An Attentional Model of Time Discounting

Zark Zijian Wang

March 13, 2024

1 Introduction

2 The Model

2.1 Definition

Assume time is discrete. Let $s_{0\to T} \equiv [s_0, s_1, ..., s_T]$ denote a reward sequence that starts delivering rewards at period 0 and ends at period T. At each period t of $s_{0\to T}$, a specific reward s_t is delivered, where $t \in \{0, 1, \dots, T\}$. Throughout this paper, we only consider non-negative rewards and finite length of sequence. Therefore, we set $s_t \in \mathbb{R}_{\geq \not\vdash}$ and $0 \leq T < \infty$. The DM's choice set is constituted by a range of alternative reward sequences which start from period 0 and end at some finite period. When making an intertemporal choice, the DM seeks to find a reward sequence $s_{0\to T}$ in her choice set, which has the highest value among all alternative reward sequences. To calculate the value of each reward sequence, we admit the additive discounted utility framework. The value of $s_{0\to T}$ is defined as $U(s_{0\to T}) \equiv \sum_{t=0}^{T} w_t u(s_t)$, where $u(.)$ is the instantaneous utility function, and w_t is the decision weight assigned to s_t. The function $u(.)$ is twice differentiable, $u' > 0$ and $u'' < 0$.

The determination of w_t is central to this paper. We believe that the formation of w_t is subjective to limited attention. Specifically, we term a decision weight w_t as an *attention-adjusted discount* (AAD) factor if it satisfies Definition 1.

Definition 1: *Let $\mathcal{W} \equiv [w_0, ..., w_T]$ denote the decision weights for all specific rewards in $s_{0\to T}$. $\mathcal{W}$ is called attention-adjusted discount factors if for any $t \in \{0, 1, \dots, T\}$*

$$w_t = \frac{d_t e^{v(s_t)}}{\sum_{\tau=0}^{T} d_\tau e^{v(s_\tau)}} \tag{1}$$

where $d_t \geq 0$, $v(.)$ is a twice-differentiable function, $v' > 0$ and $v'' < 0$.

In intuition, how Definition 1 reflects the role of attention mechanisms in decision-making can be explained with four points. First, note that the attention-adjusted discount factors follow a logistic-like distribution. This is consistent with the prediction of rational inattention theory. Second, for each t, w_t is increasing with s_t, indicating that DM tends to pay more attention to larger rewards. This is consistent with an empirical phenomenon called "value-driven attentional capture". Third, w_t is "anchored" in the initial weight d_t. We can let d_t denote the initial weight that the DM would assign to a reward delivered at period t, without knowing its realization. This indicates that reallocating attention based on the newly acquired information is costly. Fourth, note that the sum of w_t is fixed at 1, which implies the DM's capacity of information processing is limited. Being too focused on one reward will make DM insensitive to another reward in the sequence.

In popular time-discounting models, such as exponential and hyperbolic discounting, discount factors are typically assumed to be independent of how each reward is realized in $s_{0\to T}$. This type of discount factors reflects impatience, and can be viewed as initial weights assigned to each reward when the DM has no information about its value. That is, we can use them for d_t. By contrast, w_t in AAD is influenced by impatience but also reflects the attention allocated to each specific reward realized in the sequence. Notably, w_t is dependent on all rewards realized in $s_{0\to T}$. An increase in s_t would attract more attention to it, thus the DM would be more sensitive to s_t, and the attention remained for other rewards in $s_{0\to T}$ will decrease. As a result, she could reduce other decision weights, i.e. discount the value of other rewards by a larger degree.

2.2 Related Literature in Attention

1 attention bottleneck

2 attentional capture

3 costly information acquisition

2.3 Related Literature in Time Preferences

3 Axiomatic Characterizations

3.1 The Optimal Discounting Approach

The first axiomatic characterization of AAD is based on the optimal discounting model proposed by Noor and Takeoka (2022; 2023). In one version of their model, they assume that the DM has a limited capacity of attention (or in their term, "empathy"), and before encountering an intertemporal choice problem, the DM naturally focuses on the current time period. During the choice process, she needs to split attention over the time interval spanned by an alternative reward sequence in order to evaluate it. This re-allocation of attention is cognitive costly. Thus, for each alternative reward sequence, the DM seeks to maximize the value she can subjectively obtain from the reward sequence minus the cost incurred by attention re-allocation. In our setting, we relax the assumption that the DM should initially focuses all her capacity on the current time; instead, the initial distribution of decision weights can be flexible.[1] The formal definition of this optimal discounting problem is given by Definition 2.

