An Attentional Model of Time Discounting

Zark Zijian Wang

June 03, 2024

1 Introduction

decision maker (DM)

Kullback–Leibler (KL) divergence (also called relative entropy)

hard attention

information avoidance

endogenous time preferences

optimal expectation

we present an axiomatic characterization of AAD with the optimal discounting framework

2 Model Setting

Assume time is discrete. Let $s_{0\rightarrow T} \equiv [s_0, s_1, ..., s_T]$ denote a reward sequence that starts delivering rewards at period 0 and ends at period T. At each period t of $s_{0\rightarrow T}$, 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, i.e. we set $s_t \in \mathbb{R}_{\geq 0}$ and $1 \leq T < \infty$. 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 $s_{0\rightarrow T}$ 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 $s_{0\rightarrow T}$ is defined as $U(s_{0\rightarrow T}) \equiv \sum_{t=0}^{T} w_t u(s_t)$, where $u(s_t)$ is the instantaneous utility of receiving s_t, and w_t is the decision weight (sometimes called discount factors) assigned to s_t.

The determination of w_t is central to this paper. We believe that, due to the DM's limited attention and demand for information, the DM tends to overweight the large rewards and underweight the small rewards within the sequence. Specifically, we suggest w_t follow a (generalized) softmax function. We define any decision weight in this style as an *attention-adjusted discount* factors (AAD), as in Definition 1.

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

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

where $d_t > 0$, $\lambda > 0$, $u(.)$ is the utility function.

In intuition, how Definition 1 reflects the role of attention in valuating reward sequences can be explained with four points. First, each reward in a sequence could be viewed as an information source and we assume the DM allocates limited information-processing resources across those information sources. The AADs capture this notion by normalizing the discount factors, i.e. fixing the sum of w_t at 1. As a result, increasing the decision weight of one reward would reduce the decision weights of other rewards in the sequence, implying that focusing on one reward would make DM insensitive to the values of other rewards. Meanwhile, when there are more rewards in the sequence, DM needs to split attention across a wider range to process each of them, which may reduce the attention to, or decision weight of, each individual reward.

Second, w_t is strictly increasing with s_t, indicating that DM would pay more attention to larger rewards. This is consistent with many empirical studies that suggest people tend to pay more attention to information associated with larger rewards. For instance, people perform a "value-driven attentional capture" effect in visual search (Della Libera and Chelazzi, 2009; Hickey et al., 2010; Anderson et al., 2011; Chelazzi et al., 2013; Jahfari and Theeuwes, 2017). In one study (Hickey et al., 2010), researchers recruit participants to do a series of visual search trials, in each of which participants earn a reward after detecting a target object from distractors. If a target object is associated with a large reward in previous trials, it can naturally capture more attention. Therefore, in the next trial, presenting the object as a distractor slows down the target detection.[1] In addition, in financial decision making, investors usually perform an ostrich effect (Galai and Sade, 2006; Karlsson et al., 2009). One relevant evidence is that stock traders log in their brokerage accounts less frequently after market declines (Sicherman et al., 2016).

Third, w_t is "anchored" in a reference weight d_t. For a certain sequence of rewards, d_t could denote the initial weight that the DM would assign to a reward delivered at period t without knowing its realization. The determination of d_t is mediated by the difficulty to mentally represent a future event (Trope and Liberman, 2003) and the frequency of time delays in a global context (Stewart et al., 2006). The constraint on the deviation between w_t and d_t indicates that reallocating attention or acquiring new information is costly. The deviation of w_t from d_t depends on parameter λ, which as we discuss in the next section, can reflect the unit cost of information acquisition. A large λ implies a low learning rate and a high cognitive cost in adapting the decision weights to the local context.

Fourth, we adopt the idea of Gottlieb (2012) and Gottlieb et al. (2013) that attention can be understood as an active information-sampling mechanism which selects information based on the perceived utility of information. For intertemporal choices, we assume the DM would selectively sample value information from each reward (information source) when processing a reward sequence, and the AAD can represent an approximately optimal sampling strategy. Note that the AADs follow a softmax function. Matějka and McKay (2015) and Maćkowiak et al. (2023) claim that if a behavioral strategy conforms to this type of function, then it can

[1] Some scholars may classify attention into two categories: "bottom-up control" and "top-down control". However, the evidence about value-driven attentional capture does not fall into either of these categories. Thus, in this paper, we do not describe attention with this dichotomy. Instead, we view attention as a mechanism that seeks to maximize the utility of information.

be interpreted as a solution to some optimization problem under information constraints.

3 Interpretation

3.1 Information Maximizing Exploration

In this section, we provide two approaches to characterize AAD: the first is based on information maximizing exploration, and the second is based on optimal discounting. These approaches are closely related to the idea proposed by Gottlieb (2012), Gottlieb et al. (2013) and Sharot and Sunstein (2020), that people tend to pay attention to information with high *instrumental utility* (help identifying the optimal action), *cognitive utility* (satisfying curiosity), or *hedonic utility* (inducing positive feelings). It is worth mentioning that the well-known rational inattention theories are grounded in the instrumental utility of information.[2] Instead, in this paper, we draw on the cognitive and hedonic utility of information to build our theory of time discounting. Our first approach to characterizing AAD is relevant to the cognitive utility: the DM's information acquisition process is curiosity-driven. The model setting of this approach, similar with Gottlieb (2012) and Gottlieb et al. (2013), is based on a reinforcement learning framework. Specifically, we assume the DM seeks to maximize the information gain with a commonly-used exploration strategy. Our second approach is relevant to the hedonic utility: the DM consider the feelings of multiple selves and seeks to maximize their total utility under some cognitive cost. The theoretical background for the second approach is Noor and Takeoka (2022, 2024). We describe the first approach in this subsection and the second approach in Section 3.2.

For the information maximizing exploration approach, we assume that before having any information of a reward sequence, the DM perceives it has no value. Then, each reward in the sequence $s_{0\to T}$ is processed as an individual information source. The DM engages her attention to actively sample signals at each information source and update her belief about the sequence value accordingly. The signals are nosiy. For any $t \in \{0, 1, \ldots, T\}$, the signal

[2] The rational inattention theory assumes the DM learns information about different options in order to find the best option. For details, see Sims (2003), Matějka and McKay (2015), and Maćkowiak et al. (2023).

sampled at information source s_t could be represented by $x_t = u(s_t) + \epsilon_t$, where each ϵ_t is i.i.d. and $\epsilon_t \sim N(0, \sigma_\epsilon^2)$. The sampling weight for information source s_t is denoted by w_t. We assume the function $u(s_t)$ is strictly increasing with s_t and for any $s_t > 0$, we set $u(s_t) > 0$.

