Emotional Inattention

Lukas Bolte* Collin Raymond*

August 12, 2024

Abstract

We propose a framework where a decision-maker allocates attention across payoff-dimensions, such as different consumption decisions, states of the world, or time periods. Attention to a dimension is instrumentally valuable, as it enables better decisions, but also leads to an emotional response that scales with the amount of attention and the payoff of the dimension. The framework predicts novel forms of well-known biases, including optimism, subjective probability weighting, and dynamic inconsistency, and provides a unifying explanation for these and other behavioral phenomena, such as the ostrich effect, incomplete consumption smoothing, and default effects.

JEL Classification Codes: D81, D83

Keywords: Attention, attention utility, information, adaptive preferences

---

*Lukas Bolte; Department of Social and Decision Sciences, Carnegie Mellon University, Pittsburgh, PA; lukas.bolte@outlook.com. Collin Raymond; SC Johnson Graduate School of Management, Cornell University, Ithaca, NY; collinbraymond@gmail.com. We thank Roland Bénabou, Douglas Bernheim, Andrew Caplin, Gabriel Carroll, Russell Golman, Matthew Jackson, George Loewenstein, Yusufcan Masatlioglu, Muriel Niederle, Joshua Tasoff, Mu Zhang, and seminar participants at UC Berkeley, Berlin, CMU, ITAM, LMU Munich, Newcastle, Nottingham, Princeton, Stanford, Tilburg, UC Santa Barbara, Université de Montréal, University of Miami, Virginia Tech, BEAM, BRIC, Conference on Belief-Based Utility, ESA, M-BEES/M-BEPS, RUD and SABE for their helpful comments. Bolte gratefully acknowledges financial support from the Leonard W. Ely and Shirley R. Ely Graduate Student Fellowship.

[...] each of us literally chooses, by his ways of attending to things, what sort of a universe he shall appear to himself to inhabit.

William James (1890)

1 Introduction

Attention is an important input into economic decision-making, allowing individuals to reason, process information, and (consciously) take action. This instrumental role of attention has been extensively studied by psychologists (since at least James (1890); see Desimone et al. (1995) for a review) and economists (e.g., Sims (2003); Loewenstein and Wojtowicz (2023)).

However, work in psychology and cognitive science highlights an important second aspect of attention, one that is typically neglected by economists: Attention generates and regulates emotions (see Dixon et al. (2017) and Gross (1998) for reviews).[1] For instance, attending to a news article about recent stock market losses may lead to a negative visceral reaction, while focusing on an upcoming vacation can generate excitement.

This paper develops a tractable model that incorporates both the instrumental and the less attended-to emotional aspects of attention. Substantively, we make three contributions. First, we demonstrate the key implications of incorporating attention's emotional role in a general setting. Second, applying our model to different economic environments, we show that the emotional aspect of attention can serve as a unifying mechanism for a large number of well-known behavioral anomalies: Across different environments, our model predicts that agents exhibit the "ostrich effect" and avoid thinking about situations with low payoffs (such as poorly performing portfolio, Karlsson et al. (2009), or health tests, Oster et al. (2013); Ganguly and Tasoff (2017)), optimism and other forms of subjective probability weighting (Brunnermeier and Parker, 2005; Sharot, 2011; Kahneman and Tversky, 1979), incomplete consumption smoothing and memorable consumption (Hai et al., 2020), dynamic inconsistency (Laibson, 1997), and default effects (Carroll et al., 2009). Although these phenomena

[1]That is not to say economists have completely ignored attention's emotional role. Schelling (1988) highlights the role of the mind as a "pleasure machine or consuming organ, the generator of direct consumer satisfaction," as well as an "information processing and reasoning machine." Schelling also implicitly suggests that these roles should be considered jointly and writes: "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."

are often modeled as emerging from distinct psychological considerations, our results point out that a single cognitive mechanism may underpin many seemingly disparate behavioral patterns. Third, we demonstrate how emotional inattention gives rise to new forms of well-known biases. For example, our model predicts payoff-level dependent dynamic inconsistency and a novel form of attentional unraveling where future inattention generates current inattention. The model also predicts new forms of optimistic behavior, where agents asymmetrically avoid and engage with situations and payoffs, leading to novel predictions about as-if probability weighting and engagement with defaults and incentives.

We can distinguish our approach from alternative mechanisms in three ways. First, via distinct predictions within an environment: E.g., other models of dynamic inconsistency (such as quasi-hyperbolic discounting and anticipatory utility) do not feature payoff-level dependency or unraveling; similarly, alternative approaches to motivated beliefs and probability weighting do not generate the distinct interplay between the payoff levels and the instrumental value of states and lotteries that gives rise to subjective probability weights in our model. Second, our model can link behaviors across different domains—we know of no other approach that simultaneously explains anomalies in the domains of risk, time, and consumption decisions. Third, our model posits relationships between the environment, measured attention, and observed choice, which can match recent choice process data, such as financial log-ins (Quispe-Torreblanca et al., 2020) or visual attention (Bhatnagar and Orquin, 2022).

We operationalize the two aspects of attention as follows. The decision-maker (henceforth, DM—they) devotes attention across a number of dimensions, each associated with a payoff. The attention allocation determines which actions are available to the DM, and the action taken (which may be multi-dimensional) affects the payoff from each dimension. This formulation captures the instrumental role of attention in a reduced form and nests situations where attention leads to information acquisition as well as when attention is required to execute an action (e.g., adjust an investment portfolio), even absent any information acquisition.

The DM derives two kinds of utility. First, they directly value the payoffs from the different dimensions as "material utility," which is the utility the DM derives in a model without attention's emotional consequences. Second, and crucially, the DM values "attention utility," capturing the emotional role of attention. We assume that each dimension generates attention utility that is proportional to both the dimension's payoff and the amount of attention devoted to it. This may be due to feelings of anticipation; e.g., when the DM devotes attention to an upcoming vacation, they receive some additional utility because they think about the enjoyable activities they will do. It may also occur due to memory; e.g., a decision-maker may recall past vacations in order to help with planning the current one, and revisiting these memories is pleasant. In spirit, this approach is similar, albeit broader, 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 and other flow utilities from attention, such as that from thinking about a past memory, a function of the amount of attention paid. The total weight a dimension (and its payoff) takes in the DM's objective is thus determined by the attention allocation, which formalizes the sense in which individuals choose "what sort of a universe [...] to inhabit" as mentioned by William James.

We formally introduce our framework in Section 2.1, and derive general properties of the optimal attention allocation in Section 2.2. A key "standard" result carries over: 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 payoffs across dimensions: Ceteris paribus, the DM devotes more attention to dimensions with a higher payoff. 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 payoffs. Moreover, because the DM can re-allocate attention to high payoffs, attention utility implies a preference for varied payoffs across dimensions, and since increased attention, in turn, implies higher payoffs, this endogenously generates "sparse" attention allocations, as in Gabaix (2014).

Section 2.3 supposes that the dimensions correspond to different consumption decisions. In this setting, our results imply the well-documented ostrich effect: Individuals tend to be inattentive to (and possibly avoid information about) consumption decisions 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)), in the political domain (D'Amico and Tabellini, 2022), as well as in the lab (Avoyan and Schotter, 2020). We show that attention's emotional role can not only explain ostrich behavior a la information avoidance (which is what extant models have focused on) but also when attention has no impact on beliefs, which existing explanations have more difficulty rationalizing.

In Section 2.4, we let different dimensions correspond to different states of the world. In this context, the attention-dependent weight placed on a state leads to as-if subjective probability weighting, even though our DM understands the true probabilities perfectly. When attention is non-instrumental, the DM devotes all their attention to high-payoff states while ignoring the ones with low payoffs, leading to optimism (Sharot, 2011) and a preference for positively-skewed lotteries. For instance, our DM is willing to buy a lottery ticket (more generally, invest in a risky, potentially lucrative asset) because they can devote attention to the state where they win the jackpot (with evidence documented in Blume and Friend (1975); Garrett and Sobel (1999); Forrest et al. (2002)). But attention's emotional role can also lead to other forms of subjective probability weighting, as in Kahneman and Tversky (1979), due to its interaction with attention's instrumental role, differentiating our predictions from those of models with motivated beliefs (e.g., Brunnermeier and Parker (2005)). For example, when the material returns to attention to a state are concave, our DM can exhibit the widely documented "inverse-S" shaped probability weighting, where low probabilities are over- and high ones under-weighted. Thus, we provide an alternative foundation for probability weighting, and unlike existing approaches, our model links the details of the economic environment to changes in the subjective weights via the attention allocation.

In Section 3, we extend our model to dynamic settings where a dimension corresponds to a time period, and the DM chooses an attention allocation in multiple periods. 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. We also show that the emotional role of attention naturally leads to a preference for memorable consumption: The DM will intersperse periods where consumption is smoothed with occasional periods that feature high levels of consumption and devote a disproportionate share of attention to these periods.

In Section 4, we demonstrate how our model can tractably be used in a variety of standard economic environments that combine different types of dimensions discussed in Sections 2 and 3. First, in Section 4.1, we show that emotional inattention leads to dynamic inconsistency and attentional unraveling, leading to situations where both selves are worse off relative to the commitment allocation. In Section 4.2, we show how our DM responds in an asymmetric fashion to increases in incentives: Although additional bonuses can increase attention to a task, increasing penalties can have the opposite effect, implying that schemes that rely on future punishment (e.g., many commitment devices) may be counterproductive. Last, in Section 4.3, we use our model to study defaults, such as those commonly observed in savings plans or portfolio adjustment problems, and, typically, changing the default action requires some attention. Such defaults impact our DM in two ways. First, there is an asymmetric default effect: Because of the attentional cost of attending to a low-payoff decision, the default binds if the associated payoff is low but not when it is high. For instance, defaults for end-of-life medical treatments will matter, whereas they will matter less for planning a vacation. Second, because of our DM's desire to focus on high-payoff futures, they may choose defaults that are too optimistic. Thus, our DM suffers from dual distortions in low-payoff situations: They rely on the default in lower-payoff states yet tailor it to higher-payoff states.

Section 5 discusses extensions and situates our paper in the broader literature. Section 5.1 highlights potential limitations and considers extensions of our simple model, including relaxing conceptual and functional form assumptions, as well as how to endogenously group dimensions; formal discussion of these occurs in Appendices D and B. Section 5.2 highlights how one can distinguish our model from extant related approaches, including rational inattention, anticipatory utility, motivated beliefs, recursive preferences, chosen preferences, and others. Section 6 concludes.

2 Model

We consider a decision-maker (henceforth, DM—they) who allocates attention and chooses an action. We develop a model that captures the two fundamental features of attention: (i) instrumental, where attention determines which actions are available to the DM, and (ii) emotional, where attention generates attention utility. We then characterize the optimal attention allocation and explore our model's implications in two canonical decision problems: a deterministic problem with multiple consumption decisions and a problem with an uncertain state.

2.1 Setup

The DM faces a finite number of dimensions indexed by $i \in \mathcal{D}$.[2] A dimension can correspond to a dimension of consumption, a realization of an unknown state, a time period, or a combination of these; keeping the model general allows us to nest various interpretations in a wide range of applications. Each dimension i is associated with a payoff V_i. The DM chooses an (action, attention)-pair denoted by (x, α). Action x, chosen from a compact topological space $\mathcal{X}$, determines the payoff associated with a dimension i, i.e., $V_i(x)$, where $V_i(\cdot)$ is continuous. Attention $\alpha = (\alpha_i)_{i \in \mathcal{D}}$ is a measure on the set of dimensions, with α_i denoting the attention devoted to dimension i, and where we impose $\alpha_i \geq 0$ and $\sum_i \alpha_i = 1$.[3]

Attention has two implications. First, it is instrumentally valuable. To capture this, we let the available actions depend on the attention allocation: Given α, action x is chosen from a set $X(\alpha) \subseteq \mathcal{X}$, where $X(\cdot)$ is compact- and non-empty-valued and upper hemicontinuous, ensuring an optimum will exist. Second, attention directly generates utility. Specifically, attention to a dimension i generates attention utility, which we take to be proportional to the attention devoted to i and i's payoff, i.e., it is given by $\alpha_i V_i(x)$. Depending on the setting, 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, or as capturing how attention enhances contemporaneous consumption (Capra et al., 2023).

The relative importance of attention utility to the usual standard payoffs, which we call material utility, is given by a parameter λ. 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."

In general, the DM's objective is the weighted sum of material utility and attention

[2]We take the dimensions as given. In practice, the boundaries between dimensions may not always be obvious. In Appendix B, we study a meta-optimization problem in which the DM chooses how to define a dimension, and in Section 6, we provide some guidance as to how the primitives of our model can be identified from data.

[3]Alternatively, one can impose an upper bound on the measure of attention. By adding a trivial dimension with payoff 0 to $\mathcal{D}$, our model becomes equivalent to one with an upper bound.

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 future payoffs. Parameter ψ_i is used in Section 3 to capture the amount of attention the DM's future "selves" devote to a period i. In static environments, it is natural to set $\psi_i = 0$.[4]

Our approach requires that we specify how the action set varies with attention—$X(\alpha)$. In many settings, this mapping requires no additional degrees of freedom compared to standard rational inattention models. This is because our approach allows us to nest situations where action sets explicitly vary with attention, as well as, through a suitable redefinition of variables, situations where the action set is fixed, but the choice of action within the set depends on attention (potentially stochastically), as is the case when attention allows for information acquisition. Examples 1–3 in Appendix A.1 provide details of how to convert canonical settings of rational inattention, trembling, and memory recall into our framework. Moreover, only one of our results (Proposition 10) requires us to specify how the action set varies with attention, while the rest only require mild technical and monotonicity conditions, and in this one case, the mapping is quite natural.

Our model embeds a variety of other assumptions, including the functional form of utility, the fact that individuals can solve the optimization problem of Equation (1) without devoting attention, and assuming attention is fully controllable by the DM. We discuss these assumptions and how to relax them in Section 5.1 and Appendix D (which focuses on functional form assumptions).

We next study the DM's choice of attention and action when they are jointly chosen to maximize (1). Note, however, that in our model, fixing the attention allocation, the utility an action generates depends on the attention allocation (as differential attention leads to differential weighting of the payoff dimensions). Thus, our model differs from the standard one, even when our DM cannot choose attention,

[4] 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.

with our DM choosing actions that increase payoffs in dimensions that attract high attention (similar to recent "bottom-up" approaches to attention, as discussed in Section 5.1).

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—and underlie the mechanisms that drive different behavioral phenomena that we discuss in detail later. 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. We show that these results continue to hold under less restrictive functional form assumptions in Appendix D.

To strengthen some statements, we introduce the notion of a *separable environment*, which captures a natural restriction on how actions relate to payoffs. In words: An 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 (as a notational convention, for any variable that is indexed by $i \in \mathcal{D}$, e.g., x_i, we let $x_{-i} := (x_{i'})_{i' \in \mathcal{D} \setminus \{i\}}$):

Definition 1. *The environment is* separable *if: (i) 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 $X(\alpha) = \Pi_{i \in \mathcal{D}} X_i(\alpha_i)$, and (ii) X_i is monotone, i.e., $X_i(\alpha_i) \subseteq X_i(\alpha_i')$ for all $\alpha_i \leq \alpha_i'$.*

Under separability, payoffs can be written more simply as a function of the attention allocation directly. Specifically, maximizing (1) with respect to an (action, attention)-pair is equivalent to maximizing $\sum_i (\omega_i + \lambda(\alpha_i + \psi_i))\hat{V}_i(\alpha_i)$ with respect to attention only, where $\hat{V}_i(\alpha_i) := \max_{x_i \in X_i(\alpha_i)} V_i(x_i, \cdot)$. Payoff $\hat{V}_i$ is increasing in attention to dimension i because of X_i's monotonicity.

Throughout the rest of Section 2, we assume that the solution is unique. This is purely for expositional ease—we state and prove general versions of the propositions in Appendix C.

We begin our formal results by considering what happens when we vary the payoff V_i. For each i, we fix some function v_i (of action x) and define $V_i := \beta_i v_i + \gamma_i$, for scalars $\beta_i \geq 0$ and γ_i. Increasing γ_i increases the payoff level, and increasing β_i increases the payoff difference from 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 DM's (i.e., $\lambda = 0$) solution. However, the attention utility from dimension i increases in proportion to the attention devoted to it. So when the DM puts positive weight on attention utility (i.e., $\lambda > 0$), they devote more attention to the improved dimension i. If the environment is separable, this increase in attention, in turn, leads to a better action for that dimension, i.e., the value of 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 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 dimension i. If the environment is separable, this 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. Our first proposition formalizes.

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

- *If $\gamma_i' \geq \gamma_i$ and $\beta_i = \beta_i'$, then $\alpha_i' \geq \alpha_i$. If, in addition, the environment is separable, then $v_i(x') \geq v_i(x)$.*
- *If $\beta_i' \geq \beta_i$ and $\gamma_i' = \gamma_i - (\beta_i' - \beta_i)v_i(x)$, then $v_i(x') \geq v_i(x)$. If, in addition, the environment is separable, then $\alpha_i' \geq \alpha_i$.*[5]

We next turn to the relative weight on attention utility λ and show three results.

Proposition 2. *Consider a change of parameter λ to λ' with $\lambda' > \lambda$ and let x and x' denote the optimal actions, respectively. Then: (i) $\sum_i \omega_i V_i(x) \geq \sum_i \omega_i V_i(x')$, (ii) the DM's value is convex in $(\gamma_i)_{i \in \mathcal{D}}$, and (iii) if the environment is separable, for each $i \in \mathcal{D}$, if the objective given λ is convex in α_i, then it is also convex in α_i given λ'.*[6]

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

[6]It is not true that increasing λ preserves joint convexity in α.

The first part, which says material utility falls with λ, implies that the DM's actions are suboptimal if judged through the lens of the standard model with $\lambda = 0$ (or if attention utility is considered a bias). The second and third parts of the proposition imply a preference for "extreme" payoffs (i.e., payoffs are convex in the level of the payoff of a dimension) and attention allocations (i.e., increasing λ implies that utility is more likely to be convex in the attention to a given dimension). This preference is due to the complementarity between a payoff increase (exogenous or endogenous due to attention) and the weight of the associated dimension in the DM's attention utility: Increasing the payoff of a dimension is particularly valuable if that dimension is heavily weighted 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. In separable environments (third part), attention drives both, so the objective becomes "more convex" relative to the standard model with $\lambda = 0$. Thus, the DM's attention may be naturally "sparse," as in Gabaix (2014), not for the usual instrumental reasons but due to the complementarity of attention's instrumental and emotional roles.

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 (but can capture distinct features—e.g., ω_i can capture the probability of a given dimension occurring, and ψ_i the impact of exogenously determined attention). 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 $(\omega_i', \psi_i') \geq (\omega_i, \psi_i)$ element-wise. 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 $\alpha_i' \geq \alpha_i$.*

Next, we explore the implications of these general results in more specific contexts.

2.3 Attention across consumption dimensions

We now consider attention allocation when dimensions correspond to different consumption decisions. Those may be 'arranging a retirement home for a relative,' 'vacation,' '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 effect: (attentional) avoidance of low-payoff situations and, conversely, excessive attention to high-payoff ones.[7]

Evidence for such behavior has been extensively documented in several domains. In finance, retail investors' propensity to check their portfolios, an act of paying attention, generally comoves with the market (Karlsson et al., 2009; Sicherman et al., 2015).[8] Similar behavior has been documented in the domain of health, where Oster et al. (2013) and Ganguly and Tasoff (2017) document the avoidance of testing for diseases (which would require attention to negative health outcomes) with more serious diseases inducing more avoidance.

