the policy implications of our model in Section 4. Our focus is on the novel implications of our model for some typical interventions: in turn, we revisit (simple) models of taxation, dividing tasks between the government and the DM, and effort provision. First, since the DM derives attention utility from a dimension in proportion to the attention devoted to it, consumption dimensions that receive little attention should be taxed more heavily. Second, the government, relative to the DM, should devote resources to improve payoffs in low-payoff dimensions. This is because the government does not incur the emotional cost from attending to those dimensions, whereas the DM does. And third, policies that penalize bad outcomes in order to motivate effort can backfire; the DM may shy away from a problem if the possibility of a penalty reduces the expected payoff. An implication is that negative commitment devices, i.e., those that penalize deviations from the committed-to-action, may not only be ineffective but, in fact, counterproductive.

Last, we study the optimal way to group payoffs into different dimensions. Instead of viewing dimensions i and i' as distinct, the DM may bundle them together as one larger dimension, where the larger dimension simply pays off the sum of the payoffs from i and i'. For example, the DM may be able to think of their weekend chores and Saturday evening plans as separate dimensions or jointly. Given a large set of primitive dimensions, we characterize the optimal bracketing, providing a partial microfoundation for one of the key premises of our model—the set of dimensions.

Section 5 discusses how our model relates to several other literatures. We consider, in turn, models of rational inattention, Bayesian and non-Bayesian models with anticipatory utility, and other models of attention. Section 6 concludes with a discussion of some of the limitations of our approach and avenues for future research.

2 Model

We consider a decision-maker (henceforth, DM—they) who allocates attention and chooses an action. The ensuing model formally captures the two fundamental features of attention:

1. Instrumental role: attention determines which actions are available to the DM.

2. Emotional role: attention generates attention utility.

Below, we characterize the optimal attention allocation, focusing on the role of attention utility in particular, before exploring our model's implications in two simple decision problems: a deterministic problem with multiple consumption dimensions and a problem with an uncertain state (Sections 2.3 and 2.4, respectively).

2.1 Setup

The DM faces a finite number of dimensions indexed by $i \in \mathcal{D}$.[3] A dimension can correspond to a dimension of consumption, a realization of an unknown state or a time period, or a combination of these. The DM chooses an (action, attention)-pair denoted by (x, α). Dimension i leads to a material utility $V_i(x)$, where $V_i(\cdot)$ is a continuous function of the action chosen. Attention $\alpha = (\alpha_i)_{i \in \mathcal{D}}$ is a measure on the set of dimensions. In particular, α_i denotes the attention devoted to dimension i. We assume that $\alpha_i \geq 0$, and normalize the total attention devoted to have measure 1, i.e., $\sum_i \alpha_i = 1$.[4] As a notational convention, for any variable that is indexed by $i \in \mathcal{D}$, say b_i, we let $b_{-i} := (b_{i'})_{i' \in \mathcal{D} \setminus \{i\}}$.

Attention has two implications. First, attention is instrumentally valuable. To capture this, we assume that the set of available actions depends on the attention allocation α. Given α, action x is chosen from a set $X(\alpha)$, where $X(\cdot)$ is compact- and non-empty-valued and upper hemicontinuous. When $X(\alpha)$ is independent of α, attention has no instrumental value. This reduced-form formulation nests canonical settings with (cognitively) costly information acquisition (as, e.g., in the rational inattention literature; see Maćkowiak et al. (2018); Sims (2003)), attention reducing "trembles," and recall of memories (Examples 1–3 in Appendix A.1).

Second, the DM derives attention utility from the dimensions they devote attention to. Specifically, the attention utility from dimension i is proportional to the amount of attention devoted to i and material utility, i.e., it is given by $\alpha_i V_i(x)$. The relative importance of attention utility to material utility is given by parameter λ.

[3]We take here the dimensions as given. In practice, the boundaries between dimensions may not always be obvious. While we acknowledge this underspecification, we provide some guidance as to how our model can be applied in practice (Section 6) and study a meta-optimization problem in which the DM chooses how to define a dimension (Section 4.4).

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

In some settings, attention utility can be interpreted as anticipatory utility (Loewenstein, 1987; Caplin and Leahy, 2001) or memory utility (Gilboa et al., 2016; Hai et al., 2020), but one that is only generated when the DM devotes attention to (future or past) consumption.

We view the first consequence of attention—its instrumental role—as relatively standard, and for $\lambda = 0$, 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." More generally, the DM's objective is the weighted sum of material utility and attention utility:

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

where ω_i and ψ_i are nonnegative parameters. Parameter ω_i captures the weight of dimension i in the DM's material utility. When dimensions are different states, these weights can capture the probability of each state; when dimensions are different time periods, these weights can capture exogenous time discounting of material utility from future consumption. Parameter ψ_i is used in Section 3.1 to capture the amount of attention the DM's future "selves" devote to a period i. In static environments, it is natural to set $\psi_i = 0$.[5]

Comments on the model. Before analyzing the model, we pause to comment on some implicit modeling assumptions. A key tenet of our model is that some attention is voluntarily directed by the individual, a premise often referred to as "top-down attention." While involuntarily allocated attention (often referred to as "bottom-up attention") is also important, evidence suggest that at least some attention is indeed directed (e.g., Corbetta and Shulman (2002); Buschman and Miller (2007); Bronchetti et al. (2020)). Our model can be viewed as determining the use of the residual stock of attention after involuntary attentional allocations have been made. We return, in Section 6, to discussing how voluntary and involuntary attention may interact.

For tractability, we also implicitly assume that the DM understands the consequences of attention captured by the set of available actions, $X(\alpha)$, perfectly and

[5]Parameter ψ_i can also be used to nest the case where attention utility is independent of the amount of directed attention α_i as in anticipatory utility models like Loewenstein (1987); Caplin and Leahy (2001): let λ go to 0, and ψ_i go to infinity, keeping their product constant.

that they can find the maximizers of Equation (1) without devoting attention. It is straightforward to consider a dynamic attention allocation, where the DM may stop devoting attention upon learning that further attention has only limited instrumental value. We believe our qualitative results would also hold in this more complicated setting. As we mention in Section 6, similar concerns are relevant in models of attention allocation more generally.

Our approach to modeling attention utility and how it is generated abstracts away from many details. In reality, the generation of attention utility is likely a variegated process: attention to past consumption may generate memory utility, attention to future consumption may generate anticipatory utility, and attention to contemporaneous consumption may intensify the experienced utility. In our static model, we do not differentiate between these interpretations, while in practice, the type of cognitive process (remembering, anticipating, or focusing) may matter. Extending our setting, one may partially account for this by analyzing a model with different λ parameters for each cognitive process. Furthermore, attention utility may take a different functional form altogether, say, $\sum_i (F(\alpha_i) + \psi_i) V_i(x)$, where, e.g., a concave F captures decreasing returns from devoting attention for the creation of attention utility. We consider our model as a plausible and simple starting point of incorporating attention's emotional role in formal economic models.

It may also seem that many decisions require only short bursts of attention (e.g., one's portfolio choice in the context of financial decision-making mentioned in the introduction and further discussed in Section 2.3). If so, then the DM should intensely devote attention for a brief amount of time to maximize material utility and devote attention during the remaining time to high-payoff dimensions to maximize attention utility. While plausible in some situations, we believe attention's instrumental role to be generally less trivial. Even if deciding which action to take may be done with short contemplation (and little attention), executing the chosen action may take a nontrivial amount of time. For example, an individual with health concerns may decide instantaneously in a doctor's office to monitor their symptoms; however, actually monitoring will require attention allocated over a long period of time. Similarly, a student can easily (and quickly) decide to study hard for an important test, but actually doing so requires ongoing focus. It may also be the case that once attention has been devoted to a dimension, it is psychologically infeasible to then ignore it. For example, after making a portfolio choice, the investor cannot help but think about the portfolio for the rest of the day.

Lastly, we note that the objective (1) has parallels to models where the DM chooses their preferences (e.g., Bernheim et al. (2021)). Choosing an attention allocation α corresponds to choosing a particular utility function, or what Bernheim et al. (2021) would call a "world view." A key difference is that in our model, the choice of utility function (i.e., the attention allocation) indirectly influences payoffs across different dimensions by varying available actions.

2.2 Optimal attention and action

We provide multiple comparative static results (Propositions 1–3) to understand how the DM's optimal (action, attention)-pair depends on the environment. These comparative statics are general; they do not depend on whether dimensions correspond to different dimensions of consumption, realizations of an unknown state, or time periods. In turn, we consider the dependence on the payoff in a dimension, V_i, the relative weight on attention utility, λ, and parameters ω_i and ψ_i. These results describe the core mechanisms that lead to the various behavioral phenomena and biases, as we discuss later and preview here.

To strengthen some statements, we introduce the notion of a *separable environment*. In words, the environment is separable if the DM takes separate actions for each dimension and whether a dimension-specific action is available depends only on the amount of attention devoted to the dimension. Formally, the environment is separable if action x is a vector $x = (x_i)_{i \in \mathcal{D}}$, payoff $V_i(x_i, x_{-i})$ is independent of x_{-i} for all i and x_i, and $X(\alpha) = \Pi_{i \in \mathcal{D}} X_i(\alpha_i)$. Note that maximizing (1) with respect to an (action, attention)-pair is then equivalent to maximizing $\sum_i (\omega_i + \lambda(\alpha_i + \psi_i))\hat{V}_i(\alpha_i)$ with respect to attention only, where $\hat{V}_i(\alpha_i) \coloneqq \max_{x_i \in X_i(\alpha_i)} V_i(x_i, \cdot)$. We assume that X_i is monotone, i.e., $X_i(\alpha_i) \subseteq X_i(\alpha_i')$ for all $\alpha_i \leq \alpha_i'$, and so $\hat{V}_i$ is increasing.

We begin with varying payoff V_i. For each i, we fix some $\tilde{V}_i$ and define $V_i \coloneqq \beta_i \tilde{V}_i + \gamma_i$, for scalars $\beta_i \geq 0$ and γ_i. Increasing γ_i increases the payoff level, and increasing β_i increases the payoff difference induced by different actions.

An increase in the payoff level of dimension i, γ_i, does not affect which (action, attention)-pair maximizes overall material utility and hence does not affect the standard (i.e., $\lambda = 0$) DM's solution. However, the attention utility from dimension i increases in proportion to the attention devoted it. So when the DM puts positive weight on attention utility (i.e., $\lambda > 0$), they devote more attention. If the environment is separable, this increase in attention, in turn, leads to a better action for that dimension, i.e., the value of $\tilde{V}_i$ increases.

An increase in the payoff difference from different actions, β_i, increases the importance of taking an action suitable for dimension i. It may also move the payoff up or down (e.g., V_i increases everywhere if $\tilde{V}_i$ is nonnegative), inducing the DM to change their attention just as above. In the proposition below, we offset such level changes, and the DM always chooses an action better suited for i. If the environment is separable, the "more suitable" action can only be available if the DM increases their attention. Note that this comparative static does not rely on attention utility. It captures the standard intuition that the DM devotes attention where it is most instrumental.

We assume that the optimal solution is unique for the rest of this section. We state and prove a general version of the propositions that does not assume a unique solution in Appendix C. All other results in the rest of the paper are also proved there.

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

- *If $\gamma_i' \geq \gamma_i$ and $\beta_i = \beta_i'$, then $\alpha_i' \geq \alpha_i$. If, in addition, the environment is separable, then $\tilde{V}_i(x') \geq \tilde{V}_i(x)$.*
- *If $\beta_i' \geq \beta_i$ and $\gamma_i' = \gamma_i - (\beta_i' - \beta_i)\tilde{V}_i(x)$, then $\tilde{V}_i(x') \geq \tilde{V}_i(x)$. If, in addition, the environment is separable, then $\alpha_i' \geq \alpha_i$.*[6]

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

[6]Note that $\beta_i\tilde{V}_i(x) + \gamma_i = \beta_i'\tilde{V}_i(x) + \gamma_i'$, and so unless the DM changes their optimal (action, attention)-pair, there is no level change in the payoff from dimension i.

