Bounded Attention in Intertemporal Choice*

Jawwad Noor Norio Takeoka

March 15, 2016

Abstract

We behaviorally model an agent in an intertemporal context who exhibits diminishing impatience with respect to the size of consumption stakes. Our representation result describes the agent as a discounted utility maximizer, where the discount function is determined by a cognitive process that optimally uses a limited stock of cognitive resources to improve the agent's "telescopic faculties".

1 Introduction

When deciding if it is worthwhile to sacrifice immediate consumption for larger future consumption, we need to consider not just the magnitude of the future outcome, but also our willingness to wait for it. The standard Discounted Utility model of intertemporal choice views "willingness to wait" to be a fixed feature of the agent's preference, modelling it as a discount function. However, in this paper we study the possibility that our willingness to wait may have rich properties. In particular, we may be more willing to wait for larger outcomes. Our willingness to wait for a $10,000 outcome next week may be substantially different from that for a $3 outcome. In trying to

*INCOMPLETE DRAFT, PLEASE DO NOT CIRCULATE. Noor (the corresponding author) is a the Dept of Economics, Boston University, 270 Bay State Road, Boston MA 02215. Email: jnoor@bu.edu. Takeoka is at the Dept of Economics, Hitotsubashi University, Tokyo, Japan. Email: . The authors are grateful to Larry Epstein, Bart Lipman for very useful discussions, and to seminar participants at Boston University, Yokohama National University for helpful comments. The usual disclaimer applies.

understand why, we provide a theory where the difference comes from the fact that $10,000 invokes a rich mental picture, whereas $3 does not.

Taking a preference over the space of consumption streams as a primitive, we formalize the notion of "diminishing impatience with respect to scale" in an axiom that we call *Increasing Attention*. We prove a representation theorem for an agent who respects Increasing Attention along with some regularity conditions. The representation describes the agent as evaluating streams in terms of their discounted utility, but the discount function may change with the stream. The dependence is such that as any stream is scaled up, the discount function weakly increases, and moreover, it may be that a change in the outcome in one period can nontrivially impact how other periods are discounted.

The representation goes further, however, and provides a story for why the discount function may be so. Specifically, it is *as if* the discount function is the result of a *cognitive process* that operates as follows.

The cognitive process is motivated to maximize the discounted utility of the stream. The instrument available to it for doing so is *attention* (which leads to a higher discount function). Intuitively, the promise of future utility motivates the focus of attention, and the attention in turn enhances the current appreciation of it – in a sense, affect feeds cognition and cognition feeds affect. There is a pool of cognitive resources that can be used, at a cost, for producing attention.[1] These resources are *limited*, and thus there are bounds on the agent's attention. The cognitive process solves the attention allocation problem optimally.

While Pigou described discounting as arising due to "faulty telescopic faculties",[2] our model views the agent's telescopic faculties as something that can be sharpened by the mind in response to the promise of future utility.

In our interpretation of the model, the agent's focus on future outcomes is not necessarily conscious and deliberate, but rather it is the outcome of a non-conscious cognitive process. Indeed, our axioms do not carry with them any suggestion of deliberation on the part of the agent, and they are in fact consistent with some reduced form of the representation where the discount function just depends on the size of outcomes in particular ways in some unexplained way. We show nevertheless that this discount function can be

[1]See Kahneman (1973) for a model in cognitive psychology that views attention as a scarce resource that is allocated across alternatives being considered.

[2]Quoting Pigou (1920), "our telescopic faculty is defective, and we, therefore, see future pleasures, as it were, on a diminished scale".

viewed as the outcome of a cognitive optimization problem which disciplines the dependence of the discount function on the stream. The components of this cognitive optimization problem are uniquely pinned down by preferences.

The remainder of the paper proceeds as follows. We close the Introduction with a discussion of related literature. Section 2 presents our primitives and our axioms, while Section 3 presents our main representation theorem. Section 4 discusses implications for preferences over rewards, as opposed to consumption. Section 5 presents a tractable subclass. Section 6 concludes. All proofs are relegated to appendices.