The DM's belief about the sequence value $U(s_{0\to T})$ is updated as follows. At first, she holds a prior U_0, and given she perceives no value from the reward sequence, the prior could be represented by $U_0 \sim N(0, \sigma^2)$. Second, she draws a series of signals at each information source s_t. Note we define $U(s_{0\to T})$ as a weighted mean of instantaneous utilities, i.e. $U(s_{0\to T}) = \sum_{t=0}^{T} w_t u(s_t)$. Let $\bar{x}$ denote the mean sample signal and U denote a realization of $U(s_{0\to T})$. If there are k signals being sampled in total, we should have $\bar{x}|U, \sigma_\epsilon \sim N(U, \frac{\sigma_\epsilon^2}{k})$. Third, she uses the sampled signals to infer $U(s_{0\to T})$ in a Bayesian fashion. Let U_k denote the valuer's posterior about the sequence value after receiving k signals. According to the Bayes' rule, we have $U_k \sim N(\mu_k, \sigma_k^2)$ and

$$\mu_k = \frac{k^2\sigma_\epsilon^{-2}}{\sigma^{-2} + k^2\sigma_\epsilon^{-2}}\bar{x} \qquad , \qquad \sigma_k^2 = \frac{1}{\sigma^{-2} + k^2\sigma_\epsilon^{-2}}$$

We assume the DM takes μ_k as the valuation of reward sequence. It is clear that as $k \to \infty$, the sequence value will converge to the mean sample signal, i.e. $\mu_k \to \bar{x}$.

The DM's goal of sampling signals is to maximize her information gain. The information gain is defined as the KL divergence from the prior U_0 to the posterior U_k. In intuition, the KL divergence provides a measure for distance between distributions. As the DM acquires more information about $s_{0\to T}$, her posterior belief should move farther away from the prior. We let $p_0(U)$ and $p_k(U)$ denote the probability density functions of U_0 and U_k. Then, the information gain is

$$\begin{aligned} D_{KL}(U_k||U_0) &= \int_{-\infty}^{\infty} p_k(U)\log\left(p_k(U)/p_0(U)\right)dU \\ &= \frac{\sigma_k^2 + \mu_k^2}{2\sigma^2} - \log\left(\frac{\sigma_k}{\sigma}\right) - \frac{1}{2} \end{aligned} \tag{2}$$

Notably, in Equation (2), σ_k depends only on sample size k and μ_k is proportional to $\bar{x}$. Therefore, the problem of maximizing $D_{KL}(U_k||U_0)$ could be reduced to maximizing $\bar{x}$ (as each $u(s_t)$ is non-negative). The reason is that, drawing more samples can always increase the precision of the DM's estimate about $U(s_{0\to T})$, and a larger $\bar{x}$ implies more "surprises" in comparison to the DM's initial perception that $s_{0\to T}$ contains no value.

Maximizing the mean sample signal $\bar{x}$ under a limited sample size k is actually a multi-armed bandit problem (Sutton and Barto, 2018, Ch.2). On the one hand, the DM wants to draw more samples at information sources that are known to produce greater value signals (exploit). On the other hand, she wants to learn some value information from other information sources (explore). We assume the DM would take a softmax exploration strategy to solve this problem. That is,

$$w_t \propto d_t e^{\bar{x}_t/\lambda}$$

where $\bar{x}_t$ is the mean sample signal generated by information source s_t so far, $1/\lambda$ is the learning rate, and d_t is the initial sampling weight for s_t.[3] Note $\bar{x}_t$ cannot be calculated without doing simulations under a certain σ_ϵ. For researchers, modelling an intertemporal choice in this way requires conducting a series of simulations and then calibrating σ_ϵ for every choiceable option, which could be computationally expensive. Fortunately, according to the weak law of large numbers, as the sample size k gets larger, $\bar{x}_t$ is more likely to fall into a neighborhood of $u(s_t)$. Thus, the AAD which assumes $w_t \propto d_t e^{u(s_t)/\lambda}$ could be viewed as a proper approximation to the softmax exploration strategy.

Those who familiar with reinforcement learning algorithms may notice that here $u(s_t)$ is a special case of action-value function (assuming that the learner only cares about the value of current reward in each draw of sample). The AAD thus can be viewed as a specific version of the soft Q-learning or policy gradient method for solving the given multi-armed bandit problem (Haarnoja et al., 2017; Schulman et al., 2017). Such methods are widely used (and sample-efficient) in reinforcement learning. Moreover, one may argue that the applicability of softmax exploration strategy is subject to our model assumptions. Under alternative assumptions, the strategy may not be ideal. We acknowledge this limitation and suggest that researchers interested in modifying our model consider different objective functions or different families of noises. For example, if the DM aims to minimize the regret rather than

[3] Classic softmax strategy assumes the initial probability of taking an action follows an uniform distribution. We relax this assumption by importing d_t, so that the DM can hold an initial preference of sampling over the dated rewards.

maximizing $\bar{x}$, the softmax exploration strategy can produce suboptimal actions and one remedy is to use the Gumbel–softmax strategy (Cesa-Bianchi et al., 2017). If noises $\epsilon_0, ..., \epsilon_T$ do not follow an i.i.d normal distribution, the information gain $D_{KL}(U_k||U_0)$ may be complex to compute, thus one can use its variational bound as the objective (Houthooft et al., 2016). Compared to these complex settings, the model setting in this subsection aims to provides a simple benchmark for understanding the role of attention in mental valuation of a reward sequence.

Two strands of literature can justify the information maximizing approach to characterizing AAD. First, our assumption that the DM updates decision weights toward a greater $D_{KL}(U_k||U_0)$ is consistent with the finding that Bayesian surprises attract attention in vision research (Itti and Baldi, 2009). Second, the softmax exploration strategy is widely used by neuroscientists in fitting human actions in reinforcement learning tasks (Daw et al., 2006; Niv et al., 2012; FitzGerald et al., 2012; Niv et al., 2015; Leong et al., 2017). For instance, Daw et al. (2006) find the softmax strategy can characterize humans' exploration behavior better than other classic strategies (e.g. ϵ-greedy). Besides, Collins and Frank (2014) show that models based on the soft strategy exhibit a good performance in explaining behaviors of the striatal dopaminergic system (which is central in brain's sensation of pleasure and learning of rewarding actions) in reinforcement learning.