Let $\succsim$ denote the DM's preference for reward sequences. We say $\succsim$ has an optimal discounting

[1]There are another two difference between us and Noor and Takeoka (2022; 2023). First, in our setting, shifting attention to future periods may also reduce the attention to the current period, while this would never happen in their settings. Second, for any $w_t \in [0,1]$, they assume that $f'(w_t)$ could be 0 when w_t is under a lower bound, could be infinity when w_t is above a upper bound, and is strictly increasing in between. To keep simplicity, we assume $f(.)$ is strictly convex, that is, $f'(w_t)$ is always increasing. Note that our assumption is satisfied by many commonly used cost functions (such as the power cost function they discussed in their settings, and the entropy-based cost function discussed in this paper).

representation if $s_{0\to T} \succsim s'_{0\to T'}$ implies $\sum_{t=0}^{T} w_t \cdot s_t \succsim \sum_{t=0}^{T'} w'_t \cdot s'_t$, and both $\{w_t\}_{t=0}^{T}$ and $\{w'_t\}_{t=0}^{T'}$ are generated by optimal discounting problems.

Definition 2: *The following optimization problem is called optimal discounting problem:*

$$\begin{aligned} \max_{\mathcal{W}} \quad & \sum_{t=0}^{T} w_t v(s_t) - C(\mathcal{W}) \\ s.t. \quad & \sum_{t=0}^{T} w_t \leq M \\ & w_t > 0 \text{ for all } t \in \{0, 1, ..., T\} \end{aligned}$$

where $C(.)$ is the cognitive cost function. $C(\mathcal{W})$ is constituted by time-separable costs, that is, $C(\mathcal{W}) = \sum_{t=0}^{T} f_t(w_t)$, where $f_t(.)$ is a twice differentiable and strictly convex function.

We focus on a specification of $C(.)$, in which we assume that

$$C(\mathcal{W}) = \lambda \cdot \sum_{t=0}^{T} w_t \ln\left(\frac{w_t}{d_t}\right)$$

3.2 The Intertemporal Trade-Off Approach

4 Implications in Decision Making

To illustrate how ADU with Shannon cost function can account for a broad set of anomalies about time preferences, imagine that a DM receives a positive detereminstic reward in period j (and no reward in other periods). That is, she receives a sequence of rewards $X_T = [x_0, x_1, \ldots, x_T]$, where $x_j > 0$ and is certain, and $x_t = 0$ for all $t \neq j$ (both j and t are in $\{0, 1, ..., T\}$).

For the convenience of illustration, I assume the DM holds stationary time preferences before acquiring any information, that is, $d_t = \delta^t$. Meanwhile, $\delta \in (0, 1]$, where $\delta = 1$ implies the initial attention is uniformly distributed across periods. For simplicity, I define $v(x_t) = u(x_t)/\lambda$, and set $v(0) = 0$. Let $w_t(X_T)$ denote the discounting factor for period t. From the formula of ADUS we can infer that

$$w_j(X_T) = \begin{cases} \delta^j \cdot \dfrac{1}{1 + \frac{\delta}{1-\delta}(1-\delta^T)e^{-v(x_j)}} , & 0 < \delta < 1 \\ \dfrac{1}{1 + T \cdot e^{-v(x_j)}} , & \delta = 1 \end{cases}$$

Clearly, w_j is decreasing in T. This offers an account for a phenomenon called *hidden zero effect*.