Most existing explanations lean on information as a driving mechanism; either in a standard way, where the instrumental value of information, or the cost of information acquisition co-moves with the market, or in a behavioral way, due to belief-based utility, whether from anticipation (Caplin and Leahy, 2001; Brunnermeier and Parker, 2005) or news (Kőszegi and Rabin, 2009; Karlsson et al., 2009) induces avoidance (see Golman et al. (2017) for a survey). However, recent research suggests that the ostrich phenomenon is not just about information but rather the direct utility derived from attention. Using lab data that rules out informational motives, Avoyan and Schotter (2020) show that attention covaries with expected payoffs. In the field, there is also strong evidence for attention utility in the context of retail investors independent of information. Building on evidence in Sicherman et al. (2015), Quispe-Torreblanca et al. (2020) find that investors devote excessive attention to portfolios after positive information that is already known and are willing to provide more feedback about their portfolio (via a survey) when the portfolio is doing well. Similarly, Olafsson and Pagel (2017) find that individuals condition their attention to financial accounts around plausibly already known shifts in balances (e.g., individuals look more at accounts after regular pay-days). In the health domain, individuals often avoid medically recommended 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 indi-

[7] To our knowledge, 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.

[8] Individuals may also 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 payoffs relative to some reference point.

viduals experiencing serious medical conditions, individuals with worse health status tend to adhere less to medical regimes. And in the political domain, partisans post fewer responses to articles about their favorite candidate when they are unfavorable (D'Amico and Tabellini, 2022).

The absence of new information from attention in all these examples renders belief-based utility models mute, and variations in non-emotional costs and benefits seem unlikely. Therefore, our model not only provides explanations for avoidance behaviors that align with existing theories but also rationalizes other behaviors that current explanations cannot explain.

2.4 Attention across states

Next, we 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 as well as probability weighting.

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 realizing. 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.[9]

We briefly discuss some of the general implications of Propositions 1–3 in this environment before relating our model to more concrete behavioral phenomena. In particular, individuals will take actions that are better suited for states with relatively high payoffs (Proposition 1, at least in a separable environment) and for states with a relatively high chance of occurring (Proposition 3), possibly leading to low expected material utility (Proposition 2).[10] Thus, in the context of individuals devoting attention across future contingencies, agents will know what to do with a financial windfall

[9]Because attention utility is independent of the probabilities, our model allows individuals to derive attention utility from 0-probability events, and so in applications, we must be careful when specifying the set of dimensions. In Appendix D, we discuss how our results extend when attention utility from a state can also depend on the probability assigned to that state.

[10]The one nuance in applying Proposition 3 is that there is a constraint on the set of probabilities: Increasing the probability of one state means reducing the probability of another. 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.

(as they have contemplated such contingency) but not which expenses to cut when they are laid off (as this has been ignored).

Next, we highlight two implications of attention utility in situations with risk, focusing on choices over lotteries: non-standard risk attitudes and as-if probability weighting. Take a standard environment: A DM, equipped with an increasing Bernoulli utility u, chooses a lottery from set X, where lottery $x \in X$ leads to monetary payoff of x_i in state i, maximizing $\sum_i p_i u(x_i)$. Consider now our DM in this environment, i.e., our DM's action is now choosing a lottery, $V_i(x) = u(x_i)$, and they also value attention utility in addition to the expected payoff. To isolate the emotional role of attention, we suppose that attention has no instrumental role, i.e., $X(\alpha)$, the set of available lotteries, is constant. However, all parts of the ensuing proposition can be generalized.[11]

The following proposition states that attention's emotional role both reduces risk aversion and generates a preference for lotteries with a high payoff. In other words, attention utility 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.

Proposition 4. *Let DM(λ) refer to the DM with a relative weight λ on attention utility. Then the DM(λ) is more risk-averse than DM(λ') for any $\lambda' > \lambda$. Moreover, denoting $H(x) := \max_i x_i$, for any pair of lotteries x, x': if $H(x) > H(x')$, then the DM strictly prefers x to x' if λ is large enough, but if $H(x) = H(x')$, then the DM's preferences over x, x' are independent of λ.*

Proposition 4 first states that attention utility leads to an additional preference for risk. Intuitively, given a lottery x, the DM devotes attention to the high-payoff states—the "upside" of the lottery—that consequently are relatively overweighted, making the DM as-if optimistic. Such optimism has been documented in a wide range of circumstances, both in the lab and the field (e.g., see Mijović-Prelec and Prelec (2010); Sharot (2011); Mayraz (2011); Oster et al. (2013); Engelmann et al. (2019); Orhun et al. (2021) for evidence in both monetary and non-monetary domains).

The second part of the proposition states that the DM prefers lotteries with higher highest payoffs, and thus, when comparing two binary lotteries with the same mean

[11] For instance, the first part of Proposition 4 goes through allowing for an instrumental role of attention, as long as the DM is able to devote disproportionately much attention to high-payoff states: Let $n(i, x)$ denote the $n(i, x)$-th highest payoff state given some lottery $x \in X(\alpha)$ (with ties broken arbitrarily); the result goes through as long as for all (x, α) with $x \in X(\alpha)$, there exists α' such that $x \in X(\alpha')$ and $\sum_{i:n(i,x)\leq N} \alpha'_i \leq \sum_{i:n(i,x)\leq N} p_i$ for all N.

and low payoff, the DM prefers the more positively skewed (i.e., higher standardized third moment) one. Intuitively, because the DM devotes their attention exclusively to the high payoff state, if the DM puts enough weight on attention utility, the DM then prefers the lottery with the higher high payoff.[12] The third part notes that because the DM devotes attention to highest-payoff states only, attention utility does not affect the DM's preference over lotteries with the same highest payoff.

These results rationalize individuals who simultaneously gamble (e.g., buy low-probability but high payoff lottery tickets) as well as buy insurance against low-probability but high-loss outcomes, in line with extensive evidence from portfolio choice, gambling, and the lab documenting a preference for positive skewness (for examples, see Blume and Friend (1975); Golec and Tamarkin (1998); Ebert and Wiesen (2011); Dertwinkel-Kalt and Köster (2020)). Moreover, consistent with our model, Jullien and Salanié (2000) and Snowberg and Wolfers (2010) suggest that the preference for skewness is driven by subjective probabilities.

The previous proposition is reminiscent of well-known intuitions from the motivated beliefs literature (e.g., Bénabou and Tirole (2002); Brunnermeier and Parker (2005); Bracha and Brown (2012); Caplin and Leahy (2019)). However, emotional inattention makes quite distinct predictions about behavior both for lottery choices and in other situations involving risk. We next show how our model leads to a variety of patterns of as-if probability weighting that motivated reasoning models cannot accommodate. Moreover, in Section 4, we highlight another key distinction: Models of motivated beliefs predict that DMs will engage with all decisions in a way that overweights the upside; in contrast, emotional inattention predicts that DMs will typically engage only with the subset of dimensions that have high payoffs, and simply not engage (e.g., by not taking any action, or sticking with a default) with the remainder. Our model also directly speaks to emerging choice process data, which documents that visual attention is directed towards more preferred options (e.g., see Bhatnagar and Orquin (2022) for a survey, and Glöckner and Herbold (2008) for early evidence showing that within a lottery, higher payoff options attract more visual attention).

To more clearly see how our DM acts as if they use subjective probability weights, take the DM's objective, divide by $1 + \lambda$, and denote the terms in front of V_i as

[12] The preference for high payoffs can also be seen in another way: Consider a binary lottery x' with mean μ and low-payoff L. Then there exists $\bar{H}$ such that for all $\mathcal{D}$ and binary lotteries x, with mean μ, low payoff L and $H(x) > \bar{H}$, the DM prefers x to x'. (Such lottery x may not exist for some $\mathcal{D}$ if $\mathcal{D}$ is "too coarse.")

$q_i(p_i) := \frac{p_i+\lambda\alpha_i}{1+\lambda}$. Note that $q_i \in [0,1]$ for all i and $\sum_i q_i = 1$, i.e., q_i describes a probability measure. The DM, conditional on their attention allocation, behaves like a subjective expected payoff maximizer with measure q. This as-if belief distortion is a function of the attention allocation: As attention to state i increases, so does the subjective probability q_i assigned to that state, and $q_i(p_i) \geq p_i$ if and only if $\alpha_i \geq p_i$. States that attract attention in excess of (less than) their true probability are overweighted (underweighted).

We focus on the situation where there are only two states, $\mathcal{D} = \{i, i'\}$. When there is no instrumental value (as in Proposition 4), the DM devotes full attention to the higher-payoff state, which is subsequently overweighted. To generate more general forms of probability weighting, we allow for attention to be instrumentally valuable and consider separable environments. An example of separability in a lottery environment would be that in order to understand the value of any given outcome in the lottery, the DM must pay attention to it; e.g., they need to process a noisy signal about the value of that outcome, in line with evidence in Woodford (2020); Frydman and Jin (2022) and the broader literature on visual attention (Smith and Krajbich, 2018). Alternatively, the DM could be making realization-contingent plans that are independent of one another (i.e., the plan contingent on winning is of no help for planning conditional on not winning, and vice-versa).

Building on these intuitions, we will assume that the payoffs associated with each dimension feature initially large but decreasing returns to attention. Of course, in line with Proposition 4, our DM will devote their residual attention to the high-payoff state. This implies the probability weighting function is compressed; that is, small (large) probabilities are over (under)-weighted so that probability weighting takes an inverse S shape.[13] The following proposition summarizes (recall that we use $\hat{V}_i$ to denote payoffs when the environment is separable).

Proposition 5. *Suppose $\mathcal{D} = \{i, i'\}$ and that the environment is separable. 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, $q_i(\cdot) = q_{i'}(\cdot) = q(\cdot)$ and there exists some $\bar{p}$ with $0 < \bar{p} < 1/2$, such that: $q(0) = 0$, $q(1) = 1$, $q(p) > p$ if $0 < p < \bar{p}$ and $q(p) < p$ if $1 - \bar{p} < p < 1$.*[14]

[13] In Appendix D, we discuss how decreasing returns to attention in the attention utility term, rather than in determining payoffs, also leads to probability weighting.

[14] 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

Figure 1 illustrates. Panel (a) shows the optimal level of attention α_i^* devoted to state i as a function of the probability p_i of that state occurring; panel (b) shows the resulting probability weighting, $q_i(p_i)$. We choose $\hat{V}(a) = -\frac{1}{a}$ as tractable functional form, as it implies $\alpha_i^* = (p_i - \sqrt{p_i(1-p_i)})/(2p_i - 1)$ which is inverse S-shaped and hence so is q_i.

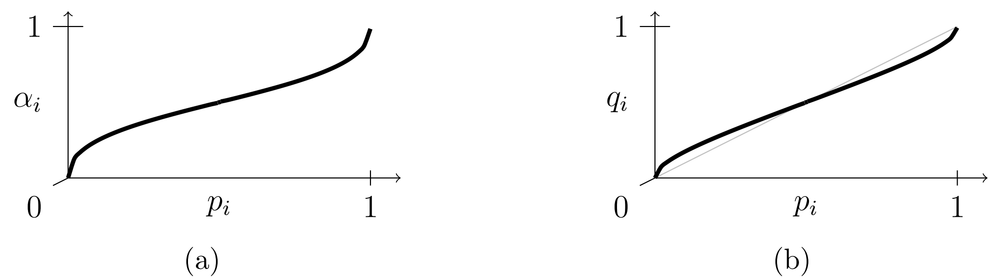

Figure 1: This figure visualizes Proposition 5. We have $\lambda = 1$ and $\hat{V}(a) = -\frac{1}{a}$. Consequently, $\alpha_i^* = (p_i - \sqrt{p_i(1-p_i)})/(2p_i - 1)$ which is inverse S-shaped (Panel (a)) and hence so is q_i (Panel (b)).

Our approach suggests that, in conjunction with concave returns to understanding, the emotional returns to attention can serve as a microfoundation for probability weighting, in particular the classic finding of an inverse S-shaped weighting function (as in Kahneman and Tversky (1979), see Wu and Gonzalez (1996) for empirical evidence). Existing models of noisy cognition can also lead to weighting functions (see Frydman and Jin (2023)), with the shape of the weighting function depending on prior beliefs about probabilities. In contrast, however, our results rely on decreasing returns to attention, which we view as empirically plausible.

Our model differs from existing approaches. Unlike cumulative prospect theory (Tverseky and Kahneman, 1992) and rank-dependent utility (Quiggin, 1982), our model can explain why subjective weights may vary with payoff differentials, not just payoff ranks. Moreover, although our model can replicate patterns of optimism arising from models of motivated beliefs, Proposition 5 shows our model can generate patterns of probability weighting that are different.

Interestingly, there are environments where our model makes predictions distinct from both those of motivated beliefs models and the standard parameterization of cumulative prospect theory: If $\hat{V}_i = \hat{V}_{i'} = \hat{V}$, where $\hat{V}$ is strictly convex, then the DM

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-shaped probability weighting everywhere.

optimally devotes attention to the more likely state, leading them to overweight high and underweight low probabilities, S-shaped probability weighting.

3 Attention across time

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 to intertemporal choice. We show that our model endogenizes temporal preferences and can rationalize non-smooth consumption paths.

The DM faces a sequence of time periods $\mathcal{D} = \{1, \ldots, T\}$, with generic period t. For simplicity, we assume that there is no exogenous discounting, i.e., $\omega_t = 1$ for all t. We do so to highlight the effect of the attention-dependent weights on the different periods (the dimensions) for the as-if time preferences; however, our results can be extended easily to allow for standard exogenous temporal preferences. In each period t, the DM chooses an (action, attention)-pair denoted by (x_t, α_t). The actions jointly determine the payoffs across periods: Given $x := (x_t)_{t=1}^T$, the payoff in period t is $V_t(x)$ (one can impose natural restrictions on how future actions impact past payoffs). 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 $X_t(\alpha_t)$, where $X_t(\cdot)$ is compact- and non-empty-valued and upper hemicontinuous. Thus, attention at t can improve payoffs at t' because it allows for a different x_t, which impacts $V_{t'}(x)$. Informally, in each period t, the allocation of attention determines the set of actions that can be taken in t (but do not impact the actions available in t'), and the action taken in period t can impact payoffs in all other periods. (An alternative way of allowing for intertemporal dependence of payoffs on attention would be to have X_t depend on the attention devoted to t in all other time periods but have the payoff in period t only depend on action x_t.)

In each period, the DM receives material utility and attention utility—what we call 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 DM's objective in any period t holding fixed (x_{-t}, α_{-t}) (their "best response function"): The DM chooses (x_t, α_t) with $x_t \in X_t(\alpha_t)$ to maximize[15]

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

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'). Thus, 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 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. Thus, time preferences—whether the DM is present- or future-focused—are endogenous and depend on the attention allocation.

In this environment, the DM is dynamically inconsistent as future selves do not value past selves' attention utility; we study the consequences in more detail in Section 4.1. As a result, we need to be careful about applying Propositions 1–3. For example, Proposition 1 suggests the DM weighs a period more if its payoff level or the instrumental value of attention to that period increases. Indeed, 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. Similarly, consistent with these intuitions, Carvalho et al. (2016) provide suggestive evidence that savings—i.e., high future payoffs to which the individual may devote attention—*cause* a preference for delayed gratification instead of the other way around. However, this is true only when fixing (action, attention)-pairs in other periods and looking at the DM's best response in the current one. Actual attentional choices are the result of an intrapersonal game (solved via backward induction) where the DM predicts their optimal future behavior and how it depends on actions today.[16] Thus, Propo-

[15] In this formulation, the DM values every future self's flow utility the same, regardless of the attention allocation. Equation (2) can be generalized by allowing the weights on period-t' flow utility (currently 1) to also depend on $\alpha_{t \to t'}$, e.g., flow utility in period t' receives weight $1 + \tilde{\lambda}\alpha_{t \to t'}$ in time t's objective, for some $\tilde{\lambda} \geq 0$. All results in this section go through with this more general formulation.

[16] 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

sitions 1–3 may 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.

Next, we study the implications of attention utility in a simple but classic environment: a consumption-saving problem. A DM receives a unit of income in every period that they 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, and consumption in period t is valued by $V(\sum_{t'=1}^{t} x_{t'\to t})$ (note that while we allow the DM to allocate consumption to past periods, the flow material utility in any given period t depends only on consumption assigned from periods prior to t, i.e., periods t' such that $t' \leq t$). The concavity of V implies that in this standard problem, the DM would smooth consumption by consuming all income in the period they receive it.

Consider now our DM in this environment. Letting $x = (x_t)_{t=1}^{T}$, our formulation captures the consumption allocation of income as action x with period-t (consumption) payoff $V_t(x) = V(\sum_{t'=1}^{t} x_{t'\to t})$. 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. In the following proposition, we suppose that for all t, we have $X_t(\alpha_t) = \{x_t : x_{t\to t'} \leq \alpha_{t\to t'} \forall t'\}$, i.e., the DM needs to allocate attention to a period in order to allocate their income to that period.

To simplify the statement of the proposition, we make some technical assumptions: First, assume that V is satiated at exactly integer 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. Second, assume that $-\frac{V''(K)}{V'(0)} > \frac{2}{K}$, where V' and V'' correspond to the first and second derivative of V. This assumption guarantees that the benefit of allocating attention to a state that currently has no attention is not too high.[17]

corresponding (action, attention)-pair optimally. 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 $x \in X(\alpha)$, where $X(\alpha) := (X_t(\alpha_t))_{t=1}^{T}$.

[17]If K is not a divisor of T, then when $\lambda > \bar{\lambda}$ (where $\bar{\lambda}$ is defined in the ensuing proposition), the last payoff in period T would be less than those in other high-payoff periods. If $-\frac{V''(K)}{V'(0)} \leq \frac{2}{K}$ then the DM's consumption in high-consumption periods only approaches K as $\lambda \to \infty$ but never reaches it.

Proposition 6. *There exist* $\underline{\lambda} > 0$ *and* $\bar{\lambda} < \infty$*, such that the DM optimally chooses* $\alpha_t = x_t$*, and (i) if* $\lambda < \underline{\lambda}$*, in each period* t*,* $\alpha_{t \to t} = 1$*, (ii) if* $\lambda > \bar{\lambda}$*, then there are exactly* $\frac{T}{K}$ *periods* t *with* $\sum_{t'=1}^{t} \alpha_{t' \to t} = K$*.*

When the weight on attention utility λ is small, the DM behaves as in the standard model, where they maximize 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. Although all attention is devoted to the present, we still have $\delta_{t \to t'} = 1$ for all t, t' with $t' \geq t$, and so there is no discounting of future payoffs. The reason is that while 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 in the DM's objective.[18]

For λ large, the DM allocates all income and attention to a subset of periods with high consumption. These high consumption periods are then exploited for attention utility. In fact, the DM never devotes attention to any period outside of the high-consumption periods set. Our model thus rationalizes non-smooth consumption paths, e.g., weddings, vacations, and other lavish celebrations.[19] Because in our environment, the DM has no intrinsic preference over the timing of consumption (it is only due to attention utility), they are indifferent between the actual timing of the high consumption periods. If we allow individuals to have a slight intrinsic preference for earlier consumption (i.e., standard discounting), the DM then desires a particular structure to attention: There are contiguous blocks of K periods, where all periods in the block pay attention to the last period. Thus, the individual has cycles of low consumption, punctuated by a single high-consumption period.

