Emotional Inattention

Lukas Bolte* Collin Raymond*

October 19, 2023

Abstract

We propose a framework where a decision-maker allocates attention across payoff-dimensions, which can be different dimensions of consumption, realizations of an unknown state, or time periods. Attention has two features: (1) it is instrumentally valuable by allowing the decision-maker to take actions, and (2) it leads to an emotional response, which is proportional to the attention devoted to a dimension and the associated payoff. The framework provides a unifying explanation for a number of behavioral phenomena, such as the ostrich effect, optimism and other forms of subjective probability weighting, incomplete consumption smoothing, dynamic inconsistency, 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 Douglas Bernheim, Gabriel Carroll, Matthew Jackson, Muriel Niederle, and seminar participants at Stanford University, BEAM, BRIC, M-BEES/M-BEPS, SABE, and ESA 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 is well recognized and studied by both psychologists (since at least James (1890); see Desimone et al. (1995) for a review) and economists (e.g., Sims (2003); Loewenstein and Wojtowicz (2023)).

Evidence from psychology and cognitive science also highlights an important second aspect of attention, one that is typically absent from economic models: 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 generates excitement.

This paper extends economists' understanding of the role of attention in decision-making by developing a tractable model that incorporates both the instrumental and the less attended-to emotional aspects of attention. We demonstrate how trading off these dual roles of attention is a unifying mechanism that can rationalize a large number of well-known behavioral anomalies: Across different economic 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, (as in Brunnermeier and Parker (2005); Sharot (2011); Kahneman and Tversky (1979)), incomplete consumption smoothing and memorable consumption (as in Hai et al. (2020)), dynamic inconsistency (Laibson, 1997), and default effects (Carroll et al., 2009). Although these phenomena are

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

often thought of as manifestations of distinct psychological considerations, our results point out that a single cognitive mechanism may underpin many seemingly disparate behavioral patterns.

We operationalize the two consequences 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 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. For example, when the DM devotes attention to an upcoming vacation, they receive some additional utility because of feelings of anticipation. In spirit, this approach is similar to models of anticipatory utility (e.g., Loewenstein (1987); Caplin and Leahy (2001)), which assume that agents derive flow utility as a function of beliefs about future payoffs. Our innovation is to make this and other flow utilities from attention, e.g., from remembering, 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.

After formally introducing this framework in Section 2.1, we derive general properties of the optimal attention allocation in Section 2.2, which apply across our different settings. Because our model extends the standard model by placing a non-zero weight on attention utility, 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. Since, in turn, payoffs are endogenously determined by attention, accounting for attention utility makes the DM's objective more likely to be convex, naturally generating "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 dimensions with low payoffs, e.g., they ignore their investment portfolio when the market is down (Karlsson et al., 2009). Such behavior has also been noted in medical decision-making (e.g., Becker and Mainman (1975) and Oster et al. (2013)), 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 existing 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, which leads to optimism (Brunnermeier and Parker, 2005; Sharot, 2011) and a preference for positively-skewed lotteries. For instance, our DM is willing to buy a lottery ticket (or invest in a risky, potentially lucrative asset) because they can devote attention to the state where they win the jackpot (as is 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 the interaction between attention's instrumental and emotional roles. 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 attentional 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.[2] We 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 because future selves allocate attention differently to what the current self would prefer. More starkly, we show that if the DM has two tasks, one of which is non-trivial and requires multiple periods of attention to complete, the equilibrium can lead to an outcome where both selves are worse off relative to the commitment allocation. In Section 4.2, we turn to understanding how standard incentive policies designed to increase attention to tasks (which we model as a consumption dimension) that feature risky payoffs can backfire. Although our DM will increase attention to the task if they receive an extra bonus when the outcome is a success, increasing penalties conditional on a failure can have the opposite effect as it reduces the expected payoff of the task. This asymmetric response implies 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 default effects. Defaults are ubiquitous—workers set up default monthly contributions to their savings, and today's investment portfolio is the default for tomorrow's modulo investment returns—and typically, changing the default action requires some attention.

[2]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.

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 should matter, whereas they should matter less for planning a vacation. Second, because of our DM 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: first, they rely too often on a default, and second, the default is tailored to higher-payoff states.

Section 5 discusses extensions and situates our paper in the broader literature. Section 5.1 highlights potential limitations of our model, including the extent to which attention allocation is a choice variable, functional form assumptions, and the focus on expected payoffs, rather than changes in payoffs, as determining attention utility. Section 5.2 formally addresses one major concern: that our model takes the dimensions as given. Although it may be clear in some settings what a dimension is, in others, we may imagine that the DM has flexibility about which dimensions to link together. For example, the DM may be able to think of their weekend chores and Saturday evening plans as separate dimensions or jointly. We close the model by showing how, given a large set of primitive dimensions, an agent would choose to optimally bracket them. Section 5.3 highlights how one can distinguish our model from other approaches that share similar foundations, including rational inattention, anticipatory utility, 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. The ensuing model formally captures the two fundamental features of attention:

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

2. Emotional role: Attention generates attention utility.

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

2.1 Setup

The DM faces a finite number of dimensions indexed by $i \in \mathcal{D}$.[3] A dimension can correspond to a dimension of consumption, a realization of an unknown state, 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 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 $\sum_i \alpha_i = 1$.[4]

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 $X(\alpha)$, where $X(\cdot)$ is compact- and non-empty-valued and upper hemicontinuous, ensuring an optimum will exist. This reduced-form formulation nests canonical settings with (cognitively) costly information acquisition (as, e.g., in the rational inattention literature; see Maćkowiak et al. (2023); Sims (2003)), attention reducing "trembles," and recall of memories (Examples 1–3 in Appendix A.1).

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.,

[3] We take the dimensions as given. In practice, the boundaries between dimensions may not always be obvious. In Section 5.2, 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.

[4] Alternatively, one can impose an upper bound on the measure of attention (with no lower bound). By adding a trivial dimension with payoff 0 to $\mathcal{D}$, our model becomes equivalent to one with an upper bound.

2023).

The relative importance of attention utility to the usual standard payoffs, which we say generate 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 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 $\psi_i = 0$.[5]

Of course, our model embeds several assumptions from those governing how the interaction between attention and payoffs generates attention utility to the fact that individuals can solve the optimization problem of Equation (1) without devoting attention to assuming attention is fully controllable the DM. We discuss these assumptions and how to relax them in Section 5.1.