4.1 Hidden Zero Effect

The most direct evidence that could support the ADUS model is likely the hidden zero effect (Magen et al., 2008). The hidden zero effect means, supposing people face a small sooner reward (SS) and a large later reward (LL), they tend to exhibit more patience when SS and LL are framed as sequences rather than being framed as single-period rewards. For instance, suppose SS is "receive £100 today" and LL is "receive £120 in 6 months", and we have

SS_0: "receive £100 today and £0 in 6 months"

LL_0: "receive £0 today and £120 in 6 months"

people will be more likely to prefer LL_0 over SS_0 than preferring LL over SS. Subsequent research (e.g. Read et al. (2017)) suggests that the hidden zero effect is asymmetric. That is, shifting SS to SS_0 and keeping LL unchanged leads to an increase in patience, whereas shifting LL to LL_0 and keeping SS unchanged cannot increase patience. ADUS assumes that, within a sequence, attention is limited and the weight assigned to each period is anchored in an initial positive weight. These properties naturally explain the hidden zero effect. To illustrate, in SS, the DM perceives the length of sequence as "today" and allocate no attention to future. Whereas, in SS_0, she perceives the length as "6 months". This makes some attention be paid to future periods with no reward, and decreases the attention paid to the only period with positive reward (given attention is limited); thus, the overall utility of sequence decreases. By contrast, shifting from LL to LL_0 does not change the length of sequence, thus does not change overall utility.

The existence of hidden zero effect also provides a hint in selection of time length T. When evaluating a reward delivered in period j, the range of T is $[j, +\infty)$. Any increase in T will reduce the overall utility. Thus, when comparing SS and LL, the DM may tend to set $T = j$ (the minimum length she can set), in order to maximize the overall utility. Any period out of this length can be perceived as irrelevant to the decision; so, she does not need to sample from the periods after j, when evaluating the given reward. Though, explicitly mentioning the periods after j will direct her attention to those periods, and lead to the hidden zero effect. By setting $T = j$, we have

$$w_T(x_T) = \frac{1}{1 + G(T)e^{-v(x_T)}}$$

where

$$G(T) = \begin{cases} \dfrac{1}{1-\delta}(\delta^{-T} - 1) \ , & 0 < \delta < 1 \\ T \ , & \delta = 1 \end{cases}$$

Given period T is now the only period with a non-zero reward within the sequence, I use x_T to directly represent the whole sequence, and let $w_T(x_T)$ denote the discounting factor for period T. Interestingly, when $\delta = 1$, $w_T(x_T)$ takes a form similar with hyperbolic discounting.

4.2 Common Difference Effect

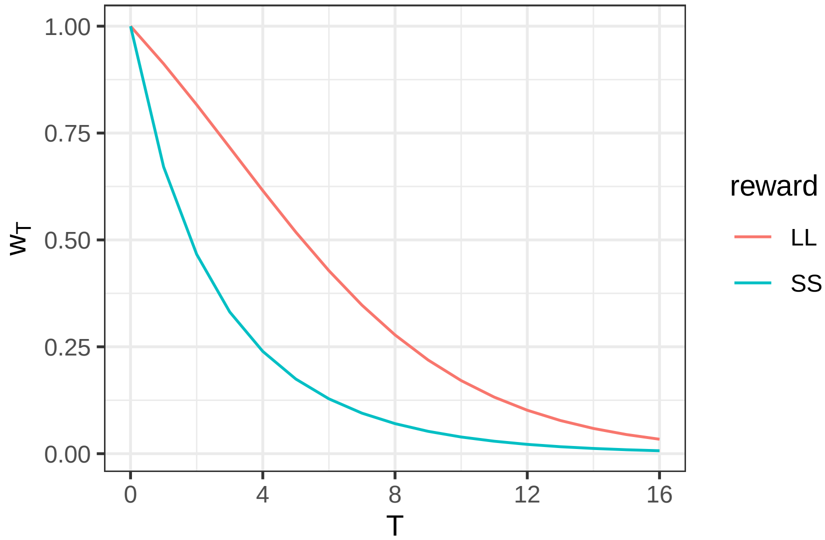
A well-known anomaly about time preferences is *common difference effect*, firstly defined by Loewenstein and Prelec (1992). Suppose there are a large later reward x_l arriving at period t_l (denoted by LL) and a small sooner reward x_s arriving at period t_s (denoted by SS), where $x_l > x_s > 0$, $t_l > t_s > 0$. Define $V(x,t) = w_t(x_t)v(x_t)$. The common difference effect means, supposing$V(x_l, t_l) = V(x_s, t_l)$, we must have $V(x_l, t_l + \Delta t) > V(x_s, t_s + \Delta t)$ for any positive integer Δt.