Panel (b) visualizes the second part and considers an increase in β_i to β_i' with an offsetting change in γ_i to γ_i'. The material utility is depicted in the top figure, and the attention utility in the bottom figure, both as functions of $\tilde{V}_i(x)$. Throughout, the optimal attention α_i^* is held fixed. This change then pivots the material utility around its initial optimal value. Already here, the DM benefits relatively more from increasing $\tilde{V}_i(x)$ than before. The same pivoting occurs for attention utility. Thus, the DM increases $\tilde{V}_i(x)$ by choosing a different action in response to the change.

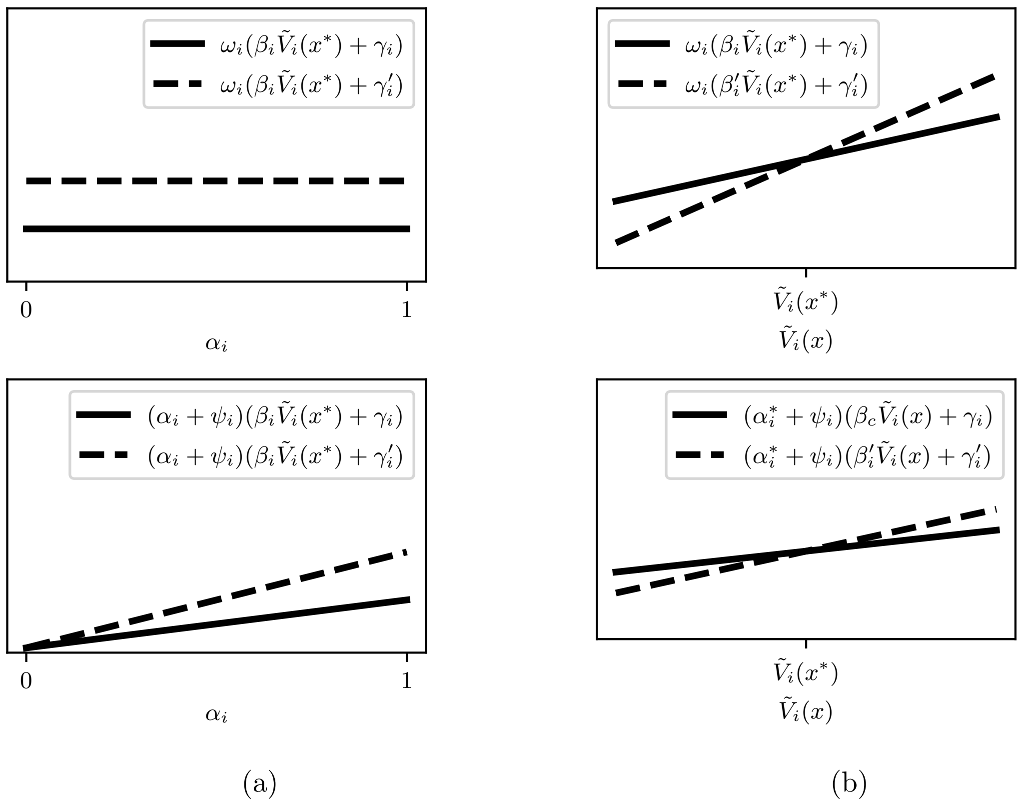


Figure 1: Panel (a) corresponds to an increase of γ_i to γ_i'. We hold the optimal action x^* fixed. The payoff V_i is shifted up, and hence so is the material utility (top figure), independently of α_i. However, increasing α_i now increases the attention utility now by more than before (bottom figure).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

We next turn to the relative weight on attention utility λ and show three results. First, increasing λ decreases the relative importance of material utility, and so it decreases through the choice of the optimal action. Second, parameterizing V_i as $V_i \coloneqq \tilde{V}_i + \gamma_i$, the DM's objective is evidently linear in payoff levels γ_i and hence the DM's value, i.e., (1) for optimal (action, attention)-pairs, is convex in γ_i. Note that in the absence of attention utility ($\lambda = 0$), it is only linear. Third, when the environment is separable and $\lambda > 0$, then increasing attention to dimension i increases both the payoff V_i but also the weight on dimension i; hence, accounting for attention utility makes the DM's objective more convex in attention. The proposition below provides the formal statements.

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

- $\sum_i \omega_i V_i(x) \geq \sum_i \omega_i V_i(x')$;
- *the DM's value is convex in $(\gamma_i)_{i \in \mathcal{D}}$;*
- *if the environment is separable, $\omega_i = 1$ and $\psi_i = 0$ for all $i \in \mathcal{D}$, and the objective given λ is convex in α, then it is also convex in α given λ'.*

The first part implies that the DM's actions are suboptimal if judged through the lens of the standard model (with $\lambda = 0$).[7] The second and third parts of the proposition imply that the DM has a preference for "extreme" payoffs so that they can hone in on high-payoff dimensions and ignore others.[8] When the environment is separable, variation in payoffs is generated through extreme attention allocations, and so the DM's attention is naturally "sparse" (Gabaix, 2014).[9] One way of interpreting these results is that the DM has a preference for specialization—it is better to be outstanding in one area and relatively poor in many others rather than doing mediocre in all.

[7]They may also be judged as suboptimal through the lens our model if we take attention utility as not normative.

[8]Devoting attention to high-payoff dimensions and ignoring others can also be understood as choosing preferences that view the high-payoff dimension as more important (see the discussion at the end of Section 2). Indeed, Bernheim et al. (2021) show a similar convexity result (Proposition 11) in their model of chosen preferences.

[9]Note that the sparse attention allocation occurs not just for instrumental reasons as in Gabaix (2014), but because of the complementarity between the instrumental and emotional role of attention: The instrumental value of attention for a dimension (i.e., increasing the associated payoff) is particularly useful if the dimension has a large weight in the DM's attention utility; conversely, increasing the weight of a dimension in the DM's attention utility is particularly useful when the associated payoff is high. Since attention drives both, the objective becomes "more convex" relative to the standard model (i.e., with $\lambda = 0$).

As we show, these results also lead to an additional preference for risk (Section 2.4), non-smooth consumption paths (Section 3.1), and intrinsic preferences for information (Section 3.3).

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

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

Suppose, e.g., ω_i represents the probability with which dimension i realizes. In that case, the comparative statics is again entirely standard: the DM chooses an action that increases the payoff of a dimension if it becomes more likely. Less standard is that the DM also chooses such action if there is some exogenous attention γ_i on that dimension.

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

2.3 Attention across consumption dimensions

We consider attention allocation across different dimensions of consumption. Those may be 'arranging a retirement home for a relative', 'vacation', 'personal health', 'financial situation', etc. The DM's overall material utility is the unweighted sum of the material utilities across these dimensions (i.e., $\omega_i = 1$). In this context, Proposition 1 rationalizes the well-known ostrich phenomenon: (attentional) avoidance of low-payoff situations and, conversely, excessive attention to high-payoff ones.[10] Evidence for such behavior has been extensively documented in the domains of finance and health.

[10] The term "ostrich effect" was coined in Galai and Sade (2006), where it describes individuals avoiding risky financial situations by pretending they do not exist—i.e., they bury their figurative heads in the sand like an ostrich. We use the term ostrich phenomenon to refer to attentional patterns due to changes in payoff levels and not information. Some readers may be interested to know that although the idea of ostriches burying their heads in the sand to avoid predators dates back to Roman times, they do not display this behavior. Instead, they put their heads into their nests (which are built on the ground) to check temperatures and rotate eggs.

For instance, retail investors' propensity to check their portfolios generally co-moves with the market (Karlsson et al., 2009; Sicherman et al., 2015).[11] Accessing one's portfolio requires attention to it. Our model suggests that doing so co-varies with the payoff associated with the portfolio. Reasonably, this payoff may be eventual consumption. A down market (for most investors) implies low future consumption. By decreasing attention to their portfolio, investors can improve their attention utility.[12] Similar behavior has been documented in the domain of health.[13] Inattention to potential health issues to manage one's emotions leads to worse material utility—the health outcomes—in line with Proposition 2.[14]

In principle, several factors can explain these attention patterns. The instrumental value of attention, e.g., via information, may vary with the market in a way that makes increased monitoring in up markets optimal. Similarly, changes in the cognitive (but non-emotional) costs of attention or the outside option of attention can also play a role. Finally, the patterns in attention could be explained by belief-based utility models, where utility is derived from anticipation (Caplin and Leahy, 2001; Brunnermeier and Parker, 2005) or from 'news' (Kőszegi and Rabin, 2009; Karlsson et al., 2009).

However, the ostrich phenomenon has also been documented in settings where all of these explanations are implausible, suggesting an important role of direct emotional utility derived from attention. For instance, Avoyan and Schotter (2020) provide

[11] Gherzi et al. (2014) find increased monitoring following market downturns.

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

[13] For instance, researchers have noted low rates of testing for severe medical illnesses (Huntington's disease (Shouldson and Young, 2011; Oster et al., 2013); sexually transmitted diseases (Ganguly and Tasoff, 2017)). Our model predicts that an individual at risk of such a disease may have a low (expected) payoff related to the consumption dimension 'health' and hence avoids any actions, such as taking a test, that require attention. Indeed, Ganguly and Tasoff (2017) document that the demand for medical testing for sexually transmitted diseases is decreasing as the expected health outcome worsens. Individuals also often fail to follow medical recommendations, both regarding information-generating activities (e.g., self-screening) and non-information-generating activities, such as taking medicine (Becker and Mainman, 1975; Sherbourne et al., 1992; Lindberg and Wellisch, 2001; DiMattero et al., 2007). For instance, DiMattero et al. (2007) find that, among individuals experiencing serious medical conditions, individuals with worse health status tend to adhere less to medical regimes.

[14] Low compliance with medical recommendations may negate the benefits from medical therapies or limited medical testing, e.g., in the context of Hunting's disease leads to worse decisions about "childbearing, retirement, education, participation in clinical research" (Oster et al., 2013).

evidence in a stylized laboratory environment where experimental participants choose to allocate time (i.e., attention) between two games (i.e., consumption dimensions). In line with our model, they find that "as payoffs in a given game increase, subjects plan more attention to the game." Variation across treatments to non-emotional benefits or costs of attention are ruled out by design, and belief-based utility models also seem implausible.[15]

There is also strong evidence for an important role of attention utility in the context of retail investors. Sicherman et al. (2015) find a positive correlation between market returns and the frequency of investors logging in to their portfolio twice during a single weekend—when markets are closed, and no new information can be revealed. Quispe-Torreblanca et al. (2020) find that investors devote excessive attention to positive information that is already known. They also invite investors to respond to a survey about their investment strategy and find that participation is positively correlated with their portfolio returns. This evidence points towards attention's emotional role as an important determinant of (investment) behavior.[16] The absence of new information from attention renders belief-based utility models mute, and variations in non-emotional costs and benefits seem unlikely.[17]

Our model's rationalization of the ostrich phenomenon connects it to adaptive preferences (Elster, 1983). Recall that the DM's attention allocation could be interpreted as them choosing their preferences (Bernheim et al. (2021); see the end of our Section 2). A decrease in the payoff associated with their portfolio leads the DM to choose that the portfolio dimension is unimportant (via devoting little attention to it). However, as eluded to in that earlier section, a key difference is that here the "chosen preferences" have implications for the actions the DM is able to take.

[15]For the results to be explained by belief-based utility models, the participants would need to consider the random payoffs from the game as separate and prefer the payoff variations induced by devoting more attention to the high-payoff game.

[16]In a related setting, Olafsson and Pagel (2017) study individuals' attention to their financial accounts and finds increased attention after they are paid and decreased attention when the account balance becomes low, in particular, when it turns negative. Arguably individuals often know about their payment dates and amounts and their overdrawn status, so information avoidance may be an implausible motive.