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, although we assume optimal allocation of attention, even if we did not, our model differs from the standard one, with our DM choosing actions that increase payoffs in dimension with high attention.[6]

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

[6]If attention is not optimally allocated, our model similar to "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. We consider the dependence on the payoff in a dimension, V_i, the relative weight on attention utility, λ, and parameters ω_i and ψ_i. These results describe the core mechanisms that lead to the different behavioral phenomena and biases, as we discuss later and preview here.

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

We begin with varying payoff V_i. For each i, we fix some v_i and define $V_i := \beta_i v_i + \gamma_i$, for scalars $\beta_i \ge 0$ 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 (i.e., $\lambda = 0$) DM's 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.

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

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

- *If $\gamma_i' \geq \gamma_i$ and $\beta_i = \beta_i'$, then $\alpha_i' \geq \alpha_i$. If, in addition, the environment is separable, then $v_i(x') \geq v_i(x)$.*
- *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 $\alpha_i' \geq \alpha_i$.*[7]

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

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

- $\sum_i \omega_i V_i(x) \geq \sum_i \omega_i V_i(x')$;
- *the DM's value is convex in $(\gamma_i)_{i \in \mathcal{D}}$;*

[7]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.

- *if the environment is separable, $\omega_i = 1$ and $\psi_i = 0$ for all $i \in \mathcal{D}$, and the objective given λ is convex in α, then it is also convex in α given λ'.*

The first part implies that the DM's actions are suboptimal if judged through the lens of the standard model with $\lambda = 0$.[8] The second and third parts of the proposition imply a preference for "extreme" payoffs and attention allocations. 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 is 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. Thus, the following proposition follows straightforwardly, and a formal proof is omitted.

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

For example, ω_i may represent the probability with which dimension i, now a state, realizes. In that case, the comparative statics is again entirely standard: The DM chooses an action that increases the payoff in a state if that state becomes more likely. Less standard is that the DM also chooses such action if there is some exogenous attention ψ_i on that state.

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

[8] They may also be judged as suboptimal through the lens of our model if we take attention utility as not normative but rather a bias that distorts decision-making.

2.3 Attention across consumption dimensions

We consider attention allocation across different dimensions of consumption. Those may be 'arranging a retirement home for a relative,' 'vacation,' '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.[9]

Evidence for such behavior has been extensively documented in the domains of finance and health. For instance, retail investors' propensity to check their portfolios generally comoves with the market (Karlsson et al., 2009; Sicherman et al., 2015).[10] To the extent that accessing one's portfolio requires attention to it and that portfolio levels are correlated with future consumption, our model suggests investors' (in)attention to their portfolios as a strategy to manage attention utility.[11] Similar behavior has been documented in the domain of health, where Shouldson and Young (2011); Oster et al. (2013) and Ganguly and Tasoff (2017) document the avoidance of testing for diseases, and the latter find that the interest in testing decreases with the seriousness of the disease. Our model predicts that an individual at risk of such a disease may have a low (expected) payoff related to the consumption dimension 'health' and hence avoid any actions, such as taking a test, that require attention.

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

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

[10]Gherzi et al. (2014) finds the opposite: Individuals increase monitoring in downturns.

[11]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 consumption payoffs relative to some reference point.

et al., 2009).

These alternative stories are typically about informational preferences; however, recent lines of research suggest that the ostrich phenomenon is not just about information but rather the direct utility derived from attention. For instance, Avoyan and Schotter (2020) provide evidence in a stylized laboratory environment where experimental participants choose to allocate time (i.e., attention) between two games (i.e., consumption dimensions). In line with our model, they find that "as payoffs in a given game increase, subjects plan more attention to the game." Variation across treatments to non-emotional benefits or costs of attention are ruled out by design, and belief-based utility models also seem implausible.[12]

There is also strong evidence of an important role of attention utility in the context of retail investors. Sicherman et al. (2015) find a positive correlation between market returns and the frequency of investors logging in to their portfolio twice during a single weekend—when markets are closed, and no new information can be revealed. Quispe-Torreblanca et al. (2020) find that investors devote excessive attention to positive information that is already known and are willing to provide more feedback about their portfolio (via a survey) when the portfolio is doing well. 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 domain of health, individuals also often fail to follow medical recommendations regarding non-information-generating activities, such as taking medicine (Becker and Mainman, 1975; Sherbourne et al., 1992; Lindberg and Wellisch, 2001; DiMattero et al., 2007). For instance, DiMattero et al. (2007) find that, among individuals experiencing serious medical conditions, individuals with worse health status tend to adhere less to medical regimes. 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

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

behaviors that align with existing theories but also rationalizes other behaviors that current explanations cannot explain.[13]

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.[14]

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).[15] Thus, in the context of individuals devoting attention across future contingencies, agents will know what to do with a financial windfall (as they have contemplated such contingency) but not which expenses to cut when they are laid off (as this has been ignored).

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

[14]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. If we want to extend the model, one could consider an alternative specification, e.g., one where the amount of attention utility a given state generates depends not only on the attention devoted to the state, α_i, but also the probability of that state, p_i: E.g., $\sum_i p_i V_i(x) + \lambda \sum g(p_i, \alpha_i) V_i(x)$, where g is increasing in both arguments. $g(p_i, \alpha_i) = \alpha_i$ returns our standard model. This model generates similar results, both in this section and with regard to the basic propositions described in Section 2.2.

[15]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.

We next 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.

The following proposition states that emotional inattention both reduces risk aversion and generates a preference for lotteries with a high payoff. In other words, attention utility drives a particular 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.*

- *DM(λ) is more risk-averse than DM(λ') for any $\lambda' > \lambda$.*

Let $H(x) \coloneqq \max_i x_i$. 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;*
- *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.,Sharot (2011); Mijović-Prelec and Prelec (2010); Mayraz (2011); Engelmann et al. (2019); Orhun et al. (2021); Oster et al. (2013) for evidence in both monetary and non-monetary domains).

The second part of the proposition states that the DM has a preference for lotteries with a high highest payoff. For instance, when comparing two binary lotteries with the same mean and low payoff, the DM prefers the more positively skewed one.[16]

[16]The skew of a lottery is defined as its third standardized moment. Fixing a low outcome and a mean for a set of binary lotteries, comparing the skewness of two lotteries is equivalent to comparing their high payoffs.

Intuitively, a particularly high payoff state always increases attention utility since the DM devotes their attention exclusively to such states. If the DM puts enough weight on attention utility, the DM then prefers the lottery with the higher high payoff.[17] The third case 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.