3.2 Optimal Discounting

The second approach to characterize AAD is based on the optimal discounting model (Noor and Takeoka, 2022, 2024). In one version of this model, the authors assume that the DM has a limited capacity of attention (or in their term, "empathy"), and before evaluating a reward sequence $s_{0\to T}$, she naturally focuses on the current period. The instantaneous utility $u(s_t)$ represents the well-being that the DM's self of period t can obtain from the reward sequence. For valuating $s_{0\to T}$, the DM needs to split attention over T time periods to consider the feeling of each self. This re-allocation of attention is cognitive costly. The DM seeks to find a balance between improving the overall well-being of multiple selves and reducing the incurred cognitive cost. Noor and Takeoka (2022, 2024) specify an optimization problem to capture this decision. In this paper, we adopt a variant of their original model. The formal definition of the optimal discounting problem is given by Definition 2. [4]

Definition 2: *Given reward sequence $s_{0\to T} = [s_0, ..., s_T]$, the following optimization problem is called an optimal discounting problem for $s_{0\to T}$:*

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

where $M > 0$, $C(\mathcal{W}) \geq 0$. For any $t \in \{0, 1, \ldots, T\}$, $u(s_t) < \infty$. $C(\mathcal{W})$ *is the cognitive cost function and is constituted by time-separable costs, i.e.* $C(\mathcal{W}) = \sum_{t=0}^{T} f_t(w_t)$, *where for all $w_t \in (0, 1)$, $f_t(w_t)$ is differentiable. $f_t'(w_t)$ is continuous and* stric*tly increasing, and* $\lim_{w_t \to 0} f_t'(w_t) = -\infty$.

Here w_t reflects the attention paid to consider the feeling of t-period self. The DM's objective function is the attention-weighted sum of utilities obtained by the multiple selves minus the cognitive cost of attention re-allocation. As is illustrated by Noor and Takeoka (2022, 2024), a key feature of this model is that decision weight w_t is increasing with s_t, indicating the DM tends to pay more attention to larger rewards. Moreover, it is easy to validate that if the following three conditions are satisfied, the solution to the optimal discounting problem will take an AAD form:

(i) The constraint on sum of decision weights is always tight. That is, $\sum_{t=0}^{T} w_t = M$. Without loss of generality, we can set $M = 1$.

[4] There are three differences between Definition 2 and the original optimal discounting model (Noor and Takeoka, 2022, 2024). First, in our setting, shifting attention to future rewards may reduce the attention to the current reward, while this would never happen in Noor and Takeoka (2022, 2024). Second, the original model assumes $f_t'(w_t)$ must be continuous at 0 and w_t must be no larger than 1. We relax these assumptions since neither $w_t = 0$ nor $w_t \geq 1$ is included our solutions. Third, the original model assumes that $f_t'(w_t)$ is left-continuous in $[0, 1]$, and there exist $\underline{w}, \bar{w} \in [0, 1]$ such that $f_t'(w_t) = 0$ when $w_t \leq \underline{w}$, $f_t'(w_t) = \infty$ when $w_t \geq \bar{w}$, and $f_t'(w_t)$ is strictly increasing when $w_t \in [\underline{w}, \bar{w}]$. We simplify this assumption by setting $f_t'(w_t)$ is continuous strictly increasing in $(0, 1)$. For behavior of $f_t'(w_t)$ near the border of $[0, 1]$, we set $\lim_{w_t \to 0} f_t'(w_t) = -\infty$ instead.

(ii) There exists a realization of decision weights $\mathcal{D} = [d_0, ..., d_T]$ such that $d_t > 0$ for all $t \in \{0, \ldots, T\}$ and the cognitive cost is proportional to the KL divergence from $\mathcal{D}$ to the DM's strategy $\mathcal{W}$ where applicable. That is, $C(\mathcal{W}) = \lambda \cdot D_{KL}(\mathcal{W}||\mathcal{D})$, where $\lambda > 0$.

Here d_t sets a reference for determining the decision weight w_t, the parameter λ indicates how costly the attention re-allocation process is, and $D_{KL}(\mathcal{W}||\mathcal{D}) = \sum_{t=0}^{T} w_t \log(\frac{w_t}{d_t})$. The solution to the optimal discounting problem under condition (i)-(ii) can be derived in the same way as Theorem 1 in Matějka and McKay (2015). Note this solution is equivalent to that of a bounded rationality model: assuming the DM wants to find a $\mathcal{W}$ that maximizes $\sum_{t=0}^{T} w_t u(s_t)$ but can only search for solutions within a KL neighborhood of $\mathcal{D}$. Related models can be found in Todorov (2009).

We interpret the implications of condition (i)-(ii) with behavioral axioms. Note if each s_t is an independent option and $\mathcal{W}$ simply represents the DM's choice strategy across options, then these condition can be characterized by rational inattention theories, e.g. Caplin et al. (2022). However, here $\mathcal{W}$ is a component of sequence value $U(s_{0\to T})$, and the DM is assumed to choose the option with highest sequence value. Thus, the behavioral implications of condition (i)-(ii) should be derived in different ways. To illustrate, let $\succsim$ denote the preference relation between two reward sequences.[5] For any reward sequence $s_{0\to T} = [s_0, ..., s_T]$, we define $s_{0\to t} = [s_0, ..., s_t]$ as a sub-sequence of it, where $1 \leq t \leq T$.[6] We first introduce two axioms for $\succsim$:

Axiom 1: $\succsim$ *has the following properties:*

(a) *(complete order)* $\succsim$ *is complete and transitive.*

(b) *(continuity) For any reward sequences* s, s' *and reward* $c \in \mathbb{R}_{\geq 0}$*, the sets* $\{\alpha \in (0,1) | \alpha \cdot s + (1-\alpha) \cdot c \succsim s'\}$ *and* $\{\alpha \in (0,1) | s' \succsim \alpha \cdot s + (1-\alpha) \cdot c\}$ *are closed.*

(c) *(state-independent) For any reward sequences* s, s' *and reward* $c \in \mathbb{R}_{\geq 0}$*,* $s \succsim s'$ *implies for any* $\alpha \in (0,1)$*,* $\alpha \cdot s + (1-\alpha) \cdot c \sim \alpha \cdot s' + (1-\alpha) \cdot c$*.*

[5] If $a \succsim b$ and $b \succsim a$, we say $a \sim b$ ("a is the same good as b"). If $a \succsim b$ does not hold, we say $b \succ a$ ("b is better than a"). $\succsim$ can also characterize the preference relation between single rewards as the single rewards can be viewed as one-period sequences.