Such consumption paths can be similarly rationalized through memory utility (Gilboa et al., 2016; Hai et al., 2020) or anticipatory utility (Loewenstein, 1987). The key innovation in our model is that the enjoyment of fond memories requires attention to be experienced. More generally, our DM controls the flow of anticipatory and memory utility via their attention allocation instead of taking it as given.

[18]When attention additionally determines the weights on the flow utilities, for instance, if the weight on period t' in period-t's objective is $1 + \tilde{\lambda}\alpha_{t \to t'}$ instead of 1 (see footnote 15), then $\delta_{t \to t'} = 1/(1 + \tilde{\lambda})$ and so the DM falls in the class of quasi-hyperbolic discounters (Laibson, 1997).

[19]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. Another example of such consumption paths is workout and diet plans featuring "cheat meals": The individual's material utility is low, other than on cheat days to which they devote much attention.

Our DM's ability to manage their anticipatory or memory utility via attention also differentiates our model from others with endogenous, and in particular payoff-dependent, discounting. Loewenstein (1987) shows how anticipatory utility can drive a DM to negatively discount a high future payoff since it creates high anticipatory utility until it is realized. Noor and Takeoka (2022) develop a model where the discount rate is chosen optimally subject to a cost and show that this also leads to payoff-dependent discounting. However, these models fail to generate one of our key predictions, which is that the observed discount factor varies not just with levels of payoffs but also with the marginal return of attention across time.

4 Applications

Our previous results show how emotional inattention can impact behavior in three classic domains of decision-making: multi-dimensional consumption, risk, and time. Of course, many important economic environments involve at least two of these simultaneously. Here, we demonstrate how our model can help us understand important behaviors in economic applications that involve multiple kinds of domains: task completion and dynamic inconsistency, incentivizing an agent to exert effort, and default effects. Incorporating multiple domains means we must be more explicit about what dimensions are salient to the DM. For example, in a domain with both dimensions of consumption and risk, does the DM focus on the cross-product of consumption decisions and states of the world, or instead, would they focus on consumption decisions and aggregate across states (or vice versa)? Thus, this section not only shows that emotional inattention is portable to more complex settings and, under reasonable assumptions, delivers testable predictions but also highlights the need to think carefully about the ways in which a DM can direct attention.

4.1 Dynamic inconsistency and attentional scarcity

The DM may be dynamically inconsistent because future selves do not value past selves' attention utility. This feature is also present in existing models, e.g., those with anticipatory utility such as Loewenstein (1987); however, in our model, dynamic inconsistency only arises for intermediate payoff levels. Furthermore, and distinct from existing models, actions across periods to increase the payoff in a particular period are complements; this is because attending to a particular period leads to higher attention utility if the DM attends to that period already during another period. As a result, provided the DM only has limited attention available (i.e., there is "attention scarcity"), the equilibrium outcome may involve full inaction and is Pareto-dominated by the commitment solution.

To formalize these issues, we consider a simple two-period model, where, in each period, the DM devotes attention to either a non-trivial or a trivial consumption decision, with payoffs in period 2, denoted by c (for consumption) and o (for outside option), respectively (thus, our primitive dimensions correspond to time-dated consumption decisions). To further simplify the model (and notation), we assume that we can write the non-trivial consumption payoff as a function of the attention allocated to it. To highlight this, we write the payoff as $\hat{V}_c(\alpha_{1\to c}, \alpha_{2\to c})$, where $\alpha_{t\to c}$ denotes the attention in period $t = 1, 2$ devoted to the non-trivial decision (although this is slightly abusing notation, as our environment does not satisfy separability as in Definition 1).

The DM's objective in period 1 is

$$\underbrace{\hat{V}_c(\alpha_{1\to c}, \alpha_{2\to c}) + \hat{V}_o}_{\text{material utility in 2}} + \lambda \underbrace{\left(\alpha_{1\to c}\hat{V}_c(\alpha_{1\to c}, \alpha_{2\to c}) + \alpha_{1\to o}\hat{V}_o\right)}_{\text{attention utility in 1}} + \lambda \underbrace{\left(\alpha_{2\to c}\hat{V}_c(\alpha_{1\to c}, \alpha_{2\to c}) + \alpha_{2\to o}\hat{V}_o\right)}_{\text{attention utility in 2}}, \quad (3)$$

where $\hat{V}_o$ denotes the fixed payoff from the trivial consumption decision, and $\alpha_{t\to c}$ is the attention in period $t = 1, 2$ devoted to it, with $\alpha_{t\to c} + \alpha_{t\to o} = 1$ for $t = 1, 2$.

A commitment solution solves (3). When the DM cannot commit to a future attention allocation, period-2 self, for a given $\alpha_1 = (\alpha_{1\to c}, \alpha_{1\to o})$, maximizes the sum of period-2 material utility and attention utility (the first and third terms in (3)). A no-commitment solution (in pure strategies) is thus a pair $\alpha_1^*, \alpha_2^*(\cdot)$, where $\alpha_2^*(\alpha_1)$ solves the DM's period-2 problem given α_1, and α_1^* solves (3) if period-2 attention is given by $\alpha_2^*(\cdot)$.

The driver of the dynamic inconsistency is that in period 2, the DM ignores the impact of their actions on period-1 attention utility. Thus, the DM in period 2 devotes too little attention to the non-trivial payoff relative to what period-1 self desires.

Proposition 7. *Suppose $\hat{V}_c$ is continuous. Then: (i) both a commitment solution* *and a no-commitment solution exist, and (ii) fixing any* α_1 *the optimal* $\alpha_{2\to c}$ *chosen by period-2 self is less than that period-1 self would choose.*

Unlike other models, like Loewenstein (1987), with dynamic inconsistency due to anticipatory utility, our model only generates time inconsistency for intermediate payoff levels. When the non-trivial payoff is sufficiently high, both selves will devote all their attention to it, and when it is sufficiently low, both selves fully ignore it.

We turn to a second distinct feature of our model: unraveling. Period-2 self devoting less attention relative to what period-1 self desires decreases the attention utility period-1 self gets from attending to the non-trivial payoff. As a result, period-1 self may, in turn, also decrease attention to the non-trivial payoff, making it less pleasant to attend to for period-2 self, and so forth. For example, consider a student needing to study over the course of two days for an unpleasant exam—there is not enough time in a single day to prepare. If the student could commit to studying on both days, they would, as the performance after studying would be high enough to offset the unpleasantness of thinking about the class. But if the student cannot commit, then on day 2, the marginal returns from studying are not enough to outweigh the emotional costs. Thus, on the first day, anticipating that studying will not happen in the future, the benefit of studying is no longer worth the emotional cost, which is higher as the anticipated grade is lower.

The ensuing proposition shows a consequence of this complementarity of attention: If the DM does not have enough attention available in any given period, they may fail to devote attention to the non-trivial payoff in either period, leading to a Pareto-dominated outcome.

Proposition 8. *Suppose (a)* $\lambda > 0$*, (b)* $\hat{V}_c(0,0) < \hat{V}_o < \hat{V}_c(1,0)$*, (c)* $\hat{V}_c$ *is continuously differentiable with positive derivatives bounded away from zero everywhere, (d) and period-2 attention is not too instrumental. Then there exist* λ *large enough and* $(\bar{\alpha}_{1\to c}, \bar{\alpha}_{2\to c}) \in [0,1]^2$ *such that if the DM's attention allocations must satisfy* $(\alpha_{1\to c}, \alpha_{2\to c}) \leq (\bar{\alpha}_{1\to c}, \bar{\alpha}_{2\to c})$*: (i) The unique no-commitment outcome is* $(\alpha_{1\to c}, \alpha_{2\to c}) = (0,0)$*, (ii) with commitment, the DM devotes attention to the non-trivial consumption decision in both periods, and (iii) in both periods, the DM strictly prefers commitment solutions to the no-commitment solution.*

Moreover, if there are no constraints on the DM's attention allocation, then the unique commitment and no-commitment outcome is $(\alpha_{1\to c}, \alpha_{2\to c}) = (1,1)$*.*

Three key intuitions drive the result. First, emotional inattention implies that for large λ, the optimization problem can be convex for each period's self. Second, because of dynamic inconsistency, we can construct situations where period-2 self wants to avoid paying attention to the non-trivial task because the marginal material gain is not worth the emotional cost. Third, if period-2 self committed to paying attention to the non-trivial task, the payoff would be high enough to justify period-1 self paying attention. But, knowing that period-2 self, in fact, will not pay attention, the convexity of the payoff function for period-1 self implies that it becomes optimal to devote no attention either—we get attentional unraveling. This is despite the fact that there are no direct material complementarities for attention across periods, making it a novel prediction of our model.

4.2 Perverse effects of negative incentive schemes

We next consider how emotionally inattentive DMs respond to changes in incentive schemes. We focus on a simple environment where there are binary outcomes ("success" and "failure"), and effort requires attention. Standard ways to induce effort are to increase the reward for success or the penalty for failure, and for a standard DM (with $\lambda = 0$), they are similarly effective. However, when $\lambda > 0$, their consequences may differ starkly: A penalty decreases the expected payoff and can thus lead to lower attention and, perversely, lower effort.

Formally, there are multiple consumption decisions—leisure (l) and work (w)—and the work dimension has multiple states associated with it. By allocating attention, the DM affects the relative likelihoods of the two states. In contrast to previous sections, we suppose that the DM, when devoting attention to one state, also devotes attention to the other state in proportion to their respective likelihoods. In other words, they bracket the two possible states together and devote attention to the expected payoff.[20] We thus consider a separable environment, and use $\hat{V}_i$ to highlight this, with only consumption decisions $\mathcal{D} = \{w, l\}$, i.e., the DM chooses α_w, α_l, with

[20] Our qualitative conclusions only depend on the fact that the DM must devote at least some attention to the failure state whenever they devote attention to the success state (and vice versa)—not that those attention levels must be proportional to the respective likelihoods. Furthermore, in many situations, the agent would want to force themselves to "bracket" the two states together so that attention to one state naturally leads to attention to the other. In Appendix B, we formalize how the DM might optimally bracket different dimensions as one or separate.

$\alpha_w + \alpha_l = 1$, to maximize

$$\underbrace{\hat{V}_w(\alpha_w) + \hat{V}_l(\alpha_l)}_{\text{material utility}} + \lambda \underbrace{(\alpha_w \hat{V}_w(\alpha_w) + \alpha_l \hat{V}_l(\alpha_l))}_{\text{attention utility}},$$

where $\hat{V}_w(\alpha_w) = p(\alpha_w)v_H + (1 - p(\alpha_w))v_L$; v_H, v_L, with $v_H > v_L$, are the payoffs in the success and failure state, respectively, and $p(\alpha_w)$ is the probability of success given attention α_w, with p increasing and continuously differentiable. Probability p captures that individual effort (i.e., attention) changes the distribution of observed payoffs. The difference in payoffs, $v_H - v_L$, may be due to nature or could be a result of a contracting problem, where effort is not directly contractible.

We now consider what happens when the incentive scheme (v_H, v_L) changes, e.g., we consider different contracts the DM faces, ignoring the participation constraint. There are two possible ways in which the stakes in the contract could be increased: Either the payoff for a success could increase (i.e., v_H increases—the "carrot"), or the payoff for a failure could decrease (i.e., v_L falls—the "stick"). For a standard DM with $\lambda = 0$, the only thing that matters for their attention allocation is $v_H - v_L$. In contrast, when $\lambda > 0$, changes in the incentive scheme may have very different consequences.

Proposition 9. *Consider the environment as introduced prior to this proposition and suppose the optimal α_w is unique. Then: (i) Increasing v_H, v_L by the same amount increases α_w, (ii) increasing v_H increases α_w, and (iii) decreasing v_L decreases α_w if $p(\alpha_w) + \alpha_w \frac{\partial}{\partial \alpha_w} p(\alpha_w) < 1$ everywhere and λ is large enough.*

The first part of the proposition points out (in line with previous results) that the DM prefers to pay attention to work where all payoffs are high. Note that neither a standard DM (with $\lambda = 0$) nor a DM who overweights the high payoff, say, the DM in Brunnermeier and Parker (2005), would adjust their attention.[21] Thus, once again, our model's prediction differs starkly from those of other related ones.

The second and third parts note an asymmetric response to incentives and suggest that the strengthening of incentives can have perverse effects under some conditions. Standard theory would predict that increasing the impact of passing an exam or

[21] It is also the case that if the probability of success is simply shifted up by a constant for any given level of attention, the standard DM with $\lambda = 0$ will not change their behavior, while those with $\lambda > 0$ would increase their attention. Thus, emotionally inattentive DMs like to exert more effort on tasks with a higher likelihood of success, fixing the return to effort.

failing (or payment conditional on a success versus a failure in a contract or the earnings of being employed versus unemployed) should all lead to increased effort. In contrast, an emotionally inattentive DM increases their effort in response to a carrot (increasing v_H); a stick (decreasing v_L), however, may decrease the expected payoff and can thus lead to lower attention. Our model thus predicts that the lower payoffs conditional on the negative outcome happening (while fixing payoffs conditional on the good outcome) can lead to lower effort and worse performance. This occurs when attention is not very effective in increasing p ($\frac{\partial}{\partial \alpha_w} p(\alpha_w)$ is low), and success is far from guaranteed ($p(\alpha_w)$ is also low).

4.3 Default effects

We now explore how emotional inattention induces asymmetric default effects and ex-ante default choice. The environment features multiple time periods, consumption decisions, and risk. We show that defaults bind only in low-payoff states but that our DM, perversely, may set the default to maximize the payoff in high-payoff states.

Formally, there are two periods—period 1 and period 2. The setup in period 2 is that of Section 2.3 with two consumption decisions, $\mathcal{D} = \{c, o\}$. One of them, c, involves the choice of a default action. Its payoff is parameterized by some state $s \in \mathcal{S}$, capturing, e.g., an income shock, where $\mathcal{S}$ is finite, which is revealed to the DM at the beginning of the second period (thus, our primitive dimensions correspond to state-contingent, time-dated consumption decisions). The DM's default action x_1 is chosen in the first period. The other consumption decision, o (for attentional outside option), is trivial, i.e., its payoff V_o is constant. We model this attentional outside option as a consumption decision in period 2; however, it can also correspond to a past or period-1 payoff. For simplicity, we assume there are no (other) period-1 payoffs to which the DM devotes attention.

If the DM does not devote sufficient attention to the non-trivial dimension, then the default binds; we capture this by assuming that $X_2(\alpha_2) = \{\tilde{x}_2\}$ if $\alpha_{2 \to c} < \eta$ (i.e., the available period-2 actions are a constant singleton). The payoff from the non-trivial dimension is given by

$$V_c(x_1, x_2|s) = v_c(x_1, x_2|s) + \beta u(x_1|s),$$

where $\beta \geq 0$. The first component, $v_c(x_1, x_2|s)$ captures x_1 serving as a default:

$v_c(x_1, x_2|s)$ depends on x_1 if and only if $x_2 = \tilde{x}_2$, i.e., "the default binds." The second component reflects the part of period-2 payoffs that is impacted permanently by the choice of the default in period 1 (e.g., due to an irreversible investment), where we assume that for some state s, $u(\cdot|s)$ is not constant. Parameter β reflects how much of an impact the default has on payoffs, conditional on a different action being taken in period 2: If $\beta = 0$, then x_1 is a pure default—it only affects payoffs if it is not altered. In the limit as β gets larger, only the action in period 1 impacts payoffs, regardless of what occurs in period 2.

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

$$\underbrace{\overbrace{v_c(x_1, x_2|s) + \beta u(x_1|s)}^{V_c(x_1,x_2|s)} + V_o}_{\text{material utility}} + \lambda \underbrace{(\alpha_{2\to c} \overbrace{(v_c(x_1, x_2|s) + \beta u(x_1|s))}^{V_c(x_1,x_2|s)} + \alpha_{2\to o} V_o)}_{\text{attention utility}}. \tag{4}$$

We let $U_2(x_1, s)$ denote the maximized value of (4) 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.

In period 1, when the default is chosen, the DM also values their current attention utility. We assume that they can devote attention across the realizations of future consumption decisions and 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 $\alpha_{1\to(c,s)}$ denote the attention in period 1 devoted to the nontrivial consumption decision in state s, and $\alpha_{1\to o}$ that to the trivial decision. The DM's period-1 attention utility is $\sum_{s\in\mathcal{S}} \alpha_{1\to(c,s)} V_c(x_1, x_2(x_1, s)|s) + \alpha_{1\to o} V_o$.

In period 1, the DM's objective is the sum of period-1 attention utility and the expected period-2 utility (recall that attention utility in either period has weight λ). The following proposition summarizes and characterizes the states in which the default binds when the weight on attention utility is large.

Proposition 10. *Let $\underset{\sim}{\mathcal{S}}(x_1) \coloneqq \{s : \alpha_{2\to c}(x_1, s) < \eta\}$, i.e., it is the set of states in which the default binds. Suppose $X_1(\alpha_1)$ is independent of α_1 and finite, and λ is large enough. Then: (i) For any x_1, $\underset{\sim}{\mathcal{S}}(x_1) = \{s : \max_{(x_2,\alpha_2), x_2\in X_2(\alpha_2)} V_c(x_1, x_2|s) < V_o\}$, (ii) if $\beta = 0$, then the optimal default action x_1 satisfies*

$$x_1 = \underset{x_1' \in X_1}{\arg\max} \sum_{s\in\underset{\sim}{\mathcal{S}}(x_1)} p_s V_c(x_1', x_2(x_1', s)|s),$$

and, (iii) if $\beta > 0$, then the optimal default action x_1 satisfies

$$x_1 = \arg\max_{x_1' \in X_1} \sum_{s \in \mathcal{S} \setminus \underset{\sim}{\mathcal{S}}(x_1)} \tilde{p}_s u(x_1'|s),$$

as long as $\underset{\sim}{\mathcal{S}}(x_1) \neq \mathcal{S}$, where $(\tilde{p}_s)_{s \in \mathcal{S} \setminus \underset{\sim}{\mathcal{S}}(x_1)}$ are some weights.

The first part of the proposition describes a default effect: The DM fails to readjust their action in some states of the world, even though it is instrumentally costless. For example, if individuals have default savings plans, it suggests that individuals will tend to adjust more often with positive income shocks rather than negative ones. The key distinction relative to other models with costly adjustment is that the default binds asymmetrically—it only matters for states with low payoffs.

The second and third parts of the proposition imply that depending on the circumstances, the initially chosen default may look very different. In the second part, the default is "pure," i.e., it has no impact on payoffs unless it is not adjusted—e.g., a consumption plan does not impact future utility unless it is not changed. In this case, the DM chooses the default in order to maximize the expected payoff conditional on when the default binds: for example, consumers anticipating future inattention in periods of low payoffs plan their default consumption level to be low as insurance.

The third part considers the case where the choice of default directly impacts payoffs in period 2, even when a different action is subsequently taken; for example, current portfolio allocations both generate current returns and serve as a default for future investing. A DM with a high λ now wants to choose a default action that generates high attention utility in both periods, and so picks a default suitable for the weighted average of the states the DM devotes attention to, which are precisely those states when the default does not bind. In other words, this DM plans for the best but fails to re-optimize when the worst happens.

5 Discussion

5.1 Limits and extensions

Top-down vs bottom-up attention. A key tenet of our model is that the individual voluntarily directs attention, a premise often referred to as "top-down" attention. Recent economic models of attribute-based choice, e.g., Bordalo et al. (2013);

Kőszegi and Szeidl (2013); Bushong et al. (2021), have shown how involuntary attentional shifts ("bottom-up" attention) can play an important role in behavior such as violations of independence and non-exponential discounting. That said, evidence suggests that at least some attention is directed (e.g., Corbetta and Shulman (2002); Buschman and Miller (2007); Bronchetti et al. (2023)). Our model is applicable in any situation where at least some of the attention is optimally chosen, even if the rest is involuntarily allocated (which we capture via parameter ψ_i).