Related Literature

The idea of diminishing impatience with respect to stake size first originated in the experimental time preference literature, where it was called a *magnitude effect*. For instance, Thaler (1981) reports that his subjects were indifferent between $3000 now and $4000 in a year, while also being indifferent between $15 now and $60 in a year, suggesting diminishing impatience with respect to stake size. Such conclusions in the literature have relied on assuming linear utility for money. For an experiment that does not rely on assumptions about utility, see Ericson and Noor (2015). For some theoretical work on the magnitude effect, see Loewenstein and Prelec (1992) and Noor (2009, 2011). In this paper we formulate an axiom that clearly captures diminishing impatience with respect to stake size, and study what it might imply about the cognitive processes underlying the agent's choices.

Becker and Mulligan (1997) hypothesize that an agent can change his discount function through deliberate physical investment in certain commodities (such as education). In our model, the agent changes his discount function through non-deliberative subjective investment of cognitive resources.

Brunnermier and Parker (2005) describe optimistic expectations as the outcome of some underlying process that wishes to increase the utility from anticipation through beliefs while anticipated real costs of consequent suboptimal future decisions discipline these beliefs. The model deals with a very specialized process, one that requires a special kind of awareness of optimism (the agent believes in his optimal beliefs, but also possesses different 'true' beliefs that discipline his optimism). Our model views the agent as simply maximizing his utility when it comes to choice, and the utility being determined by a separate non-conscious cognitive process.

In the decision theory literature, Ellis (2015) studies the behavioral foundations of models in the inattention literature where the agent subjectively chooses how much information to accumulate. Unlike our model, his model carries with it the presumption of a deliberative action (choice of information) on the part of the agent, and the optimality of the action expresses itself behaviorally. Moreover, while his model features a cost function φ where the cognitive constraint may be implicitly defined via the effective domain $\{D : \varphi(D) < \infty\}$, in our model the analysis of cognitive resource constraints is more nuanced and does not coincide with the effective domain of the cognitive cost function.

2 Axioms

The consumption space is given by a metric space C that is also a mixture space, and endowed with the metric topology. As we will see, this space cannot be assumed compact. The time horizon is $T + 1 < \infty$. Consider the space of consumption streams $X = C^{T+1}$, endowed with the product topology. The primitive of our model is a binary relation $\succsim$ over X.

2.1 Basic Axioms

The first two axioms are entirely standard.

Axiom 1 (Order) $\succsim$ *is complete and transitive.*

Axiom 2 (Continuity) *For all* $x \in X$, $\{y \in X : y \succsim x\}$ *and* $\{y \in X : x \succsim y\}$ *are closed.*

It is conceivable that an agent with limited cognitive resources may possibly exhibit intransitivities; for example the agent may exhibit indifference to barely noticeable improvements but non-indifference to noticeable improvements. However, with an eye on applications, our modelling choice is to write a model that can be described by a utility function.

Axiom 3 (Present Equivalents) *(i) There exists* $0 \in C$ *s.t.* $x \succsim (0, 0, .., 0)$ *for all* x.

(ii) For any x *there exists* $c \in C$ *s.t.* $(c, 0, ..., 0) \succsim x$..

The first part of the axiom implies that there is a "worst consumption", denoted (and interpreted as) 0. The second states that for any stream x, there is an immediate consumption level (with 0 in every subsequent period) that is better than x. Given other axioms, this ensures the existence of present equivalents, that is, for any stream x there exists c such that $(c, 0, ..., 0) \sim x$. Present equivalents will be needed to formulate our novel axioms in the next subsection.

One consequence of the Present Equivalents axiom is that C cannot be bounded. Imagine that C was bounded and there was a "best consumption" c. Then our Monotonicity axiom below would imply that $(c, 0, ..., 0) \prec (c, c, ..., c)$, that is, getting c today would not be better than getting c in every period. Indeed, no present equivalent exists for the latter stream. This is ruled out by Present Equivalents.

When the context is clear, we use c to denote both an element in C and the stream $(c, 0, .., 0)$. Denote by c^t the stream that pays c at time t and 0 in all other periods. The standard Impatience axiom can be stated as:

$$\text{If } c \succ 0 \text{ then } t < t' \Longrightarrow c^t \succ c^{t'}. \text{ If } c \sim 0 \text{ then } c^t \sim 0 \text{ for all } t.$$

This states that desirable consumption is always preferred sooner, and that the agent does not care when he receives "zero" consumption. We relax the first part so that consumption in the future is never more attractive than immediate consumption.