[6] Notably, every sub-sequence starts with period 0.

(d) *(reduction of compound alternatives) For any reward sequences s, s', y and rewards $c_1, c_2 \in \mathbb{R}_{\geq 0}$, if there exist $\alpha, \beta \in (0,1)$ such that $s \sim \alpha \cdot y + (1-\alpha) \cdot c_1$, then $s' \sim \beta \cdot s + (1-\beta) \cdot c_2$ implies $s' \sim \beta\alpha \cdot y + \beta(1-\alpha) \cdot c_1 + (1-\beta) \cdot c_2$.*

Axiom 2: *For any $s_{0\to T}$ and any $\alpha_1, \alpha_2 \in (0,1)$, there exists $c \in \mathbb{R}_{\geq 0}$ such that $\alpha_1 \cdot s_{0\to T-1} + \alpha_2 \cdot s_T \sim c$.*

The two axioms are almost standard in decision theories. The assumption of complete order implies preferences between reward sequences can be characterized by an utility function. Continuity and state-independence ensure that in a stochastic setting where the DM can receive one reward sequence under some states and receive a single reward under other states, her preference can be characterized by expected utility (Herstein and Milnor, 1953). Reduction of compound alternatives ensures that the DM's valuation on a reward sequence is constant across states. Axiom 2 is an extension of the Constant-Equivalence assumption in Bleichrodt et al. (2008). It implies there always exists a constant that can represent the value of a linear combination of sub-sequence $s_{0\to T}$ and the end-period reward s_T, as long as the weights lie in $(0,1)$.

For a given $s_{0\to T}$, the optimal discounting model can generate a sequence of decision weights $[w_0, ..., w_T]$. Furthermore, the model assumes the DM's preference for $s_{0\to T}$ can be characterized by the preference for $w_0 \cdot s_0 + w_1 \cdot s_1 + ... + w_T \cdot s_T$. We use Definition 3 to capture this assumption.[7]

Definition 3: *Given reward sequence $s_{0\to T} = [s_0, ..., s_T]$ and $s'_{0\to T'} = [s'_0, ..., s'_{T'}]$, the preference relation $\succsim$ has an optimal discounting representation if*

$$s_{0\to T} \succsim s'_{0\to T'} \iff \sum_{t=0}^{T} w_t \cdot s_t \succsim \sum_{t=0}^{T'} w'_t \cdot s'_t$$

where $\{w_t\}_{t=0}^{T}$ and $\{w'_t\}_{t=0}^{T'}$ are solutions to the optimal discounting problems for $s_{0\to T}$ and $s'_{0\to T'}$ respectively.

Furthermore, if Definition 3 is satisfied and $\{w_t\}_{t=0}^{T}$ as well as $\{w'_t\}_{t=0}^{T'}$ takes the AAD

[7] Noor and Takeoka (2022) refer the "optimal discounting representation" in Definition 3 as Costly Empathy representation.

form, we say $\succsim$ has an *AAD representation.* Now we specify two behavioral axioms that are key to characterize the AAD functions.

Axiom 3 (sequential outcome-betweenness): *For any $s_{0\to T}$, there exists $\alpha \in (0,1)$ such that $s_{0\to T} \sim \alpha \cdot s_{0\to T-1} + (1-\alpha) \cdot s_T$.*

Axiom 4 (sequential bracket-independence): *Suppose $T \geq 2$. For any $s_{0\to T}$, if there exist $\alpha_1, \alpha_2, \beta_0, \beta_1, \beta_2 \in (0,1)$ such that $s_{0\to T} \sim \alpha_1 \cdot s_{0\to T-1} + \alpha_2 \cdot s_T$ and $s_{0\to T} \sim \beta_0 \cdot s_{0\to T-2} + \beta_1 \cdot s_{T-1} + \beta_2 \cdot s_T$, then we must have $\alpha_2 = \beta_2$.*

Axiom 3 implies that for a reward sequence $s_{0\to T-1}$, if we add a new reward s_T at the end of the sequence, then the value of the new sequence should lie between the original sequence $s_{0\to T-1}$ and the newly added reward s_T. This characterizes condition (i), i.e. the sum of decision weights is bounded tightly at 1. Notably, Axiom 3 is consistent with the empirical evidence about *violation of dominance* (Scholten and Read, 2014; Jiang et al., 2017) in intertemporal choice. Suppose the DM is indifferent between a small-sooner reward (SS) "receive \$75 today" and a large-later reward (LL) "receive \$100 in 52 weeks". Scholten and Read (2014) find when we add a tiny reward after the payment in SS, e.g. changing SS to "receive \$75 today and \$5 in 52 weeks", the DM would be more likely to prefer LL over SS. Jiang et al. (2017) find the same effect can apply to LL. That is, if we add a tiny reward after the payment in LL, e.g. changing LL to "receive \$100 in 52 weeks and \$5 in 53 weeks", the DM may be more likely to prefer SS over LL.

Axiom 4 implies that no matter how the DM brackets the rewards into sub-sequences (or how the sub-sequences get further decomposed), the decision weights for rewards outside them should not be affected. Specifically, suppose we decompose reward sequence $s_{0\to T}$ and find its value is equivalent to a linear combination of $s_{0\to T-1}$ and s_T. We also can further decompose $s_{0\to T-1}$ to a linear combination of $s_{0\to T-2}$ and s_{T-1}. But no matter how we operate, as long as the decomposition is carried out inside $s_{0\to T-1}$, the weight of s_T in the valuation of $s_{0\to T}$ will always remain the same. This axiom is an analog to independence of irrelevant alternatives in discrete choice problems (which is a key feature of softmax choice function).

We show in Proposition 1 that the optimal discounting model plus Axiom 1-4 can exactly produce AAD.