[17]We note that the model in Karlsson et al. (2009) also has a direct emotional effect of attention, what they call "impact effect." Even without further uncertainty after the initial news, this impact effect can still lead to the observed attentional patterns (in their model, set $\theta = 0$ and $SD(r_d) = 0$). But the impact effect is crucial and not a feature of news utility models per se.

2.4 Attention across states

We next consider attention allocation across possible realizations of an uncertain state. The attention-dependent weights on different states lead to as-if belief distortions (characterized by Propositions 1–3) and, with them, to implications for the DM's attitude towards risk and probability weighting.

A state i is weighted in the DM's material utility by $\omega_i = p_i$, where p_i denotes the objective probability of state i occurring. For simplicity, we suppose that $\psi_i = 0$, i.e., there is no exogenous attention outside the DM's control. The DM's objective is then to choose (x, α) with $x \in X(\alpha)$ to maximize $\sum_i p_i V_i(x) + \lambda \sum_i \alpha_i V_i(x)$, i.e., the expected material utility plus attention utility.

It is useful to divide the DM's objective by $1 + \lambda$ and denote the terms in front of V_i as $q_i := \frac{p_i + \lambda\alpha_i}{1+\lambda}$. Note that $q_i \in [0, 1]$ for all i and $\sum_i q_i = 1$, i.e., q_i describes a probability distribution. The DM, conditional on their attention allocation, behaves like a subjective expected payoff maximizer choosing action x who assigns probability q_i to state i. Note that as attention to state i increases, so does the subjective probability q_i assigned to that state. Thus, Propositions 1–3 have direct implications for q_i.

We briefly discuss some of the general implications of Propositions 1–3 in this environment before relating our model to more concrete behavioral phenomena. Proposition 1 implies that individuals devote more attention to high-payoff states and thus (at least in a separable environment) will take actions suited for those states relative to those with a low payoff—an ostrich phenomenon in environments with uncertainty. For instance, in the context of individuals devoting attention across future 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 has been ignored). Proposition 2 indicates that these plans lead to a lower expected material utility (given the objective probabilities). Daydreaming, where individuals devote attention to a pleasant state with a very low probability or zero probability, is an example, although empirical tests of these predictions are warranted.[18] Proposition 3 implies a standard intuition: increasing the objective probability of a state

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

p_i leads to an action better suited for that state.[19]

We next consider the implications of incorporating our model into a standard framework for choosing lotteries. A DM, equipped with an increasing Bernoulli utility u, chooses a lottery from a set X. Lottery $x \in X$ has monetary payoff x_i in state i (and so is an act), and so its expected payoff is $\sum_i p_i u(x_i)$.

Consider now our DM in this environment. Assume first that attention has no instrumental role, $X(\alpha)$ is constant. We then can let action x capture the choice of lottery, with $V_i(x) = u(x_i)$. In addition to maximizing the expected payoff (or material utility in the language of our model), our DM also values attention utility. Note that the DM in the standard environment described in the preceding paragraph is nested in our model with $\lambda = 0$.

The following proposition states the implications of our model in this setting. For parts of it, we consider binary lotteries, those where any state either pays a low payoff $L(x)$ or a high payoff $H(x)$. Also, let $X(\mu, L)$ denote the set of binary lotteries with mean μ and low payoff L and $Y(\mu, H)$ those with mean μ and high payoff H.

Proposition 4.

- *Let DM(λ) refer to the DM with a relative weight λ on attention utility. DM(λ) is more risk-averse than DM(λ') for any $\lambda' > \lambda$.*[20]

- *Suppose u is unbounded and $\lambda > 0$. For any μ, L and $x \in X(\mu, L)$, there exists a high payoff $\bar{H}$ so that if a lottery $x' \in X(\mu, L)$ has high payoff $H(x') > \bar{H}$ then the DM strictly prefers x' to x.*

- *For any μ, L and $x, x' \in X(\mu, L)$ with $H(x) > H(x')$, the DM strictly prefers x to x' if λ is large enough.*

- *For any μ and H, the DM's preferences over $Y(\mu, H)$ are independent of λ.*

Proposition 4 first states that the DM with attention utility has an additional preference for risk. Intuitively, given a lottery x, the DM devotes attention to the

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

[20] 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.

high-payoff states—the "upside" of the lottery—resulting in those states having a higher subjective probability q_i. Notice that attention utility then drives a wedge between risk preferences elicited via choice data (as in the proposition) and those derived from the curvature on the Bernoulli utility u. The second and third cases of the proposition state that the DM has a preference for positively skewed lotteries.[21] Intuitively, positive skew always increases attention utility since the DM devotes their attention exclusively to high-payoff states. If the high payoff is large enough (second case) or the DM puts enough weight on attention utility (third case), the DM prefers the more positively skewed lottery. The fourth case states that no equivalent preferences exist for negatively skewed lotteries.

An implication of Proposition 4 is that individuals appear more optimistic than objective probabilities justify. Such optimism has been documented in a wide range of circumstances. Sharot (2011) summarizes: "we underrate our chances of getting divorced, being in a car accident, or suffering from cancer. We also expect to live longer than objective measures warrant, overestimate our success in the job market, and believe that our children will be especially talented." In our model, the DM devotes little attention to low-payoff states, thus acting as if they "underrate" them and, conversely, overweight the high-payoff ones. There is laboratory evidence consistent with optimism, e.g., Mayraz (2011).[22] Participants guess the realization of a random variable, and some are rewarded for high and others for low realizations. Our model predicts that participants whose payoff is high for high realizations devote attention to those realizations, which then have a large weight in their objective, leading them to guess a high realization. Indeed, this type of "wishful thinking" is what Mayraz (2011) finds.

There is also extensive evidence for preferences for positively skewed lotteries.[23] Furthermore, consistent with our model, Jullien and Salanié (2000) and Snowberg and Wolfers (2010) suggest that the preference for skewness is driven by subjective

[21] 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.

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

[23] Evidence comes from a variety of contexts: 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).

probabilities, as in our model, rather than the Bernoulli utility u.

Our model can also decouple large-stakes from small-stakes risk aversion. Hansson (1988) and Rabin (2000) show that the risk aversion over small-stakes lotteries that is typically observed in individuals implies unreasonable risk preferences over large-stakes lotteries. One way to resolve this puzzle is through preferences where lotteries are evaluated relative to some reference point and with loss aversion (Kahneman and Tversky, 1979). We provide an alternative explanation. Suppose the DM can also devote attention to some outside option, such as an additional consumption dimension o. Attention α_o yields attention utility $\alpha_o V_o$.[24] When facing a lottery with small stakes, none of the potential payoffs may be high enough to warrant attention; specifically, $u(x_s) \leq V_0$ for all s. In this case, the DM's preferences are independent of the weight on attention utility since all attention utility is derived from the outside option. When the lottery has high stakes instead, then the highest payoff may exceed the value of the attentional outside option, and so the DM's preferences become attention-utility dependent (as in the first case of Proposition 4). Thus, attention utility, and hence its impact on risk preferences (which we have shown reduces risk aversion), may only matter for large-stakes lotteries.

We next study how our model leads to as-if probability weighting and consider the mapping of objective probabilities p_i to subjective probabilities q_i, which we denote by $q_i(p_i)$. The setting is as when studying preferences over lotteries, but focus on environments with only two states $\mathcal{D} = \{i, i'\}$. Notice that when there is no instrumental value, the DM devotes full attention to the higher-payoff state, which is subsequently always overweighted (as in the first part of Proposition 4. To generate other forms of probability weighting, we thus need to allow for attention to be instrumentally valuable. We illustrate using separable environments.

First, suppose that at least $\underline{\alpha}$ attention must be devoted to each state but attention beyond $\underline{\alpha}$ is of no instrumental use. The DM devotes all residual attention to the high-payoff state. The probability weighting function is compressed; that is, small probabilities are over-, and large probabilities are underweighted.

We then drop the hard constraint on attention and replace it with large and decreasing initial returns to attention. A similar compression of the probability weighting occurs, and probability weighting now takes a smooth inverse S shape. The

[24]Formally, this is included in the model by adding state o with $p_o = 0$ and payoff V_o.

following proposition summarizes.[25]

Proposition 5. *Suppose $\mathcal{D} = \{i, i'\}$ and that the environment is separable.*

- *If $\hat{V}_i(\alpha_i) = -\infty$ for $\alpha_i < \underline{\alpha}$ and $\bar{V}$ otherwise, and $\hat{V}_{i'}(a) = \hat{V}_i(a) - \Delta$ for all $a \in [0,1]$ with $\Delta > 0$, then $q_i(p_i) = \frac{p_i + \lambda(1-\underline{\alpha})}{1+\lambda}$ for $p_i < 1$ and $q_i(1) = 1$ and $q_{i'}(p_{i'}) = 1 - q_i(p_i)$.*
- *If $\hat{V}_i = \hat{V}_{i'} = \hat{V}$, $\hat{V}$ 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, $q_i = q_{i'} = q$ and there exists some $\bar{p}$ with $0 < \bar{p} < 1/2$, such that*[26]
$$q(p) \begin{cases} = 0 & \text{if } p = 0 \\ > p & \text{if } 0 < p < \bar{p} \\ < p & \text{if } 1 - \bar{p} < p < 1 \\ = 1 & \text{if } p = 1. \end{cases}$$

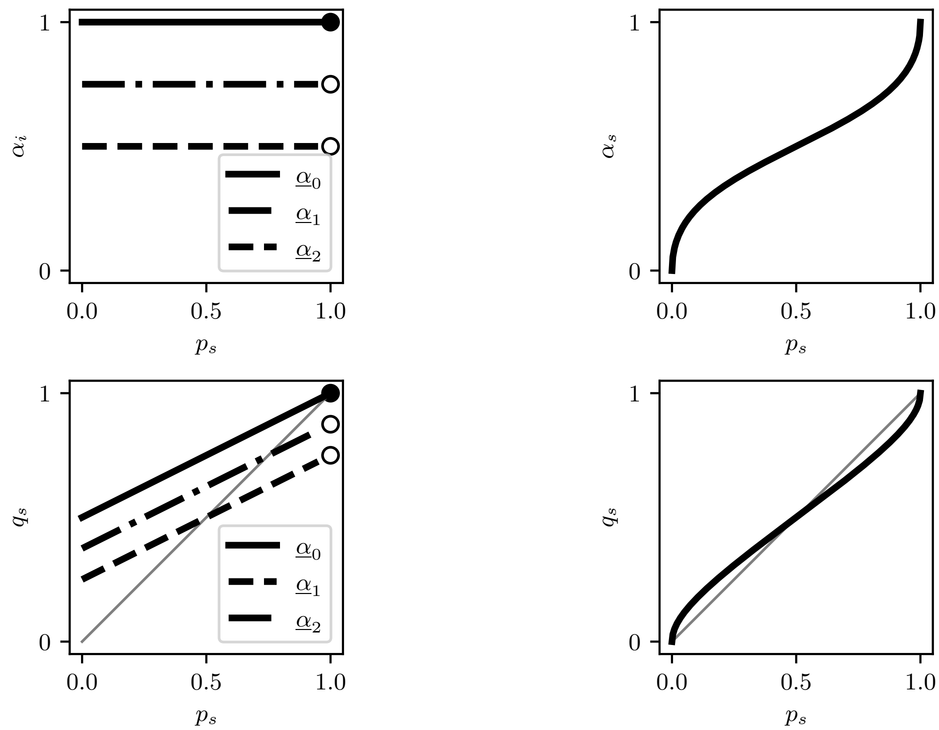
Figure 2 provides an illustration of Proposition 5. In each panel, the top figure shows optimal attention to state i as a function of the probability p_i, and the bottom figure shows the resulting probability weighting, $q_i(p_i)$. Panel (a) corresponds to the first case with forced attention, for three levels of $\underline{\alpha}$: $\underline{\alpha}_0, \underline{\alpha}_1, \underline{\alpha}_2$, with $\underline{\alpha}_0 < \underline{\alpha}_1 < \underline{\alpha}_2$. Panel (b) corresponds to the second case, where we choose $\hat{V}(a) = -\frac{1}{a}$ as tractable functional form.[27]

Probability weighting has been extensively studied since first discussed in Kahneman and Tversky (1979), and there is now a voluminous literature analyzing and empirically estimating prospect-theory models, which are typically descriptive (see Wakker (2010); Barberis (2013) for two surveys). Our model provides attention as a

[25] Probability weighting, in particular, overweighting of small and underweighting of large probabilities, can also be generated by allowing attention utility to be concave in attention. Specifically, suppose that the attention utility form dimension i is given by $F(\alpha_i)\hat{V}_i$ ($\hat{V}_i(x)$ is constant), where F is concave and continuously differentiable. If $\frac{\partial}{\partial a}F(1) = 0$, and $\hat{V}_i \geq \hat{V}_{i'} > 0$, then the DM devotes more attention to state i than i', but not all. We consider this type of extension plausible (see the comments on the model setup in Section 2.1); however, this particular result is outside the model presented here and hence we do not pursue it further.