Axiom 4 (Weak Impatience) *(i)* $c \succsim c^t$ *for any* c *and* t.
(ii) If $c \sim 0$ *then* $c^t \sim 0$ *for all* t.

We impose this axiom despite the fact that we can prove results without it. "Anticipatory utility" has been modelled in the literature (Loewenstein 1987) as a violation of (the first part of) Weak Impatience. In the interpretation of our model, appreciation of future utility plays an important role, and we maintain Weak Impatience to show that behaviorally this not the same phenomenon as "anticipatory utility" as it is discussed in the literature.

Finally, we require that the agent's preferences towards risk is standard.

Axiom 5 (Risk Preference) *For any* $c, c', c'' \in C$, *and* $\alpha \in [0, 1]$,

$$c \succ c' \Longrightarrow \alpha c + (1 - \alpha)c'' \succ \alpha c' + (1 - \alpha)c''.$$

Since these are fairly standard axioms we will say that:

Definition 1 (Regularity) $\succsim$ *satisfies **Regularity** if it satisfies Order, Continuity, Present Equivalents, Weak Impatience and Risk Preference.*

2.2 Main Axioms

In this section we formulate key axioms that describe an agent who exhibits diminishing impatience with respect to stake size.

For any c and $\alpha \in [0,1]$ define the mixture $\alpha c := \alpha c + (1-\alpha)0$. For any stream $x = (x_0, ..., x_T)$ define

$$\alpha x := (\alpha x_0, ..., \alpha x_T).$$

Intuitively, the stream αx uniformly "scales down" the consumption offered by x in every period. Abusing notation, write αc for the stream $(\alpha c, 0, ...0)$. For any x define its present equivalent $c_x = (c, 0, ...0)$ by

$$c_x \sim x.$$

The agent's evaluation of c_x is based only on his evaluation of (immediate) consumption. In contrast, his evaluation of x is based on the consumption sequence and also on his level of patience, that is, his willingness to wait for the future outcomes. A key question is: what behavior reveals changes in impatience in response to scaling down consumption?

We argue that if impatience does not change in response to scaling down x by α then it must be that:

$$\alpha c_x \sim \alpha x.$$

This is because impatience is held constant, and the scaling down affects the evaluation of consumption equally for both the immediate reward and the stream (this relies on the fact that we assume Independence for evaluation of consumption c in any period – a consequence of Risk Preference and Monotonicity). On the other hand, if impatience changes in response to scaling down by α, we should observe

$$\alpha c_x \succ \alpha x.$$

Although the value of both reduce the same way due to scaled down consumption, the value of the stream x reduces further due to the additional reason that the agent becomes less willing to wait for the small future outcomes. Thus we capture diminishing impatience with respect to stakes. This suggests the following axiom.[3]

[3] We use the term "Increasing Attention" instead of "Diminishing Impatience" because the latter is commonly used in the literature to describe diminishing impatience with respect to delay.

Axiom 6 (Increasing Attention) *For any x, c_x and $\alpha \in [0,1]$,*

$$c_x \sim x \Longrightarrow \alpha c_x \succsim \alpha x.$$

While Risk Preference states that immediate consumption becomes worse by mixing with 0, Increasing Attention extends this property to future consumption as well: if $x \neq 0$ and $\alpha < 1$ then $x \sim c_x \succ \alpha c \succsim \alpha x$, that is, $x \succ \alpha x$.

We have seen how a comparison of αc_x and αx for different α allows us to capture change in behavior towards changing stakes arising from changes in impatience. We hypothesize further that impatience may not diminish indefinitely with stake size and that the agent might hit a lower bound. How do we tell when the agent is at his minimum impatience and when he is not? We argue that when the agent hits a lower bound, his impatience should not vary with small changes in the size of the stakes, and in particular:

$$c_x \sim x \Longrightarrow \alpha c_x \sim \alpha x \text{ for all } \alpha \text{ close to } 1.$$

When this condition is satisfied, we say that x is a *large stream*.