Proposition 1: *Suppose $\succsim$ has an optimal discounting representation, then it has an AAD representation if and only if it satisfies Axiom 1-4.*

The proof of Proposition 1 is in Appendix A.

4 Implications for Decision Making

hidden-zero effect

common-difference effect

concavity of discount function

S-shaped value function

Intertemporal correlation aversion

Learning and inconsistent planning

`{# {r child='Implications.Rmd'}`

5 Discussion

5.1 Relation to Other Models of Intertemporal Choices

The theory most similar to AAD is the salience theory (Bordalo et al., 2012, 2013, 2020).

rational inattention

focus-weighted utility

bayesian updating and discounting

optimal precision

Relation with money/delay trade-off

5.2 Other relevant phenomena

5.3 Limitation

attention biases learning: learning rate is high for attended reward

6 Conclusion

Reference

Anderson, B. A., Laurent, P. A., and Yantis, S. (2011). Value-driven attentional capture. *Proceedings of the National Academy of Sciences*, 108(25):10367–10371.

Bleichrodt, H., Rohde, K. I., and Wakker, P. P. (2008). Koopmans' constant discounting for intertemporal choice: A simplification and a generalization. *Journal of Mathematical Psychology*, 52(6):341–347.

Bordalo, P., Gennaioli, N., and Shleifer, A. (2012). Salience theory of choice under risk. *The Quarterly journal of economics*, 127(3):1243–1285.

Bordalo, P., Gennaioli, N., and Shleifer, A. (2013). Salience and consumer choice. *Journal of Political Economy*, 121(5):803–843.

Bordalo, P., Gennaioli, N., and Shleifer, A. (2020). Memory, attention, and choice. *The Quarterly journal of economics*, 135(3):1399–1442.

Caplin, A., Dean, M., and Leahy, J. (2022). Rationally inattentive behavior: Characterizing and generalizing shannon entropy. *Journal of Political Economy*, 130(6):1676–1715.

Cesa-Bianchi, N., Gentile, C., Lugosi, G., and Neu, G. (2017). Boltzmann exploration done right. *Advances in neural information processing systems*, 30.

Chelazzi, L., Perlato, A., Santandrea, E., and Della Libera, C. (2013). Rewards teach visual selective attention. *Vision research*, 85:58–72.

Collins, A. G. and Frank, M. J. (2014). Opponent actor learning (opal): modeling interactive effects of striatal dopamine on reinforcement learning and choice incentive. *Psychological review*, 121(3):337.

Daw, N. D., O'doherty, J. P., Dayan, P., Seymour, B., and Dolan, R. J. (2006). Cortical substrates for exploratory decisions in humans. *Nature*, 441(7095):876–879.

Della Libera, C. and Chelazzi, L. (2009). Learning to attend and to ignore is a matter of gains and losses. *Psychological science*, 20(6):778–784.

FitzGerald, T. H., Friston, K. J., and Dolan, R. J. (2012). Action-specific value signals in reward-related regions of the human brain. *Journal of Neuroscience*, 32(46):16417–16423.

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

Gottlieb, J. (2012). Attention, learning, and the value of information. *Neuron*, 76(2):281–295.

Gottlieb, J., Oudeyer, P.-Y., Lopes, M., and Baranes, A. (2013). Information-seeking, curiosity, and attention: computational and neural mechanisms. *Trends in cognitive sciences*, 17(11):585–593.

Haarnoja, T., Tang, H., Abbeel, P., and Levine, S. (2017). Reinforcement learning with deep energy-based policies. In *International conference on machine learning*, pages 1352–1361. PMLR.

Herstein, I. N. and Milnor, J. (1953). An axiomatic approach to measurable utility. *Econometrica, Journal of the Econometric Society*, pages 291–297.

Hickey, C., Chelazzi, L., and Theeuwes, J. (2010). Reward changes salience in human vision via the anterior cingulate. *Journal of Neuroscience*, 30(33):11096–11103.

Houthooft, R., Chen, X., Duan, Y., Schulman, J., De Turck, F., and Abbeel, P. (2016). Vime: Variational information maximizing exploration. *Advances in neural information processing systems*, 29.

Itti, L. and Baldi, P. (2009). Bayesian surprise attracts human attention. *Vision research*, 49(10):1295–1306.

Jahfari, S. and Theeuwes, J. (2017). Sensitivity to value-driven attention is predicted by how we learn from value. *Psychonomic bulletin & review*, 24(2):408–415.

Jiang, C.-M., Sun, H.-M., Zhu, L.-F., Zhao, L., Liu, H.-Z., and Sun, H.-Y. (2017). Better is worse, worse is better: Reexamination of violations of dominance in intertemporal choice. *Judgment and Decision Making*, 12(3):253–259.

Karlsson, N., Loewenstein, G., and Seppi, D. (2009). The ostrich effect: Selective attention to information. *Journal of Risk and uncertainty*, 38:95–115.

Leong, Y. C., Radulescu, A., Daniel, R., DeWoskin, V., and Niv, Y. (2017). Dynamic interaction between reinforcement learning and attention in multidimensional environments. *Neuron*, 93(2):451–463.

Maćkowiak, B., Matějka, F., and Wiederholt, M. (2023). Rational inattention: A review. *Journal of Economic Literature*, 61(1):226–273.

Matějka, F. and McKay, A. (2015). Rational inattention to discrete choices: A new foundation for the multinomial logit model. *American Economic Review*, 105(1):272–298.

Niv, Y., Daniel, R., Geana, A., Gershman, S. J., Leong, Y. C., Radulescu, A., and Wilson, R. C. (2015). Reinforcement learning in multidimensional environments relies on attention mechanisms. *Journal of Neuroscience*, 35(21):8145–8157.

Niv, Y., Edlund, J. A., Dayan, P., and O'Doherty, J. P. (2012). Neural prediction errors reveal a risk-sensitive reinforcement-learning process in the human brain. *Journal of Neuroscience*, 32(2):551–562.

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

Noor, J. and Takeoka, N. (2024). Constrained optimal discounting. *Available at SSRN 4703748*.

Scholten, M. and Read, D. (2014). Better is worse, worse is better: Violations of dominance in intertemporal choice. *Decision*, 1(3):215.

Schulman, J., Chen, X., and Abbeel, P. (2017). Equivalence between policy gradients and soft q-learning. *arXiv preprint arXiv:1704.06440*.