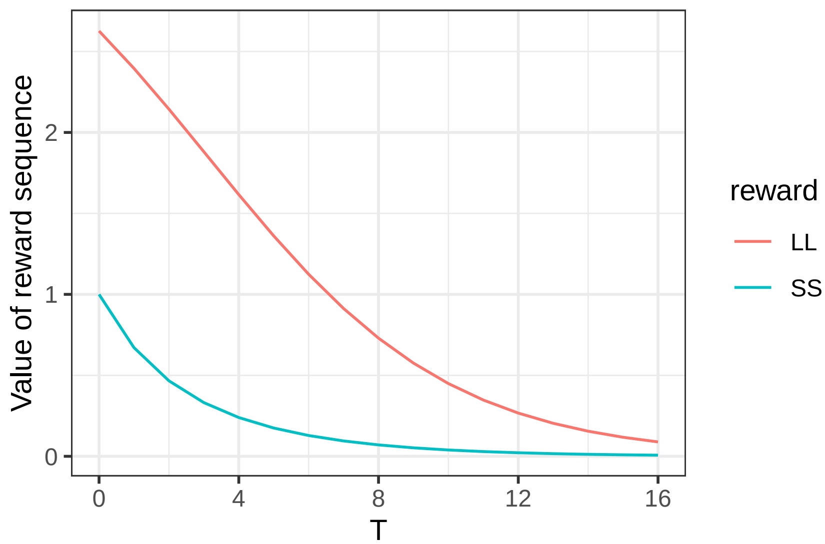
ADUS predicts that, if people are impatient, to observe the common difference effect, the difference between SS and LL in reward level must be set significantly larger than the difference in time delay. This is shown in Proposition 2.

Proposition 2: *In ADUS, if the initial weights are uniformly distributed, then the common difference effect always holds; if the initial weights exponentially declines over time, the common difference effect holds when* $v(x_l) - v(x_s) + \ln \frac{v(x_l)}{v(x_s)} > -(t_l - t_s) \ln \delta$.

Proposition 2 is interpreted as follows. When $\delta = 1$, ADUS predicts the DM always performs the common difference effect. This is obvious because discounting factor $w_T(x_T)$ takes a hyperbolic-like form. When $\delta < 1$, there are four factors jointly deciding whether we could observe the common difference effect or not. First, without considering attentional mechanism, when we extend time delay, each of $w_{t_l}(x_l)$ and $w_{t_s}(x_s)$, i.e. the discounting factor for (and attention paid to) the only period with positive reward, declines in an exponential fashion. Second, without considering newly added time interval, due to the decline of $w_{t_l}(x_l)$ and $w_{t_s}(x_s)$, the DM frees up some attention and can reallocate it across periods. Given that in LL, the DM has to wait longer for reward, the periods where she wait can grab more attention from the released capacity of attention, compared with those in SS. In other words, an extension of delay makes she focus more on the waiting time in LL than in SS, which decreases the preference for LL. Third, the newly added time interval also grabs some attention from other periods. Note the time delay is extended by $[t_l, t_l + \Delta t]$ in LL and by $[t_s, t_s + \Delta t]$ in SS; given $t_l > t_s$, if people are impatient, the newly added time interval will receive less attention in LL than in SS, without considering other factors. This increases the preference for LL. Fourth, ADU generally assumes that the DM tends to pay more attention to periods with larger rewards. Given $x_l > x_s$, the newly added interval grabs less attention from the period where x_l is positioned (in LL) than from the period where x_s is positioned (in SS). That is, the DM focuses comparatively more on reward level in LL than in SS, which mitigates the impact of discounting factor declining. This also increases the preference for LL. When the impact of the later two factors succeeds that of the second factor, the DM will perform the common difference effect.

Notably, if we explicit mention the zeros in LL and SS, extending time delay always lead to the common difference effect.





4.3 Magnitude Effect

The *magnitude effect* is another well-known anormaly about time preferences. Assuming we have t_l, t_s, x_s fixed, and want to find a x_l such that $V(x_l, t_l) \equiv V(x_s, t_s)$, the magnitude effect implies that, if we increase x_s, then the x_l/x_s that makes the equality valid will decrease.