Thus, our results can rationalize individuals who are simultaneously willing to gamble (e.g., buying low probability but high payoff lottery tickets) as well as purchase insurance to prevent low probability, but high loss outcomes. In fact, there is extensive evidence supporting such a preference for positively skewed payoffs, coming from portfolio choice (Blume and Friend, 1975), real-world gambling markets (Golec and Tamarkin, 1998; Garrett and Sobel, 1999; Forrest et al., 2002) as well as in laboratory settings (Ebert and Wiesen, 2011; Grossman and Eckel, 2015; Ebert, 2015; Åstebro et al., 2015; Dertwinkel-Kalt and Köster, 2020). Furthermore, consistent with our model, Jullien and Salanié (2000) and Snowberg and Wolfers (2010) suggest that the preference for skewness is driven by subjective probabilities, as in our model, rather than the Bernoulli utility u.

As the previous proposition suggests, our DM acts as if they assign subjective weights to states of the world that differ from the objective probabilities. We now study more generally how our model leads to as-if probability weighting. In particular, take the DM's objective and divide by $1 + \lambda$ (which does not change behavior) and denote the terms in front of V_i as $q_i(p_i) := \frac{p_i+\lambda\alpha_i}{1+\lambda}$. 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 choosing action x who assigns probability q_i to state i. And 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 their true probability are overweighted (with the opposite also being true).

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

[17]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.")

the higher-payoff state, which is subsequently always overweighted. To generate more general forms of probability weighting, we allow for attention to be instrumentally valuable. We illustrate this using separable environments.

By way of example, suppose that the payoffs associated with each dimension feature initially large but decreasing returns to attention. Devoting at least a little attention to each state is valuable, but conditional on that, the residual attention is devoted to the high payoff state. Then, the probability weighting function is compressed; that is, small probabilities are over- and large probabilities are under-weighted, and probability weighting takes an inverse S shape. The following proposition summarizes.

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*[18]

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

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.

Probability weighting has been extensively studied since first discussed in Kahneman and Tversky (1979). Our model suggests attention as a microfoundation for such weighting, and, in particular, excess attention to low-probability states as leading to the classic finding of an inverse S-shaped weighting function (Wu and Gonzalez, 1996).

We conclude this subsection highlighting some differences between our model and other theories, leading to testable implications. The most widely used other ap-

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

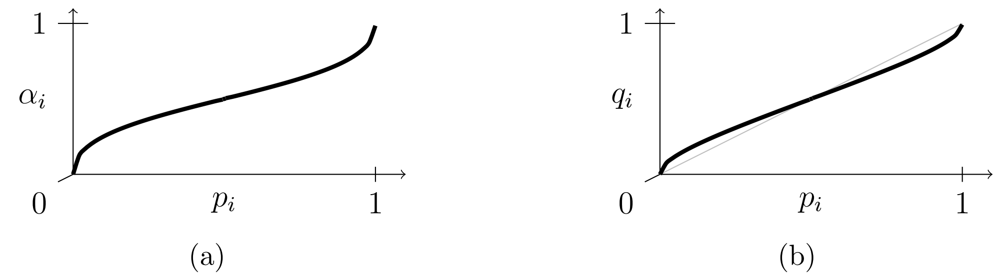

Figure 1: This figure visualizes Proposition 5. We have $\lambda = 1$ and $\hat{V}(a) = -\frac{1}{a}$. Consequently, $\alpha_i^* = (p_i - \sqrt{p_i(1-p_i)})/(2p_i - 1)$ which is inverse S-shaped (Panel (a)) and hence so is q_i (Panel (b)).

proaches to probability weighting—cumulative prospect theory (Tverseky and Kahneman, 1992) and rank dependent utility (Quiggin, 1982)—predict that the probability assigned to a state (q_i) depends only on the ranking of the states and the objective probabilities. In fact, with binary outcomes, most well known models of non-expected utility, such as Gul (1991) and Kőszegi and Rabin (2007), are equivalent to a rank-dependent approach (although this ceases to hold for lotteries with more than two outcomes in the support). Unlike these models, our model can explain why subjective weights may vary with payoff differentials, not just payoff ranks.

Of course, models of motivated beliefs, including those meant to explain overconfidence and optimism (e.g., Bénabou and Tirole (2002); Brunnermeier and Parker (2005); Bracha and Brown (2012); Caplin and Leahy (2019)), where individuals optimally choose their beliefs, can, like us, capture that belief distortions vary with payoff differentials. However, our model, fixing a DM, can generate other patterns of probability weighting relative to either of these alternatives, as the subjective weights also depend on the instrumental value of attention. For instance, if each state requires a fixed investment of $\underline{\alpha} < 0.5$ units of attention in order to guarantee a non-$-\infty$ payoff, then we get a linear weighting function that overweights low probabilities (and underweights high). And if $\hat{V}_i = \hat{V}_{i'} = \hat{V}$, where $\hat{V}$ is strictly convex, then the DM optimally devotes attention to the more likely state, leading them to overweight high and underweight low probabilities, i.e., we get 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 $x \coloneqq (x_t)_{t=1}^T$, the payoff in period t is $V_t(x)$ (one can impose natural restrictions about 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 $V_{t'}(x)$.

In each period, the DM receives a 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[19]

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

[19]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.

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 may be 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. 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.[20] Thus, propositions 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,

[20] Formally, let $\mathcal{H}_t := (x_{t'}, \alpha_{t'})_{t'=1}^{t-1}$ denote the (action, attention)-pairs the DM chose up to (and excluding) period t. Let $\Gamma_t(\mathcal{H}_t)$ denote the set of credible (x, α) when the DM has chosen $\mathcal{H}_t$ so far and now chooses (x_t, α_t), where credibility requires that the DM in each future period chooses their corresponding (action, attention)-pair optimally. 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 $X(\alpha) := (X_t(\alpha_t))_{t=1}^T$.

and consumption in period t is valued by $V(\sum_{t'=1}^{t} x_{t'\to t})$. The concavity of V implies that in this standard problem, the DM would smooth consumption and consume all income in the period they receive it.

Consider now our DM in this environment. 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 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. In particular, this assumption guarantees that the benefit of allocating attention to a state that currently has no attention is not too high.[21]

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

- *if $\lambda < \underline{\lambda}$, in each period t, $\alpha_{t\to t} = 1$; and*
- *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. When attention additionally determines the weights on the flow utilities, for instance, if the weight on period t' in

[21]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.

period-t's objective is $1+\tilde{\lambda}\alpha_{t\to t'}$ instead of 1 (see footnote 19), then $\delta_{t\to t'} = 1/(1+\tilde{\lambda})$ and so the DM falls in the class of quasi-hyperbolic discounters (Laibson, 1997).