Sharot, T. and Sunstein, C. R. (2020). How people decide what they want to know. *Nature Human Behaviour*, 4(1):14–19.

Sicherman, N., Loewenstein, G., Seppi, D. J., and Utkus, S. P. (2016). Financial attention. *The Review of Financial Studies*, 29(4):863–897.

Sims, C. A. (2003). Implications of rational inattention. *Journal of monetary Economics*, 50(3):665–690.

Stewart, N., Chater, N., and Brown, G. D. (2006). Decision by sampling. *Cognitive psychology*, 53(1):1–26.

Sutton, R. S. and Barto, A. G. (2018). *Reinforcement learning: An introduction*. MIT press.

Todorov, E. (2009). Efficient computation of optimal actions. *Proceedings of the national academy of sciences*, 106(28):11478–11483.

Trope, Y. and Liberman, N. (2003). Temporal construal. *Psychological review*, 110(3):403.

7 Appendix

7.1 A. Proof of Proposition 1

The sufficiency is easy to validate. We present the proof of necessity here. That is, if $\succsim$ has an optimal discounting representation and satisfies Axiom 1-4, then it has an AAD representation.

Lemma 1: *If Axiom 1 and 3 hold, for any $s_{0\to T}$, there exist $w_0, w_1, \ldots, w_T > 0$ such that $s_{0\to T} \sim w_0 \cdot s_0 + ... + w_T \cdot s_T$, where $\sum_{t=0}^{T} w_t = 1$.*

Proof: If $T = 1$, Lemma 1 is a direct application of Axiom 3. If $T \geq 2$, for any $2 \leq t \leq T$, there should exist $\alpha_t \in (0, 1)$ such that $s_{0\to t} \sim \alpha_t \cdot s_{0\to t-1} + (1-\alpha_t) \cdot s_t$. By state-independence and reduction of compound alternatives, we can recursively apply this equivalence relation as follows:

$$\begin{aligned} s_{0\to T} &\sim \alpha_{T-1} \cdot s_{0\to T-1} + (1 - \alpha_{T-1}) \cdot s_T \\ &\sim \alpha_{T-1}\alpha_{T-2} \cdot s_{0\to T-2} + \alpha_{T-1}(1 - \alpha_{T-2}) \cdot s_{T-1} + (1 - \alpha_{T-1}) \cdot s_T \\ &\sim ... \\ &\sim w_0 \cdot s_0 + w_1 \cdot s_1 + ... + w_T \cdot s_T \end{aligned}$$

where $w_0 = \prod_{t=0}^{T-1} \alpha_t$, $w_T = 1 - \alpha_{T-1}$, and for $0 < t < T$, $w_t = (1 - \alpha_{t-1}) \prod_{\tau=t}^{T-1} \alpha_\tau$. It is easy to show the sum of $w_0, \ldots, w_T$ is equal to 1. *QED.*

Therefore, if Axiom 1 and 3 hold, for any reward sequence $s_{0\to T}$, we can always find a convex combination of all its elements, such that the DM is indifferent between the reward sequence and this convex combination. If $s_{0\to T}$ is a constant sequence, i.e. all its elements are constant, then we can directly assume $\mathcal{W}$ is AAD-style. Henceforth, we discuss whether AAD can also apply to non-constant sequences.

By Lemma 2, we show adding a new reward to the end of $s_{0\to T}$ has no impact on the relative decision weights of rewards in the original reward sequence.

Lemma 2: *For any $s_{0\to T+1}$, if $s_{0\to T} \sim \sum_{t=0}^{T} w_t \cdot s_t$ and $s_{0\to T+1} \sim \sum_{t=0}^{T+1} w'_t \cdot s_t$, where*

$w_t, w'_t > 0$ and $\sum_{t=0}^{T} w_t = 1$, $\sum_{t=0}^{T+1} w'_t = 1$, *then when Axiom 1-4 hold, we can obtain* $\frac{w'_0}{w_0} = \frac{w'_1}{w_1} = \ldots = \frac{w'_T}{w_T}$.

Proof: According to Axiom 3, for any $s_{0\to T+1}$, there exist $\alpha, \zeta \in (0,1)$ such that

$$\begin{aligned} s_{0\to T} &\sim \alpha \cdot s_{0\to T-1} + (1-\alpha) \cdot s_T \\ s_{0\to T+1} &\sim \zeta \cdot s_{0\to T} + (1-\zeta) \cdot s_{T+1} \end{aligned} \tag{A1}$$

On the other hand, we drawn on Lemma 1 and set

$$s_{0\to T+1} \sim \beta_0 \cdot s_{0\to T-1} + \beta_1 \cdot s_T + (1-\beta_0-\beta_1) \cdot s_{T+1} \tag{A2}$$

where $\beta_0, \beta_1 > 0$. According to Axiom 4, $1-\zeta = 1-\beta_0-\beta_1$. So, $\beta_1 = \zeta - \beta_0$. This also implies $\zeta > \beta_0$.

According to Axiom 2, we suppose there exists a reward sequence s such that $s \sim \frac{\beta_0}{\zeta} \cdot s_{0\to T-1} + (1-\frac{\beta_0}{\zeta}) \cdot s_1$. By Equation (A2) and reduction of compound alternatives, we have $s_{0\to T+1} \sim \zeta \cdot s + (1-\zeta) \cdot s_{T+1}$. Combining Equation (A2) with the second line of Equation (A1) and applying transitivity and state-independence, we obtain $s_{0\to T} \sim \frac{\beta_0}{\zeta} \cdot s_{0\to T-1} + (1-\frac{\beta_0}{\zeta}) \cdot s_1$.

We aim to prove that for any $s_{0\to T+1}$, we can obtain $\alpha = \frac{\beta_0}{\zeta}$. To do this, we first assume (without loss of generality) that $\alpha > \frac{\beta_0}{\zeta}$.