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

[27] One can show that the optimal attention is $\alpha_i = (p_i - \sqrt{p_i(1-p_i)})/(2p_i - 1)$ which is inverse S-shaped and hence so is q_i.



(a) (b)

Figure 2: Panels (a) and (b) correspond to the two cases discussed in Proposition 5. The weight on attention utility is always $\lambda = 1$. For the first case, $\underline{\alpha}$ takes values $\underline{\alpha}_0 = 0, \underline{\alpha}_1 = 1/4$, and $\underline{\alpha}_2 = 1/2$. For the second case, the payoff as a function of attention is given by $\hat{V}(a) = -\frac{1}{a}$.

mechanism leading to probability weighting. The classic finding is that individuals' probability weighting follows an inverse S-shape (Wu and Gonzalez, 1996)—which our model provides a microfoundation for under the conditions described in Proposition 5.[28] Our model can also provide conditions for other forms of probability weighting, e.g., an S-shaped probability weighting.[29]

[28]Our model has implications for probability weighting that distinguishes it from other models. For example, cumulative prospect theory (Tverseky and Kahneman, 1992) or rank dependent utility (Quiggin, 1982) predict that the probability assigned to a state (q_i) depends only on the ranking of the states ($V_s > V_{i'}$ or $V_i < V_{i'}$) and the objective probabilities of each state occurring (p_i). In contrast, in our model, q_i additionally depends on the difference in payoffs, $V_i - V_{i'}$, and the instrumental value of attention.

[29]For instance, suppose that $\hat{V}_i = \hat{V}_{i'} = \hat{V}$ and that $\hat{V}$ is convex and not constant. The DM optimally devotes attention to the more likely state leading them to overweight high and underweight low probabilities.

3 Extension: Multi-period attention allocations

The previous section explored some implications of our model in static environments where attention is allocated once. In this section, we consider time as a type of dimension, and so we extend our model dynamically. Doing so allows us to study attention allocation across time periods, which endogenizes time preferences. It also allows us to study our model's implications in two dynamic settings: choice over a default action and dynamic resolution of uncertainty.

But first, we note a complication of our analyses that arises due to repeated attentional choice: The DM is generally time-inconsistent. This is because past attention utility, e.g., from devoting attention to a future period, may depend on today's action and its impact on the future payoff, but the DM today does not incorporate this past attention utility. Because the DM cannot commit their future selves to a particular attention allocation, the solution to the DM's problem is the equilibrium of an intrapersonal game played between each period's self. One can treat the results in Propositions 1–3 as defining the best response of the DM in a single period, given the choices of the DM in all other periods.

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

3.1 Attention across time periods

We first focus on attention across time periods. The DM now faces a sequence of time periods $\mathcal{D} = \{1, \dots, T\}$, with generic period t. For simplicity, we assume that there is no exogenous discounting, i.e., $\omega_t = 1$ for all t. However, the attention-dependent weights on the different periods (the dimensions) can be interpreted as endogenous time preferences.

In each period t, the DM chooses an (action, attention)-pair denoted by (x_t, α_t). The actions jointly determine the payoff in each period: given $x := (x_t)_{t=1}^{T}$, the payoff in period t is $V_t(x)$.[30] Attention is a measure on the set of time periods, i.e., $\alpha_t = (\alpha_{t \to t'})_{t' \in \mathcal{D}}$, where $\alpha_{t \to t'}$ denotes the attention (in period t) devoted to period t' with $\alpha_{t \to t'} \geq 0$, and we normalize the total attention devoted (in each period) to have measure 1, i.e., $\sum_{t'} \alpha_{t \to t'} = 1$; we also let $\alpha = (\alpha_t)_{t \in \mathcal{D}}$. We assume that the available actions in period t only depend on attention in period t, i.e., x_t must be in $X_t(\alpha_t)$ which is compact- and non-empty-valued and upper hemicontinuous.

In each period, the DM receives a material utility and attention utility—the DM's flow utility in period t—just as in the static model. As a natural first step, we assume that the DM maximizes the sum of such flow utilities across periods. We first consider the best response function in any period t holding fixed (x_{-t}, α_{-t}): The DM chooses (x_t, α_t) with $x_t \in X_t(\alpha_t)$ to maximize[31]

$$\sum_{t'=t}^{T} \Bigg(\underbrace{V_{t'}(x_t, x_{-t})}_{\text{material utility in } t'} + \lambda \underbrace{\sum_{t''=1}^{T} \alpha_{t' \to t''} V_{t''}(x_t, x_{-t})}_{\text{attention utility in } t'} \Bigg). \tag{2}$$

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

From period t's perspective, the weight on period t' is given by $1 + \lambda(\alpha_{t \to t'} + \psi_{t'})$. These weights across periods t' can be interpreted as discounting: Fixing (x, α) the DM behaves like a standard DM (with $\lambda = 0$) who discounts period t' (relative to period t) by $\delta_{t \to t'} := \frac{1+\lambda(\alpha_{t \to t'} + \psi_{t'})}{1+\lambda \alpha_{t \to t}}$. For instance, as attention to the present period increases, the DM discounts future periods by more. Time preferences—whether the DM is present- or future-focused—are endogenous and depend on circumstances.

As in Section 2.4, we begin by discussing some of the general implications of Propositions 1–3 in this environment, and then relate our model to a more concrete behavioral phenomenon. Here, Propositions 1–3 provide predictions on time preferences. In particular, Proposition 1 suggests that the DM may weigh a period more if its payoff level or the instrumental value of attention to that period increases.

[30] One may make natural assumptions on future actions' impact on past payoff.

[31] Note that in this formulation, the DM values every future self's flow utility the same. Equation (2) can be generalized by allowing the weights on period-t' flow utility (currently 1) to also depend on $\alpha_{t \to t'}$. For instance, when this weight is given by $1 + \tilde{\lambda} \alpha_{t \to t'}$ and $\lambda = 0$, the DM may alternatively maximize $\sum_{t'=t}^{T} (1 + \tilde{\lambda} \alpha_{t \to t'}) V_{t'}(x_t, x_{-t})$. Our preferred formulation is Equation (2) as it captures concerns for future attention utility which will be crucial in Sections 3.2 and 3.3, while not further complicating the objective (implicitly $\tilde{\lambda} = 0$). However, all results in this section go through for any $\tilde{\lambda} \geq 0$.

Magnitude-dependent discounting is a well-known empirical regularity (e.g., Green et al. (1997) is an early paper), although it has not been directly linked to attention. Proposition 3 implies that some (exogenous) attention devoted to period t' leads the DM to increase the payoff in period t'.

Holding (action, attention)-pairs in other periods fixed and looking at the DM's best response provides potentially useful intuition. Actual attentional choices are the result of an intrapersonal game where the DM predicts their optimal future behavior and how it depends on actions today. The set of solutions is found via backward induction.[32] Propositions 1–3 cease to hold due to coordination motives in the DM's problem. Example 4 in Appendix A.2 shows that increasing a future payoff can lead to less attention to that period; Example 5 shows that varying λ can affect the material utility non-monotonically.

We next consider the implications of incorporating our model into a consumption-saving problem. A DM receives a unit of income in every period that they can irreversibly allocate for consumption across current and future periods. In each period, they value consumption according to some strictly concave function V. Formally, in period t, the DM chooses $x_t = (x_{t\to t'})_{t'=1}^T$, where $x_{t\to t'}$ denotes the amount of period-t income allocated for consumption in period t', and $\sum_{t'=1}^T x_{t\to t'} \leq 1$ for all t; letting $x = (x_t)_{t=1}^T$, consumption in period t is valued by $V(\sum_{t'=1}^t x_{t'\to t})$. The concavity of V implies that in this standard problem, the DM would smooth consumption and consume all income in the period they receive it.

Consider now our DM in this environment. Our formulation captures the consumption allocation of income as the action x. We need to make an assumption about the instrumental value of attention, i.e., how the feasible consumption allocations x_t, depend on the attention allocation α_t. Here are two: 1) $X(\alpha) = \{x : x_{t\to t'} \leq \alpha_{t\to t'} \forall t, t'\}$, i.e., the DM needs to allocate attention to a period in order to allocate their income to that period, or 2) $X(\alpha) = \{x : x_{t\to t'} \leq 1 \forall t, t'\}$, i.e., attention has no instrumental value and the only constraint is the available income, which is 1. For the following result, the exact formulation makes no difference; we state it for the

[32]Formally, 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 (2) 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 (2) over (x, α), with $(x, \alpha) \in \{\mathcal{H}_T, (x_T, \alpha_T)\}$ and $x \in X(\alpha)$.

first case.

To simplify the statement of the following proposition we also assume that 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.[33]

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

- *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 case where the weight on attention utility λ is small is essentially the standard model where the DM maximizes their material utility by smoothing consumption. Since attention goes hand in hand with the action (allocation of income), the DM devotes all their attention to the present period. Note that although the DM devotes full attention to the present, we still have $\delta_{t \to t'} = 1$ for all t, t' with $t' \geq t$, and so no discounting of future payoffs. The reason is that while the attention utility in period t depends on V_t only, attention utility in period t' similarly depends on $V_{t'}$ only, and so both payoffs receive the same weight. When attention additionally determines the weights on the flow utilities, for instance, weight $1 + \tilde{\lambda}\alpha_{t \to t'}$ instead of weight 1 (see footnote 31), then $\delta_{t \to t'} = 1/(1+\tilde{\lambda})$ and so the DM falls in the class of quasi-hyperbolic discounters (Laibson, 1997).

When λ is large, the DM allocates their income and attention to create periods with particularly high consumption, which are exploited for attention utility. In between those periods, the payoff is low; however, the DM never devotes attention to these low-payoff periods, so their attention utility is unaffected.

Our model thus rationalizes consumption paths that are purposefully lumpy because of, e.g., weddings, vacations, and other lavish celebrations.[34] Such paths can be similarly rationalized through memory utility (Gilboa et al., 2016; Hai et al., 2020) or anticipatory utility (Loewenstein, 1987). For instance, Gilboa et al. (2016) write: "One may enjoy fond memories of a vacation, wedding, or special night out long after they have occurred." The key innovation in our model is that the enjoyment

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

[34] Hai et al. (2020) notes that the average expenditure on weddings is about USD 20,000 and that the average annual household income of a newly married couple is USD 55,000.

of fond memories requires attention to be experienced. Similarly, low-payoff periods, say, weekly chores or an unpleasant memory, need not lead to anticipatory or remembrance utility and instead can be ignored.

3.2 Implications for default effects

We next explore how incorporating attention utility can lead to particular kinds of default effects. To this end, consider the following environment. There are two periods—period 1 and period 2. The setup in period 2 is that of Section 2.3 with two consumption dimensions, $\mathcal{D} = \{c_{nt}, c_t\}$. One of them, c_{nt} (the nontrivial dimension), involves the choice of a default action. Its payoff is parameterized by some state $s \in \mathcal{S}$, where $\mathcal{S}$ is finite, and the DM's default action x_1 chosen in the first period. The other one, c_t (the trivial dimension) serves as an attentional "outside option" and its payoff V_{c_t} is constant.