Functional form of attention utility. Our approach 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 enhance consumption utility. Extending our setting, one may partially account for this by analyzing a model with different λ parameters for each cognitive process.

Furthermore, attention utility likely does not have the simple functional form we assume. In Appendix D, we consider more general functional forms and alternative specifications and assess the robustness of Propositions 1–3; we show that material payoffs and attention being complements is the key driving force behind many of our results.

How focused is attention? 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 may be completed within minutes). This might imply that emotional inattention has a second-order impact.[22] 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, actual monitoring will require attention allocated over a long period of time.

[22]We note that one needs to be careful about interpreting the fact that many individuals may seem to take these decisions using a short period of as equivalent to the fact that these decisions, if made to maximize material utility, would take only a short amount of time. In fact, our model predicts that we would see (too) short decision times for unpleasant tasks.

Levels vs changes. We model attention utility as a function of attention to a dimension and the payoff in that dimension. Of course, comparison to reference points can be an important driver of utility (as in Kahneman and Tversky (1979); Kőszegi and Rabin (2006), among others). Our model can straightforwardly be extended so that attention utility and/or material utility are based on payoffs relative to some reference points. In fact, our model already captures the case where both material utility and attention utility are evaluated relative to the same fixed referent: replace payoff $V_i(x)$ with $V_i(x) - r_i$ everywhere and for all dimensions i, where $r = (r_i)_i$ is the referent. That said, models that allow attention utility to only depend on concerns about changes in (expected) payoffs would fail to match some of the predictions of our model, such as optimism and non-smooth consumption.

What requires attention? For tractability, we assume that the DM (as-if) understands and solves the optimization problem, i.e., Equation (1), without devoting attention. Although such an approach is tractable, it does beg the question of how the implications of the model might be changed if the act of optimization itself generates attention utility (this issue of recursion, see Lipman (1991) for a discussion, arises in many other models of optimal attention). One could construct a dynamic model of gradual learning about the problem, and we believe our qualitative results would also hold in this more complicated setting.

What is a dimension? A primitive of our model is a set of dimensions. Although in many environments, the set of dimensions may be obvious, in others, it may not be clear. This can provide the modeler with degrees of freedom in specifying the environment. However, we can eliminate this by extending the model so that the decision-maker optimally partitions their environment into dimensions in a way that will maximize their utility, given the optimal attentional allocation across dimensions. Appendix B demonstrates how to do this.

5.2 Relation to existing models

This section compares our model to related approaches. Further discussion of how our model relates to alternative explanations of (as-if) belief distortions, such as optimism and probability weighting (Section 2.4), or temporal discounting (Section 3), can be found in the respective sections.

Cognitive costs and rational inattention. The burgeoning area of rational inattention (Maćkowiak et al., 2023; Sims, 2003; Caplin and Dean, 2015) also studies the allocation of attention. In these models, individuals allocate attention to gain information, subject to a cost (often mental) that depends on the amount of information gained (i.e., the total amount of attention used). We believe our approach is complementary to the rational inattention literature for two reasons. First, while rational inattention models primarily (albeit with some exceptions, such as (Gabaix, 2014)) focus on how much attention should be used for a particular problem, we focus on how to allocate a fixed amount of attention across problems. Second, our definition of V_i is general enough so that it can capture both the utility benefits as well as the utility costs of acquiring information in dimensions i, allowing us to capture rational inattention concerns (Example 1 in Section A.1 shows how a canonical rational inattention setup is nested in our approach).

Although, like our model, rational inattention can explain why individuals may deliberately choose not to fully acquire information, emotional inattention makes several novel predictions. Our model predicts that the "cost of attention" falls for a particular dimension if the relative payoff of that dimension increases. Thus, in contrast to the predictions of rational inattention, emotionally inattentive individuals will pay less attention to low-payoff situations, even if the returns to additional information may be high (e.g., as when individuals avoid inexpensive tests for Huntington's disease (Oster et al., 2013)), and may devote large amounts of attention to positive situations even when there are no obvious material benefits (as in Quispe-Torreblanca et al. (2020)).[23] In the domain of risk, rational inattention fails to predict violations of expected utility, while in dynamic settings, rationally inattentive agents will be dynamically consistent, unlike emotionally inattentive agents.

Anticipatory utility and motivated beliefs. Our model can be seen as extending models of anticipatory utility, where agents gain flow utility from their beliefs about future outcomes (see Loewenstein (1987); Loewenstein and Elster (1992) for early contributions, and recent efforts of Caplin and Leahy (2001); Brunnermeier and Parker (2005); Bénabou and Tirole (2002)). In contrast to that literature, we suppose that the flow of anticipatory utility is mediated via the allocation of attention.

[23]Chambers et al. (2020) model rational inattention where the cost of attention can exhibit wealth effects; however, there, the costs depend on absolute, not relative, payoffs.

Anticipatory utility models can explain some of the same behavioral anomalies as our model—e.g., they predict violations of independence in choices over risk, dynamic inconsistency in temporal choice, incomplete consumption smoothing, and ostrich effects.

However, emotional inattention makes distinct predictions in many situations. For example, because anticipatory models are driven by future expectations, they fail to predict individuals avoiding situations when there is no information to be gained (or alternatively, paying attention when there is nothing to be learned), unlike our model, where agents may avoid taking actions related to low-payoff situations (such as preventative health actions) even in the absence of learning. In addition, because in anticipatory models, utility occurs due to beliefs regardless of attentional allocation, they fail to predict that changing the set of situations that the agent could pay attention to will alter their information acquisition (as in Falk and Zimmermann (2016), where the availability of a distractor affects informational preferences).

Our model of emotional inattention also makes distinct predictions from a well-known subclass of anticipatory utility models—those that feature motivated beliefs. Within an environment featuring uncertainty, our model predicts that attention's instrumental role may lead to increased focus on, and subsequent overweighting of, low-payoff states, leading to a wide variety of as-if probability weighting patterns, including pessimism and inverse S shapes (Proposition 5), that are outside of the predictions of motivated beliefs models. Furthermore, an emotionally inattentive DM may completely avoid an environment with uncertainty. For instance, in Section 4.2, we discussed a DM whose choice of effort stochastically determines a payoff. A motivated-beliefs agent would always exert more effort than a standard DM (one with $\lambda = 0$). Our DM, instead, reduces their effort if payoffs are shifted down (first part of Proposition 9), and this recusal may look like pessimistic attitudes toward the expected payoff.

Recursive preferences. A distinct approach to capturing non-standard attitudes towards both risk and time is the recursive preferences introduced by Kreps and Porteus (1978), including the widely used functional form of Epstein and Zin (1989). These models posit that individuals do not fully reduce compound risk and, like our model, have been developed to explain both behavior with respect to risk and consumption smoothing (by decoupling the coefficient of risk aversion and the intertemporal elasticity of substitution).

Although these models can accommodate non-expected utility risk attitudes as well as information aversion, they differ from emotional inattention in three key dimensions already mentioned with respect to other models. First, they fail to explain "action aversion" in the absence of information, and second, most of the models' implied attitudes towards risk, with binary outcomes, reduce to rank-dependent utility, i.e., only rank, rather than payoff differentials, matter for implied subjective probability distortions. Third, in a world without risk, recursive models predict that individuals should fully consumption smooth—i.e., these models cannot easily account for memorable consumption events.

Chosen preferences and beliefs. There is a small literature modeling agents who can optimally choose or adapt their future preferences (Bernheim et al. (2021); Elster (1983); related are models of motivated beliefs and optimal discounting, both of which we discuss elsewhere). Similarly, our DM can endogenously adjust the weights applied to various dimensions of the environment, changing their utility function via the choice of α, which Bernheim et al. (2021) call choosing a "world view."

However, our model has a distinctive feature relative to, e.g., Bernheim et al. (2021): The choice of the utility function (i.e., the weights on different dimensions) impacts the set of available actions, and so there is a tradeoff. As a result, our model makes predictions such as present-focus when attention to the present is particularly useful, inverse-S-shaped probability weighting if attention has decreasing (instrumental) returns, and, more generally, utility functions that emphasize dimensions where attention's instrumental value is high.

Other models of attention. We know of two other papers in economics that simultaneously model the instrumental and emotional roles of attention in a way similar to ours. The DM in Tasoff and Madarasz (2009) faces a decision problem with multiple consumption dimensions and receives anticipatory utility from each as a function of its payoff and the attention devoted to it. Attention increases if the (expected) payoff of a dimension changes, either because the DM takes an action or acquires information. Although some of their results overlap with ours, we develop a general model that is also applicable in the domains of risk and time.[24] Within the domain of contemporaneous consumption, unlike them, we allow attention to affect payoffs even if there is no instrumental consequence. Thus, our DM, unlike theirs, avoids low-payoff dimensions, such as their investment portfolio, even if there is no information to acquire and action to take, as in Quispe-Torreblanca et al. (2020).

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. Attention has two effects: It increases the impact of gain-loss utility, and it speeds up a reference point adjustment. Under some conditions, the DM pays additional attention to a situation only when there is positive initial news. Our model is similar in that attention also increases the impact (or weight) of a payoff. However, in our model, attention's instrumental value may also come from actions requiring attention. Moreover, we explore the implications of attention in novel environments (such as risk and time) as well as in applications, developing entirely new predictions.

Distinctly, Golman and Loewenstein (2018); Golman et al. (2021) and Golman et al. (2022) develop a different approach to attention. Like in our model, individuals weigh expected outcomes by attention when computing current utility. Unlike in our model, the attention weights are not determined by a joint optimization over material payoffs and attentional utility. Rather, attention is an exogenous function of the material value of information in that dimension, the salience of a dimension, and the expected reduction of entropy of beliefs given information. Thus, they aim to explain a different set of facts about attention using very different foundations for its allocation.

6 Conclusion

This paper has presented a model where 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 low-payoff situations, states, and time periods, even if doing so is instrumentally harmful, to decrease its weight in their objective

[24]Their main application is on how information provision can increase consumption, even when the DM learns their marginal payoff is less than what they expected—this intuition can be expressed in our framework as we show in Example 6 in Appendix A.3.

function (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.

We recognize, of course, that there are situations where individuals seem to freely allocate attention to 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 social media feeds that induce negative feelings. However, we believe that the large body of empirical findings discussed throughout the paper provides strong evidence that, in many situations, individuals exhibit a desire to focus on the positive aspects of their environment.[25]

This paper has focused on what kinds of novel behavior the emotional inattention framework can generate rather than the extent to which the model parameters can be identified from the data. However, in many situations, the novel primitives of our model, the set of dimensions and λ, can be recovered from the data. The details would vary by the environment, but here, we provide the intuition for a situation where the dimensions are states. The choices over lotteries allow us then to identify the degree of overweighting of the high-payoff state(s) and thus λ. In particular, recall that the subjective probability weight assigned to the high outcome in a binary lottery is equal to $\frac{p_i+\lambda}{1+\lambda}$, where p_i is the true probability. Once the probability weight on the high outcome is known (which can be derived from standard methods, see Gonzalez and Wu (1999)), recovering λ is simple algebra.

References

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

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

[25]Moreover, it may be that successfully allocating attention in the way our model prescribes is a skill that needs to be acquired and trained: Mamat and Anderson (2023) report on an intervention teaching individuals how to suppress unwanted (negative) thoughts and document persistent improvements in mental well-being.

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

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.

Bhatnagar, Roopali and Jacob L Orquin, "A meta-analysis on the effect of visual attention on choice.," *Journal of Experimental Psychology: General*, 2022, *151* (10), 2265.

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," *Econometrica*, 2023, *91* (2), 669–707.

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*, 2001, *116* (1), 55–79.

_ **and John V Leahy**, "Wishful thinking," Working Paper 25707, NBER 2019.

_ **and Mark Dean**, "Revealed Preference, Rational Inattention, and Costly Information Acquisition," *American Economic Review*, July 2015, *105* (7), 2183–2203.

Capra, C. Monica, Joshua Tasoff, and Jin Xu, "To Savor Consumption or To Confront Dread: The Hedonic Opportunity Cost of Attention," Working Paper, SSRN September 2023.

Carroll, Gabriel D., James J. Choi, David Laibson, Brigitte C. Madrian, and Andrew Metrick, "Optimal Defaults and Active Decisions," *The Quarterly Journal of Economics*, 2009, *124* (4), 1639–1674.

Carvalho, Leandro S., Silvia Prina, and Justin Sydnor, "The effect of saving on risk attitudes and intertemporal choices," *Journal of Development Economics*, 2016, *120*, 41–52.

Chambers, Christopher P., Ce Liu, and John Rehbeck, "Costly information acquisition," *Journal of Economic Theory*, 2020, *186*, 104979.

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.

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.

D'Amico, Leonardo and Guido Tabellini, "Disengaging from Reality: Online Behavior and Unpleasant Political News," Working Paper 687, IGIER 12 2022.

Ebert, Sebastian 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.

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

Falk, Armin and Florian Zimmermann, "Beliefs and Utility: Experimental Evidence on Preferences for Information," Discussion Paper 10172, IZA August 2016.

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

Frydman, Cary and Lawrence J Jin, "Efficient coding and risky choice," *The Quarterly Journal of Economics*, 2022, *137* (1), 161–213.

_ and _ , "On the Source and Instability of Probability Weighting," Technical Report, National Bureau of Economic Research 2023.

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.

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

Glöckner, Andreas and Ann-Katrin Herbold, "Information processing in decisions under risk: Evidence for compensatory strategies based on automatic processes," *MPI collective goods preprint*, 2008, (2008/42).

Golec, Joseph and Maurry Tamarkin, "Bettors Love Skewness, Not Risk, at the Horse Track," *Journal of Political Economy*, 1998, *106* (1), 205–225.

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

_ , David Hagmann, and George Loewenstein, "Information avoidance," *Journal of economic literature*, 2017, *55* (1), 96–135.

_ , George Loewenstein, Andras Molnar, and Silvia Saccardo, "The Demand for, and Avoidance of, Information," *Management Science*, 2022, *68* (9), 6454–6476.

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

Gonzalez, Richard and George Wu, "On the Shape of the Probability Weighting Function," *Cognitive Psychology*, 1999, *38* (1), 129–166.

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.

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.

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.

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.

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

Kőszegi, Botond and Adam Szeidl, "A model of Focusing in Economic Choice," *The Quarterly journal of economics*, 2013, *128* (1), 53–104.

_ **and Matthew Rabin**, "A Model of Reference-Dependent Preferences," *The Quarterly Journal of Economics*, 2006, *121* (4), 1133–1165.

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

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

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

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

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

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

_ **and Zachary Wojtowicz**, "The Economics of Attention," Working Paper 4368304, SSRN 2023.

Maćkowiak, Bartosz, Filip Matějka, and Mirko Wiederholt, "Rational Inattention: A Review," *Journal of Economic Literature*, March 2023, *61* (1), 226–73.

Mamat, Zulkayda and Michael C. Anderson, "Improving mental health by training the suppression of unwanted thoughts," *Science Advances*, 2023, *9* (38), eadh5292.

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.

Noor, Jawwad and Norio Takeoka, "Optimal Discounting," *Econometrica*, 2022, *90* (2), 585–623.

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

Orhun, A Yesim, Alain Cohn, and Collin Raymond, "Motivated Optimism and Workplace Risk," Working Paper 3966686, 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.

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.

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.

Smith, Stephanie M and Ian Krajbich, "Attention and choice across domains.," *Journal of Experimental Psychology: General*, 2018, *147* (12), 1810.

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.

Woodford, Michael, "Modeling imprecision in perception, valuation, and choice," *Annual Review of Economics*, 2020, *12*, 579–601.

Wu, George and Richard Gonzalez, "Curvature of the Probability Weighting Function," *Management Science*, 1996, *42* (12), 1676–1690.

Online Supplementary Material

A Additional examples

A.1 Examples of canonical problems

Examples 1–3 show how canonical settings with (cognitively) costly information acquisition (Maćkowiak et al., 2018; Sims, 2003), attention reducing "trembles" (Fudenberg et al., 2015), and recall of memories (Kahana, 2012) are nested in our framework. "Actions" and "payoffs" shall refer to those in the now explicitly modeled dimension i corresponding to the respective canonical choice problem.

Example 1 (Costly information acquisition)**.** *We first present a standard rational inattention model, then construct an equivalent problem using our notation, and last show they lead to the same behavior and payoffs.*

We begin by setting up a standard rational inattention problem. The environment consists of:

- *A finite set of possible action A, with cardinality N and typical element j;*
- *a finite set of states Θ, with typical element θ;*
- *state contingent utilities for each action $u(j, \theta)$;*
- *a prior $\rho_0 \in \Delta(\Theta)$;*
- *information structures M, where each information structure is a finite distribution over posterior beliefs with mean ρ_0, i.e., $M \subseteq \Delta(\Delta(\Theta))$ and for all $\mu \in M$, Bayes' plausibility holds, $\sum_{\rho_1 \in \text{support}(\mu)} \mu(\rho_1)\rho_1 = \rho_0$, where ρ_1 denotes a posterior; and*
- *a function c that associates with each $\mu \in M$ a cost $c(\mu, \rho_0)$, where $c(\mu, \rho_0)$ is increasing in the Blackwell ordering of μ.*

The DM's problem is as follows. For a given posterior ρ_1, the DM chooses an action $j^*(\rho_1)$ to maximize $E_{\theta \sim \rho_1}[u(j^*(\rho_1), \theta)]$. Thus, they choose an information structure $\mu \in M$ to maximize

$$U(\mu) = \sum_{\rho_1 \in \text{support}(\mu)} \mu(\rho_1) E_{\theta \sim \rho_1}[u(j^*(\rho_1), \theta)] - c(\mu, \rho_0).$$

We show how to reformulate this rational inattention model into our setting. The setup is that of Section 2.3 with two consumption decisions, $\mathcal{D} = \{c, o\}$ and no emotional consequence of attention, i.e., $\lambda = 0$. One of the consumption dimensions, o, is an attentional outside option that is trivial, i.e., its payoff V_o is constant. The other, c, will capture the rational inattention problem from above. In particular, the rational inattention problem provides a micro-foundation for the functional form of V_c, otherwise modeled in reduced form.

We first describe the three primitives for emotional inattention: A set of actions, a function mapping attention to subsets of actions, and a payoff function. An action x in the emotional inattention model is now a pair (μ, j), where μ is an information structure and j a function prescribing an action (in the rational inattention problem) to each posterior obtained under μ. Thus, implicitly (as it is not formally part of the model), the DM also faces an uncertain state θ, can acquire an information structure μ, and takes some action. We thus interpret x as a "contingent plan." Without loss, we assume the DM takes optimal actions j^*, and so j may be suppressed.

We now turn to how the set of available actions depends on attention. Consider the function $f : \mathbb{R} \to [0, 1]$, which is a strictly monotone function mapping each cost level $c(\mu, \rho)$ in the rational inattention set-up to an attention level; available actions as a function of attention is then defined as $X(\alpha_c, \alpha_o) := \{(\mu, j^*) | f(c(\mu, \rho_0) \leq \alpha_c)\}$. In other words, for any given level of attention, an attention level induces a possible set of information acquisition strategies, plus the associated ex-post optimal actions.