Consider the case that $s_{0\to T-1} \succ s_T$. By state-independence, for any $c \in \mathbb{R}_{\geq 0}$, we have $(\alpha - \frac{\beta_0}{\zeta}) \cdot s_{0\to T-1} + (1-\alpha+\frac{\beta_0}{\zeta}) \cdot c \succ (\alpha - \frac{\beta_0}{\zeta}) \cdot s_T + (1-\alpha+\frac{\beta_0}{\zeta}) \cdot c$. By Axiom 2, there exists $z \in \mathbb{R}_{\geq 0}$ such that $(1-\alpha) \cdot s_T + \frac{\beta_0}{\zeta} \cdot s_{0\to T-1} \sim z$. Given c is arbitrary, we set $(1-\alpha+\frac{\beta_0}{\zeta}) \cdot c \sim z$. By reduction of compound alternatives, we can derive that

$$(\alpha - \frac{\beta_0}{\zeta}) \cdot s_{0\to T-1} + (1-\alpha) \cdot s_T + \frac{\beta_0}{\zeta} \cdot s_{0\to T-1} \succ (\alpha - \frac{\beta_0}{\zeta}) \cdot s_T + (1-\alpha) \cdot s_T + \frac{\beta_0}{\zeta} \cdot s_{0\to T-1}$$

where the LHS can be rearranged to $\alpha \cdot s_{0\to T-1} + (1-\alpha) \cdot s_T$, and the RHS can be rearranged to $\frac{\beta_0}{\zeta} \cdot s_{0\to T-1} + (1-\frac{\beta_0}{\zeta}) \cdot s_1$. They both should be indifferent from $s_{0\to T}$. This results in a contradiction. Similarly, in the case that $s_T \succ s_{0\to T-1}$, we can also derive a contradiction. Meanwhile, when $s_{0\to T} \sim s_T$, α and $\frac{\beta_0}{\zeta}$ can be any number within $(0,1)$. So, we can directly set $\alpha = \frac{\beta_0}{\zeta}$.

Thus, we have $\alpha = \frac{\beta_0}{\zeta}$ for any $s_{0\to T+1}$, which indicates $\frac{\beta_0}{\alpha} = \frac{\beta_1}{1-\alpha} = \zeta$. We can recursively apply this equality to any sub-sequence $s_{0\to t}$ $(t \leq T)$ of $s_{0\to T+1}$, so that the lemma will be proved. *QED.*

Now we move on to prove Proposition 1. The proof contains six steps.

First, we add the constraints $\sum_{t=0}^{T} w_t = 1$ and $w_t > 0$ to the optimal discounting problem for $s_{0\to T}$ so that the problem is compatible with Lemma 1. According to the FOC of its solution, for all $t = 0, 1, \ldots., T$, we have

$$f_t'(w_t) = u(s_t) + \theta \tag{A3}$$

where θ is the Lagrange multiplier. Given that $f_t'(w_t)$ is strictly increasing, w_t is increasing with $u(s_t) + \theta$. We define the solution as $w_t = \phi_t(u(s_t) + \theta)$.

Second, we add a new reward s_{T+1} to the end of $s_{0\to T}$ and apply Lemma 2 as a constraint to optimal discounting problem. Look at the optimal discounting problem for $s_{0\to T+1}$. For all $t \leq T$, the FOC of its solution will take the same form as Equation (A3). So, if importing s_{T+1} changes some w_t to w_t' ($w_t' \neq w_t$, where w_t is the solution to optimal discounting problem for $s_{0\to T}$), the only way is through changing the multiplier θ. Suppose importing s_{T+1} changes θ to $\theta - \Delta\theta$, we have $w_t' = \phi_t(u(s_t) + \theta - \Delta\theta)$.

By Lemma 2, we know $\frac{w_0}{w_0'} = \frac{w_1}{w_1'} = \ldots = \frac{w_T}{w_T'}$. In other words, for $t = 0, 1, \ldots, T$, we have $w_t \propto \phi_t(u(s_t) + \theta - \Delta\theta)$. We can rewrite w_t as

$$w_t = \frac{\phi_t(u(s_t) + \theta - \Delta\theta)}{\sum_{\tau=0}^{T} \phi_\tau(u(s_\tau) + \theta - \Delta\theta)} \tag{A4}$$

Third, we show that in $s_{0\to T}$, if we change each s_t to z_t such that $u(z_t) = u(s_t) + \Delta u$, the decision weights $w_0, \ldots, w_T$ will remain the same. Note $\sum_{t=0}^{T} \phi_t(u(s_t) + \theta) = 1$. It is clear that $\sum_{t=0}^{T} \phi_t(u(z_t) + \theta - \Delta u) = 1$. Suppose changing every s_t to z_t moves θ to θ' and $\theta' < \theta - \Delta u$. Then, we must have $\phi_t(u(z_t) + \theta') < \phi_t(u(z_t) + \theta - \Delta u)$ since $\phi_t(.)$ is strictly increasing. This results in $\sum_{t=0}^{T} \phi_t(u(z_t) + \theta') < 1$, which contradicts with the constraint that the sum of all decision weights is 1. The same contradiction can apply to the case that $\theta' > \theta - \Delta u$. Therefore, changing every s_t to z_t must move θ to $\theta - \Delta u$, and each w_t can only be moved to $\phi_t(u(z_t) + \theta - \Delta u)$, which is exactly the same as the original decision weight.

A natural corollary of this step is that, subtracting or adding a common number to all intantaneous utilities in a reward sequence has no effect on decision weights. What actually matters for determining the decision weights is the difference between instantaneous utilities. This indicates, for convenience, we can subtract or add an arbitrary number to the utility function.

In other words, for a given $s_{0\to T}$ and s_{T+1}, we can define a new utility function $v(.)$ such that $v(s_t) = u(s_t) + \theta - \Delta\theta$. So, Equation (A4) can be re-written as

$$w_t = \frac{\phi_t(v(s_t))}{\sum_{\tau=0}^{T} \phi_\tau(v(s_\tau))} \tag{A5}$$

If w_t takes the AAD form under the utility function $v(.)$, i.e. $w_t \propto d_t e^{v(s_t)/\lambda}$, then it should also take the AAD form under the original utility function $u(.)$.

Fourth, we show that in Equation (A4), $\Delta\theta$ has two properties: (i) $\Delta\theta$ is strictly increasing with $u(s_{T+1})$; (ii) suppose $\Delta\theta = \underline{\theta}$ when $u(s_{T+1}) = \underline{u}$ and $\Delta\theta = \bar{\theta}$ when $u(s_{T+1}) = \bar{u}$, where $\underline{u} < \bar{u}$, then for any $l \in (\underline{\theta}, \bar{\theta})$, there exists $u(s_{T+1}) \in (\underline{u}, \bar{u})$ such that $\Delta\theta = l$.