Similarly we can consider the case where the agent has not hit his lower bound. For such streams, any change in stakes should should alter his impatience. In particular,

$$c_x \sim x \Longrightarrow \alpha c_x \succ \alpha x \text{ for all } \alpha \in (0,1).$$

When this is satisfied, we call x a *small stream*. Intuitively, for small rewards, if there is a cognitive process determining impatience through the optimal use of limited resources, then this process is not hitting the cognitive resource constraint.

Define the set of small streams $X_s \subset X$ by

$$X_s = \{x \in X : \alpha c_x \succ \alpha x \text{ for all } \alpha \in (0,1)\}.$$

The preceding discussion motivates our next axiom.

Axiom 7 (Small Stakes Regularity) *The set X_s of small stakes is*
(i) closed,
(ii) star-shaped: $x \in X_s \Longrightarrow \alpha x \in X_s$ for all $\alpha \in (0,1]$,
(iii) absorbing: $x \in X \Longrightarrow \alpha x \in X_s$ for some $\alpha \in (0,1]$.

The axiom states that a (i) a sequence of small rewards converges to a small reward, (ii) if a stream x is small, then any scaled down version of it (αx for any $\alpha \in (0, 1]$) is also small, and finally, (iii) any arbitrary stream can be made small by scaling it down sufficiently.

Our last axiom rules out the possibility where if the agent becomes less impatient toward an outcome in period t then his impatience towards outcomes in adjacent periods is affected.[4] That is, when he is not hitting a lower bound on impatience, his impatience towards each part of the stream will be assumed to be independent of this impatience towards other parts.[5] This motivates a form of *separability* when dealing with small streams.

To formulate an axiom that expresses this, let xty denote the stream that pays according to x at t and according to y otherwise. As above, c_x denotes the present equivalent of x.

Axiom 8 (Small-Stakes Separability) *For all $x \in X_s$ and all t,*

$$\frac{1}{2}c_{xt0} + \frac{1}{2}c_{0tx} \sim \frac{1}{2}c_x + \frac{1}{2}c_0.$$

Given that preferences over immediate consumption satisfy Independence and that the axiom considers lotteries over present equivalents, we interpret the above axiom via its analogy with the following condition defined for a hypothetical preference that is defined over lotteries over streams that satisfied independence:

$$\frac{1}{2} \circ xt0 + \frac{1}{2} \circ 0tx \sim \frac{1}{2} \circ x + \frac{1}{2} \circ 0.$$

For instance, in a three period setting,

$$\frac{1}{2} \circ (0, c', 0) + \frac{1}{2} \circ (c, 0, c'') \sim \frac{1}{2} \circ (c, c', c'') + \frac{1}{2} \circ (0, 0, 0).$$

This says that the agent only cares about the distribution of consumption across periods, and not the possible correlation across periods.

An axiom that is conspicuously missing from our list is Monotonicity, that is, the condition that a stream that yield more preferred consumption

[4] Alternatively, we require that the duration of a period is sufficiently "long" that such effects disappear.

[5] While conceptually not necessary for the model, the axiom will allow us to provide a representation with uniqueness properties.

in each period than another must also be preferred. We will revisit this later. For now we just note that our axioms imply that preferences must be monotone on certain subdomains. First, given separability, monotonicity will hold on X_s. Second, preferences will be monotone along rays: For any $0 \neq x \in X$ and $\alpha < 1$, $x \succ \alpha x$. This is implied by Increasing Attention and Risk Preferences.[6]

3 Representation Theorem

Consider a tuple $(u, (\varphi_t), K)$ where

(i) $u : C \to \mathbb{R}$ is a continuous utility index with $\min_{c \in C} u(c) = 0$,

(ii) $\varphi_t : [0,1] \to \mathbb{R}_+ \cup \{\infty\}$ is a increasing convex cost function that satisfies $\varphi_t(0) = 0$ and is strictly increasing, strictly convex and differentiable on the effective domain,[7] and,

(iii) $K : X \to \mathbb{R} \cup \{\infty\}$ is a continuous function satisfying

$$K_x = K_{\lambda x} \text{ for any } x \text{ and } \lambda > 0.$$

Here u represents the agent's instantaneous utility from consumption, and φ_t is the cognitive cost function. The function K_x describes the stock of cognitive resources available to the agent when he is facing stream x. This stock is invariant to the scale of the stream – this will be interpreted below.