In standard discounted utility model, the magnitude effect requires the elasticity of utility function to increase with the reward level (Loewenstein and Prelec, 1992). This requirement might be too restrictive, so that many commonly used utility functions (such as power or CARA utility function) does not satisfy it. By contrast, in ADU model, DM is generally assumed to attend more to periods with larger rewards. This implies that when comparing SS and LL, she exhibits more patience towards larger reward level, which is naturally compatible with the magnitude effect (Noor, 2011; Noor and Takeoka, 2022). By Proposition 3, I focus on ADU with Shannon cost function, and show how this requirement for curvature of utility function can be relaxed in this setting.

Proposition 3: *Define $v(x) \equiv u(x)/\lambda$ as the utility function. In ADUS, the magnitude effect always holds true when function $v(x)$ satisfies*

$$RRA_v(x) \leq 1 - \frac{e_v(x)}{v(x)+1}$$

where $RRA_v(x)$ is the relative risk aversion coefficient of $v(x)$, $e_v(x)$ is the elasticity of $v(x)$ to x.

Note that Proposition 3 is a very broad condition. In Corollary 1 and Corollary 2, I show that power utility function and CARA utility function both satisfy this condition in most cases.

Corollary 1: Suppose $v(x) = x^\gamma/\lambda$, where $0 < \gamma < 1$ and $\lambda > 0$. Then magnitude effect holds true for any $x \in \mathbb{R}_{>0}$.

Corollary 2: Suppose $v(x) = (1 - e^{-\gamma x})/\lambda$, where $\gamma > 0$ and $\lambda > 0$. The magnitude effect holds true for any $x \geq \frac{1+\eta}{\gamma}$, where $\eta > 0$ and $\eta e^{1+\eta} - \eta = 1$ (it can be calculated that $\eta \approx 0.35$).

4.4 Concavity of Time Discounting

Many time discounting models assumes discount function is convex in time delay, e.g. exponential and hyperbolic discounting. This style of discount function predicts DM is *risk seeking over time lotteries.* That is, suppose a deterministic reward of level x is delivered in period t_l with probability π and delivered in period t_s with probability $1-\pi$ ($0 < \pi < 1$, $c > 0$); while another deterministic reward, of the same level, is delivered in a certain period t_m, where $t_m = \pi t_l + (1-\pi)t_s$. The DM should prefer the former reward to the latter reward.

However, some experimental studies, such as Onay and Öncüler (2007) and DeJarnette et al. (2020), suggest that people are often *risk averse over time lotteries*, i.e. preferring the reward delivered in a certain period.

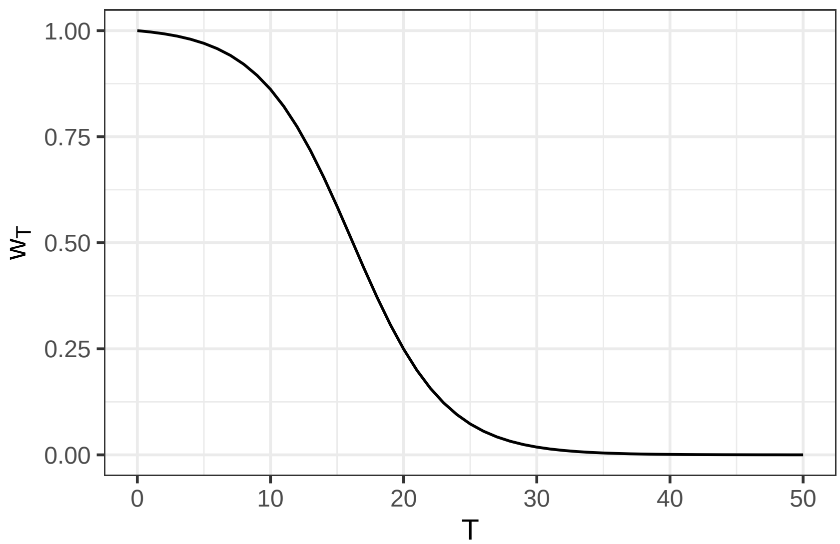
One way to accommodate the evidence about risk aversion over time lotteries, as is suggested by DeJarnette et al. (2020), is to modify the convexity (concavity) of discount function. Under a general EDU framework, DM is risk averse over time lotteries when $\pi w_{t_l}(x) + (1-\pi)w_{t_s}(x) < w_{t_m}(x)$. Fixing t_s and t_l, the inequality suggests $w_{t_m}(c)$ is concave in t_m. In reverse, being risk seeking over time lotteries suggests $w_{t_m}(x)$ is convex in t_m. Notably, Onay and Öncüler (2007) find that people are more likely to be risk averse over time lotteries when π is small, and to be risk seeking over time lotteries when π is large. Given that when π gets larger, t_m is also larger, we can conclude that the discount function may be concave in delay for the near future but convex for the far future. Moreover, Takeuchi (2011) also find evidence that support this shape of discount function.