Last, define the payoff function for any given action as $V_c(\mu, j^*) := U(\mu)$; that is, the emotionally inattentive payoff from an action is the expected payoff from an experiment that generates that set of posteriors, and the associated optimal actions for each posterior, less the cost of that experiment. Note that the choice of f does not impact the solution to the optimization of V_c.

The two problems yield the same behavior. Indeed, the set of solutions is essentially the same mutatis mutandis: Consider $\mu \in M$ maximizing $U(\mu)$. Since μ is optimal, for all (x, α) with $x \in X(\alpha)$ and $x = (\mu', j^*)$, we must have $V_c(x) \leq U(\mu)$; furthermore, $\alpha_c = f(c(\mu, \rho_0))$ and $x = (\mu, j^*)$ achieves $U(\mu)$. Thus, the values of the rational and emotional inattention problems are the same.

Introducing attention utility, i.e., $\lambda > 0$, allows the modeler then to assess the effects of attention's emotional role in a standard rational inattention model.

Example 2 (Trembling)**.** *Dimension i is the reduced form of a canonical choice* *problem with trembles (see, e.g., Fudenberg et al. (2015) for an example).*

The DM chooses an action j from set $A = \{1, \dots, N\}$. The vector $v \in \mathbb{R}^N$ where v_j is the payoff of action $j \in A$ is known. The DM's choice is random—they "tremble"—and attention α_i to dimension i is useful because it allows the DM to reduce trembling. Specifically, suppose that $X(\alpha) = \prod_i X_i(\alpha_i)$, where $x_i \in X_i(\alpha_i)$ denotes the "recuction in trembling." The DM can then choose $B \in \mathcal{F}(x_i) \subseteq \Delta(A)$, where $\mathcal{F}$ is compact-valued, increasing and non-empty. The DM's payoff from dimension i given $x = (x_i, x_{-i}) \in X(\alpha)$ is

$$V_i(x_i) := E_B[v_j].$$

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

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

for some $\kappa \geq 0$, where $H(B)$ denotes the entropy of belief B 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.[26]

Example 3 (Memory recall)**.** *The goal is to model memory recall, as discussed in Kahana (2012). We interpret the memory recall model as a special case of rational inattention (Example 1), where information acquisition is now recalling signal realizations from memory.*

Thus, consider the setup from Example 1 and construct a particular set of information structures M. The DM is endowed with a memory base, modeled as an infinite set of independent signals $\{s_1, s_2, \dots\}$ each independently drawn from some $G(\theta)$, where G is finite. Let μ_d correspond to the information structure resulting from d draws. The information structures corresponding to a standard memory recall model are then given by $M = \{\mu_d : d = 0, \dots, \infty\}$.

A.2 Examples for Section 3

Example 4. *This example shows that increasing a future payoff can lead to less attention to the associated period. There are three time periods, $T = 3$. The payoffs*

[26]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 $-\int_v f(v) \log(f(v)) dv$.

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 $\alpha_{2\to 3} \geq \underline{\alpha}_{2\to 3}$, with $0 < \underline{\alpha}_{2\to 3} < \eta_2$ (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 $\alpha_{2\to 2} = 1 - \underline{\alpha}_{2\to 3}$ and $\alpha_{2\to 3} = \underline{\alpha}_{2\to 3}$. Otherwise, they take action x^* (and $\alpha_{2\to 2} = 1 - \eta_2$ and $\alpha_{2\to 3} = \eta_2$) over $\underline{x}$ (and $\alpha_{2\to 2} = 1 - \underline{\alpha}_{2\to 3}$ and $\alpha_{2\to 3} = \underline{\alpha}_{2\to 3}$) if

$$(1 + \lambda(1 - \eta_2))\bar{V} + (1 + \lambda\eta_2)\bar{V}_3 \geq (1 + \lambda(1 - \underline{\alpha}_{2\to 3}))\bar{V} + (1 + \lambda\underline{\alpha}_{2\to 3})\underline{V}_3. \quad (5)$$

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

$$(1+\lambda(1+(1-\eta_2))\bar{V}+(1+\lambda\eta_2)\bar{V}_3 \geq (1+\lambda((1-\eta_1)+1)\bar{V}+(1+\lambda\eta_1)\bar{V}_3 \iff \eta_1 \geq \eta_2. \quad (6)$$

Finally, also 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

$$(1+\lambda(1-\eta_1))\bar{V} + (1+\lambda(\eta_1+\underline{\alpha}_{2\to3}))\bar{V}_3 \geq (1+\lambda)\bar{V} + (1+\lambda\underline{\alpha}_{2\to3})\underline{V}_3. \tag{7}$$

Since $\underline{V}_3 < \bar{V}_3 < \bar{V}$*, there exists* $\lambda > 0$ *such that* (7) *holds with equality. For such* λ*, since* $\underline{\alpha}_{2\to3} > 0$*, there exists* $\eta_2 < \eta_1$ *(i.e.,* (6) *holds) so that* (5) *does not hold. Furthermore,* λ *can be slightly decreased so that* (7) *now holds strictly and* (5) *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* (5) *holds (and* (6) *and* (7) *remain to hold), so that the unique no-commitment 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. *Return to the setting of Example 4; the construction of an example showing* λ *can affect the material utility non-monotonically proceeds almost identically.*

Since $\underline{V}_3 < \bar{V}_3 < \bar{V}$*, there exists* $\lambda > 0$ *such that* (7) *holds with equality. For such* λ*, since* $\underline{\alpha}_{2\to3} > 0$*, there exists* $\eta_2 < \eta_1$ *(i.e.,* (6) *holds) so that* (5) *does not hold. Furthermore,* λ *can be slightly decreased so that* (7) *now holds strictly and* (5) *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* (5) *holds. As before, the DM now takes action* x^* *in period 2. Of course, material utility (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 material utility.*

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 decision 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.

We consider two cases: The DM devotes full attention $\alpha_c = 1$, and the DM devotes no attention $\alpha_c = 0$. These cases may be the result of the DM optimally choosing their attention allocation or due to advertising by the producer of the consumption good.

The DM learns the value of θ if $\alpha_c = 1$ (formally, such attention allows for some action x that corresponds to learning the value of θ). For $\alpha_c = 0$, the DM decides k before knowing θ and receives the expected payoff from consumption.

Suppose the DM learns the value of θ, i.e., $\alpha_c = 1$. Then, they choose c to satisfy

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

If they do not learn θ, i.e., $\alpha_c = 0$, 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 any information.

B Closing the model: optimal bracketing

Our model requires carefully specifying the environment: A key component is partitioning the environment into a set of dimensions. (In this way, our model is similar to prospect theory, which also requires an additional theory—that of the reference point.) In many real economic environments, natural partitions exist. However, what defines a dimension may be less obvious in other situations.

If natural partitions do not exist, one way of "closing" the model is to assume the DM themselves partitions the environment into dimensions and does so optimally. The DM may be able to do such partitioning by associating one dimension with another, either through some purely cognitive process or with the help of physical cues they install.

Formally, consider the setup 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, 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

$$\underbrace{\sum_i V_i(x)}_{\text{material utility}} + \lambda \underbrace{\sum_i \alpha_i \bar{V}_B(i)(x)}_{\text{attention utility}}, \tag{8}$$

where $\bar{V}_D(x) := \frac{\sum_{i\in D} V_i(x)}{|D|}$ for $D \subseteq \mathcal{D}$. For the ensuing proposition also let $\bar{\alpha}_D(x) := \frac{\sum_{i\in D} \alpha_i}{|D|}$ 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 $\lambda = 0$.

Proposition 11 states a necessary condition for bracketing to be optimal: The average amount of attention devoted to a bracket must be monotone in the average payoff of a bracket. 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. (This condition is not sufficient as, e.g., $B = \{\mathcal{D}\}$ trivially satisfies it but need not be optimal.)

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 $D, D' \in B$.*

C Proofs of the results in the main body

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*. *Consider 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 $\min_{(x,\alpha)\in\Gamma(\gamma_i',\beta_i)} v_i(x) \geq \max_{(x,\alpha)\in\Gamma(\gamma_i,\beta_i)} v_i(x)$.*
- *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 $\gamma_i' = \gamma_i - (\beta_i' - \beta_i)v_i(x)$, where $(x, \alpha) \in \Gamma(\gamma_i, \beta_i)$, we have $\min_{(x,\alpha)\in\Gamma(\gamma_i',\beta_i')} v_i(x) \geq \max_{(x,\alpha)\in\Gamma(\gamma_i,\beta_i)} v_i(x)$. If, in addition, the environment is separable, then $\min_{(x,\alpha)\in\Gamma(\beta_i',\gamma_i')} \alpha_i \geq \max_{(x,\alpha)\in\Gamma(\beta_i,\gamma_i)} \alpha_i$.*

It is immediate that Proposition 1* implies Proposition 1.

Proof of Proposition 1.* Take any 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

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

Combining the above gives

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

The outer inequality implies

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

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

If the environment is separable, then 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 $\gamma_i^{'} = \gamma_i - (\beta_i^{'} - \beta_i)v_i(x)$, 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

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

Combining the above and substituting for γ_i' gives

$$\begin{aligned}
&-((\omega_i + \lambda(\alpha_i + \psi_i))(\beta_i v_i(x) + \gamma_i) - (\omega_i + \lambda(\alpha_i' + \psi_i))(\beta_i' v_i(x') + \gamma_i - (\beta_i' - \beta_i)v_i(x))) \\
\geq &\kappa_2 - \kappa_3 \\
\geq &-((\omega_i + \lambda(\alpha_i + \psi_i))(\beta_i v_i(x) + \gamma_i) - (\omega_i + \lambda(\alpha_i' + \psi_i))(\beta_i v_i(x') + \gamma_i)).
\end{aligned}$$

The outer inequality implies

$$-(\omega_i + \lambda(\alpha_i' + \psi_i))(v_i(x) - v_i(x'))(\beta_i' - \beta_i) \geq 0,$$

and thus, it must be that $v_i(x') \geq v_i(x)$.

If the environment is separable, then 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 λ', λ 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

$$\begin{aligned}
\sum_i \omega_i V_i(x) + \lambda \sum_i (\alpha_i + \psi_i) V_i(x) &\geq \sum_i \omega_i V_i(x') + \lambda \sum_i (\alpha_i' + \psi_i) V_i(x'), \quad \text{and} \\
\sum_i \omega_i V_i(x') + \lambda' \sum_i (\alpha_i' + \psi_i) V_i(x') &\geq \sum_i \omega_i V_i(x) + \lambda' \sum_i (\alpha_i + \psi_i) V_i(x).
\end{aligned}$$

Combining the above gives

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

To reach a contradiction, suppose the expression in the middle is strictly negative. Then, the expression on the right must also be strictly negative; but then it is strictly larger than the left one as $\lambda' > \lambda$—a contradiction. 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

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

where the inequality is due to optimality, and so the second claim follows.

Now, suppose the environment is separable; consider dimension $i \in \mathcal{D}$ and suppose the objective given λ is convex in α_i. Since (1) is additively separable across dimensions, it suffices to only show that if $\omega_i \hat{V}_i(\alpha_i) + \lambda(\alpha_i + \psi_i)\hat{V}_i(\alpha_i)$ is convex in α_i, it remains so as λ is increased to λ'. Furthermore, since $\omega_i \hat{V}_i(\alpha_i) + \lambda(\alpha_i + \psi_i)\hat{V}_i(\alpha_i)$ is a linear combination of $\hat{V}_i(\alpha_i)$ and $\alpha_i\hat{V}_i(\alpha_i)$ with the relative weight on the latter increasing in λ, it suffices to show that if $\hat{V}_i(\alpha_i)$ in convex in α_i, so is $\alpha_i\hat{V}_i(\alpha_i)$. To this end, assume $\hat{V}_i(\alpha_i)$ in convex in α_i and take any $\chi \in [0,1]$ and α_i, α_i' with $\alpha_i < \alpha_i'$. Then

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

where the first inequality follows since $\hat{V}_i(\alpha_i)$ in convex in α_i, and the second as $\hat{V}_i$ is increasing; hence, the third claim follows. □

C.3 Proof of Proposition 4

Proof of Proposition 4. Take any lottery x, and suppose that the DM(λ) prefers x to δ_y for some payoff y, i.e., $\sum_i p_i u(x_i) + \lambda u(H(x)) \geq (1+\lambda)u(y)$, where the DM optimally devotes full attention to the states with the highest payoff $H(x)$. Since $u(H(x)) \geq \sum_i p_i u(x_i)$, by definition of $H(x)$, we must have $u(H(x)) \geq u(y)$ for the inequality to hold. Thus, the inequality continues to hold when λ is increased to λ', and so DM(λ') also prefers x to δ_y.

For the second and third claims, take any x, x'. The DM's strictly prefers x to x' if and only if

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

For the second claim, note that if $H(x) > H(x')$, the above is satisfied for large enough λ since the left and right sides converge to $u(H(x))$ and $u(H(x'))$, respectively; for the third claim, note that if $H(x) = H(x')$, then the above is logically equivalent to $\sum_i p_i u(x_i) > \sum_i p_i u(x_i')$, and so the DM's preferences are indeed independent of λ. □

C.4 Proof of Proposition 5

Proof of Proposition 5. 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 $\frac{\partial}{\partial a}\hat{V}(1) < \infty$. $q_i(\cdot) = q_{i'}(\cdot)$ 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, $q(0) = 0$.

We next show that for $p_i > 0$ small enough, the optimal α_i exceeds p_i, which implies $q(p_i) > p_i$. Consider the derivative of the DM's overall payoff,

$$\frac{(p_i + \lambda\alpha_i)\frac{\partial}{\partial a}\hat{V}(\alpha_i) - ((1-p_i) + \lambda(1-\alpha_i))\frac{\partial}{\partial a}\hat{V}(1-\alpha_i)}{1+\lambda} + \frac{\lambda(\hat{V}(\alpha_i) - \hat{V}(1-\alpha_i))}{1+\lambda}. \quad (9)$$

Note that 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 (9) is strictly positive at $\alpha_i = 0$. Now, consider the limit of (9) as $\alpha_i \to 0$ for some fixed $\tilde{p}_i$ with $0 < \tilde{p}_i < 1/2$; the limit is infinite since we assumed $\lim_{a\to 0} a\frac{\partial}{\partial a}\hat{V}(a) = \infty$, implying that $(\tilde{p}_i + \lambda\alpha_i)\frac{\partial}{\partial a}\hat{V}(\alpha_i) \to \infty$, and all other terms in (9) are finite. Thus, there exists a $\bar{\alpha}_i$ with $0 < \bar{\alpha}_i \le \tilde{p}_i$ such that for all $\alpha_i \le \bar{\alpha}_i$, (9) evaluated at α_i (given $\tilde{p}_i$) is strictly positive. Let $\bar{p} = \bar{\alpha}_i$. Since (9) is decreasing in p_i (provided $p_i < 1/2$) for all $\alpha_i \in [0, 1/2]$ and $\bar{p} \le \tilde{p}_i$, we have for all p_i with $0 < p_i < \bar{p}$, (9) is strictly positive for all $\alpha_i \in [0, \bar{p}]$. This implies that the optimal α_i is strictly greater than p_i. Since $q_i(p_i) = \frac{p_i+\lambda\alpha_i}{1+\lambda}$, it follows that $q(p_i) > p_i$ for such

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 $q_i(\cdot)$. $\square$

C.5 Proof of Proposition 6

Proof of Proposition 6. For both λ small and λ large, we consider the solution to an auxiliary problem and under commitment. We then show that the solution to this problem is also the solution to the original problem without commitment.

In particular, in the auxiliary problem, we will suppose that income allocations from all (both past and future) affect any given period's payoff, i.e., consumption in period t is valued by $V(\sum_{t'=1}^{T} x_{t' \to t})$.

This auxiliary problem simplifies our analysis: Given commitment, it is easy to see that the DM optimally chooses $x_{t \to t'} = \alpha_{t \to t'}$ for all t, t'; thus, at time 1, DM then maximizes

$$\sum_{t=1}^{T} (1 + \lambda \alpha_{\to t}) V(\alpha_{\to t}), \tag{10}$$

where $\alpha_{\to t} := \sum_{t'=1}^{T} \alpha_{t' \to t}$ is the amount of attention devoted to period t, such that $\sum_{t=1}^{T} \alpha_{\to t} = T$.

Consider the case of small λ. Let $\underline{\lambda} := \max_{\alpha_{\to t} \in [0,T]} \frac{V''(\alpha_{\to t})}{2V'(\alpha_{\to t}) + \alpha_{\to t} V(\alpha_{\to t})}$; we next show that for $\lambda < \underline{\lambda}$, (10) is strictly concave. Note that if $(1 + \lambda \alpha_{\to t}) V(\alpha_{\to t})$ is strictly concave for $\alpha \in [0, T]$, then (10) is strictly concave in $(\alpha_{\to t})_{t=1}^{T}$.[27]

Thus, consider

$$\begin{aligned} \frac{\partial^2}{\partial \alpha_{\to t}^2} (1 + \lambda \alpha_{\to t}) V(\alpha_{\to t}) &= 2V'(\alpha_{\to t}) + (1 + \lambda \alpha_{\to t}) V''(\alpha_{\to t}) \\ &= V''(\alpha_{\to t}) + \lambda (2V'(\alpha_{\to t}) + \alpha_{\to t} V''(\alpha_{\to t})). \end{aligned}$$

If $2V'(\alpha_{\to t}) + \alpha_{\to t} V''(\alpha_{\to t}) < 0$, then, since V is strictly concave, the above is also strictly negative for any λ; otherwise, the above is bounded by $V''(\alpha_{\to t}) + \underline{\lambda}(2V'(\alpha_{\to t}) + \alpha_{\to t} V''(\alpha_{\to t}))$, which is strictly negative by definition of $\underline{\lambda}$. Thus, the above is strictly negative for all $\alpha_{\to t}$.

It is easy to check that $\alpha_{\to t} = 1$ for all t satisfies the Karush-Kuhn-Tucker con-

[27] I.e., if for all $\chi \in (0,1)$ and $\alpha_{\to t}, \alpha'_{\to t}$, $\chi(1 + \lambda \alpha_{\to t}) V(\alpha_{\to t}) + (1 - \chi)(1 + \lambda \alpha'_{\to t}) V(\alpha'_{\to t}) < (1 + \lambda(\chi \alpha_{\to t} + (1-\chi)\alpha'_{\to t})) V(\chi \alpha_{\to t} + (1-\chi)\alpha'_{\to t})$, then for all $\chi \in (0,1)$ and $(\alpha_{\to t})_{t=1}^{T}, (\alpha'_{\to t})_{t=1}^{T}$, $\chi \sum_{t=1}^{T} (1 + \lambda \alpha_{\to t}) V(\alpha_{\to t}) + (1-\chi) \sum_{t=1}^{T} (1 + \lambda \alpha'_{\to t}) V(\alpha'_{\to t}) < \sum_{t=1}^{T} (1 + \lambda(\chi \alpha_{\to t} + (1-\chi)\alpha'_{\to t})) V(\chi \alpha_{\to t} + (1-\chi)\alpha'_{\to t})$.

ditions of the associated Lagrangian. Since the unconstrained problem is strictly concave for $\lambda < \underline{\lambda}$, $\alpha_t = 1$ for all t is the unique global solution.

Moreover, note that the only way of achieving this optimum in the original problem, where only past and current income allocations increase the consumption payoff, is by choosing $\alpha_{t\to t} = x_{t\to t} = 1$ for all t. What remains to be shown is that the DM can implement these attention allocations and actions without commitment. Consider the DM at time t and suppose $\alpha_{t'\to t'} = x_{t'\to t'} = 1$ for all $t' < t$. Note that this DM's objective is maximized with $\alpha_{t'\to t'} = x_{t'\to t'} = 1$ for all $t' \geq t$. Since this holds for all t, $\alpha_{t\to t} = x_{t\to t} = 1$ for all t is indeed credible and hence chosen by the DM.