Changing the action in period 2 from the default requires at least η attention. Formally, let $\alpha_2 := (\alpha_{2\to c_{nt}}, \alpha_{2\to c_t})$ denote the attention in period 2. For any state s, if $\alpha_{2\to c_{nt}} < \eta$, then $X_2(\alpha_2)$ is a singleton (i.e., the default binds and the DM cannot change the payoff), and if $\alpha_{2\to c_{nt}} \geq \eta$, then $\max_{x_2 \in X_2(\alpha_2)} V_{c_{nt}}(x_1, x_2|s)$ is independent of x_1 for all x_1 (i.e., the default is irrelevant).

In period 2, the DM chooses (x_2, α_2) with $x_2 \in X_2(\alpha_2)$ to maximize

$$\underbrace{V_{c_{nt}}(x_1, x_2|s) + V_{c_t}}_{\text{material utility}} + \lambda(\underbrace{\alpha_{2\to c_{nt}} V_{c_{nt}}(x_1, x_2|s) + \alpha_{2\to c_t} V_{c_t}}_{\text{attention utility}}). \tag{3}$$

We let $U_2(x_1, s)$ denote the maximized value of (3) and the corresponding action and attention by $x_2(x_1, s)$ and $\alpha_2(x_1, s)$, respectively, where we suppose that the solution is unique to simplify notation.

Define $\underset{\sim}{\mathcal{S}}$ as those states in which the DM is inattentive to the nontrivial dimension, i.e., $\underset{\sim}{\mathcal{S}} := \{s : \alpha_{2\to c_{nt}}(x_1, s) < \eta\}$. If $x_1 \in \arg\max_{x_1 \in X_1} \sum_{s \in \mathcal{S}} p_s U_2(x_1, s)$ then $x_1 \in \arg\max_{x_1 \in X_1} \sum_{s \in \underset{\sim}{\mathcal{S}}} p_s U_2(x_1, s)$ (taking $\underset{\sim}{\mathcal{S}}$ as fixed). In words, the optimal default for period-2 conditions only on states in which it binds.

In period 1, when the default is chosen, the DM also values their current attention utility. We assume that they can devote attention to across the realizations of future consumption dimensions but that this attention is non-instrumental (i.e., the set of available default actions $X_1(\alpha_1)$ is independent of attention α_1). Let $\alpha_{1\to(c_{nt},s)}$ denote the attention in period 1 devoted to the nontrivial consumption dimension in state s, and $\alpha_{1\to c_t}$ that to the trivial dimension. The DM's period-1 attention utility is $\sum_{s\in\mathcal{S}} \alpha_{1\to(c_{nt},s)} V_{c_{nt}}(x_1, x_2(x_1,s)|s) + \alpha_{1\to c_1} V_{c_1}$.

In period 1, the DM's objective is the sum of period-1 attention utility and the expected period-2 utility, where the former has relative weight λ. However, the DM optimally devotes all period-1 attention to the highest payoff state, or the trivial dimension, whichever has a higher payoff, and so their period-1 attention utility is independent of the default chosen. The following proposition summarizes and characterizes the states when the default binds as the weight on attention utility increases.

Proposition 7. *Let* $\underline{\mathcal{S}}(x_1) \coloneqq \{s : \alpha_{2\to c_{nt}}(x_1, s) < \eta\}$.

- *Suppose* $X_1(\alpha_1)$ *is independent of* α_1. *Then the optimal default action* x_1 *satisfies*
$$x_1 = \arg\max_{x_1'\in X_1} \sum_{s\in\underline{\mathcal{S}}(x_1)} p_s V_{c_{nt}}(x_1', x_2(x_1', s)|s).$$
- *For any* x_1, *if* λ *is large enough, then*
$$\underline{\mathcal{S}}(x_1) = \{s : \max_{(x_2,\alpha_2), x_2\in X_2(\alpha_2)} V_{c_{nt}}(x_1, x_2|s) < V_t\}.$$

$X_1(\alpha_1)$ independent of α_1 is required as otherwise the DM's choice of x_1 may additionally be driven to maximize current attention utility.

The choice of default is in part similar to how a default is chosen in models of, e.g., rational inattention. If changing the default is costly, the DM may keep it and the optimal default conditions on states where it binds. A key difference in our model is that the default binds asymmetrically, as the second part of the proposition states. It identifies set $\underline{\mathcal{S}}(x_1)$ of states where the default binds for large λ as those states s with a low payoff. In turn, the optimal default is chosen conditional on these low-payoff states.

Our model thus predicts asymmetric reactions to changes in the environment. For instance, individuals may adjust their planned consumption only in response to positive income shocks ignoring negative ones. In anticipation, they plan for low consumption (the default) to insure against their inattention. Similarly, individuals may not want to revisit previously negotiated business or wage contracts in response to some bad events occurring. This may add to firm inertia in economic downturns.[35]

We lastly note that Proposition 7 relies on the default being a "pure default." To illustrate, modify the setting as described above: the payoff of dimension c_{nt} is now given by $V_{c_{nt}}(x_1, x_2|s) = \tilde{V}_{c_{nt}}(x_1, x_2|s) + \beta F(x_1)$, where $\tilde{V}_{c_{nt}}$ has the "default property" from above, $\beta \geq 0$, and F is some function through which action x_1 has irreversible payoff consequences. Thus, the previous setting is nested with $\beta = 0$.

Proposition 8. *Suppose $\beta > 0$ and that $X_1(\alpha_1)$ is independent of α_1 and finite. Let $(s, x_1) = \arg\max_{s' \in \mathcal{S}, x_1' \in X_1} V_{c_{nt}}(x_1', x_2(x_1', s')|s')$. If $V_{c_{nt}}(x_1, x_2(x_1, s)|s) > V_{c_t}$ and λ is large enough, then $x_1 = \arg\max_{x_1' \in X_1} F(x_1')$.*

For large λ, the DM's utility is primarily determined by their attention utility (in both periods). Thus, they choose x_1 to maximize the payoff in the high-payoff states they devote attention to (in either period). For $\beta > 0$, this is done by maximizing F since the default does not bind for those states (Proposition 8). In contrast, when $\beta = 0$, x_1 has no effect on the high-payoffs and so the DM chooses x_1 to maximize their material utility and chooses it optimally for the low-payoff states (Proposition 7).

The fact that the DM can freely devote attention in period 1 is a strong assumption. Note, however, that the same reasoning outlined above goes through if period-1 attention utility is ignored. Thus, the propositions also describe how period-2 utility is maximized, hence isolating an additional consideration when the environment is dynamic.

3.3 Implications for information acquisition

Our model leads to intrinsic preferences over information as we explore next. We consider a setting in which there is no instrumental value of information. Although the standard DM (with $\lambda = 0$) is indifferent to information, our DM has strict preferences over whether and when they acquire information and what type. As we show, our model can rationalize many preferences relating to information observed in laboratory settings (CHEW and HO, 1994; Masatlioglu et al., 2017; Möbius et al., 2022; Nielsen, 2020).

[35]More generally, our model suggests individuals fall back on heuristics and decision processes that can be implemented automatically ("System 1") when the situation at hand has a low payoff, whereas they engage with the situation and use System 2 when its payoff is high.

There are again two periods, with two consumption dimensions, $\mathcal{D} = \{c_{nt}, c_t\}$ in period 2. The payoff from the trivial dimensions V_{c_t} is constant; the payoff from the nontrivial dimension is parameterized by the DM belief p, $V_{c_{nt}}(p) = pV_H + (1-p)V_L$, where $V_L < V_{c_t} < V_H$. The DM's belief in period t is denoted by p_t.

Attention is allocated across the two consumption dimensions. It also leads to information acquisition when devoted to the nontrivial dimension. Specifically, the DM's period-1 action x_1 is a distribution of posteriors and we assume that more attention allows for more informative posteriors: $X_1(\alpha_{1\to c_{nt}}) = \{x_1 \in \Delta([0,1]) : \mathrm{Var}(x_1) \leq \beta\alpha_{1\to c_{nt}}, E[x_1] = p_1\}$. Parameter $\beta \geq 0$ governs how much information can be acquired. Period-2 attention may also lead to information, although optimal attention will not depend on it.

In period 2, the DM chooses $\alpha_2 \coloneqq (\alpha_{2\to c_{nt}, c_t})$ to maximize

$$\underbrace{V_{c_{nt}}(p_2) + V_{c_t}}_{\text{material utility}} + \underbrace{\alpha 2 \to c_{nt} V_{c_{nt}} + \alpha 2 \to c_t V_{c_t}}_{\text{attention utility}}. \tag{4}$$

The DM devotes devotes full attention whichever dimension has the higher (expected) payoff, and so their period-2 attention utility is $\max\{V_{c_{nt}}(p_2), V_{c_t}\}$. Since $V_{c_{nt}}(p_2)$ is linear in p_2, the DM attention utility and hence total utility in period 2 is convex in p_2.

In period 1, the DM maximizes period-2 utility and their current attention utility given by $\alpha_{1\to c_{nt}} V_{c_{nt}}(p_1) + \alpha_{1\to c_t} V_{c_n}$. The following proposition characterizes the DM's optimal attention and information acquisition.

Proposition 9. *Let* $\bar{p} \coloneqq \frac{v - v_L}{v_H - v_L}$.

1. *In period 2 the DM optimally chooses* $\alpha_{2\to c_{nt}} = 0$ *if* $p_2 < \bar{p}$ *and* $\alpha_{2\to c_{nt}} = 1$ *otherwise.*

2. *There exists* $\tilde{p} \leq \bar{p}$ *such that in period 1 the DM optimally chooses* $\alpha_{1\to c_{nt}} = 0$ *if* $p_1 < \tilde{p}$, $\alpha_{1\to c_{nt}} > 0$ *if* $\tilde{p} \leq p_1 < \bar{p}$, *and* $\alpha_{1\to c_{nt}} = 1$ *otherwise. Furthermore,* $\tilde{p}$ *is decreasing in* β.

3. *For any* p_1, *if* $\alpha_{c_{nt}}\beta$ *is small enough, then if* $p_1 > \bar{p}$, x_1 *is negatively skewed, and if* $p_1 \in (\tilde{p}, \bar{p})$, x_1 *is positively skewed.*

The proposition highlights three preferences. First, the DM devotes attention and acquires information when the payoff is likely to be high (first case). This is an example of the ostrich phenomenon discussed in Section 2.3. Möbius et al. (2022) provide experimental evidence where participants' beliefs about a payoff-relevant state are exogenously varied. High beliefs increase the willingness to pay to learn about the state, as our model rationalizes.[36]

Second, the DM has a preference for early resolution of uncertainty. The available information in period 1, β, increases the DM utility as it relaxes a constraint. The DM's utility is independent of how much information is available in period 2 (it is not formally modeled for this reason). Thus, the DM prefers the same information to be received early. The reason is that it allows the DM to condition their period-2 attention on the information received, which makes the period-2 utility convex in beliefs. It is this additional benefit of information that causes the DM to acquire more information early on, holding their belief fixed ($\tilde{p} \leq \bar{p}$; second case). Experimental evidence has been consistent with a preference for early resolution (Masatlioglu et al., 2017; Nielsen, 2020).

Third, the DM prefers positively skewed information when their belief is low and negatively skewed information otherwise (third case). As explained before, early information is useful because it allows the DM to condition future attention on its realization. But this requires sufficient movement in the DM's belief. When the prior is low, this means that good news needs to be unlikely in order to change the DM's attention (and similarly for a high prior). The restriction to a small amount of information acquisition is needed as, e.g., full information mechanically implies positive skew if $p_1 \leq \frac{1}{2}$ and negative skew otherwise.