In Proposition 4, I show that ADUS can produce such a shape of discount function as long as the reward level x is large enough.

Proposition 4: In ADUS, *if* $\delta = 1$, *then the discount function is convex in* t. *If* $0 < \delta < 1$, *then there are a reward threshold* $\underline{x}$ *and a time threshold* $\underline{t}$ *such that*

1) *when* $x \leq \underline{x}$, *the discount function is convex in* t;
2) *when* $x > \underline{x}$, *the discount function is convex in* t *given* $t \geq \underline{t}$, *and it is concave in* t *given* $t < \underline{t}$.

It can be derived that $v(\underline{x}) = \ln(\frac{2}{1-\delta})$, *and* $\underline{t} = \frac{\ln[(1-\delta)e^{v(x)}-1]}{-\ln\delta}$.



4.5 S-Shaped Value Function

In prospect theory, Kahneman and Tversky (1979) propose an S-shaped value function that is convex for losses and concave for gains. Since that, S-shaped value functions have been widely embraced by behavioral economists. More recent theories have provided further justifications for it, including reference-dependent utility in a broad sense (Koszegi and Rabin, 2006), and efficient coding of values (Frydman and Jin, 2021). Here, I provide an account based on selective attention to time periods.

Suppose a DM is faced with a choice between a risky lottery and a fixed amount of money. When making this choice, she does not obtain any money from either option. Thus, she perceives the outcome of each option as something that will happen in the future. She allocate her attention between the present period and the period when she may receive the money. Assume that she perceives the outcome will be realized in period t, and in a certain state, the option she chooses yields reward x, then we can use the attentional discounted utility $V(x,t)$ to represent the value function. I derive the conditions in which ADUS can produce a S-shaped value function in Proposition 5.

Proposition 5: *Suppose* $t \geq 1$, $\frac{d}{dx}\left(\frac{1}{v'(x)}\right)$ *is continuous in* $(0,+\infty)$*, in ADUS,*

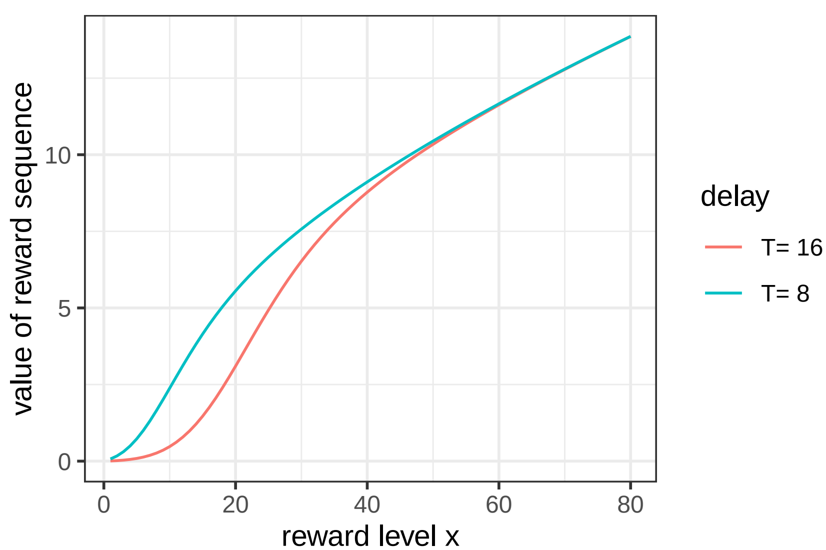
1) *there exists a threshold* $\bar{x}$ *in* $(0,+\infty)$ *such that* $V(x,t)$ *is strictly concave in* x *when*

$x \in [\bar{x}, +\infty)$;