We define:

Definition 2 (Bounded Attention Representation) *A Bounded Attention representation for $\succsim$ is a tuple $(u, (\varphi_t), K)$ satisfying (i)-(iii) such that $\succsim$ is represented by the function $U : X \to \mathbb{R}$ defined by*

$$U(x) = u(x_0) + \sum_{t=1}^{T} D_x(t) \cdot u(x_t),$$

$$s.t. \;\; D_x = \arg\max_{D \in \Lambda_x} \{\sum_{t=1}^{T} D(t) \cdot u(x_t) - \varphi_t(D(t))\}$$

$$\Lambda_x = \{D : \sum_{t=1}^{T} \varphi_t(D(t)) \leq K_x\}.$$

[6]By these axioms, for any $x \neq 0$ and $\alpha < 1$, $x \sim c_x \succ \alpha c_x \succsim \alpha x$.

[7]The effective domain of φ_t is the set $\{d \in [0,1] : \varphi_t(d) < \infty\}$.

The cognitive constraint set Λ_x is defined by the set of discount functions whose cost does not exceed the stock K_x. Therefore it is as if, when faced with x, the agent has available resources of amount K_x, and he draws on this at a cost to pay attention to the future. As discussed in the Introduction, it is as if the cognitive process is motivated to appreciate future rewards, and it does so by focusing attention. The cognitive process participates in producing affect.

The dependence of the stock K_x on the stream x can be interpreted as follows: while there may be an overall stock K, the agent may incur some fixed costs to understanding the structure of a given stream (the stream may be an action, such as buying an investment property, and determining the implied consumption stream may require some thinking), leaving him with $K_x \leq K$ resources to devote to appreciating the stream.

The stock K_x is assumed to be invariant along the ray $\{\lambda x : \lambda > 0\}$, that is, K_x is constant for a given stream x and any scaling of it. Note that K_x could equivalently be written as a function of the normalized distribution of utilities offered by the stream $(\frac{u(x_t)}{\sum_{i=0}^{T} u(x_i)})_{t\in\{0,..,T\}}$. This observation further supports the interpretation that some fixed cost is incurred to understand the structure of the stream. While not imposed on the model, one would presume that constant streams are the easiest to understand.

The sum in the cognitive constraint is taken over $t > 0$, since discounting refers to the future. However, the cognitive resource K_x depends on the entire stream including the value of immediate consumpion. One case where dependence on immediate consumption may arise is in *temptation*: the possibility of higher immediate consumption may shift the agent's attention to the present and reduce the amount of cognitive resources available to evaluate the future. We leave it to future research to explore this.

A priori one might imagine that a cognitive cost function φ would be sufficient to describe cognitive constraints via the effective domain of φ, i.e. $\{D : \varphi(D) < \infty\}$. However our axioms deliver a model where the cognitive constraint Λ_x is generically a strict subset of the effective domain.[8] The behavioral significance of this remains to be understood.

It is worth noting that in the representation, while the cognitive costs impact the cognitive choice of D_x, these costs are not deducted from the

[8]When $K < \infty$ is constant, it is always a strict subset. We conjecture, in fact, that it is only in trivial cases (such when there is no cognitive constraint, $X_s = X$) where it equals the effective domain .

utility of the stream.[9] This is intuitive. The cognitive costs are *sunk* by the time that the agent makes his choices. Therefore the impact of these costs must be limited only to the role they played in the evaluation of the alternatives.

3.1 Results

The main result of this paper is the following representation theorem:

Theorem 1 *A preference $\succsim$ satisfies Regularity, Increasing Attention, Small-Stakes Regularity and Small-Stakes Separability if and only if it admits a Bounded Attention Representation.*