When λ is large, the DM allocates all their income and attention to a subset of periods that feature particularly 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 period set. Our model thus rationalizes non-smooth consumption paths, e.g., weddings, vacations, and other lavish celebrations.[22] 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). For instance, Gilboa et al. (2016) write: "One may enjoy fond memories of a vacation, wedding, or special night out long after they have occurred." The key innovation in our model is that the enjoyment 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.

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 to attention across time.

[22] 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.

4 Applications

Our previous results show how emotional inattention can impact behavior in three classic domains of decision-making: multidimensional consumption, risk, and time. Of course, many important economic environments involve at least two of these simultaneously. Thus, in this section, 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 choice and 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 dimensions and states of the world, or instead, would they focus on consumption dimensions and aggregate across states? (Or vice versa?) Thus, this section serves as a proof of concept showing that emotional inattention is portable to more complex settings and, under reasonable assumptions, delivers testable predictions; at the same time, it also highlights the need to think carefully about the ways in which a DM can direct attention. In Section 5.2, we provide a formal model that endogenizes the set of dimensions.

4.1 Dynamic inconsistency and attentional scarcity

The DM may be dynamically inconsistent because future selves do not value past selves' attention utility. In particular, future actions can affect a payoff a past self devotes attention to, and so it affects that past self's overall utility without this effect being internalized. To formalize the dynamic inconsistency, we next consider a simple two-period model, where, in each period, the DM devotes attention to either a non-trivial or a trivial consumption dimension, with payoffs in period 2, denoted by c (for consumption) and o (for outside option), respectively. We assume that if both selves devote sufficient attention to the non-trivial payoff, the payoff becomes higher than the outside option, and so there is no inconsistency issue. However, we then show that constraints on the DM's attentional budget, i.e., "attentional scarcity," can lead to Pareto-dominated outcomes, where the DM prefers the solution under commitment in each period to the equilibrium outcome.

To further simplify the model (and notation), we also assume that we can write the non-trivial consumption payoff as a function of the attention allocated to it, i.e., $\hat{V}_c(\alpha_{1\to c}, \alpha_{2\to c})$, where $\alpha_{t\to i}$ denotes the attention in period $t = 1, 2$ devoted to consumption dimension $i \in \{c, o\}$.

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 dimension, and $\alpha_{t\to c} + \alpha_{t\to o} = 1$ 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 material utility in period 2 and attention utility in period 2 (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 anticipatory utility. Thus, if period-1 self devotes at least some attention to the non-trivial payoff, 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:*

- *A commitment solution and a no-commitment solution exist.*
- *Fix any α_1. The optimal $\alpha_{2\to c}$ chosen by period-2 self is less than that period-1 self would choose.*

Although the DM (in period 1) is generally hurt due to their inability to commit, period-2 self may be better off without commitment. It is thus natural to ask when the lack of commitment reduces the overall utility in both periods; we provide a partial answer related to "attentional scarcity": 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.

Proposition 8. *Suppose $\lambda > 0$, $\hat{V}_c(0,0) < \hat{V}_o < \hat{V}_c(1,0)$, and $\hat{V}_c$ is continuously differentiable with positive derivatives bounded away from zero everywhere.*

As long as period-2 attention is not too instrumental, 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 $(\alpha_{1\to c}, \alpha_{2\to c}) \leq (\bar{\alpha}_{1\to c}, \bar{\alpha}_{2\to c})$, we have:[23]

- *The unique no-commitment solution is $(\alpha_{1\to c}, \alpha_{2\to c}) = (0,0)$.*
- *With commitment, the DM devotes attention to the non-trivial consumption dimension in both periods.*
- *The DM strictly prefers commitment solutions to the no-commitment solution in each period.*

Moreover, if there are no constraints on the DM's attention allocation, then:

- *The unique commitment and no-commitment solution is $(\alpha_{1\to c}, \alpha_{2\to c}) = (1,1)$.*

The proof is constructive, and we briefly sketch it out here. For large λ, devoting no attention in either period is a local maximum for period-1 self; this is because increasing attention benefits instrumental utility and hurts attention utility, and the latter is weighted by more. Consider points $(\alpha_{1\to c}, \alpha_{2\to c})$ such that period-1 self is indifferent between those attention allocations and devoting no attention. Take one such point where period-1 attention to the non-trivial payoff is large, and period-2 attention is small. At this point, the benefit of increasing the non-trivial payoff benefits period-1 self more relative to period-2 self, as they devote a lot of attention. As a result, we can show that period-1 self wants to increase period-2 self's attention to the non-trivial payoff, whereas period-2 self may only want to do so if the instrumental value of attention is large enough ((4) provides the condition). In other words, period-2 self may decrease their attention, and period-1 self, anticipating this, chooses no attention in either period as the equilibrium outcome when there is no commitment. With commitment, period-1 self then forces period-2 self to devote the desired amount

[23] A formal condition on period-2 attention's instrumental value is this: Let $\tilde{\alpha}_{1\to c}$ be defined by $\hat{V}_c(\tilde{\alpha}_{1\to c}, 0) = \hat{V}_o$; period-2 attention is sufficiently non-instrumental if

$$\frac{\partial}{\partial \alpha_{2\to c}} \hat{V}_c(\tilde{\alpha}_{1\to c}, 0) < \frac{\hat{V}_o - \hat{V}_c(0,0)}{\tilde{\alpha}_{1\to c}}. \tag{4}$$

of attention, and both selves devote attention. The attentional bounds ensure that no attentional point with large attention levels, e.g., those that make the non-trivial payoff higher than the trivial one, is possible; and without the attentional bounds and since $\hat{V}_o < \hat{V}_c(1,0)$, there is no misalignment of preferences, and so the DM simply devotes all attention to the non-trivial consumption problem.

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 a 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 both multiple consumption dimensions—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. We can thus consider an environment with only consumption dimensions and with $\hat{V}_w(\alpha_w) = p(\alpha_w)v_H + (1 - p(\alpha_w))v_L$, where 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, where p is increasing and continuously differentiable.[24] 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.,

[24] Our 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 on the other. In Section 5.2, we formalize how the DM might optimally bracket different dimensions as one or separate.

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.*

- *Increasing v_H, v_L by the same amount increases α_w.*
- *Increasing v_H increases α_w.*
- *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 a standard DM (with $\lambda = 0$) would not adjust their attention.[25]

The second and third parts note an asymmetric response to incentives. Although DMs increase their effort in response to a carrot (increasing v_H), a stick (decreasing v_L) decreases the expected payoff and can thus lead to lower attention and, perversely, worse actions. 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). The stick lowers the payoff, and in order to avoid thinking about the reduced expected payoff, the agent allocates attention away from work. The proposition thus suggests the DM may avoid contracts with particularly bad downsides and commitment contracts involving penalties (while those with rewards may be too expensive).