2) *if* $\frac{d}{dx}\left(\frac{1}{v'(x)}\right)$ *is right-continuous at* $x = 0$, *and* $\frac{d}{dx}\left(\frac{1}{v'(0)}\right) < 1$, *then there exists a threshold* x^* *in* $(0, \bar{x})$ *such that, for any* $x \in (0, x^*)$, $V(x,t)$ *is strictly convex in* x;

3) *there exist a hyper-parameter* λ^* *and an interval* (x_1, x_2) *such that, if* $\lambda < \lambda^*$, *for any* $x \in (x_1, x_2)$, $V(x,t)$ *is strictly convex in* x, *where* $\lambda^* > 0$ *and* $(x_1, x_2) \subset (0, \bar{x})$.

Proposition 5 implies, if the derivative of $\frac{1}{v'(x)}$ converges to a small number when $x \to 0^+$, or the unit cost of information λ is small enough, value function $V(x,t)$ will perform an S shape in some interval of x. At the intuition level, note that $V(x,t) = w_t(x)v(x)$. When the level of reward x grows, both the instantaneous utility of it, i.e. $v(x)$, and the discounting factor assigned to it, i.e. $w_t(x)$, can increase. These functions are both concave in x: when the level of reward is small, they both grow fast. So, it is possible that their product is convex in this case. By contrast, when the level of reward is large, they grow slowly, so their product keeps concave.



4.6 Inseparability of Sequences

Let x and y denote two 2-period risky reward sequences. For x, the realized sequence is [£100,£100] with probability 1/2, and is [£3,£3] with probability 1/2. For y, the realized sequence is [£3,£100] with probability 1/2, and is [£100,£3] with probability 1/2. Classical models of intertemporal choice typically assume the separability of potentially realized sequences. This implies that the DM is indifferent between x and y. However, Andersen et al. (2018) find evidence of *intertemporal correlation aversion*, that is, people often prefer y to x.

ADU can naturally yield intertemporal correlation aversion. For simplicity, suppose the initial attention is uniformly distributed across the two periods. For x, under each potentially realized sequence, the DM equally weights each period. For y, DM tends to assign more weight to the period with a reward of £100 (suppose that weight is w). Then the value of x is $\frac{1}{2}u(100) + \frac{1}{2}u(3)$ and the value of y is $w \cdot u(100) + (1 - w) \cdot u(3)$. Given that $x > \frac{1}{2}$, the DMs should strictly prefer y to x.

- Other evidence related to inseparability: common sequence effect, (reverse) mere token effect, magnitude-increasing temporal sensitivity

5 The Role of Attention in Inconsistent Planning

5.1 Attention Grabbing and Updating

Suppose a DM has budget m ($m > 0$) and is considering how to spend it over different time periods. We can use a reward sequence x to represent this decision problem, where the DM's spending in period t is x_t. In period 0, she wants to find a x such that

$$\max_{x} \sum_{t=0}^{T} w_t u(x_t) \quad s.t. \sum_{t=0}^{T} x_t = m \tag{3}$$

where w_t is the attention-adjusted discounting factor in period t. I assume $w_t = \delta^t e^{u(x_t)/\lambda} / \sum_{t=\tau}^{T} \delta^\tau e^{u(x_\tau)/\lambda}$ and there is no risk under this setting.

In models like exponential and hyperbolic discounting, the discounting factor of a future period is consistently smaller than that of the current period. Thus, the DM should spend more at the present than in the future. By contrast, in ADU, when increasing the spending in a certain period, the discounting factor corresponding to that period should also increase. So it is possible that the DM spends more in the future and that a future period has a greater discounting factor than the current period. This is consistent with Loewenstein and Prelec (1993) that find people sometimes prefer improving sequences to declining sequences.

ADU suggests there are two mechanisms that can help explain why people may perform dynamically inconsistent behavior. The first is *attention-grabbing effect*, that is, keeping the others equal, when we increase x_t (which lead to an increase in w_t), the discounting factor in any other period should decrease due to limited attention. After omitting a previous period from the decision problem in Equation (3), the DM can assign more weights to remaining periods; thus, the attention-grabbing effect is enhanced. The increased attention-grabbing effect will offset some benefit of increasing spending toward a certain period. Therefore, when the DM prefers improving sequences, the attention-grabbing effect will make her perform a present bias-like behavior (always feeling that she should spend more at the present than the original plan); when the DM prefers declining sequences, this effect will maker her perform a future bias-like behavior (always feeling she should spend more in the future).