The construction of the utility representation from our axioms is as follows. Regularity and Small-Stakes Separability yields an additively separable representation on the space of small rewards X_s. This representation can be rewritten in the obvious way so that it looks like a discounted utility as in the desired representation, with the discount function D_x dependent on the stream. Since u and D_x is given, we can using the first order condition $u(x_t) = \varphi'(D_x(t))$ for each t to obtain an additive cost function $\varphi = \sum \varphi_t$ for which D_x is optimal. This yields a representation close to the desired one on the space of small rewards X_s (the constraint remains to be defined) The remaining problem is as follows. Increasing Attention requires that along a ray $\{\lambda x : \lambda > 0\}$, as λ increases, $D_{\lambda x}$ should increase and eventually become constant once λx crosses the boundary of X_s. The nontrivial step in proving the theorem was to find a cognitive constraint Λ_x for which the optimal D has this property. An arbitrary closed and convex set $\Lambda_x \subset [0, 1]^T$ satisfying $\Lambda_x = \Lambda_{\lambda x}$ for all λ does *not* define a model that is consistent with Increasing Attention.[10] We find that the cognitive constraint in the Bounded Attention representation does the job.

[9] Such a model would take the form

$$U(x) = \max_{D \in \Lambda} \{ \sum_{t=0}^{T} D(t) \cdot u(x_t) - \varphi_t(D(t)) \},$$

which is related to the model of ambiguity due to Maccheroni, Marinacci and Rustichini (2006).

[10] Fix x and let $\Lambda_x = \Lambda_{\lambda x} = \Lambda$. Increasing Attention requires that as stakes are increased, the discount function eventually ceases to change. But without additional structure on Λ, this property may not obtain. To see this suppose D_x is optimal for x, that is, it satisfies

The model has strong uniqueness properties, inherited from the separability of the representation on the subdomain X_s and because $\min_{c\in C} u(c) = 0$ is assumed in the representation. Say that $(u, (\varphi_t), K)$ has "maximal cost" if for any other representation of the form $(u, (\widehat{\varphi}_t), K)$ we have $\varphi_t \geq \widehat{\varphi}_t$ for all t.[11] Then

Theorem 2 *If there are two representations $(u^i, (\varphi_t^i), K^i)$, $i = 1, 2$ of the same preference $\succsim$, then there exists $\alpha > 0$ such that (i) $u^2 = \alpha u^1$, (ii) $\varphi_t^2 = \alpha\varphi_t^1$, and (iii) $K^2 = \alpha K^1$.*

This theorem implies that for all x,

$$\Lambda_x^2 = \{D : \alpha \sum_{t=1}^{T} \varphi_t^1(D(t)) \leq \alpha K^1(x)\} = \Lambda_x^1.$$

Therefore,

$$\arg\max_{D\in\Lambda_x^1}\{\sum_{t=1}^{T} D(t)u^1(x_t) - \varphi_t^1(D(t))\} = \arg\max_{D\in\Lambda_x^2}\{\sum_{t=1}^{T} D(t)u^2(x_t) - \varphi_t^2(D(t))\}$$

That is, D_x is uniquely determined from preference. Moreover, the theorem also ensures that the curvature or elasticity of φ_t is uniquely derived from preference.

3.2 Monotonicity

Axiom 9 (Monotonicity) *For any $x, x' \in X$,*

$$(x_t, 0, .., 0) \succsim (x'_t, 0, .., 0) \text{ for all } t \Longrightarrow x \succsim x'.$$

$D_x \cdot u(x) - \varphi(D_x) > D \cdot u(x) - \varphi(D)$ for all $D \in \Lambda$ (the strict inequality comes from the strict convexity of the cost function, which yields a unique maximizer). This can be rewritten as

$$D_x \cdot u(x) - D \cdot u(x) > \varphi(D_x) - \varphi(D)$$

for all $D \in \Lambda$. However, suppose $D_x \cdot u(x) - D \cdot u(x) < 0$ for some $D \in \Lambda$. Exploiting the linearity of u, it is readily seen that scaling up x to λx for $\lambda > 1$ can lead to the inequality

$$D_x \cdot u(\lambda x) - D \cdot u(\lambda x) < \varphi(D_x) - \varphi(D).$$

Consequently, even if D_x is on the boundary of Λ, scaling up rewards may change the agent's discount function in a way inconsistent with Increasing Attention.

[11] The cost function is pinned down by preference only on some interval $[0, d] \subset [0, 1]$ and in a maximal representation the cost is set to infinity outside this interval.

Moreover, if $(x_t, 0, .., 0) \succ (x'_t, 0, .., 0)$ *for some* t *then* $x \succ x'$.