4 Implications for (self-imposed) policies

In this section, we study our model's implications for some important policy questions. Our focus is on three: in turn, we revisit (simple) models of taxation, allocating tasks between the government and the DM, and effort provision. We also consider how the DM or policymaker should optimally construe the dimensions providing a microfoundation for a key ingredient of our model. For simplicity, we set $\omega_i = 1$ and $\psi_i = 0$.

[36] An implication 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.

4.1 Taxation

Suppose the policymaker is a government that needs to raise revenue via a linear tax scheme. The DM's action x is multi-dimensional with x_i corresponding to the amount consumed in dimension i. Furthermore, the DM brackets their payoff from consumption together with the expenditure associated with a dimension, i.e., $V_i(x) = v_i(x_i) - x_i(p_i + \tau_i)$, where p_i is the price per unit of consumption in period i, and τ_i is a linear tax imposed. Other assumptions regarding how the DM brackets their consumption are possible; we consider this one a plausible starting point.

Note that the government revenue from taxing dimension i, $\tau_i x_i$, comes at a utility cost to the DM of $(1 + \lambda\alpha_i)x_i\tau_i$. Hence, when consumption and attention are held fixed, the government should levy high taxes dimensions to which the DM devotes little attention. The mechanism is different from that in Farhi and Gabaix (2020), who also find that taxes should be high when attention is low; there low attention implies low elasticity of consumption Ramsey (1927). [37] When the DM's attention is endogenous, and in particular may depend on the taxes, we suspect that the DM would devote low attention to high-tax dimensions, reinforcing the inverse relationship between attention and taxes.

4.2 Which tasks should the government do, and which the DM?

We next equip the policymaker, say, the government, with the same technology "attention" as the DM and ask which dimensions they devote attention to, and which should be left to the DM? The policymaker devotes attention ζ_i to dimension i, with $\zeta_i \geq 0$ and $\sum_i \zeta_i = \zeta$, and so given α the payoff in dimension i is $\hat{V}_i(\alpha_i + \zeta_i)$, i.e., the environment is separable and the two kinds of attention are perfect substitutes. Thus, the standard consideration of assigning tasks by comparative advantage is ruled out. Crucially, the weight on the different dimensions in the DM's attention utility is still

[37] In fact, with $v_i(x_i) = \frac{x^{1-1/\psi_i}-1}{1-1/\psi_i}$ for $i \neq 0$, and $v_0(x_0) = x_0$, for simplicity, one can show that the Ramsey Inverse Elasticity Rule becomes: $\frac{\tau_i}{p_i} = \frac{\Lambda_i}{\psi_i - \Lambda_i}$, where $\Lambda_i = 1 - (1 + \lambda\alpha_i)\frac{\mu_{\text{DM}}}{\mu_{\text{revenue}}}$, where μ_{DM} and μ_{revenue} are the welfare weights on the DM's utility and the government's revenue, respectively.

determined by α only.[38] The optimal (α, ζ)-pair then maximizes

$$\underbrace{\sum_i \tilde{V}_i(\alpha_i + \zeta_i)}_{\text{material utility}} + \lambda \underbrace{\sum_i (\alpha_i + \psi_i)\tilde{V}_i(\alpha_i + \zeta_i)}_{\text{attention utility}} .$$

For an optimal attention pair the following must hold: for all $i, j \in \mathcal{D}$, if $\alpha_i > \alpha_j$ and $\zeta_j \geq \zeta_i, \zeta_j > 0$, then $\tilde{V}_i(\alpha_i + \zeta_i) \geq \tilde{V}_j(\alpha_j + \zeta_j)$. In words, if the DM devotes more attention to dimension i and the policymaker to dimension j, then dimension i must yield a higher payoff. If not, then the DM attention utility can be increased by shifting the DM's attention to j and the policymaker's to i.

4.3 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." For a standard DM (with $\lambda = 0$), their effects are similar, but when $\lambda > 0$, they may have very different consequences. A penalty decreases the expected payoff and can thus lead to lower attention and, perversely, worse actions.

Consider a separable environment. Attention α_i to dimension i, which may be interpreted as effort, leads to "success" with probability $p(\alpha_i)$, where p is increasing and continuously differentiable, and "failure" otherwise. The expected payoff of dimension i is $\hat{V}_i(\alpha_i) = p(\alpha_i)V_H + (1 - p(\alpha_i))V_L$, with $V_H > V_L$.

The standard DM's ($\lambda = 0$) optimal attention is unchanged when V_H and V_L are increased by the same amount. They increase attention α_i in response to an increase in $V_H - V_L$; that is, they respond by increasing attention to both "carrots" (an increase in V_H) and "sticks" (a decrease in V_L).

In contrast, when $\lambda > 0$, the DM increases α_i when both the low and high payoff is shifted up because this increases the expected payoff holding fixed the instrumental value of attention (see Proposition 1). They also increase attention in response to carrots: increasing V_H increases α_i. However, increasing the stick can decrease attention.

[38] For example, α_i and ζ_i could represent amounts of information, where the DM acquires the former, and the latter is provided by the policymaker. Here, the DM can process attention ζ_i without devoting attention.

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

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

2. *Increasing V_H increases α_i.*

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

In the third part of Proposition 10, attention is not very effective in increasing p ($\frac{\partial}{\partial \alpha_i} p(\alpha_i)$ is low), and success is far from guaranteed ($p(\alpha_i)$ is also low). In these circumstances, the stick may induce the DM to shy away from dimension i 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).

4.4 Optimal bracketing of dimensions

Lastly, we ask when different dimensions should be considered separate and when they should be thought of as one dimension. For instance, the DM may be able to learn to associate one problem with another, either through some purely cognitive process or with the help of, say, physical cues that the policymaker installs. Such optimal bracketing serves as a microfoundation for the set of dimensions problems in Section 2.1, which can be viewed as the optimally bracketed set of smaller dimensions.

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 $B \in \mathcal{P}(\mathcal{D})$. Let $B(i)$ be defined by $i \in B(i) \in B$. Whenever the DM devotes attention to i and all dimensions $i' \in B(i)$ "come to mind." As multiple dimensions come to mind, the DM's attention is diluted uniformly among them. Thus, given (x, α) and B, the DM utility is

$$\underbrace{\sum_i V_i(x)}_{\text{material utility}} + \underbrace{\lambda \sum_i \alpha_i \bar{V}_B(i)(x)}_{\text{attention utility}}, \tag{5}$$

where $\bar{V}_D(x) := \frac{\sum_{i \in D} V_i(x)}{|D|}$ for $D \subseteq \mathcal{D}$.

Note that the model in Section 2 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 (with $\lambda = 0$).

Also let $\bar{\alpha}_D(x) := \frac{\sum_{i \in D} \alpha_i}{|D|}$ for $D \subseteq \mathcal{D}$.

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

In words, Proposition 11 states that the ordering of brackets by attention is optimally the same as ordering them by their material utility. If this were not the case, the DM could combine a low-attention but high-payoff bracket with a high-attention but low-payoff bracket increasing their attention utility as the high payoffs take a larger weight.

5 Relation to existing models

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

Rational inattention: In models of rational inattention (Sims (2003); Mackowiak et al., 2022), attention serves an instrumental role as in ours. Additionally, attention is costly, e.g., because of deploying cognitive resources. While we do not model these costs explicitly other than via a total attention constraint, they can be captured in the functional form of the payoffs V_i. The key difference is thus that in our model, unlike in models of rational inattention, attention generates emotional utility, what we call attention utility.

Our model rationalizes behaviors that are at odds with rational inattention as monetary or cognitive, but non-emotional, costs do not seem sufficient in many important situations to justify individuals' behavior. For instance, genetic tests for Hunting's disease cost no more than \$300 (Oster et al., 2013). Furthermore, information avoidance (inattention) varies with the level of future payoffs (Karlsson et al. (2009); Sicherman et al. (2015) in the context of investors; and Ganguly and Tasoff (2017) in the context of health) with no (obvious) corresponding change in rational costs or benefits. Even more basic, there is no reason in models of rational inattention to devote attention to already known information (Quispe-Torreblanca et al., 2020). These examples suggest that there is a "cost" (or benefit) of attention missing from the consideration. In this project, we model this cost (see also the discussion in Section 2.3).

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

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

Just as models of rational inattention, models where attention is allocated to induce changes in anticipatory utility or gain utility (via information acquisition) rely on the presence of uncertainty. Such models thus also fail to make predictions in situations where information is unlikely to play a major role, such as in much of the evidence presented in Section 2.3.

Belief based utility with chosen beliefs: Our model, in Section 2.4, also relates to those where subjective beliefs are optimally chosen to increase anticipatory utility as in Bénabou and Tirole (2002); Brunnermeier and Parker (2005); Bracha and Brown (2012); Caplin and Leahy (2019)[39] While our model is, of course, conceptually very different (beliefs do not feature in Sections 2.3 and 3.1, and in Section 2.4, the DM chooses an attention allocation that leads to weights that 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,

[39]For a recent summary of the larger literature see Bénabou and Tirole (2016).

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 Proposition 5 for an example).

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

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

Other Models of Attention: A couple of other papers directly model the two fundamental features of attention in ways similar to ours. The model of Tasoff and Madarasz (2009) is perhaps the closest. A DM faces a decision problem with multiple dimensions and receives anticipatory utility from each as a function of its payoff and the attention devoted to it. Attention to a dimension increases when its payoff changes because the DM chooses an action different from a default or receives payoff-relevant information. Similar to Proposition 1, the DM is more likely to take a non-default action or acquire information if the payoff is high. Such formulation is, in some sense, nested in ours: Let $x_d \in X(\alpha)$ for all α be a default action that is always available, and let acquiring information be encoded as some action x (providing payoff-relevant information for an underlying problem) and suppose x is only available for some attention allocations. A difference between the two formulations is that we allow for attention allocation with no instrumental consequence. More broadly, in our model, attention is chosen to enable non-default actions and information acquisition, whereas, in theirs, the order is reversed.

Their subsequent focus is on how information provision (as requested by the DM or forced by an advertiser) can increase consumption, even when the DM learns their marginal payoff is less than what they expected (this follows from the increase in attention and hence the importance in the DM's objective; this intuition can be expressed in our framework as we show in Example 6 in Appendix A.3). Instead, our focus is on attention allocation across uncertain states and time and the behavioral phenomena our model rationalizes. We also consider dynamic extensions allowing us to explore additional questions such as preferences over dynamic information acquisition and the implications for policymaking.

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

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

6 Conclusion

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

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

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

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

Our model also requires carefully specifying the environment: a key component is a way of partitioning the environment into sets of dimensions. In many real economic environments, natural partitions exist. However, in many situations, it may not be as obvious what the correct sets are. Although Section 4.4 provides some guidance given a finer partition, there are also likely situations where the environment is determined differently.

Thankfully, the novel primitives of our model, the set of dimensions and λ, can be identified from the data. The details would vary by the environment, but here we provide the intuition for a situation where the dimensions are states. We first can identify whether two states are considered jointly (they are in the same "bracket," Section 4.4) by reducing the payoff for one state and increasing the other the same amount and seeing whether the (action, attention)-pair changes. If we can find some shift such that it does, then the two states are not part of the same bracket. The choices over lotteries allow us then to identify the degree of overweighting of the high payoff state(s) and thus λ.

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

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

References

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

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

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

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.

Bernheim, B. Douglas, Luca Braghieri, Alejandro Martínez-Marquina, and David Zuckerman, "A Theory of Chosen Preferences," *American Economic Review*, February 2021, *111* (2), 720–54.

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

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

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

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

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

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

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

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.

CHEW, SOO HONG and JOANNA L. HO, "Hope: An Empirical Study of Attitude Toward the Timing of Uncertainty Resolution," *Journal of Risk and Uncertainty*, 1994, *8* (3), 267–288.

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

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

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

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

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

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

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

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

Elster, Jon, *Sour Grapes: Studies in the Subversion of Rationality*, Editions De La Maison des Sciences De L'Homme, 1983.

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

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.

Farhi, Emmanuel and Xavier Gabaix, "Optimal Taxation with Behavioral Agents," *American Economic Review*, January 2020, *110* (1), 298–336.