4.3 Default effects

We now explore how emotional inattention induces particular default effects and impacts the ex-ante default choice. We consider an environment that features multiple

[25] 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 to work tasks with a higher level of success, fixing the return to effort.

time periods, consumption dimensions, 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 dimensions, $\mathcal{D} = \{c, o\}$. One of them, c, involves the choice of a default action. Its payoff is parameterized by some state $s \in \mathcal{S}$, e.g., capturing an income shock, where $\mathcal{S}$ is finite, which is revealed to the DM at the beginning of the second period. The DM's default action x_1 is chosen in the first period. The other consumption dimension, o (for attentional outside option), is trivial, i.e., its payoff V_o is constant. For simplicity, the DM never devotes any attention to period 1, and all attention is directed at period-2 payoffs, so we drop period-1 payoffs from the DM's consideration.

If the DM does not devote sufficient attention to the non-trivial dimension, then the default binds; we capture this by assuming that $X_2(\alpha_2) = \{\tilde{x}_2\}$ (i.e., it is a constant singleton) if $\alpha_{2\to c} < \eta$. 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 the fact that x_1 serves as a default: unless $x_2 = \tilde{x}_2$, i.e., "the default binds," $v_c(x_1, x_2|s)$ is independent of x_1. 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 $u(\cdot|s)$ is not constant for all s. 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

$$\underbrace{\overbrace{v_c(x_1, x_2|s) + \beta u(x_1|s)}^{V_c(x_1,x_2|s)} + V_o}_{\text{material utility}} + \lambda\big(\underbrace{\alpha_{2\to c}\big(\overbrace{v_c(x_1, x_2|s) + \beta u(x_1|s)}^{V_c(x_1,x_2|s)}\big) + \alpha_{2\to o} V_o}_{\text{attention utility}}\big). \tag{5}$$

We let $U_2(x_1, s)$ denote the maximized value of (5) 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 dimensions 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 dimension in state s, and $\alpha_{1\to o}$ that to the trivial dimension. 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 when 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:*

- *For any x_1, $\underset{\sim}{\mathcal{S}}(x_1) = \{s : \max_{(x_2,\alpha_2), x_2\in X_2(\alpha_2)} V_c(x_1, x_2|s) < V_o\}$.*
- *If $\beta = 0$, then the optimal default action x_1 satisfies*

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

- *If $\beta > 0$, then the optimal default action x_1 satisfies*

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

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 notes a default effect: the DM fails to readjust their action in some states of the world, even though it is instrumentally costless. A key difference relative to other models with costly adjustment (e.g., rational inattention or fixed cost of action) is that in our model, the default binds asymmetrically: The set of states where the default binds for large λ are those states s with a low payoff. For instance, individuals may adjust their planned consumption only in response to positive income shocks, ignoring negative ones. Similarly, individuals may not want to revisit previously negotiated business or wage contracts in response to some bad events occurring.

The second and third parts of the proposition provide insight into how the default is chosen. In the second part, the default is "pure," i.e., it does not matter in states in which the DM changes it. In this case, the DM chooses the default in order to maximize the expected payoff conditional on those states in which the default binds. For instance, consumers may rationally anticipate their future inattention in periods of low payoffs, and in anticipation, they plan their default consumption level to be low to insure against their inattention.

Part three considers the case where the choice of the default directly impacts payoffs in period 2, even when a different action is subsequently taken. A DM with a high λ now wants to take a default action in period 1 that generates high attention utility in both periods. Thus, they choose 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

5.1.1 Top-down vs bottom up-attention

A key tenet of our model is that attention is voluntarily directed by the individual, 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)). Although in this paper we focused on situations where all attention can freely be allocated, our model extends to situations where some attention is involuntarily allocated (captured via ϕ_i), and the residual is optimally allocated.

5.1.2 Functional form of attention utility

Our approach to modeling attention utility and how it is generated abstracts away from many details. In reality, the generation of attention utility is likely a variegated process: attention to past consumption may generate memory utility, attention to future consumption may generate anticipatory utility and attention to contemporaneous consumption may enhance consumption utility. In our static model, we do not differentiate between these interpretations, while in practice, the type of cognitive process (remembering, anticipating, or focusing) may matter. Extending our setting, one may partially account for this by analyzing a model with different λ parameters for each cognitive process.

Furthermore, attention utility likely does not have the simple functional form we assume, where it is the product of a dimension's payoff and attention. A more general form for attention utility is $\sum_i F(\alpha_i + \psi_i, V_i(x))$, with F having increasing differences in $(\alpha_i + \psi_i, V_i(x))$ to capture that attention to higher-payoff dimensions leads to higher attention utility. Propositions 1–3, characterizing the optimal attention allocation, continue to hold except for Proposition 2 part two. Additionally, this form can capture decreasing returns from devoting attention, e.g., the marginal excitement generated from thinking about a future vacation decreases in the amount of attention devoted, by assuming that F is concave in the first argument.[26]

5.1.3 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). If so, then the DM should intensely devote attention for a brief amount of time to maximize material utility and devote attention during the remaining time to high-payoff dimensions to maximize attention utility.[27] While plausible in some situations,

[26]As an example, inverse-S-shaped probability weighting can be generated not from instrumental returns to attention, as is demonstrated in Proposition 5, but rather from attention utility being concave in attention. Specifically, suppose that the attention utility form dimension i is given by $f(\alpha_i)\hat{V}_i$ ($\hat{V}_i(\alpha_i)$ is constant), where f is concave and continuously differentiable. If $\frac{\partial}{\partial a}F(1) = 0$, and $\hat{V}_i \geq \hat{V}_{i'} > 0$, then the DM devotes more attention to state i than i', but not all.

[27]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.

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. Similarly, a student can easily (and quickly) decide to study hard for an important test, but actually doing so requires ongoing focus. It may also be the case that once attention has been devoted to a dimension, it is psychologically infeasible to then ignore it. For example, after making a portfolio choice, the investor cannot help but think about the portfolio for the rest of the day.

5.1.4 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, probability weighting, and non-smooth consumption.