This condition may fail around the boundary of X_s. Imagine that x dominates y but also comes with a much lower K_x. Intuitively, x is a better stream but understanding the stream (the fixed cognitive cost) requires a very high cognitive cost, leaving little cognitive resources to appreciate the stream.

[Characterization in progress].

4 Attention to "Bad" Outcomes

In this section we clarify some matters related to discounting, attention and 'bad' outcomes.

Consider the intuition that people pay more attention not just to larger gains but also to larger losses. A key point to note is that this intuition is based on preferences over *changes* in consumption, while our model is one of preferences over consumption (where consumption is always more desirable than 0). Consequently, the proper application of the model to study "bad" outcomes is by looking at the induced preferences over streams of changes relative to a given consumption profile.

Our model implies that, for a given base-line consumption stream b, a stream Δc of changes in consumption is accepted if and only if[12]

$$\sum_{t=0}^{T} D_{b+\Delta c}(t)u(b_t + \Delta c_t) - D_b(t)u(b_t) \geq 0.$$

That is, the agent considers not just the change in u but also the change in D associated with Δc.

How is the change in D associated with attention? Because D is increasing with consumption, it is intuitive that higher gains $\Delta c > 0$ will get higher levels of attention. To the extent that "attention is increasing in losses" means that "the loss is deemed worse as the magnitude of the loss increases", this in fact still requires D to be increasing in consumption, as in our model. For instance, without a change in D (intuitively, holding attention fixed), losing $-\Delta c$ at time t (and no change in any other period) yields a loss of

$$D_b(t)u(b_t - \Delta c) - D_b(t)u(b_t) < 0,$$

[12] We have not defined additive changes in consumption but this can be viewed as notation for proportional changes for each period, which is well-defined.

but the loss is *greater* in magnitude only with a decrease in D due to lower consumption, that is, if $D_{b-\Delta c}(t) < D_b(t)$:

$$D_{b-\Delta c}(t)u(b_t - \Delta c) - D_b(t)u(b_t) < D_b(t)u(b_t - \Delta c) - D_b(t)u(b_t) < 0,$$

as was to be shown. Intuitively, increases in D are increases in appreciation, not increases in attention. If attention towards losses means diminished appreciation of the consumption offered, then this requires D to decreasing with $u(c)$, as in our model.

One may wish further to impose a property analogous to "loss aversion":

$$D_b(t)u(b_t) - D_{b-\Delta c}(t)u(b_t - \Delta c) > D_{b+\Delta c}(t)u(b_t + \Delta c) - D_b(t)u(b_t).$$

This can be captured by imposing concavity of $D_c(t)u(c_t)$.

5 A Special Case

5.1 Axioms

By Increasing Attention (along with Order and Continuity), we know that for any $x \in X_s$ and α, there exists $\beta_{\alpha,x} < \alpha$ s.t.

$$\beta_{\alpha,x} c_x \sim \alpha x.$$

In order to place more structure on the preference, we impose that $\beta_{\alpha,x}$ must be independent of x. Intuitively, a worsening of each component of the stream by (say) 10% should lead to a worsening of the entire stream by 20%, regardless of the stream being faced. While this does not preclude the possibility that discounting depends on the outcome, it requires that discounting be homogenous with respect to changes in the scale of the outcome.

Axiom 10 (Homogeneous Attention) *For any* $\alpha \in (0,1)$ *there is* β *such that for any* $0 \neq x \in X_s$,

$$\beta c_x \sim \alpha x.$$

Turning to the next axiom, recall that larger streams will induce higher levels of attention from the agent. Thus the agent hits her cognitive constraint for streams that are higher up her preference ranking. A simple condition that can be used to define the boundary between small and large rewards is that the boundary is in fact an indifference curve when looking at streams with constant immediate consumption (recall that immediate consumption does not contribute to cognitive costs). We view this not as a descriptive assumption, but rather one that is very attractive from the perspective of building a tractable model. The value of this will be demonstrated in the application below.

Axiom 11 (Restricted Boundary) *For any $x, y \in X_s$ wih $x_0 = y_0$,*

$$x, y \in bd(X_s) \Longrightarrow x \sim y.$$

5.2 Representation Result

Our main result in this section is:

Theorem 3 *Consider a preference $\succsim$ that admits a Bounded Attention representation. Then the following statements hold:*

(i) $\succsim$ satisfies Homogeneous Attention if and only if

$$\varphi_t(d) = \delta_t \cdot d^m, \qquad d \in [0, 1],$$

for some $\delta_t > 0$ that is increasing in t and some scalar $m > 1$.