Next, consider the case of large λ. We return to the auxiliary problem with commitment, where the DM chooses $\alpha_{\to t}$ for $t = 1, \ldots, T$ such that $\sum_{t=1}^{T} \alpha_{\to t} = T$ to maximize (10). We show that i) the solutions to (10) converge to those attention allocations where $\alpha_{\to t} = K$ for $\frac{T}{K}$ periods (and zero otherwise), ii) for λ large enough, each of such attention allocations is a strict local maximum, and iii) the neighborhood in which each allocation is strictly optimal increases in λ. Together, these three facts then imply that $\alpha_{\to t} = K$ for $\frac{T}{K}$ periods (and zero otherwise) are the global solutions for λ large enough.

For i), note that, in the limit, each attention allocation that maximizes $\sum_{t=1}^{T} \alpha_{\to t} V(\alpha_{\to t})$ is strictly better than any that does not. In particular, it must be that each optimal allocation converges to one that achieves the maximum amount of attention utility, $TV(K)$. Of those attention allocations, the elements in the claimed set of solutions in the limit uniquely maximize the DM objective for any λ, as they maximize material utility, and the result follows.

For ii), we again consider the Karush-Kuhn-Tucker conditions of the associated Lagrangian. Because we are considering T periods, we have T associated first-order Karush-Kuhn-Tucker conditions. Given our proposed solutions, these conditions fall into one of two categories: conditions for periods that have 0 attention devoted to them and periods that have K units of attention devoted to them.

In particular, it is easy to check that both the latter set of periods (those where $\alpha_{\to t} = K$) and the former (where $\alpha_{\to t} = 0$) will satisfy the Karush-Kuhn-Tucker conditions if there exist $\kappa \geq 0$ (i.e., the Lagrange multiplier on total attention) and

$\mu \geq 0$ (i.e., the Lagrange multiplier on non-negativity of attention) such that

$$(1 + \lambda K)V'(K) + \lambda V(K) - \kappa = 0$$
$$V'(0) + \lambda V(0) - \kappa + \mu = 0.$$

Recall that $V'(K) = 0$. Moreover, since $V(K) > V(0)$, such $\kappa \geq 0$ and $\mu \geq 0$ exist for large enough λ.

We now turn to checking second-order conditions for sufficiency. In particular, we first compute the Hessian of the Lagrangian H (i.e., the matrix of cross partials). It is easy to verify that the cross partials with respect to t and t' are not equal to 0 if and only if $t = t'$.

Sufficiency is satisfied if for all vectors s of length T, where $\sum_t s_t = 0$ and $s_t + \alpha_{\to t} \geq 0$ for all t, it is the case that $s^T H s > 0$, where $\alpha_{\to t}$ is an optimal attention allocated to period t. We can provide a lower bound on $s^T H s$ by considering a shift of attention from a single period t, which currently is receiving K units of attention, to a single other period t' that is currently receiving 0 units of attention and devoting all of its attention to period t.

Algebra then shows that we need it to be the case that $(1+\lambda K)V''(K)+2\lambda V'(0)+V''(0) < 0$, which is the condition assumed prior to the proposition for large λ.

For iii), note that since each such allocation first (in a lexicographic fashion) maximizes the sum of attention utility across time and then the sum of instrumental utility, increasing λ only improves it relative to all others, and hence, the neighborhood in which it is strictly optimal increases in the weight on attention utility, λ.

Thus, for λ large enough, (10) is maximized by $\alpha_{\to t}$ for $t = 1, \ldots, T$, such that $\sum_{t=1}^{T} \alpha_{\to t} = T$.

Lastly, note that such optima can be achieved in the original problem by, e.g., choosing $\alpha_{t \to K(t)} = 1$ where $K(t) \equiv \lceil \frac{t}{K} \rceil K$ for all t; other attention allocations can also achieve such optima. To show that they can be implemented without commitment, consider the aforementioned attention allocation and the DM at time t and suppose $\alpha_{t' \to K(t')} = x_{t' \to K(t')} = 1$ for all $t' < t$. Note that this DM's objective is maximized with $\alpha_{t' \to K(t')} = x_{t' \to K(t')} = 1$ for all $t' \geq t$. Since this holds for all t, choosing $\alpha_{t \to K(t)} = x_{t \to K(t)} = 1$ is indeed an equilibrium outcome. □

C.6 Proof of Proposition 7

Proof of Proposition 7. Since players (the DM in different periods) take actions sequentially, their choice set is compact, and their payoffs are continuous, the existence of a subgame-perfect equilibrium (the no-commitment solution) follows from, e.g., Hellwig et al. (1990). Continuity of the DM's objective and a compact choice set guarantee the existence of a commitment solution.

Fix any $\alpha_1 = (\alpha_{1\to c}, \alpha_{1\to o})$ and consider

$$\hat{V}_c(\alpha_{1\to c}, \alpha_{2\to c}) + \hat{V}_o + \xi\lambda(\alpha_{1\to c}\hat{V}_o(\alpha_{1\to c}, \alpha_{2\to c}) + \alpha_{1\to o}\hat{V}_o) \\ + \lambda(\alpha_{2\to c}\hat{V}_c(\alpha_{1\to c}, \alpha_{2\to c}) + \alpha_{2\to o}\hat{V}_o).$$

This expression equals (3) for $\xi = 1$ and for $\xi = 0$, it is period-2 self's objective. Fix any α_1. Since $\lambda(\alpha_{1\to c}\hat{V}_o(\alpha_{1\to c}, \alpha_{2\to c}) + \alpha_{1\to o}\hat{V}_o)$ is increasing in $\alpha_{2\to c}$, the above expression has increasing differences in $(\alpha_{2\to c}, \xi)$ and so the result follows from Topkis's Theorem. □

C.7 Proof of Proposition 8

Proof of Proposition 8. The proof is constructive. We first construct a point for large λ such that

- (Step 1) period-1 self is indifferent between that point and devoting no attention in either period,
- (Step 2) period-2 self strictly prefers this point to devoting no attention,
- (Step 3) period-1 self wants period-2 self to increase attention to the non-trivial payoff, while period-2 self wants to decrease it.
- (Step 4) We then construct a nearby point that both selves prefer to devoting no attention, and such that period-2 self still prefers to deviate and not devote attention, fixing period-1 self's attention; which leads period-1 self to choose $(0, 0)$ when there is no commitment.

Step 1. We first note that $(0, 0)$ is a strict local maximum of (3) for large enough λ, which we eventually use to ensure that the no-commitment outcome has no attention devoted to the non-trivial payoff in both periods.

Let the value of (3) for $(\alpha_{1\to c}, \alpha_{2\to c})$ be denoted by $V_1(\alpha_{1\to c}, \alpha_{2\to c})$.

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

For ease of exposition, the proofs of the claims are at the end of this section.

Given that $(0,0)$ is a local maximum and that $\hat{V}_o < \hat{V}_c(1,0)$, there exists a smallest strictly positive level of attention devoted by period-1 self to the non-trivial payoff, $\underline{\alpha}_{1\to c}$, such that $V_1(\underline{\alpha}_{1\to c}, 0) = V_1(0,0)$.

Claim 2. *Let $A_1(\lambda) \equiv \{\alpha_{1\to c} : V_1(\alpha_{1\to c}, 0) = V_1(0,0), \alpha_{1\to c} > 0\}$. For $\lambda > \underline{\lambda}$, where $\underline{\lambda}$ is defined in the proof of Claim 1, we have:*

i) $\underline{\alpha}_{1\to c}(\lambda) \equiv \min_{\alpha_{1\to c}\in A_1(\lambda)} \alpha_{1\to c}$ *exists,*

ii) $\underline{\alpha}_{1\to c}(\lambda) < \tilde{\alpha}_{1\to c}$, *where* $\tilde{\alpha}_{1\to c}$ *is implicitly defined by* $\hat{V}_c(\tilde{\alpha}_{1\to c}, 0) = \hat{V}_o$,

iii) $\underline{\alpha}_{1\to c}(\lambda)$ *is increasing in* λ, *and*

iv) $\lim_{\lambda\to\infty} \underline{\alpha}_{1\to c}(\lambda) = \tilde{\alpha}_{1\to c}$

By construction, period-1 self is indifferent between $(0,0)$ and $(\underline{\alpha}_{1\to c}(\lambda), 0)$ and $(0,0)$, provided λ is large enough.

Step 2. We next note that period-2 self strictly prefers $(\underline{\alpha}_{1\to c}(\lambda), 0)$. Intuitively, period-1 self's indifference implies that there is a cost in terms of attention utility when comparing $(0,0)$ and $(\alpha_{1\to c}, \alpha_{2\to c})$, and since period-2 self does not value period-1 self's attention utility, it strictly prefers the attention allocation that is more costly, in terms of attention utility.

Let $V_2(\alpha_{1\to c}, \alpha_{2\to c}) = \hat{V}_c(\alpha_{1\to c}, \alpha_{2\to c}) + \hat{V}_o + \lambda(\alpha_{2\to c}\hat{V}_c(\alpha_{1\to c}, \alpha_{2\to c}) + \alpha_{2\to o}\hat{V}_o)$ denote the DM's period-2 objective.

Claim 3. *If $\lambda > 0, V_1(0,0) = V_1(\alpha_{1\to c}, \alpha_{2\to c})$ and $\alpha_{1\to c} > 0$, then there exists $\delta > 0$ so that $V_2(\alpha_{1\to c}+y_1, \alpha_{2\to c}+y_2) > V_2(0,0)$ for all $(y_1, y_2) \in [-\delta, \delta]^2$, with $(\alpha_1+y_1, \alpha_2+y_2) \in [0,1]^2$.*

Step 3. Next, we consider the incentives at attention allocation at $(\alpha_{1\to c}, \alpha_{2\to c}) = (\underline{\alpha}_{1\to c}(\lambda), 0)$. In particular, note that since the derivatives of $\hat{V}_c$ are bounded away from zero and $\underline{\alpha}_{1\to c}(\lambda)$ is positive and increasing, for large enough λ, period-1 self strictly prefers to increase $\alpha_{2\to c}$, i.e.,

$$\frac{\partial V_1}{\alpha_{2\to c}}(\underline{\alpha}_{1\to c}(\lambda),0) = (1+\lambda\alpha_{2\to c})\frac{\partial \hat{V}_c}{\partial \alpha_{2\to c}}(\underline{\alpha}_{1\to c}(\lambda),0) - \lambda(\hat{V}_o - \hat{V}_c(\alpha_{1\to c}(\lambda),0) > 0.$$

For period-2 self's incentives, since $V_1(\underline{\alpha}_{1\to c}(\lambda),0) = V_1(0,0)$, some simple algebra steps imply $\lambda(\hat{V}_o - \hat{V}_c(\alpha_{1\to c}(\lambda),0) = \frac{\hat{V}(\underline{\alpha}_{1\to c}(\lambda),0)-\hat{V}_c(0,0)}{\underline{\alpha}_{1\to c}(\lambda)}$ and so we have

$$\begin{aligned}\frac{\partial V_2}{\alpha_{2\to c}}(\underline{\alpha}_{1\to c}(\lambda),0) &= \frac{\partial \hat{V}_c}{\partial \alpha_{2\to c}}(\underline{\alpha}_{1\to c}(\lambda),0) - \lambda(\hat{V}_o - \hat{V}_c(\alpha_{1\to c}(\lambda),0) \\ &= \frac{\partial \hat{V}_c}{\partial \alpha_{2\to c}}(\underline{\alpha}_{1\to c}(\lambda),0) - \frac{\hat{V}(\underline{\alpha}_{1\to c}(\lambda),0) - \hat{V}_c(0,0)}{\underline{\alpha}_{1\to c}(\lambda)}.\end{aligned} \tag{11}$$

By Claim 2, $\underline{\alpha}_{1\to c}(\lambda) = \tilde{\alpha}_{1\to c}$ as $\lambda \to \infty$. Thus, using as long as period-2 attention is sufficiently non-instrumental, i.e.,

$$\frac{\partial}{\partial \alpha_{2\to c}}\hat{V}_c(\tilde{\alpha}_{1\to c},0) < \frac{\hat{V}_o - \hat{V}_c(0,0)}{\tilde{\alpha}_{1\to c}},$$

it follows that (11) is negative, for large λ.

Step 4. Recall, the next step is to construct a point near $(\underline{\alpha}_{1\to c}(\lambda),0)$ that *both* selves prefer to devoting no attention, and such that period-2 self still prefers to deviate from it and not devote attention, fixing period-1 self's attention; which leads period-1 self to choose $(0,0)$ when there is no commitment.

Take λ large enough, so that $(0,0)$ is a strict local maximum, $\frac{\partial V_1}{\alpha_{2\to c}}(\underline{\alpha}_{1\to c}(\lambda),0) > 0$, and $\frac{\partial V_2}{\alpha_{2\to c}}(\underline{\alpha}_{1\to c}(\lambda),0) < 0$. Then, there exists $\bar{\delta} > 0$ such that for all $0 < \delta, \delta' < \bar{\delta}$,

i) there exists a point in $[\underline{\alpha}_{1\to c}(\lambda) - \delta, \underline{\alpha}_{1\to c}(\lambda)] \times [0,\delta']$ that period-1 self strictly prefers to $(0,0)$ (since $\hat{V}_c$ is continuously differentiable and $\frac{\partial V_1}{\partial \alpha_{2\to c}}(\underline{\alpha}_{1\to c}(\lambda),0) > 0$, period-1 self strictly prefers, e.g., $(\underline{\alpha}_{1\to c}(\lambda),\delta')$ to $(0,0)$ for δ' small enough),

ii) period-2 self strictly prefers all attention allocations in $[\underline{\alpha}_{1\to c} - \delta, \underline{\alpha}_{1\to c}] \times [0,\delta']$ to $(0,0)$ (follows from Claim 3), and

iii) period-2 self strictly prefers $(\alpha_{1\to c},0)$ to $(\alpha_{1\to c},\alpha_{2\to c})$ for any $(\alpha_{1\to c},\alpha_{2\to c}) \in [\underline{\alpha}_{1\to c}(\lambda) - \delta, \underline{\alpha}_{1\to c}(\lambda)] \times (0,\delta']$ (again, since $\hat{V}_c$ is continuously differentiable, and since $\frac{\partial V_2}{\partial \alpha_{2\to c}}(\underline{\alpha}_{1\to c}(\lambda),0) < 0$).

Since $(0,0)$ is a strict local maximum, there exists $\tilde{\delta} > 0$ such that for every $0 < \delta < \tilde{\delta}$, period-1 self strictly prefers $(0,0)$ to $(\alpha_{1\to c}, \alpha_{2\to c})$ for any $(\alpha_{1\to c}, \alpha_{2\to c}) \in [0,\delta]^2, (\alpha_{1\to c}, \alpha_{2\to c}) \neq (0,0)$.

Since period-1 self strictly prefers $(0,0)$ to $(\alpha_{1\to c}, 0)$ for any $\alpha_{1\to c} \in [\tilde{\delta}, \underline{\alpha}_{1\to c}(\lambda) - \bar{\delta}]$ (by definition of $\underline{\alpha}_{1\to c}(\lambda)$) and since V_1 is continuous, there exists $\epsilon > 0$ such that period-1 self strictly prefers $(0,0)$ to $(\alpha_{1\to c}, \alpha_{2\to c})$ for any $(\alpha_{1\to c}, \alpha_{2\to c}) \in [\tilde{\delta}, \underline{\alpha}_{1\to c}(\lambda) - \bar{\delta}] \times [0, \epsilon]$.

Finally, let $(\bar{\alpha}_{1\to c}, \bar{\alpha}_{2\to c}) = (\underline{\alpha}_{1\to c}(\lambda), \min\{\bar{\delta}, \tilde{\delta}, \epsilon, \underline{\alpha}_{1\to c}(\lambda)\})$.

By construction, there exists a point in $[\bar{\alpha}_{1\to c} - \bar{\delta}, \bar{\alpha}_{1\to c}] \times [0, \bar{\alpha}_{2\to c}]$ that both period-1 self and period-2 self strictly prefer to $(0,0)$. Furthermore, again by construction, period-1 self prefers $(0,0)$ to any $(\alpha_{1\to c}, \alpha_{2\to c}) \in [0, \bar{\alpha}_{1\to c} - \bar{\delta}] \times [0, \bar{\alpha}_{2\to c}]$; thus, the commitment solutions is preferred by the DM in both periods.

However, given any period-1 attention in $\bar{\alpha}_{1\to c} - \bar{\delta}$, period-2 self chooses not to devote any attention, which leads to an attention allocation period-1 self strictly disprefers to $(0,0)$. Since all attention allocation in $[0, \bar{\alpha}_{1\to c} - \bar{\delta}] \times [0, \bar{\alpha}_{2\to c}]$ are dispreferred to $(0,0)$ by period-1 self (and strictly so if the attention allocation is not $(0,0)$), the unique no-commitment outcome is $(\alpha_{1\to c}, \alpha_{2\to c}) = (0,0)$.

Note that with commitment, the DM devotes attention to the non-trivial payoff in both periods. If $\alpha_{1\to c} = 0$, then the DM is time-consistent, and so they could implement the commitment solution without commitment, a contradiction; if $\alpha_{1\to c} = 0$, then period-2 self cannot reduce their attention to the non-trivial payoff, and so again the commitment solution can be implemented without commitment, a contradiction.

Lastly, without the attention bounds and since $\hat{V}_0 < \hat{V}_c(1,0)$, the unique outcome with and without commitment is $(\alpha_{1\to c}, \alpha_{2\to c}) = (1,1)$.

Proof of Claim 1. We have

$$\frac{\partial V_1}{\partial \alpha_{1\to c}}(0,0) = \frac{\partial \hat{V}_c}{\partial \alpha_{1\to c}}(0,0) - \lambda(\hat{V}_o - \hat{V}_c(0,0)), \quad \text{and}$$
$$\frac{\partial V_1}{\partial \alpha_{2\to c}}(0,0) = \frac{\partial \hat{V}_c}{\partial \alpha_{2\to c}}(0,0) - \lambda(\hat{V}_o - \hat{V}_c(0,0)).$$

Since $\hat{V}_c$ is strictly increasing and $\hat{V}_c(0,0) < \hat{V}_o$, for $\lambda > \underline{\lambda} \equiv \max\{\frac{\frac{\partial}{\partial \alpha_{1\to c}}\hat{V}_c(0,0)}{\hat{V}_o - \hat{V}_c(0,0)}, \frac{\frac{\partial}{\partial \alpha_{2\to c}}\hat{V}_c(0,0)}{\hat{V}_o - \hat{V}_c(0,0)}\}$, both of the derivatives above are strictly negative and so $(0,0)$ is a strict local maximum of V_1. □

Proof of Claim 2. For the first subclaim, note that $A_1(\lambda)$ for $\lambda > \underline{\lambda}$ is non-empty since $(0,0)$ is a strict local maximum of V_1 and $\hat{V}_o < \hat{V}_c(1,0)$ implies $V_1(1,0) > V(0,0)$, and that it is compact by continuity of V_1. Thus, the minimum exists.

The second subclaim follows since, by definition of $\tilde{\alpha}_{1\to c}$, we have $V_1(\tilde{\alpha}_{1\to c}, 0) > V_1(0,0)$, and by continuity of V_1.