5.1.5 What requires attention?

For tractability, we assume that the DM understands the material consequences of attention, captured by the set of available actions and parameter values, perfectly and that they can find the maximizers of the respective optimization problems (e.g., 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 even the act of optimization itself—during which the DM arguably devotes attention to different payoffs—generates attention utility. One can embed higher-level learning by the DM about the parameters of the model and consider a dynamic model attention allocation with learning, where the DM may stop devoting attention upon learning that further attention has only limited instrumental value. We believe our qualitative results would also hold in this more complicated setting.[28]

5.2 Closing the model: optimal bracketing

Our model requires carefully specifying the environment: A key component is partitioning the environment into sets 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 many 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

$$\underbrace{\sum_i V_i(x)}_{\text{material utility}} + \lambda \underbrace{\sum_i \alpha_i \bar{V}_B(i)(x)}_{\text{attention utility}}, \tag{6}$$

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

Note that the model in Section 2 is recovered when B consists of singleton sets and that a DM who uses one bracket, i.e., B is a singleton, is equivalent to the standard DM with $\lambda = 0$.

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

[28]The question arises how the DM knows how to allocate attention to learn how to allocate attention, and so forth. This recursion problem (see Lipman (1991) for a discussion about infinite regress issues in economic models) arises not just in our model but in other models of optimal attention.

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

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

5.3 Relation to existing models

This section compares our model to related approaches, focusing on conceptual similarities and differences. 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.

5.3.1 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 individual allocate attention in order to gain information, subject to a cost (often mental) that depends on the amount 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 the extensive margin of much attention to acquire for a particular problem, we focus on the intensive margin of 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.

Although, like our model, rational inattention can explain why individuals may deliberately choose to not fully acquire information, emotional inattention makes several novel predictions. These demonstrate how considering both the emotional and the mental costs of attention can help improve the explanatory power of economic models. Our model predicts that the "cost of attention" falls for a particular dimension if the relative payoff of that dimension increases. Thus, in contrast the the predictions of rational inattention, that 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)).[29] In the domain or risk, rational inattention fails to predict violations of expected utility, while in dynamic rationally inattentive agents will still be dynamically consistent, unlike emotionally inattentive agents.

5.3.2 Anticipatory utility

Our model can be seen as extending models of anticipatory utility, where agents gain flow from their (rational) beliefs about future outcomes (see Loewenstein (1987); Loewenstein and Elster (1992) for early contributions, and recent efforts of Caplin and Leahy (2001); Kőszegi (2010); Dillenberger and Raymond (2020)). In contrast to that literature, we suppose that the flow anticipatory utility is mediated via the allocation of attention. 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)).

[29] 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.

5.3.3 Recursive preferences

A distinct approach to capturing both non-standard attitudes towards 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 as well as 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 can not easily account for memorable consumption events.

5.3.4 Chosen preferences and beliefs

There is a small literature modeling agents who can optimally choose (Bernheim et al., 2021) or adapt their future preferences (Elster, 1983).[30] Similarly, our agent 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) might 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

[30] Related are models where DMs alter their beliefs in order to maximize anticipatory utility (e.g., Bénabou and Tirole (2002); Brunnermeier and Parker (2005); Bracha and Brown (2012); Caplin and Leahy (2019)); we discuss these in more detail in Section 2.4, where we compare their predictions to ours.

attention's instrumental value is high.

5.3.5 Other models of attention

We know of two other papers in economics that simultaneously model the instrumental and emotional roles of attention. 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 to a dimension 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.[31] 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. Devoting attention to some initial news and acquiring further information has two effects: it increases the impact of gain-loss utility, and it speeds up a reference point adjustment. Under some conditions, the DM acquires additional information 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.

[31] Their main application is on how information provision (as requested by the DM or forced by an advertiser) can increase consumption, even when the DM learns their marginal payoff is less than what they expected (this follows from the increase in attention and hence the importance in the DM's objective; this intuition can be expressed in our framework as we show in Example 6 in Appendix A.3).

6 Conclusion

This paper has presented a model of attention allocation 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 (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. 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.

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 identified from the data. The details would vary by the environment, but here, we provide the intuition for a situation where the dimensions are states. We first can identify whether two states are considered jointly (they are in the same "bracket," Section 5.2) by reducing the payoff for one state and increasing the other the same amount and seeing whether the (action, attention)-pair changes. If we can find some shift such that it does, then the two states are not part of the same bracket. The choices over lotteries allow us then to identify the degree of overweighting of the high payoff state(s) and thus λ.

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

References

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

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

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

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

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

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

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

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

Bronchetti, Erin T., Judd B. Kessler, Ellen B. Magenheim, Dmitry Taubinsky, and Eric Zwick, "Is Attention Produced Optimally? Theory and Evidence From Experiments With Bandwidth Enhancements," *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.

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

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

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

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

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

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

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

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

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.

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

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

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

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

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

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

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

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

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

Gross, James J, "The Emerging Field of Emotion Regulation: An Integrative Review," *Review of General Psychology*, 1998, *2* (3), 271–299.

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

Gul, Faruk, "A theory of disappointment aversion," *Econometrica: Journal of the Econometric Society*, 1991, pp. 667–686.

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.

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.

James, William, *The Principles of Psychology*, Vol. 1, Henry Holt and Company, 1890.

Jullien, Bruno and Bernard Salanié, "Estimating Preferences under Risk: The Case of Racetrack Bettors," *Journal of Political Economy*, 2000, *108* (3), 503–530.

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

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

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

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

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

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

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

_ **and** _ , "Reference-Dependent Risk Attitudes," *American Economic Review*, September 2007, *97* (4), 1047–1073.

_ **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.

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

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

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

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.

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

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

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

Snowberg, Erik and Justin Wolfers, "Explaining the Favorite–Long Shot Bias: Is it Risk-Love or Misperceptions?," *Journal of Political Economy*, 2010, *118* (4), 723–746.

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

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

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

A Additional examples

A.1 Examples of canonical problems

In Examples 1–3, we consider a separable environment (see Section 2.2 for a definition). "Actions" and "payoffs" shall refer to those in the now explicitly modeled dimension i.

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

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

The DM's payoff from dimension i given α_i is

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

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

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

for some $\kappa \geq 0$ and where $H(B)$ denotes the entropy of belief B.[32]

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

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

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

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

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

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

for some $\kappa \geq 0$, where $H(B)$ denotes the entropy of belief B (see footnote 32 for the definition) and $\mathcal{U}$ the uniform distribution on A; i.e., if the DM devotes no attention, they will make each choice with equal probability.