(ii) $\succsim$ satisfies Homogenous Attention and Restricted Boundary if and only if it is also the case that K_x is independent of x.

Thus Homogenous Attention imposes a lot of structure on the preference, which gets expressed in the representation by the functional form imposed on the cost function with one variable and one fixed parameter. With this structure there is a connection between the utility of streams and the boundary of X_s, which is exploited by Restricted Boundary to obtain a constant cognitive resource K that does not depend on the stream. Intuitively, there is no cognitive cost of learning the stream, and all costs are associated with evaluating it.

Say that a preference admits a *Homogeneous Attention representation* if it admits a Bounded Attention representation, and satisfies Homogeneous Attention and Restricted Boundary.

5.3 Reduced Form

It is instructive to derive the reduced form of the Homogeneous Attention model.

Proposition 1 *If $\succsim$ admits a Homogeneous Attention representation then there exist parameters $m > 1$, $\kappa > 0$ and a function $D : \{0, .., T+1\} \rightarrow [0, (\frac{m-1}{m\kappa})^{m-1}]$ such that $\succsim$ is represented by $U : X \rightarrow \mathbb{R}$ where*

$$\begin{aligned} U(x) &\leq \kappa \Longrightarrow U(x) = u(x_0) + \sum_{t>0} D(t) u(x_t)^m, \\ U(x) &> \kappa \Longrightarrow U(x) = u(x_0) + \kappa^{\frac{m-1}{m}} \left[\sum_{t>0} D(t) u(x_t)^m \right]^{\frac{1}{m}}. \end{aligned}$$

The utility level κ defines an indifference curve, and any stream in the lower contour set is a small stream, while anything in its complement is a large stream. If the cost function is parametrized as $\varphi(d) = \delta_t \frac{1}{\theta} d^\theta$ then

$$D_c(t) = \left(\frac{u(c)}{\delta_t} \right)^{\frac{1}{\theta - 1}} := D(t) u(c)^{m-1},$$

that is, the discount function is a power transformation of the utility of the reward. This is used to compute discounted utility for small rewards, yielding the expression in the proposition.

We find that for small stakes, the utility function is additively separable, and future utility from lotteries is a power transformation of the immediate utility from lotteries. The latter implies that risk aversion is the same between periods. However, the parameter m will affect intertemporal substitution.

For large stakes we find that the utility function is not longer additively separable. In fact future utility is evaluated using a concave aggregator.

6 Concluding Remarks

This paper suggests several substantive avenues for future research.

Dynamic extension: [To be added]

Temptation: An interesting observation about the model is that it contains elements of a story of temptation and self-control in the context of intertemporal choice. The agent normatively wishes to understand his options fully, that is, to evaluate a stream without discounting the future $\sum_{t=0}^{T} u(x_t)$. However, since thinking about the future requires effort, he is "tempted" to evaluate it only in terms of immediate consumption $u(x_0)$. The cognitive cost of attention to the future is related to the notion of self-control cost. Intuitively, one way that an agent may exert self-control is by influencing choice via thinking about the virtue of the more virtuous action.

A richer domain is required to provide more rigorous behavioral foundations for such an interpretation of the model. For instance, if the agent has a normative preference to pay maximal attention in his intertemporal choices, then ex ante the agent may prefer menus that encourage him to do so ex post. For instance, the agent may ex ante prefer that small rewards spread across different periods be lumped together as one large reward in one period. We leave such a study for future research.

Welfare: While this extends beyond revealed preference, one can hypothesize that our agent's welfare is determined by the preference he would exhibit if he did not have any cognitive costs. In our model this is given by

$$x \mapsto \sum_{t=0}^{T} u(x_t),$$

that is, the undiscounted sum of utility. Whether this is compelling or not depends on the result of a thought experiment: would we discriminate between rewards in the present versus the future if we could view the future as clearly as we could view the present?

APPENDIX [To be added]

BIBLIOGRAPHY [To be added]