The third subclaim follows since $V_1(\alpha_{1\to c}, 0) - V_1(0,0) = (\hat{V}_c(\alpha_{1\to c}, 0) - \hat{V}_c(0,0)) - \lambda\alpha_{1\to c}(\hat{V}_o - \hat{V}_c(\alpha_{1\to c}, 0))$ is decreasing in λ for all $\alpha_{1\to c} < \tilde{\alpha}_{1\to c}$ because $\hat{V}_c(\tilde{\alpha}_{1\to c}, 0) = \hat{V}_o$ and $\hat{V}_c$ is increasing. Thus, for any λ, λ' with $\lambda' > \lambda$ and any $\alpha_{1\to c} \in (0, \underline{\alpha}_{1\to c}(\lambda))$, since $(\hat{V}_c(\alpha_{1\to c}, 0) - \hat{V}_c(0,0)) - \lambda\alpha_{1\to c}(\hat{V}_o - \hat{V}_c(\alpha_{1\to c}, 0)) < 0$, we also have $(\hat{V}_c(\alpha_{1\to c}, 0) - \hat{V}_c(0,0)) - \lambda'\alpha_{1\to c}(\hat{V}_o - \hat{V}_c(\alpha_{1\to c}, 0)) < 0$ and the result follows.

For the final subclaim, since $\underline{\alpha}_{1\to c}(\lambda)$ is increasing in λ (third subclaim) and bounded above (second subclaim), it must converge as $\lambda \to \infty$. Suppose $\underline{\alpha}_{1\to c}(\lambda) \to y < \tilde{\alpha}_{1\to c}$. Since $\hat{V}_o > \hat{V}_c(y, 0)$, as $y < \tilde{\alpha}_{1\to c}$, we have $V_1(y,0) < V_1(0,0)$ for large λ. But since V_1 is continuous and $V_1(\underline{\alpha}_{1\to c}(\lambda), 0) = V_1(0,0)$ for all λ large enough, it must be that $V_1(y,0) = V_1(0,0)$, i.e., we have a contradiction. □

Proof of Claim 3.

$$
\begin{aligned}
V_2(0,0) &= \hat{V}_c(0,0) + \hat{V}_o + \lambda\hat{V}_o \\
&= V_1(0,0) - \lambda\hat{V}_o \\
&= V_1(\alpha_{1\to c}, \alpha_{2\to c}) - \lambda\hat{V}_o \\
&= (1 + \lambda(\alpha_{1\to c} + \alpha_{2\to c}))\hat{V}_c(\alpha_{1\to c}, \alpha_{2\to c}) + (1 + \lambda(1 - \alpha_{1\to c} + 1 - \alpha_{2\to c}))\hat{V}_o - \lambda\hat{V}_o \\
&= V_2(\alpha_{1\to c}, \alpha_{2\to c}) - \lambda\alpha_{1\to c}(\hat{V}_o - \hat{V}_c(\alpha_{1\to c}, \alpha_{2\to c})) \\
&< V_2(\alpha_{1\to c}, \alpha_{2\to c}),
\end{aligned}
$$

where the inequality follows as $V_1(0,0) = V_1(\alpha_{1\to c}, \alpha_{2\to c})$ implies $\hat{V}_o > \hat{V}_c(\alpha_{1\to c}, \alpha_{2\to c})$. The claim follows by continuity. □

□

C.8 Proof of Proposition 9

Proof of Proposition 9. The first claim follows from Proposition 1, the second from Topkis's Theorem since the DM's objective has increasing differences in (α_w, v_H). For the third, note that the cross-partial derivative of the DM's objective with respect to

α_w and v_L is given by

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

If $p(\alpha_w) + \alpha_w \frac{\partial}{\partial\alpha_w} p(\alpha_w) < 1$ everywhere, then the above becomes positive for large enough λ and the claim follows again from Topkis's Theorem.[28] □

C.9 Proof of Proposition 10

Proof of Proposition 10. For the first claim, fix x_1 and consider a state $s \in \mathcal{S}$. Clearly, if $\max_{(x_2,\alpha_2),x_2 \in X_2(\alpha_2)} V_c(x_1, x_2|s) \geq V_o$, solving (4) gives $\alpha_{2\to c} = 1$ as the optimal attention allocation (for any λ), i.e., indeed $s \notin \underline{\mathcal{S}}(x_1)$. If $\max_{(x_2,\alpha_2),x_2 \in X_2(\alpha_2)} V_c(x_1, x_2|s) < V_o$, then for λ large enough, (4) is maximized by $\alpha_{2\to c} = 0$. Since $\mathcal{S}$ is finite, taking the maximum lower bound on λ (over finite $x_1 \in X_1(\alpha_1)$ and $s \in \mathcal{S}$) implies the result.

For the second claim, note that since $X_1(\alpha_1)$ is independent of α_1 and $\beta = 0$, period-1 attention utility is independent of x_1. This is because period-1 self cannot affect the highest payoff, which it devotes all attention to: If the highest payoff is the non-trivial payoff in some state s, then period-2 self will devote all attention to it, and so the default does not bind and since $\beta = 0$, there is no direct impact; if the highest payoff is the trivial payoff, then, of course, period-1 self does not affect it.

Thus, the DM only maximizes (4) in both periods. But then the claim follows immediately since it is as if the DM jointly chooses the states in which they devote attention to the non-trivial payoff as well as x_1; if the claim were not true, the DM could improve their overall utility by choosing x_1^* equal to the argmax given $\underline{\mathcal{S}}(x_1)$, since the choice of x_1 only affects payoffs in those states.

For the final claim, not that from the first claim, we know for any x_1', $\underline{\mathcal{S}}(x_1')$ is independent of λ for large λ, i.e., $x_2(x_1', s)$ and $\alpha_2(x_1', s)$ are independent of λ. Thus,

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

the sum of period-1 attention utility and expected period-2 attention utility, i.e.,

$$\underbrace{\sum_{s\in\mathcal{S}}\alpha_{1\to(c,s)}(v_c(x_1',x_2(x_1',s)|s)+\beta u(x_1'|s))+\alpha_{1\to o}V_o}_{\text{period-1 attention utility}} + \underbrace{\sum_{s\in\mathcal{S}}p_s(\alpha_{2\to c}(x_1',s)(v_c(x_1',x_2(x_1',s)|s)+\beta u(x_1'|s))+\alpha_{2\to o}(x_1',s)V_o)}_{\text{expected period-2 attention utility}},$$

is independent of λ, for large λ. Thus, the optimal action taken by period-1 self must maximize the expression above; otherwise, it cannot be optimal for large λ (and since $X_1(\alpha_1)$ is finite).

Also, note that period-2 choice is optimal for period-1 self for large λ: Period-1 attention utility is determined by either the trivial payoff, in which case period-2 choice does not affect it, or the non-trivial payoff in some state s, in which case period-2 self devotes attention if and only if the payoff is at least the attentional outside option, as period-1 self would also dictate. Thus, x_1, the optimal period-1 action, must also maximize the above when $\underline{\mathcal{S}}(x_1')$ is held fixed at $\underline{\mathcal{S}}(x_1)$. Since $\underline{\mathcal{S}}(x_1)\neq\mathcal{S}$, x_1 maximizing the above means that it also maximizes

$$\underbrace{V_c(x_1',x_2(x_1',s^*)|s^*)}_{\text{period-1 attention utility}}+\underbrace{\sum_{s\in\mathcal{S}\setminus\underline{\mathcal{S}}(x_1)}p_s(v_c(x_1',x_2(x_1',s)|s)+\beta u(x_1'|s))+\sum_{s\in\underline{\mathcal{S}}(x_1)}p_sV_o}_{\text{expected period-2 attention utility}}$$

where $s^*\in\arg\max_{s\in\mathcal{S}}V_c(x_1',x_2(x_1',s^*)|s^*)$. But then the claim immediately follows for $\tilde{p}_s=p_s$ for $s\in\mathcal{S}\setminus\underline{\mathcal{S}}(x_1)$ such that $s\neq s^*$ and $\tilde{p}_{s^*}=p_{s^*}+1$, since $v_c(x_1',x_2(x_1'|s)$ is independent of x_1' for $s\in\mathcal{S}\setminus\underline{\mathcal{S}}(x_1)$ and V_o is also independent of x_1'. □

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 $B'\coloneqq(B\cup\{D\cup D'\})\setminus\{D,D'\}$, i.e., an alternative bracketing where sets D and D' are considered jointly. Evaluate (8) 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 then implies that the above is non-positive, i.e., if $\bar{V}_D(x) > \bar{V}_{D'}(x)$, then $\bar{\alpha}_D \geq \bar{\alpha}_{D'}$. □

D General model and robustness

As is true for every model, the approach pursued in the body of the paper makes several functional form assumptions. Although natural, these assumptions raise the question of whether the intuitions in the paper are robust to either more general functions or alternative specifications. The goal of this appendix is three-fold. First, in Section D.1, we discuss specific functional form assumptions that we make and suggest alternatives and generalizations. Then, in Section D.2, we provide a more general model that encompasses all the alternatives we suggest. Lastly, in Section D.3, we discuss what conditions on the general model still ensure that our main intuitions from the body of the paper still hold.

D.1 Relaxing specific functional form assumptions

We first discuss specific function form assumptions, as well as generalizations and alternatives to the assumptions and how they matter for behavior. Here, we focus on exploring each assumption in isolation; in Section D.2, we consider a functional form that relaxes all of them at once.

D.1.1 Functional dependence on attention and payoff

We assumed that attention utility in each dimension is linear in the product of attention to that dimension and the payoff in that dimension. Doing so allows us to highlight the novel behaviors that attention utility can accommodate even under these restrictive assumptions while ruling out alternative drivers of behavior (like nonlinear anticipatory utility). We can generalize this, though: Let attention utility for dimension i be

$$\psi(\alpha_i, V_i(x)),$$

where ψ is a continuous function that maps from $[0, 1] \times \mathbb{R}$ to a real scalar and is increasing in both arguments. (To rationalize the motivating evidence of (in)attention to (low) high payoffs, such as the ostrich effect, one may impose payoff $V_i(x)$ and attention α_i to be complements, i.e., ψ to have increasing differences.) We next discuss two particular examples of functions that lie between the model in the body of the paper and the general model presented here.

One such example is $\psi(\alpha_i, V_i(x)) = \alpha_i h(V_i(x))$, where h is a (potentially nonlinear) continuous, increasing function that maps from $\mathbb{R}$ to $\mathbb{R}$. Thus, payoffs may enter attention utility nonlinearly. Nonlinearity of anticipated payoffs is a key feature of models of anticipatory utility and intrinsic preferences for information, such as Caplin and Leahy (2001), Kreps and Porteus (1978), or Epstein and Zin (1989), and drives preferences for earlier or later resolution of information. Importantly, our base model in the body of the paper can generate "non-standard" behavior even absent this additional curvature.

Alternatively, attention may enter attention utility nonlinearly: $\psi(\alpha_i, V_i(x)) = f(\alpha_i)V_i(x)$, where $f : [0,1] \to [0,1]$ is continuous and increasing. Thus, attention utility may exhibit increasing or decreasing returns to attention (conditional on action x). One implication is an additional mechanism that may drive probability weighting when attention is allocated across states (Section 2.4). To illustrate, suppose that $f(\alpha_i) = \min\{\alpha_i, l\}$, where $l > 1/2$. Suppose that attention is not instrumental, so the mechanism that generates inverse-S-shaped probability weighting in the body of the paper is eliminated. If $l = 1$, we are back to the standard model where this setting leads to optimism—the high-payoff state is overweighted. However, if $l < 1$, the DM always devotes some attention to all states, namely l attention to the high-payoff state and $1 - l$ to the low-payoff state. Thus, states with a probability less than $1 - l$ are again overweighted, and those with a probability more than l are underweighted.

Notice also that a concave f leads to a preference for spreading attention across many dimensions—in this case, individuals may prefer to defer the resolution of uncertainty if doing so allows them to devote attention to multiple (not yet realized) states.

D.1.2 Attentional spillovers

A second assumption we make is that attention utility from dimension i is a function of α_i only (conditional on action x) rather than the entire attention vector α. Such an assumption may be unnatural; e.g., in Appendix B, we discuss how individuals may want to "bracket" multiple dimensions together so that if they think about dimension i (e.g., schoolwork), they must also think about dimension j (e.g., getting ice cream after finishing schoolwork).

Such attentional spillovers may also be a natural feature of the environment. For example, consider a situation where we measure time in minutes. In this case, thinking about what occurs at 12:01 PM may necessitate also thinking about what happens at 12:00 PM and 12:02 PM. More generally, there are a variety of situations where attention to one dimension naturally leads one to think about other dimensions simply due to temporal or mental adjacency.

To accommodate these attentional spillovers, one may write attention utility from dimension i as

$$g_i(\alpha)V_i(x),$$

where $g_i : [0,1]^{|\mathcal{D}|} \to [0,1]$ is continuous with $g_i(\alpha)$ as the "effective attention" dimension i receives as a function of the entire attentional vector. For instance, dimensions i and j may then form a bracket if $g_i(\alpha) = g_j(\alpha) = \frac{\alpha_i+\alpha_j}{2}$ (see Appendix B, where g_i is a choice variable).

D.1.3 Dependence on the weight assigned to a dimension in the material utility

A third assumption is that the attention utility from dimension i is independent of the weight ω_i assigned to i in the material utility. This assumption, we feel, more than the previous ones (and the following ones), is a substantive restriction on the psychology of attention.

In some situations, the assumption seems reasonable. For example, someone may get the same utility from thinking about what their life would be like if their $100 million lottery ticket was a winner, regardless of the odds; intuitively, the individual imagines their life conditional on a state, and so the likelihood of that state does not matter. On the other hand, in many applications, this assumption may feel odd. If we think of dimensions as having a temporal dimension, but certain dimensions stretch longer than others (e.g., one covers 5 minutes, the other a day), then it is potentially incorrect to assume that they loom equally large in attention utility. Even more starkly, one might ask how an individual could gain utility from thinking about situations that are not just improbable—but impossible.

Such concerns may be particularly pertinent in our applications that suppose different dimensions correspond to different states of the world. Here, the approach taken in the body of the paper requires us to distinguish between states that have zero weight in material utility (i.e., $p_i = 0$) but can attract attention versus those that have zero weight and cannot attract attention. To address such concerns, we allow for the weights on dimensions in material utility to impact attention utility.

The first, and perhaps most obvious, way to proceed is to add the weight of dimension i, ω_i as a multiplicative factor in the attention utility for dimensions i. That is, attention utility from dimension i becomes

$$\alpha_i \omega_i V_i(x).$$

Immediately then, if a dimension has 0 weight in the material utility, it cannot generate non-zero attention utility.

Although such a model parsimoniously addresses the concerns raised above, it introduced a new, more subtle, concern related to event splitting, which we call "dimension splitting." Event splitting has been discussed extensively when considering choice over risk, e.g., Starmer and Sugden (1993); Humphrey (1995); Birnbaum and Navarrete (1998), and in riskless domains, e.g., Weber et al. (1988); Bateman et al. (1997). In order to understand the concerns, we first formally define what we mean by dimension splitting in our framework.

Take a dimension i and remove it, but add two dimension i' and i'', such that $V_i = V_{i'} = V_{i''}$, with $\omega_{i'} + \omega_{i''} = \omega_i$. In the domain of risk, this is equivalent to the notion of event splitting. With the weights on material utility determining attention utility, the optimal attention allocation (and action) may change: α_i^* need not equal $\alpha_{i'}^* + \alpha_{i''}^*$. This is because ω_i and α_i are complements in attention utility. Consequently, the DM's behavior may differ; for example, in the domain of risk, it can lead to a different pattern of probability weighting across the two problems.

Consistent with the predictions of this theory, behavior changing due to event splitting has been documented as a robust empirical phenomenon in experimental settings (see previous references as well as more recent work like Birnbaum et al. (2017)). That said, it is not clear whether we always want a model that predicts behavior should change in response to dimension splitting. Thus, we would like to have a generalized approach that can either make behavior invariant to or dependent on dimension splitting, depending on the specific assumptions.

We do so by assuming attention utility from dimension i is

$$g_i(\alpha)\Omega_i(\omega)V_i,$$

where Ω_i is a continuous function that typically maps from the vector of weights to a weakly positive real scalar (although for one example in Section D.2, we assume Ω_i maps to $\mathbb{R}^+ \times \mathbb{R}^+$). When $g_i(\alpha) = \alpha_i$ and $\Omega_i(\omega) = \omega_i$, we recover the "naive" model proposed above, where behavior can change when a dimension is split.

However, we can also impose assumptions such that behavior is invariant to dimension splitting. To this end, add a new primitive of the model, a set of "mirror" dimensions: Define a mirroring $J \in \mathcal{P}(\mathcal{D})$, i.e., a partition of the set of dimensions, and denote $J(i)$ as the mirror that contains dimension i. Two dimensions are in the same mirror if they come from a split dimension. We then make Ω_i functions of J. In particular, we can let $g_i(\alpha) = \alpha_i$ and $\Omega_i(\omega) = \sum_{j \in J(i)} \omega_j$. Now, dimension splitting need not affect behavior; e.g., if there is no instrumental role of attention, the DM's attention allocation is essentially unchanged. Of course, as discussed, whether or not we want to impose assumptions that make behavior invariant, or not, to dimensions splitting is an empirical question.

D.1.4 Additive separability of material and attention utility

We also assumed material utility is additively separable from attention utility. This separation between material utility, which we conceive of as the utility that "standard" approaches analyze, and attention utility, which is our novel contribution, helps lay out what features of our model generate novel behavioral predictions and, more generally, also allows to conduct comparative static exercises with respect to the weight on attention utility.

Nevertheless, generalizing our approach is, in principle, relatively easy. By combining (and extending) the generalizations already proposed in the previous points, we naturally generate a model that both i) nests the model in the body of the paper where material and attention utility are additively separable but ii) does not require that this be the case. Let overall utility—and not just attention utility—(from dimension i) be defined as

$$\varsigma(\alpha_i, \omega_i, V_i(x)),$$

where $\varsigma : [0,1] \times \mathbb{R}^+ \times \mathbb{R} \to \mathbb{R}$ is continuous and increasing in all its arguments. If

$\varsigma(\alpha_i, \omega_i, V_i(x)) = \omega_i V_i(x) + \lambda \alpha_i V_i(x)$, then we have recovered the model used in the body of the paper. But, of course, this functional form can also naturally allow for other interactions between the attributes and attention of a dimension.

D.1.5 Additive separability across dimensions

Lastly, we assumed utility across different dimensions is additively separable. Although, as discussed, the previous assumption can be generalized in various ways while maintaining key intuitions, we do not believe this is the case for this assumption. In particular, the most natural generalization is to relax the additive separability across dimensions to just separability. As it turns out, some of our main results cannot be maintained with functions that are simply separable (rather than additively separable) in attention utility across different dimensions.

By way of example, consider a multiplicative version of our model where there is no material utility and where total attention utility is the product of the attention utility from any given dimension. To keep things simple, suppose that there are only two dimensions. Then, attention utility is

$$(\alpha_i V_i)(\alpha_j V_j).$$

This immediately introduces some problems in recreating the intuitions from the body of the paper. In particular, notice that in this functional form, returns to α_i and α_j are exactly the same. This means that changes in V_i (as in Proposition 1) affect the incentive to increase α_i and α_j equally. In fact, the model implies that the DM should want to set $\alpha_i = \alpha_j$ so long as $V_i, V_j > 0$, which bears little resemblance to the nuanced implications for attentional allocation that our model (which embeds additive separability) generates.