The second mechanism is *initial attention updating*. As is assumed above, in period 0, prior to evaluating each reward sequence, the DM's initial weight on period t is proportional to δ^t; after evaluation, the weight becomes being proportional to $\delta^t e^{u(x_t)/\lambda}$. In period 1, if she implements the evaluation based on the information attained in period 0, the initial weight should be updated to being proportional $\delta^t e^{u(x_t)/\lambda}$; thus, the weight after evaluation should become being proportional to $\delta e^{2u(x_t)/\lambda}$. As a result, the benefit of increasing spending toward a certain period gets strengthened. The updated initial attention can make those who prefer improving sequences perform present bias and those who prefer declining sequences perform future bias.

Both the attention-grabbing effect and initial attention updating are affected by the curvature of utility function. They jointly decide which behavior pattern that people should perform in dynamics.

Proposition 6 (*spread-consistency correlation*) Suppose $\succsim$ has a ADU representation and satisfies Axiom 2-4. If there exist b and S_T such that, for any b' and S'_T, $bS_T \succsim b'S'_T$, where

$b + \sum_{t=0}^{T} s_t = b' + \sum_{t=0}^{T} s'_t$, then for any S'_T, we have

$$S_T \succsim S'_T \Longleftrightarrow b \sim S_T$$

where $\sum_{t=0}^{T} s_t = \sum_{t=0}^{T} s'_t$.

Proposition 6 implies that, when allocating a consumption budget across time periods, the DM keeps her choice dynamically consistent if and only if she performs a strong preference for spread. Given that people are typically assumed to be impatient (preferring a declining sequence), one intuitive interpretation of Lemma 2 is that the less impatient a DM is in the present, the less inclined she is to deviate from the original choice in the future.

6 Discussion

7 Conclusion

Reference

Andersen, S., Harrison, G. W., Lau, M. I., and Rutström, E. E. (2018). Multiattribute Utility Theory, Intertemporal Utility, and Correlation Aversion. *International Economic Review*, 59(2):537–555.

DeJarnette, P., Dillenberger, D., Gottlieb, D., and Ortoleva, P. (2020). Time Lotteries and Stochastic Impatience. *Econometrica*, 88(2):619–656.

Frydman, C. and Jin, L. J. (2021). Efficient Coding and Risky Choice. *The Quarterly Journal of Economics*, 137(1):161–213.

Kahneman, D. and Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, 47(2):263.

Koszegi, B. and Rabin, M. (2006). A Model of Reference-Dependent Preferences. *The Quarterly Journal of Economics*, 121(4):1133–1165.

Loewenstein, G. and Prelec, D. (1992). Anomalies in Intertemporal Choice: Evidence and an Interpretation. *The Quarterly Journal of Economics*, 107(2):573–597.

Loewenstein, G. F. and Prelec, D. (1993). Preferences for sequences of outcomes. *Psychological Review*, 100(1):91–108.

Magen, E., Dweck, C. S., and Gross, J. J. (2008). The Hidden-Zero Effect: Representing a Single Choice as an Extended Sequence Reduces Impulsive Choice. *Psychological Science*, 19(7):648–649.

Noor, J. (2011). Intertemporal choice and the magnitude effect. *Games and Economic Behavior*, 72(1):255–270.

Noor, J. and Takeoka, N. (2022). Optimal Discounting. *Econometrica*, 90(2):585–623.

Noor, J. and Takeoka, N. (2023). Constrained Optimal Discounting.

Onay, S. and Öncüler, A. (2007). Intertemporal choice under timing risk: An experimental approach. *Journal of Risk and Uncertainty*, 34(2):99–121.

Read, D., Olivola, C. Y., and Hardisty, D. J. (2017). The Value of Nothing: Asymmetric Attention to Opportunity Costs Drives Intertemporal Decision Making. *Management Science*, 63(12):4277–4297.

Takeuchi, K. (2011). Non-parametric test of time consistency: Present bias and future bias. *Games and Economic Behavior*, 71(2):456–478.