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.

Fudenberg, Drew, "Advancing Beyond Advances in Behavioral Economics," *Journal of Economic Literature*, September 2006, *44* (3), 694–711.

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

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

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

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

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

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

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

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

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

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

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

Hansson, Bengt, "Risk aversion as a problem of conjoint measurement," in Peter Gärdenfors and Nils-EricEditors Sahlin, eds., *Decision, Probability and Utility: Selected Readings*, Cambridge University Press, 1988, pp. 136–158.

James, William, *The principles of psychology*, Vol. 1, Henry Holt and Company, 1890.

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

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

Kahneman, Daniel and Amos Tversky, "Prospect Theory: An Analysis of Decision under Risk," *Econometrica*, 1979, *47* (2), 263–291.

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

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

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

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

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

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

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

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

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

Maćkowiak, Bartosz, Filip Matějka, and Mirko Wiederholt, "Survey: Rational inattention, a disciplined behavioral model," *Available at SSRN 3266436*, 2018.

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

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

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

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

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

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

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.

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

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

Rabin, Matthew, “Risk Aversion and Expected-Utility Theory: A Calibration Theorem,” *Econometrica*, 2000, *68* (5), 1281–1292.

Ramsey, F. P., “A Contribution to the Theory of Taxation,” *The Economic Journal*, 1927, *37* (145), 47–61.

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.

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.

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 Addition examples

A.1 Examples of canonical problems

In Examples 1–3, we consider a separable environment and a particular problem $i \in \mathcal{D}$. "Actions" and "payoffs" shall refer to those in the now explicitly modeled dimension i.

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

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

The DM's payoff from dimension i given α_i is

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

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

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

for some $\kappa \geq 0$ and where $H(B)$ denotes the entropy of belief B.[40]

Example 2. *Dimension i is the reduced form of a canonical choice problem with trembles.*

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

$$\hat{V}_i(\alpha_i) \coloneqq E_B[v_j].$$

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

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

for some $\kappa \geq 0$, where $H(B)$ denotes the entropy of belief B (see footnote 40 for the definition) and $\mathcal{U}$ the uniform distribution on A; i.e., if the DM devotes no attention, they will make each choice with equal probability.

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

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

Finally, let $\mathcal{F}$ be

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

As the DM devotes more attention to i, they make more draws from their memory,

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

a form of information acquisition.

A.2 Examples for Section 3.1

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

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

in particular, taking the action x^* requires attention devoted to period 3. The payoff in period 3 is high if the DM takes action x^* in at least one period; otherwise, it is low. We also force $\alpha_{2\to3} \geq \underline{\alpha}_2$, with $0 < \underline{\alpha}_2 < \eta_2$ (formally, this is modeled by assuming any payoff is negative infinity if the DM's attention differs).

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

In period 3, the DM devotes all their attention to $\bar{V}$ (from either of the other periods) and takes a degenerate action. If the DM took action x^* in period 1, then in period 2, they choose $\alpha_{2\to2} = 1 - \underline{\alpha}_2$ and $\alpha_{2\to3} = \underline{\alpha}_2$. Otherwise, they take action x^* (and $\alpha_{2\to2} = 1 - \eta_2$ and $\alpha_{2\to3} = \eta_2$) over $\underline{x}$ (and $\alpha_{2\to2} = 1 - \underline{\alpha}_2$ and $\alpha_{2\to3} = \underline{\alpha}_2$) if

$$(1 + \lambda(1 - \eta_2))\bar{V} + (1 + \lambda\eta_2)\bar{V}_3 \geq (1 + \lambda(1 - \underline{\alpha}_2))\bar{V} + (1 + \lambda\underline{\alpha}_2)\underline{V}_3. \tag{6}$$

In period 1, the DM prefers to take action $\underline{x}$ (and $\alpha_{1\to1} = 1$) and the DM in period 2 taking action x^* (with aforementioned attention) over taking action x^* (and $\alpha_{1\to1} = 1-\eta_1$ and $\alpha_{1\to3} = \eta_1$) and the DM in period 2 taking $\underline{x}$ (with aforementioned attention) *if*

$$(1+\lambda(1+(1-\eta_2))\bar{V}+(1+\lambda\eta_2)\bar{V}_3 \geq (1+\lambda((1-\eta_1)+1)\bar{V}+(1+\lambda\eta_1)\bar{V}_3 \iff \eta_1 \geq \eta_2. \quad (7)$$

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

$$(1+\lambda(1-\eta_1))\bar{V}+(1+\lambda(\eta_1+\underline{\alpha}_2))\bar{V}_3 \geq (1+\lambda)\bar{V}+(1+\lambda\underline{\alpha}_2)\underline{V}_3. \quad (8)$$

Since $\underline{V}_3 < \bar{V}_3 < \bar{V}$, there exists $\lambda > 0$ such that (8) *holds with equality. For such λ, since $\underline{\alpha}_2 > 0$, there exists $\eta_2 < \eta_1$ (i.e.,* (7) *holds) so that* (6) *does not hold. Furthermore, λ can be slightly decreased so that* (8) *now holds strictly and* (6) *still does not hold. Thus, for these parameter values, the unique solution is for the DM to take action x^* in period 1 and hence devote $\eta_1 > 0$ to period 3.*

Now, increase both $\underline{V}_3$ and $\bar{V}_3$ by γ. If γ is large enough (but still $\bar{V}_3 + \gamma < \bar{V}$), then (6) *holds (and* (7) *and* (8) *remain to hold), so that the unique solution is for the DM to take action x^* in period 2 only, i.e., the DM reduces their attention to period 3 in period 1.*

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

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

Since $\underline{V}_3 < \bar{V}_3 < \bar{V}$, there exists $\lambda > 0$ such that (8) *holds with equality. For such λ, since $\underline{\alpha}_2 > 0$, there exists $\eta_2 < \eta_1$ (i.e.,* (7) *holds) so that* (6) *does not hold. Furthermore, λ can be slightly decreased so that* (8) *now holds strictly and* (6) *still does not hold. Thus, for these parameter values, the unique solution is for the DM to take action x^* in period 1 and hence devote $\eta_1 > 0$ to period 3.*

Now decrease λ to something still strictly positive, but so that (6) *holds. As before, the DM now takes action x^* in period 2. Of course, the unweighted consumption payoff is unchanged. However, all comparisons in our constructions are strict; thus, assuming that taking action x^* in period 2 only leads to a payoff of $\bar{V}_3 - \epsilon$ in period 3 does not change the construction for $\epsilon > 0$ small enough. In this case, decreasing λ*

leads to a decrease in the unweighted consumption payoff.

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

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

Consider the setup of Section 2.3 and suppose $\mathcal{D} = \{c, m\}$. Consumption dimension c corresponds to the DM purchasing a quantity of a consumption good at a unit price of 1. Their valuation of quantity k is $\theta u(k)$, where u is strictly concave and continuously differentiable, and $\theta \in \{\theta_L, \theta_H\}$ with $P(\theta = \theta_H) = p \in (0, 1)$. The DM has wealth 1 available, and whatever amount they do not consume, $1 - k$, leads to payoff $1 - k$ as part of dimension m (the "money" problem).

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

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

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

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

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

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

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

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

B Attention utility in a strategic environment

In this section, we extend the environment of Section 2.4 to allow for strategic interaction. We use a setup similar to that of Brunnermeier and Parker (2005) Section III, suitably adjusted to our model. That is, there is a unit mass of agents with the same continuously differentiable and increasing Bernoulli utility function u situated in an exchange economy with no aggregate risk. Each agent is initially endowed with one unit of a safe and can purchase a risky asset that is in zero net supply. The price of the risky asset is P and determined in equilibrium. The risky asset's net return is random and denoted by x^r with payoff x^r_i in state i, where $x^r_s \neq x^r_{s'}$ for all $s \neq s'$. We denote a generic state and agent by s and i, respectively,

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

$$\sum_s p_s u(c_s) \tag{9}$$

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

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

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

Suppose $|\mathcal{D}| > 1$. Let $\bar{s} = \arg\max_{s\in\mathcal{D}} x^r_s$ and $\underline{s} = \arg\min_{s\in\mathcal{D}} x^r_s$. First note that $\xi^i \neq 0$ as thus it must be that $\alpha^i_{\bar{s}} = 1$ or $\alpha^i_{\underline{s}} = 1$. To see this, suppose $\xi^i = 0$. Given $\xi^i = 0$, both $\alpha^i_{\bar{s}} = 1$ and $\alpha^i_{\underline{s}} = 1$ are optimal, but $\xi^i = 0$ cannot be optimal for both $\alpha^i_{\bar{s}} = 1$ and $\alpha^i_{\underline{s}} = 1$, a contradiction. Conditional on $\alpha^i_{\bar{s}} = 1$ ($\alpha^i_{\underline{s}} = 1$), the payoff (9) is continuously decreasing (increasing) in P. Furthermore, for large enough P, the payoff given $\alpha^i_{\bar{s}} = 1$ is less than that given $\alpha^i_{\underline{s}} = 1$ with optimal ξ^i. Thus, there exists a unique P^* such that each agent i is indifferent between $\alpha^i_{\bar{s}} = 1$ and $\alpha^i_{\underline{s}} = 1$. Furthermore, at P^*, it must be that $\alpha^i_{\bar{s}} = 1$ implies be that $\xi^i > 0$ and $\alpha^i_{\underline{s}} = 1$ implies be that $\xi^i < 0$ (otherwise, the associated payoffs cannot be the same). But then

$\int_i \xi^i = 0$ is easily achieved by assigning the appropriate masses of agents to either $\alpha^i_{\bar{s}} = 1$ or $\alpha^i_{\underline{s}} = 1$ with the corresponding optimal ξ^i. The heterogeneity in attention implies the heterogeneity in subjective beliefs. $\square$

C Proofs

C.1 Proposition 1

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

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

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

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

Proof of Proposition 1.* Take any γ_i', γ_i with $\gamma_i' > \gamma_i$ and β_i. Let (x, α) and (x', α')

denote a solution given γ_i and γ_i', respectively. Optimality of (x, α) and (x', α') implies

$$
\begin{aligned}
&\underbrace{\sum_{j \in \mathcal{D}\setminus\{i\}} (\omega_j + \lambda(\alpha_j + \psi_j))V_j(x)}_{:=\kappa_0} + (\omega_i + \lambda(\alpha_i + \psi_i))(\beta_i \tilde{V}_i(x) + \gamma_i) \\
\geq &\underbrace{\sum_{j \in \mathcal{D}\setminus\{i\}} (\omega_j + \lambda(\alpha_j' + \psi_j))V_j(x')}_{:=\kappa_1} + (\omega_i + \lambda(\alpha_i' + \psi_i))(\beta_i \tilde{V}_i(x') + \gamma_i) \quad \text{and} \\
&\underbrace{\sum_{j \in \mathcal{D}\setminus\{i\}} (\omega_j + \lambda(\alpha_j' + \psi_j))V_j(x')}_{=\kappa_1} + (\omega_i + \lambda(\alpha_i' + \psi_i))(\beta_i \tilde{V}_i(x') + \gamma_i') \\
\geq &\underbrace{\sum_{j \in \mathcal{D}\setminus\{i\}} (\omega_j + \lambda(\alpha_j + \psi_j))V_j(x)}_{=\kappa_0} + (\omega_i + \lambda(\alpha_i + \psi_i))(\beta_i \tilde{V}_i(x) + \gamma_i').
\end{aligned}
$$

Combining the above gives

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

The outer inequality implies

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

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

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

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

(x, α) and (x', α') implies

$$\begin{aligned}
&\underbrace{\sum_{j \in \mathcal{D} \setminus \{i\}} (\omega_j + \lambda(\alpha_j + \psi_j))V_j(x)}_{:=\kappa_2} + (\omega_i + \lambda(\alpha_i + \psi_i))(\beta_i \tilde{V}_i(x) + \gamma_i) \\
\geq &\underbrace{\sum_{j \in \mathcal{D} \setminus \{i\}} (\omega_j + \lambda(\alpha'_j + \psi_j))V_j(x')}_{:=\kappa_3} + (\omega_i + \lambda(\alpha'_i + \psi_i))(\beta_i \tilde{V}_i(x') + \gamma_i) \\
&\underbrace{\sum_{j \in \mathcal{D} \setminus \{i\}} (\omega_j + \lambda(\alpha'_j + \psi_j))V_j(x')}_{=\kappa_3} + (\omega_i + \lambda(\alpha'_i + \psi_i))(\beta'_i \tilde{V}_i(x') + \gamma'_i) \quad \text{and} \\
\geq &\underbrace{\sum_{j \in \mathcal{D} \setminus \{i\}} (\omega_j + \lambda(\alpha_j + \psi_j))V_j(x)}_{=\kappa_2} + (\omega_i + \lambda(\alpha_i + \psi_i))(\beta'_i \tilde{V}_i(x) + \gamma'_i).
\end{aligned}$$

Combining the above and substituting for γ'_i gives

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

The outer inequality implies

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

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

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

C.2 Proof of Proposition 2

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

$$\sum_i \omega_i V_i(x) + \lambda \sum_i (\alpha_i + \psi_i) V_i(x) \geq \sum_i \omega_i V_i(x') + \lambda \sum_i (\alpha'_i + \psi_i) V_i(x'), \quad \text{and}$$
$$\sum_i \omega_i V_i(x') + \lambda' \sum_i (\alpha'_i + \psi_i) V_i(x') \geq \sum_i \omega_i V_i(x) + \lambda' \sum_i (\alpha_i + \psi_i) V_i(x).$$

Combining the above gives

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

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

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

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

and so the second claim follows.