D.2 A general functional form

In the previous subsection, we discussed how to relax four key functional form assumptions (and how one—additive separability across dimensions—could not be relaxed). Combining the insights and generalizations suggested across those four discussions leads us to a general functional form for the overall utility given by

$$\sum_i \phi(g_i(\alpha), \Omega_i(\omega), V_i(x)).$$

We assume ϕ, g_i and Ω_i have the properties described in Section D.1, and ϕ_i is a continuous function mapping from $[0,1] \times \mathbb{R}^+ \times \mathbb{R}$ to $\mathbb{R}$ (although in one case, see the third bullet point below, we consider a slightly mapping for both Ω_i and ϕ_i). This general function nests the formulations suggested in the previous subsections:

- Section D.1.1: $g_i(\alpha) = \alpha_i$, $\Omega_i(\omega) = \omega_i$, and $\phi(g_i, \Omega_i, V_i) = \Omega_i V_i + \lambda\psi(g_i, V_i)$.
- Section D.1.2: $\Omega_i(\omega) = \omega_i$, and $\phi(g_i, \Omega_i, V_i) = \Omega_i V_i + \lambda g_i V_i$.
- Section D.1.3: Consider a slightly modified version of Ω_i and ϕ so that Ω_i is a continuous mapping from a vector of weights, to a doubleton of scalars (ω_i^M, ω_i^A) (for weight on material and attention utility, respectively), and $\phi(g_i, \Omega_i, V_i) = \omega_i^M V_i + \lambda g_i \omega_i^A V_i$.
- Section D.1.4: $g_i(\alpha) = \alpha_i$, $\Omega_i(\omega) = \omega_i$.

D.3 Generalizing Propositions 1– 3

One concern is that the general model introduced in the previous subsection may be "too general" in that it may not always generate the key intuitions that drive the behavior described in the paper; in this section, we show under which restrictions Propositions 1–3 generalize.

To begin, we parameterize payoff V_i in the same way as in the body of the paper: $V_i(x) = \gamma_i + \beta v_i(x)$. We also provide a definition that generalizes the notion of separability.

Definition 2. *The environment is* separable-g *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 $X(\alpha) = \Pi_{i \in \mathcal{D}} X_i(g_i(\alpha))$.*
- *X_i is monotone, i.e., $X_i(g_i) \subseteq X_i(g_i')$ for all $g_i \leq g_i'$.*

Recall that Proposition 1 made two claims: i) if the payoff level of dimension i goes up, attention to i goes up, and if the environment is separable, then the part of the payoff determined by actions also goes up; and ii) if the slope (with respect to the action) of the payoff function increases in dimension i (controlling for the level) then the part of the payoff in dimension i determined by the action goes up, and if the environment is separable, attention to i also goes up.

The following results say that the second finding immediately generalizes, while the first one generalizes under the assumption that ϕ has increasing differences in $(g_i(\alpha), V_i(x))$. All proofs of the propositions in this section are in Section D.4.

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

- *Suppose ϕ has strictly increasing differences in $(g_i(\alpha), V_i(x))$. If $\gamma_i' > \gamma_i$, then $\min_{(x,\alpha)\in\Gamma(\gamma_i',\beta_i)} g_i(\alpha') \geq \max_{(x,\alpha)\in\Gamma(\gamma_i',\beta_i)} g_i(\alpha)$. If, in addition, the environment is g-separable, then $\min_{(x,\alpha)\in\Gamma(\gamma_i',\beta_i)} v_i(x) \geq \max_{(x,\alpha)\in\Gamma(\gamma_i,\beta_i)} v_i(x)$.*

- *If 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 $\gamma_i^{'} = \gamma_i - (\beta_i^{'} - \beta_i)v_i(x)$, where $(x,\alpha) \in \Gamma(\gamma_i, \beta_i)$, we have $\min_{(x,\alpha)\in\Gamma(\gamma_i',\beta_i')} v_i(x) \geq \max_{(x,\alpha)\in\Gamma(\gamma_i,\beta_i)} v_i(x)$. If, in addition, the environment is g-separable, then $\min_{(x,\alpha)\in\Gamma(\beta_i',\gamma_i')} g_i(\alpha) \geq \max_{(x,\alpha)\in\Gamma(\beta_i,\gamma_i)} g_i(\alpha)$.*

Proposition 2 considered comparative statics in the weight on attention utility, λ. It thus requires a clear distinction between material and attention utility, parameterized by a relative weight. To generalize Proposition 2, we first formulate a version of the general functional form that allows for a similar distinction. In particular, material utility is defined as utility that does not directly depend on attention; attention utility, instead, may depend on attention directly.

Proposition 2-g. *Suppose $\phi(g_i(\alpha), \Omega_i(\omega), V_i(x)) = U^M(\Omega_i^M(\omega), V_i(x)) + \lambda U^A(g_i(\alpha), \Omega_i^A(\omega), V_i(x))$ for all $i \in \mathcal{D}$. Consider a change of parameter λ to λ' with $\lambda' > \lambda$ and let x and x' denote the optimal actions, respectively. We have:*

- *$\sum_i U^M(\Omega_i^M(\omega), V_i(x)) \geq U^M(\Omega_i^M(\omega), V_i(x'))$;*

- *if ϕ is convex in $V_i(x)$, then the DM's value is convex in $(\gamma_i)_{i\in\mathcal{D}}$;*

- *if the environment is g-separable, for each* $i \in \mathcal{D}$*, if* $U^A(g_i(\alpha), \Omega_i^A(\omega), V_i(x)) = g_i(\alpha)U^M(\Omega_i^M(\omega), V_i(x))$*,* $g_i(\alpha) = f_i(\alpha_i)$ *for some strictly increasing* f_i*, and the objective given* λ *is convex in* α_i*, then the objective is also convex in* α_i *given* λ'*.*

Proposition 3 considered exogenous weight on material and attention utility in a given dimension; thus, the generalization of Proposition 3 again requires a distinction between material utility and attention utility. As it turns out, the multiplicative structure of the weights in the base model can be relaxed to increasing differences.

Proposition 3-g. *Suppose* $\phi(g_i(\alpha), \Omega_i(\omega), V_i(x)) = U^M(\Omega_i^M(\omega), V_i(x)) + \lambda U^A(g_i(\alpha), \Omega_i^A(\omega), V_i(x))$*, where* $\Omega_i^M = \tilde{\Omega}_i^A + \eta_i^M$ *and* $\Omega_i^A = \tilde{\Omega}_i^A + \eta_i^A$ *for all* $i \in \mathcal{D}$*. Consider dimension* $i \in \mathcal{D}$*. Fix* $g_{-i}, \Omega_{-i}^M, \Omega_{-i}^A$*, and* V_{-i}*. Let* $\Gamma(\eta_i^M, \eta_i^A)$ *denote the set of optimal (action, attention)-pairs.*

- *Suppose* U^M *has increasing differences in* $(V_i(x), \Omega_i^M(\omega))$*. If* $\eta_i^{M'} > \eta_i^M$*, then* $\min_{(x,\alpha)\in\Gamma(\eta_i^{M'}, \eta_i^A)} V_i(x) \geq \max_{(x,\alpha)\in\Gamma(\eta_i^M, \eta_i^A)} V_i(x)$*. If, in addition, the environment is g-separable, then* $\min_{(x,\alpha)\in\Gamma(\eta_i^{M'}, \eta_i^A)} g_i(\alpha) \geq \max_{(x,\alpha)\in\Gamma(\eta_i^M, \eta_i^A)} g_i(\alpha)$*.*

- *Suppose* U^A *has increasing differences in* $(g_i(\alpha), \Omega_i^A(\omega))$ *as well as in* $(V_i(x), \Omega_i^A(\omega))$*. If the environment is g-separable then, if* $\eta_i^{A'} > \eta_i^A$*, we have*

$$\min_{(x,\alpha)\in\Gamma(\eta_i^M, \eta_i^{A'})} V_i(x) \geq \max_{(x,\alpha)\in\Gamma(\eta_i^M, \eta_i^A)} V_i(x), \quad \textit{and}$$
$$\min_{(x,\alpha)\in\Gamma(\eta_i^M, \eta_i^{A'})} g_i(\alpha) \geq \max_{(x,\alpha)\in\Gamma(\eta_i^M, \eta_i^A)} g_i(\alpha).$$

D.4 Proofs of Section D.3

Proof of Proposition 1-g. 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

$$\begin{aligned}
&\underbrace{\sum_{j \in \mathcal{D}\setminus\{i\}} \phi(g_j(\alpha), \Omega_j(\omega), V_j(x))}_{:=\kappa_0} + \phi(g_i(\alpha), \Omega_i(\omega), \beta_i v_i(x) + \gamma_i) \\
\geq &\underbrace{\sum_{j \in \mathcal{D}\setminus\{i\}} \phi(g_j(\alpha'), \Omega_j(\omega), V_j(x'))}_{:=\kappa_1} + \phi(g_i(\alpha'), \Omega_i(\omega), \beta_i v_i(x') + \gamma_i) \quad \text{and} \\
&\underbrace{\sum_{j \in \mathcal{D}\setminus\{i\}} \phi(g_j(\alpha'), \Omega_j(\omega), V_j(x'))}_{:=\kappa_1} + \phi(g_i(\alpha'), \Omega_i(\omega), \beta_i v_i(x') + \gamma_i') \\
\geq &\underbrace{\sum_{j \in \mathcal{D}\setminus\{i\}} \phi(g_j(\alpha), \Omega_j(\omega), V_j(x))}_{:=\kappa_0} + \phi(g_i(\alpha), \Omega_i(\omega), \beta_i v_i(x) + \gamma_i').
\end{aligned}$$

Combining the above gives

$$\begin{aligned}
&-\left(\phi(g_i(\alpha), \Omega_i(\omega), \beta_i v_i(x) + \gamma_i') - \phi(g_i(\alpha'), \Omega_i(\omega), \beta_i v_i(x') + \gamma_i')\right) \\
\geq &\kappa_0 - \kappa_1 \\
\geq &-\left(\phi(g_i(\alpha), \Omega_i(\omega), \beta_i v_i(x) + \gamma_i) - \phi(g_i(\alpha'), \Omega_i(\omega), \beta_i v_i(x') + \gamma_i)\right).
\end{aligned}$$

Since ϕ has strictly increasing differences in $(g_i(\alpha), V_i(x))$, it must be that $g_i(\alpha') \geq g_i(\alpha)$.

If the environment is g-separable, then v_i is increasing in $g_i(\alpha_i)$, 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 $\gamma_i' = \gamma_i - (\beta_i' - \beta_i)v_i(x)$, where $(x, \alpha) \in \Gamma(\gamma_i, \beta_i)$. Let (x, α) and (x', α') denote a solution given (β_i, γ_i) and (β_i', γ_i'), respectively. Optimality of

(x, α) and (x', α') implies

$$\underbrace{\sum_{j \in \mathcal{D} \setminus \{i\}} \phi(g_j(\alpha), \Omega_j(\omega), V_j(x))}_{:=\kappa_2} + \phi(g_i(\alpha), \Omega_i(\omega), \beta_i v_i(x) + \gamma_i)$$
$$\geq \underbrace{\sum_{j \in \mathcal{D} \setminus \{i\}} \phi(g_j(\alpha'), \Omega_j(\omega), V_j(x'))}_{:=\kappa_3} + \phi(g_i(\alpha'), \Omega_i(\omega), \beta_i v_i(x') + \gamma_i) \quad \text{and}$$
$$\underbrace{\sum_{j \in \mathcal{D} \setminus \{i\}} \phi(g_j(\alpha'), \Omega_j(\omega), V_j(x'))}_{:=\kappa_3} + \phi(g_i(\alpha'), \Omega_i(\omega), \beta_i' v_i(x') + \gamma_i')$$
$$\geq \underbrace{\sum_{j \in \mathcal{D} \setminus \{i\}} \phi(g_j(\alpha), \Omega_j(\omega), V_j(x))}_{:=\kappa_2} + \phi(g_i(\alpha), \Omega_i(\omega), \beta_i' v_i(x) + \gamma_i').$$

Combining the above and substituting for γ_i' gives

$$\begin{aligned}
&-(\phi(g_i(\alpha), \Omega_i(\omega), \beta_i v_i(x) + \gamma_i) - \phi(g_i(\alpha'), \Omega_i(\omega), \beta_i' v_i(x') + \gamma_i - (\beta_i' - \beta_i) v_i(x))) \\
\geq &\kappa_2 - \kappa_3 \\
\geq &-(\phi(g_i(\alpha), \Omega_i(\omega), \beta_i v_i(x) + \gamma_i) - \phi(g_i(\alpha'), \Omega_i(\omega), \beta_i v_i(x') + \gamma_i)).
\end{aligned}$$

Since ϕ is strictly increasing in $V_i(x)$, the outer inequality implies

$$-(v_i(x) - v_i(x'))(\beta_i' - \beta_i) \geq 0,$$

and thus, it must be that $v_i(x') \geq v_i(x)$.

If the environment is g-separable, then v_i is increasing in $g_i(\alpha)$, and the result follows. □

Proof of Proposition 2-g. Take any λ', λ 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

$$\begin{aligned}
&\sum_i U^M(\Omega_i^M(\omega), V_i(x)) + \lambda \sum_i U^A(g_i(\alpha), \Omega_i^A(\omega), V_i(x)) \\
\geq &\sum_i U^M(\Omega_i^M(\omega), V_i(x')) + \lambda \sum_i U^A(g_i(\alpha'), \Omega_i^A(\omega), V_i(x')), \quad \text{and} \\
&\sum_i U^M(\Omega_i^M(\omega), V_i(x')) + \lambda' \sum_i U^A(g_i(\alpha'), \Omega_i^A(\omega), V_i(x')) \\
\geq &\sum_i U^M(\Omega_i^M(\omega), V_i(x)) + \lambda' \sum_i U^A(g_i(\alpha), \Omega_i^A(\omega), V_i(x)).
\end{aligned}$$

Combining the above gives

$$\begin{aligned}
&- \lambda' \left(\sum_i U^A(g_i(\alpha), \Omega_i^A(\omega), V_i(x)) - \sum_i U^A(g_i(\alpha'), \Omega_i^A(\omega), V_i(x')) \right) \\
\geq &\sum_i U^M(\Omega_i^M(\omega), V_i(x)) - U^M(\Omega_i^M(\omega), V_i(x')) \\
\geq &- \lambda \left(\sum_i U^A(g_i(\alpha), \Omega_i^A(\omega), V_i(x)) - \sum_i U^A(g_i(\alpha'), \Omega_i^A(\omega), V_i(x')) \right).
\end{aligned}$$

To reach a contradiction, suppose the expression in the middle is strictly negative. Then, the expression on the right must also be strictly negative; but then it is strictly larger than the left one as $\lambda' > \lambda$—a contradiction. 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 $\chi \in [0, 1]$. Let

$$(\alpha^*, x^*) \in \arg\max_{\alpha, x \in X(\alpha)} \sum_i \phi(g_i(\alpha), \Omega_i(\omega), \beta_i v_i(x) + \chi \gamma_i + (1 - \chi)\gamma'_i).$$

Then

$$\begin{aligned}
&\sum_i \phi(g_i(\alpha^*), \Omega_i(\omega), \beta_i v_i(x^*) + \chi \gamma_i + (1 - \chi)\gamma'_i) \\
\leq &\chi \sum_i \phi(g_i(\alpha^*), \Omega_i(\omega), \beta_i v_i(x^*) + \gamma_i) + (1 - \chi) \sum_i \phi(g_i(\alpha^*), \Omega_i(\omega), \beta_i v_i(x^*) + \gamma'_i) \\
\leq &\chi \max_{\alpha, x \in X(\alpha)} \sum_i \phi(g_i(\alpha), \Omega_i(\omega), \beta_i v_i(x) + \gamma_i) + (1 - \chi) \max_{\alpha, x \in X(\alpha)} \sum_i \phi(g_i(\alpha), \Omega_i(\omega), \beta_i v_i(x) + \gamma'_i),
\end{aligned}$$

where the first inequality follows from the convexity of ϕ in $V_i(x)$ and the second inequality from optimality, and so the second claim follows.

Now suppose the environment is g-separable; consider dimension $i \in \mathcal{D}$ and suppose $U^A(g_i(\alpha), \Omega_i^A(\omega), V_i(x)) = g_i(\alpha)U^M(\Omega_i^M(\omega), V_i(x))$, $g_i(\alpha) = f_i(\alpha_i)$, and the objective given λ is convex in α_i.

As in the proof of the analogous part in Proposition 2, it suffices to show that if $U^M(\Omega_i^M(\omega), \hat{V}_i(\alpha_i))$ is convex in α_i, so is $f_i(\alpha_i)U^M(\Omega_i^M(\omega), \hat{V}_i(\alpha_i))$, where we substituted $\hat{V}_i$ for V_i and f_i for g_i.

Next, denote $f_i(\alpha_i)$ by $\tilde{\alpha}_i$. Then, considering $\tilde{\alpha}_i$ as the choice variable, we have to show that if $U^M(\Omega_i^M(\omega), \hat{V}_i(f_i^{-1}(\tilde{\alpha}_i)))$ is convex in $\tilde{\alpha}_i$, so is $\tilde{\alpha}_i U^M(\Omega_i^M(\omega), \hat{V}_i(f_i^{-1}(\tilde{\alpha}_i)))$, where f_i^{-1} denotes the inverse of f_i and is well defined as f_i is strictly increasing. But this is exactly what the analogous part in Proposition 2 states, and so the third claim follows. □

Proof of Proposition 3-g. Both parts of the propositions are direct implications of Topkis's Theorem. □

Appendix References

Bateman, Ian, Alistair Munro, Bruce Rhodes, Chris Starmer, and Robert Sugden, "Does part–whole bias exist? An experimental investigation," *The Economic Journal*, 1997, *107* (441), 322–332.

Birnbaum, Michael H and Juan B Navarrete, "Testing descriptive utility theories: Violations of stochastic dominance and cumulative independence," *Journal of risk and uncertainty*, 1998, *17*, 49–79.

_ **, Ulrich Schmidt, and Miriam D Schneider**, "Testing independence conditions in the presence of errors and splitting effects," *Journal of Risk and Uncertainty*, 2017, *54*, 61–85.

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

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

Fudenberg, Drew, Ryota Iijima, and Tomasz Strzalecki, "STOCHASTIC CHOICE AND REVEALED PERTURBED UTILITY," *Econometrica*, 2015, *83* (6), 2371–2409.

Hellwig, Martin, Wolfgang Leininger, Philip J Reny, and Arthur J Robson, "Subgame perfect equilibrium in continuous games of perfect information: An elementary approach to existence and approximation by discrete games," *Journal of Economic Theory*, 1990, *52* (2), 406–422.

Humphrey, Steven J, "Regret aversion or event-splitting effects? More evidence under risk and uncertainty," *Journal of risk and uncertainty*, 1995, *11*, 263–274.

Kahana, Michael Jacob, *Foundations of Human Memory*, Oxford University Press, 2012.

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

Maćkowiak, Bartosz, Filip Matějka, and Mirko Wiederholt, "Survey: Rational Inattention, a Disciplined Behavioral Model," Discussion Paper 13243, CEPR 2018.

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

Starmer, Chris and Robert Sugden, "Testing for juxtaposition and event-splitting effects," *Journal of risk and uncertainty*, 1993, *6* (3), 235–254.

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

Weber, Martin, Franz Eisenführ, and Detlof Von Winterfeldt, "The effects of splitting attributes on weights in multiattribute utility measurement," *Management Science*, 1988, *34* (4), 431–445.