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

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

Finally, let $\mathcal{F}$ be

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

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

A.2 Examples for Section 3

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

$$X_t(\alpha_t) = \begin{cases} \{\underline{x}\} & \text{if } < \eta_t \\ \{\underline{x}, x^*\} & \text{if } \alpha_{t\to3} \geq \eta_t, \end{cases}$$

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

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

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

$$(1 + \lambda(1 - \eta_2))\bar{V} + (1 + \lambda\eta_2)\bar{V}_3 \geq (1 + \lambda(1 - \underline{\alpha}_{2\to3}))\bar{V} + (1 + \lambda\underline{\alpha}_{2\to3})\underline{V}_3. \quad (7)$$

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

$$(1+\lambda(1+(1-\eta_2))\bar{V}+(1+\lambda\eta_2)\bar{V}_3 \geq (1+\lambda((1-\eta_1)+1)\bar{V}+(1+\lambda\eta_1)\bar{V}_3 \iff \eta_1 \geq \eta_2. \quad (8)$$

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

$$(1 + \lambda(1 - \eta_1))\bar{V} + (1 + \lambda(\eta_1 + \underline{\alpha}_{2\to3}))\bar{V}_3 \geq (1 + \lambda)\bar{V} + (1 + \lambda\underline{\alpha}_{2\to3})\underline{V}_3. \qquad (9)$$

Since $\underline{V}_3 < \bar{V}_3 < \bar{V}$, there exists $\lambda > 0$ such that (9) *holds with equality. For such λ, since $\underline{\alpha}_{2\to3} > 0$, there exists $\eta_2 < \eta_1$ (i.e.,* (8) *holds) so that* (7) *does not hold. Furthermore, λ can be slightly decreased so that* (9) *now holds strictly and* (7) *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 (7) *holds (and* (8) *and* (9) *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 proceeds almost identically.*

Since $\underline{V}_3 < \bar{V}_3 < \bar{V}$, there exists $\lambda > 0$ such that (9) *holds with equality. For such λ, since $\underline{\alpha}_{2\to3} > 0$, there exists $\eta_2 < \eta_1$ (i.e.,* (8) *holds) so that* (7) *does not hold. Furthermore, λ can be slightly decreased so that* (9) *now holds strictly and* (7) *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 (7) *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 dimension* c *corresponds to the DM purchasing a quantity of a consumption good at a unit price of* 1*. Their valuation of quantity* k *is* $\theta u(k)$*, where* u *is strictly concave and continuously differentiable, and* $\theta \in \{\theta_L, \theta_H\}$ *with* $P(\theta = \theta_H) = p \in (0, 1)$*. The DM has wealth* 1 *available, and whatever amount they do not consume,* $1 - k$*, leads to payoff* $1 - k$ *as part of dimension* m *(the "money" problem).*

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

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 produce 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 Attention utility in a strategic environment

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

Agent i acquiring an amount ξ^i of the risky assets leads to monetary payoff of $c^i_s = 1 - \xi^i + \xi^i \frac{1+x^r_s}{P}$ in state s. An equilibrium is a price P, an attention allocation α^i and amount ξ^i purchased of the risky asset for each agent i, such that each agent maximizes

$$\sum_s p_s u(c^i_s) + \lambda \sum_s \alpha^i_s u(c^i_s), \tag{10}$$

with respect to α^i and ξ^i, and $\int_i \xi^i = 0$.

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

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

Suppose $|\mathcal{D}| > 1$. Let $\bar{s} = \arg\max_{s\in\mathcal{D}} x^r_s$ and $\underline{s} = \arg\min_{s\in\mathcal{D}} x^r_s$. First note that $\xi^i \neq 0$ as thus it must be that $\alpha^i_{\bar{s}} = 1$ or $\alpha^i_{\underline{s}} = 1$. To see this, suppose $\xi^i = 0$. Given $\xi^i = 0$, both $\alpha^i_{\bar{s}} = 1$ and $\alpha^i_{\underline{s}} = 1$ are optimal, but $\xi^i = 0$ cannot be optimal for both $\alpha^i_{\bar{s}} = 1$ and $\alpha^i_{\underline{s}} = 1$, a contradiction. Conditional on $\alpha^i_{\bar{s}} = 1$ ($\alpha^i_{\underline{s}} = 1$), the payoff (10) is continuously decreasing (increasing) in P. Furthermore, for large enough P, the payoff given $\alpha^i_{\bar{s}} = 1$ is less than that given $\alpha^i_{\underline{s}} = 1$ with optimal ξ^i. Thus, there exists a unique P^* such that each agent i is indifferent between $\alpha^i_{\bar{s}} = 1$ and $\alpha^i_{\underline{s}} = 1$. Furthermore, at P^*, it must be that $\alpha^i_{\bar{s}} = 1$ implies that $\xi^i > 0$ and $\alpha^i_{\underline{s}} = 1$ implies that $\xi^i < 0$ (otherwise, the associated payoffs cannot be the same). But then $\int_i \xi^i = 0$ is easily achieved by assigning the appropriate masses of agents to either $\alpha^i_{\bar{s}} = 1$ or

$\alpha^i_{\underline{s}} = 1$ with the corresponding optimal ξ^i. The heterogeneity in attention implies the heterogeneity in subjective beliefs. $\square$

C Proofs

C.1 Proposition 1

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

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

- *If $\lambda > 0$: If $\gamma_i' > \gamma_i$ then $\min_{(x,\alpha)\in\Gamma(\gamma_i',\beta_i)} \alpha_i \geq \max_{(x,\alpha)\in\Gamma(\gamma_i,\beta_i)} \alpha_i$. If, in addition, the environment is separable, then $\min_{(x,\alpha)\in\Gamma(\gamma_i',\beta_i)} v_i(x) \geq \max_{(x,\alpha)\in\Gamma(\gamma_i,\beta_i)} v_i(x)$.*
- *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

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

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

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

The outer inequality implies

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

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

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

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

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

Then

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

and so the second claim follows.