Now suppose the environment is separable, $\omega_i = 1$ and $\psi_i = 0$ for all $i \in \mathcal{D}$, and that the objective given λ is convex in α. We show that if $\sum_i \hat{V}_i(\alpha_i)$ is convex in α_i, then so is $\sum_i \alpha_i \hat{V}_i(\alpha_i)$. Take any $\chi \in [0,1]$ and α_i, α'_i with $\alpha_i < \alpha'_i$. Then

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

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

C.3 Proof of Proposition 4

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

$$\frac{p_{\bar{D}_x} + \lambda}{1+\lambda} u(H(x)) + \sum_{i \in \underline{D}} \frac{p_i}{1+\lambda} u(x_i) \geq u(y),$$

where the DM optimally devotes full attention to the states with the highest payoff, $\bar{D}(x) \coloneqq \{i \in \mathcal{D} : x_i = H(x)\}$, and $\underline{D}(x) = \mathcal{D} \setminus \bar{D}$. We rewrite the above as the expected material utility plus attention utility, each divided by $1+\lambda$, i.e.,

$$\frac{1}{1+\lambda} \sum_i p_i u(x_i) + \frac{\lambda}{1+\lambda} u(H(x)).$$

As $u(H(x)) \geq \sum_i u(x_i)$, the above is increasing in λ and so DM(λ') also prefers x to δ_y.

For the second claim, suppose $\lambda > 0$ and take any μ, L and $x \in X(\mu, L)$. Consider lottery $x' \in X(\mu, L)$; we can bound the DM's utility from x' as

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

Since u is unbounded and $\lambda > 0$, the above goes to infinity as $H(x')$ goes to infinity. Thus, there exists some $\bar{H}$ such that for all x' with $H(x') > \bar{H}$, the DM prefers x' to x.

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

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

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 claim follows.

Lastly, take any μ, H and $x, x' \in Y(\mu, H)$. For any λ, the DM devotes full attention to the states with the high payoff. Thus, they prefer x to x' if and only if

$$\begin{aligned}\frac{p_{\bar{D}(x)} + \lambda}{1+\lambda} u(H) + \frac{p_{\underline{D}(x)}}{1+\lambda} u(L(x)) \geq & \frac{p_{\bar{D}(x')} + \lambda}{1+\lambda} u(H) + \frac{p_{\underline{D}(x')}}{1+\lambda} u(L(x')) \\ p_{\bar{D}(x)} u(H) + p_{\underline{D}(x)} u(L(x)) \geq & p_{\bar{D}(x')} u(H) + p_{\underline{D}(x')} u(L(x'));\end{aligned}$$

i.e., independently of λ. □

C.4 Proof of Proposition 5

Proof of Proposition 5. The first case is proved in the text leading up to the proposition.

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

$$\frac{p_i + \lambda \alpha_i}{1+\lambda} \hat{V}(\alpha_i) + \frac{1 - p_i + \lambda(1 - \alpha_i)}{1+\lambda} \hat{V}(1 - \alpha_i).$$

Differentiating the above gives

$$\frac{(p_i+\lambda\alpha_i)\frac{\partial}{\partial a}\hat{V}(\alpha_i)-((1-p_i)+\lambda(1-\alpha_i))\frac{\partial}{\partial a}\hat{V}(1-\alpha_i)}{1+\lambda}+\frac{\lambda(\hat{V}(\alpha_i)-\hat{V}(1-\alpha_i))}{1+\lambda}.$$

For any $p_i > 0$, the DM chooses $\alpha_i > 0$ since $\frac{\partial}{\partial a}\hat{V}(0) = \infty$ and $\frac{\partial}{\partial a}\tilde{V}(1) < \infty$ and so the above is strictly positive. Also note that the above is decreasing in p_i and for $p_i = 0$ and as $\alpha_i \to 0$ it tends to infinity. Thus, there exists $\bar{\alpha} > 0$ such that the above is strictly positive for all $\alpha_i \in (0, \bar{\alpha})$ for any p_i and so the DM chooses $\alpha_i \geq \bar{\alpha}$ for all p_i. Thus, for $0 < p_i < \bar{\alpha}$, we have $\alpha_i > p_i$ and thus $q_i(p_i) > p_i$. (If $q_i(p_i)$ is a set, then the comparison applies to each element of $q_i(p_i)$.) The remaining comparisons follow from the symmetry of q_s. □

C.5 Proof of Proposition 6

Proof of Proposition 6. Notice that when $\lambda = 0$, the DM, in each period t, maximizes the sum of 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 utility continuously; hence, for λ small enough, the above still maximizes the DM's overall utility in each period. Furthermore, this attention allocation is implementable in equilibrium. Hence, the first claim follows.

Normalizing (2) by $1 + \lambda$, when $\lambda = \infty$, the DM's overall utility 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 material utility 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 utility continuously, increasing the weight on material utility, the claim follows. □

C.6 Proof of Proposition 7

Proof of Proposition 7. The first claim is obvious.

For the second claim, fix x_1 and consider a realized s. Clearly, if $\max_{x_2 \in X_2(\alpha_2)} V_{c_{nt}}(x_1, x_2|s) \geq V_{c_t}$, solving (3) gives $\alpha_{2\to c_{nt}} = 1$ (for any λ). If $\max_{x_2 \in X_2(\alpha_2)} V_{c_{nt}}(x_1, x_2|s) < V_{c_t}$, then (3) for $\alpha_{2\to c_{nt}} = 0$ (and some finite $V_{c_{nt}}(x_1, x_2|s)$) is larger than (3) for any $\alpha_{2\to c_{nt}} \geq \eta$ for $\max_{x_2 \in X_2(\alpha_2)} V_{c_{nt}}(x_1, x_2|s)$ when λ is large enough. Since $\mathcal{D}$ is finite, taking the max λ implies the result. □

C.7 Proof of Proposition 8

Proof of Proposition 8. Consider $x_1' \notin \arg\max_{x_1' \in X_1} F(x_1')$. For λ large enough, by Proposition 7, and since $\max_{(x_2,\alpha_2), x_2 \in X_2(\alpha_2)} \tilde{V}_{c_{nt}}(x_1', x_2|s) + \beta F(x_1') < \max_{(x_2,\alpha_2), x_2 \in X_2(\alpha_2)} \tilde{V}_{c_{nt}}(x_1, x_2|s) + \beta F(x_1)$, we have $B(x_1') \subseteq B(x_1) \neq$ (where non-emptiness follows from $V_{c_{nt}}(x_1, x_2(x_1, s)) > V_{c_t}$) and hence, again for λ large, the DM's period-2 utility (i.e., (3)) is strictly larger with x_1 than with x_1'.

The DM's objective in period 1 can be written as the sum of (3) and their period-1 attention utility. But the latter is also increasing when choosing x_1 instead of x_1', for large enough λ, and so the result follows. □

C.8 Proof of Proposition 9

Proof of Proposition 9. The first claim follows immediately from the DM's period-2 objective(4).

For the second claim, if $p_1 \geq \bar{p}$, then $V_{c_{nt}} \geq V_{c_t}$, and so devoting full attention $\alpha_{1\to c_{nt}}$ maximizes the DM's period-1 attention utility. Furthermore, period-2 utility is increasing in $\alpha_{1\to c_{nt}}$ (for optimally chosen x_1) as it is convex in p_2. Hence, the DM devotes full attention.

Note that the DM's utility is continuous in p_1, and that hey not devote attention to c_{nt} in either period if $p_1 = 0$. Thus, there exists some $\tilde{p} \leq \bar{p}$ such that it is optimal to devote attention to c_{nt} 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.[41] Let

[41] Any posterior distribution can be replaced with a binary distribution with values $E[p|p_2 \geq \bar{p}]$

$p_H > p_1$ and $p_L < p_1$ be the posteriors that result from the acquired information given prior p_1. Let the DM with prior p_1' choose a distribution of posteriors with mass points at p_H, p_L and p_1' such that the probability of the high posterior occurring is held constant. Note that this distribution has a lower variance than that given p_1. It leads to the same increase in expected period-2 attention utility. Lastly, since $V_{c_{nt}}(p_1) < V_{c_{nt}(p_1')}$, period-1 attention to c_{nt} leads to a smaller decrease in period-1 attention utility given p_1' than p_1. Combining all these observations, it is optimal to acquire some information given p_1'. $\tilde{p}$ is decreasing in β since the DM's objective has increasing differences in $(\alpha_{1\to c_{nt}}, \beta)$

We next prove the skewness result. Suppose that $p_1 > \bar{p}$. Then, $\alpha_{1\to c_{nt}} = 1$. It is uniquely optimal to acquire a binary signal. Let p_L and p_H denote the posteriors. The variance of such posteriors is given by $P(p_2 = p_L)P(p_2 = p_H)(p_H - p_L)^2$. 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 β small enough, it must be that either the high or low posterior is likely. 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. Thus, $P(p_2 = p_H) > 1/2$, and so the distribution of posteriors is negatively skewed.

Now suppose that $p_1 \in (\tilde{p}, \bar{p})$. For optimal x_1, it must be again 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 is likely. But the high posterior cannot be likely since it is bounded away from p_1 (as $p_1 < \bar{p} < p_H$). Thus, $P(p_2 = p_L) > 1/2$, and so the distribution of posteriors is positively skewed. □

C.9 Proof of Proposition 10

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

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

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 leads to the period-2 utility.

If $p(\alpha_i)+\alpha_i \frac{\partial}{\partial \alpha_i} p(\alpha_i) < 1$ everywhere, then the above becomes positive for large enough λ and the claim follows from Topkis's theorem.[42] $\square$

C.10 Proof of Proposition 11

Proof of Proposition 11. Fix any (x, α) with $x \in X(\alpha)$. Take any $D, D' \in B$ and consider $B' := (B \cup \{D \cup D'\}) \setminus \{D, D'\}$. Evaluate (5) at (x, α) and B' and subtract its value given (x, α) and B; after some simplifications, we have

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

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

[42] E.g., take $\lambda > \frac{\max_{\alpha_i} \frac{\partial}{\partial \alpha_i} p(\alpha_i)}{\min_{\alpha_i}(1-p(\alpha_i)+\alpha_i \frac{\partial}{\partial \alpha_i} p(\alpha_i))}$.