Now suppose the environment is separable, $\omega_i = 1$ and $\psi_i = 0$ for all $i \in \mathcal{D}$, and that the objective given λ is convex in α. Take any $\chi \in [0,1]$ and α_i, α_i' with $\alpha_i < \alpha_i'$. Then

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

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

C.3 Proof of Proposition 4

Proof of Proposition 4. Take any 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 $H(x)$, with $H(x)$

defined in the proposition. Since $u(H(x)) \geq \sum_i u(x_i)$, 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,

$$\frac{(p_i + \lambda\alpha_i)\frac{\partial}{\partial a}\hat{V}(\alpha_i) - ((1-p_i) + \lambda(1-\alpha_i))\frac{\partial}{\partial a}\hat{V}(1-\alpha_i)}{1+\lambda} + \frac{\lambda(\hat{V}(\alpha_i) - \hat{V}(1-\alpha_i))}{1+\lambda}. \tag{11}$$

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 (11) is strictly positive at $\alpha_i = 0$. Now, consider the limit of (11) 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 (11) are finite. Thus, there exists a $\bar{\alpha}_i$ with $0 < \bar{\alpha}_i \leq \tilde{p}_i$ such that for all $\alpha_i \leq \bar{\alpha}_i$, (11) evaluated at α_i (given $\tilde{p}_i$) is strictly positive. Moreover, notice that (11) is decreasing in p_i (provided $p_i < 1/2$) for all $\alpha_i \in [0, 1/2]$. Thus, let $\bar{p} = \bar{\alpha}_i$. Then for all p_i with $0 < p_i < \bar{p}$, (11) is strictly positive for all $\alpha_i \in [0, \bar{p}]$, implying 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 associated 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, 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

$$\sum_{t=1}^{T}(1+\lambda\alpha_{\to t})V(\alpha_{\to t}), \tag{12}$$

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 $\lambda < \underline{\lambda}$, (12) is strictly concave. Note that if $(1+\lambda\alpha_{\to t})V(\alpha_{\to t})$ is strictly concave for $\alpha \in [0,T]$, then (12) is strictly concave in $(\alpha_{\to t})_{t=1}^{T}$.[33]

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 $\alpha_{\to t}$.

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

It is easy to check that $\alpha_{\to t} = 1$ for all t satisfies the Karush-Kuhn-Tucker conditions 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, $\alpha_{t\to t} = x_{t\to t} = 1$ is indeed the unique equilibrium outcome.

Next, consider the case of large λ. We return to the associated 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 (12). We show that i) the solutions to (12) converge to those 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 the first fact, note that, in the limit, each such solution to (12) is strictly better than any attention allocation that does not maximize $\sum_{t=1}^{T} \alpha_{\to t} V(\alpha_{\to t})$. In particular, it must be that the optimal allocation converges to one that achieves $TV(K)$ in attention utility. Of those attention allocations, the claimed set of solutions in the limit limit uniquely maximize the DM objective for any λ, as they maximize material utility, and the result follows.

For the second fact, 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

$$
\begin{aligned}
(1+\lambda K)V'(K) + \lambda V(K) - \kappa = 0 \\
V'(0) + \lambda V(0) - \kappa + \mu = 0.
\end{aligned}
$$

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 the third fact, note that each such allocation 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 λ.

Thus, for λ large enough, (12) 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. $\square$

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 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 solution 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.*

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. □

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 defined in Footnote 23,*

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

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

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.

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 as $\hat{V}_c$ 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, note that the second and third subclaims imply that $\underline{\alpha}_{1\to c}(\lambda)$ must converge as $\lambda \to \infty$. Suppose $\underline{\alpha}_{1\to c}(\lambda) \to x < \tilde{\alpha}_{1\to c}$. Since $\hat{V}_o > \hat{V}_c(x,0)$, as $x < \tilde{\alpha}_{1\to c}$, we have $V_1(x,0) < V_1(0,0)$ 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 $V_1(x,0) = V_1(0,0)$, i.e., we have a contradiction. $\square$

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. We next note that this implies 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 $V_2(\alpha_{1\to c} + x, \alpha_{2\to c} + y) > V_2(0,0)$ for all $(x,y) \in [-\delta, \delta]^2$, with $(\alpha_1 + x, \alpha_2 + y) \in [0,1]^2$.*

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 $\hat{V}_o > \hat{V}_c(\alpha_{1\to c}, \alpha_{2\to c})$.

□

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.,

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

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

$$
\begin{aligned}
\frac{\partial V_2}{\alpha_{2\to c}}(\underline{\alpha}_{1\to c}(\lambda), 0) &= \frac{\partial \hat{V}_c}{\partial \alpha_{2\to c}}(\underline{\alpha}_{1\to c}(\lambda), 0) - \lambda(\hat{V}_o - \hat{V}_c(\alpha_{1\to c}(\lambda), 0) \\
&= \frac{\partial \hat{V}_c}{\partial \alpha_{2\to c}}(\underline{\alpha}_{1\to c}(\lambda), 0) - \frac{\hat{V}(\underline{\alpha}_{1\to c}(\lambda), 0) - \hat{V}_c(0,0)}{\underline{\alpha}_{1\to c}(\lambda)}.
\end{aligned}
\tag{13}
$$

By Claim 2, $\underline{\alpha}_{1\to c}(\lambda) = \tilde{\alpha}_{1\to c}$ as $\lambda \to \infty$. Thus, using (4), it follows that (13) is negative, for large λ.

To recap, for large λ, we have constructed a point $(\underline{\alpha}_{1\to c}(\lambda), 0)$ such that period-1 self is indifferent between that point and devoting no attention in either period, period-2 self strictly prefers this point to devoting no attention, period-1 self wants period-2 self to increase attention to the non-trivial payoff, while period-2 self wants to decrease it.

We next construct a nearby point that both selves prefer to devoting no attention, but such that period-2 self still prefers to deviate and not devote 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 solution 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 $(\alpha_{1\to c}, \alpha_{2\to c}) = (1,1)$. □

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.[34] □

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 (5) gives $\alpha_{2\to c} = 1$ as the optimal attention allocation (for any λ), i.e., indeed $s \notin \underset{\sim}{\mathcal{S}}(x_1)$. 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, (5) is maximized by $\alpha_{2\to c} = 0$. Since $\mathcal{S}$ is finite, taking the max λ 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

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

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 (5) 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 $\underset{\sim}{\mathcal{S}}(x_1)$, since the choice of x_1 only affects payoffs in those states.

For the final claim, recall that for any x_1', $\underset{\sim}{\mathcal{S}}(x_1')$ 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, 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 $\underset{\sim}{\mathcal{S}}(x_1')$ is held fixed at $\underset{\sim}{\mathcal{S}}(x_1)$. Since $\underset{\sim}{\mathcal{S}}(x_1) \neq \mathcal{S}$, x_1 maximizing the above means that it also maximizes

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

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 $s \in \mathcal{S} \setminus \underset{\sim}{\mathcal{S}}(x_1)$ 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 $s \in \mathcal{S} \setminus \underset{\sim}{\mathcal{S}}(x_1)$ and V_o is also independent of x_1'. $\square$

C.10 Proof of Proposition 11

Proof of Proposition 11. Fix any (x, α) with $x \in X(\alpha)$. Take any $D, D' \in B$ and consider $B' := (B \cup \{D \cup D'\}) \setminus \{D, D'\}$. Evaluate (6) 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 $\bar{\alpha}_D \geq \bar{\alpha}_{D'}$. $\square$