The property (i) can be shown by contradiction. Given $w_0, \ldots, w_{T+1}$ a sequence of decision weights for $s_{0\to T+1}$. Suppose $u(s_{T+1})$ is increased but $\Delta\theta$ is constant. In this case, each of $w'_0, \ldots, w'_T$ should also be constant. However, w'_{T+1} must increase as it is strictly increasing with $u(s_{T+1}) + \theta - \Delta\theta$ (θ is determined by the optimal discounting problem for $s_{0\to T}$; thus, any operations on s_{T+1} should have no effect on θ). This contradicts with the constraint that $\sum_{t=0}^{T+1} w'_t = 1$. The only way to avoid such contradictions is to set $\Delta\theta$ strictly increasing with s_{T+1}, so that $w'_0, \ldots, w'_T$ are decreasing with $u(s_{T+1})$.

For property (ii), note that for any reward sequence $s_{0\to T+1}$ and a given θ, $\Delta\theta$ is defined as the solution to $\sum_{t=0}^{T+1} \phi_t(u(s_t) + \theta - \Delta\theta) = 1$. Given an arbitrary number $l \in (\underline{\theta}, \bar{\theta})$, the proof of property (ii) consists of two stages. First, for $t = 0, 1, \ldots, T$, we need to show that $u(s_t) + \theta - l$ is still in the domain of $\phi_t(.)$. Second, for period $T+1$, we need to show for any $\omega \in (0,1)$, there exists $u(s_{T+1}) \in \mathbb{R}$ such that $\phi_{T+1}(u(s_{T+1}) + \theta - l) = \omega$.

For the first stage, note $\phi_t(.)$ is the inverse function of $f'_t(.)$. Suppose when $\Delta\theta = \bar{\theta}$, we have $f'_t(w_t^a) = u(s_t) + \theta - \bar{\theta}$, and when $\Delta\theta = \underline{\theta}$, we have $f'_t(w_t^b) = u(s_t) + \theta - \underline{\theta}$. For any $l \in (\underline{\theta}, \bar{\theta})$, we have $u(s_t) + \theta - l \in (f'_t(w_t^a), f'_t(w_t^b))$. Given that $f'_t(.)$ is continuous and strictly increasing, there must be $w_t \in (w_t^a, w_t^b)$ such that $f'_t(w_t) = u(s_t) + \theta - l$. So, $u(s_t) + \theta - l$ is in the domain of $\theta_t(.)$. For the second stage, given an arbitrary $\omega \in (0,1)$, we can set $u(s_{T+1}) = f'(\omega) - \theta + l$, so that the desired condition is satisfied.

A corollary of this step is that we can manipulate $\Delta\theta$ in Equation (A4) at any level in $[\underline{\theta}, \bar{\theta}]$ by changing a hypothetical s_{T+1}.

Fifth, we show $\ln \phi_t(.)$ is linear under some conditions. To do this, let us add a hypothetical s_{T+1} to the end of s_T and let $w'_t = \phi_t(v(s_t))$ denote the decision weights for $t = 0, 1, \ldots, T+1$ in $s_{0\to T+1}$. We change the hypothetical s_{T+1} within the set $\{s_{T+1} | v(s_{T+1}) \in [\underline{v}, \bar{v}]\}$ and see what will happen to the decision weights from period 0 to period T. Suppose this changes each w'_t to $\phi_t(v(s_t) - \eta)$. Set $\eta = \underline{\eta}$ when $u(s_{T+1}) = \underline{v}$ and $\eta = \bar{\eta}$ when $u(s_{T+1}) = \bar{v}$. By Equation (A5), we have

$$\frac{\phi_t(v(s_t))}{\sum_{\tau=0}^{T} \phi_\tau(v(s_\tau))} = \frac{\phi_t(v(s_t) - \eta)}{\sum_{\tau=0}^{T} \phi_\tau(v(s_\tau) - \eta)} \tag{A6}$$

For each $t = 0, 1, ..., T$, we can rewrite $\phi_t(v(s_t))$ as $e^{\ln \phi_t(v(s_t))}$. For the LHS of Equation (A6), multiplying both the numerator and the denominator by a same number will not affect the value. Therefore, Equation (A6) can be rewritten as

$$\frac{e^{\ln \phi_t(v(s_t)) - \kappa\eta}}{\sum_{\tau=0}^{T} e^{\ln \phi_\tau(v(s_\tau)) - \kappa\eta}} = \frac{e^{\ln \phi_t(v(s_t) - \eta)}}{\sum_{\tau=0}^{T} e^{\ln \phi_\tau(v(s_\tau) - \eta)}}$$

where κ can be any real-valued constant. By properly selecting κ, we for all $t = 0, 1, ..., T$, can obtain

$$\ln \phi_t(v(s_t)) - \kappa\eta = \ln \phi_t(v(s_t) - \eta) \tag{A7}$$

as long as $\eta \in [\underline{\eta}, \bar{\eta}]$. And given $\ln \phi_t(.)$ is strictly increasing, for any $\eta \neq 0$, we have $\kappa > 0$.

Finally, we denote the maximum and minimum of $\{v(s_t)\}_{t=0}^{T}$ by $v_{\max}$ and $v_{\min}$, and show that Equation (A7) can hold if $\eta = v_{\max} - v_{\min}$. That is, $v_{\max} - v_{\min} \in [\underline{\eta}, \bar{\eta}]$, where $\underline{\eta}, \bar{\eta}$ are the realizations of η when $v(s_{T+1}) = \underline{v}$ and $v(s_{T+1}) = \bar{v}$. Obviously, $\underline{\eta}$ can take the value $\underline{\eta} = 0$. Thus, we focus on whether $\bar{\eta}$ can take a value $\bar{\eta} \geq v_{\max} - v_{\min}$.

The proof is similar with the fourth step and consists of two stages. First, we show that for $t = 0, 1, \ldots, T$, $v(s_t) - v_{\max} + v_{\min}$ is in the domain of $\phi_t(.)$. That is, under some w_t, we have $f_t'(w_t) = v(s_t) - v_{\max} + v_{\min}$. Note $v_{\max} - v_{\min} \in [0, +\infty)$. On the one hand, there exists $w_t \in (0,1)$ for $f_t'(w_t) = v(s_t)$, which is the solution to Equation (A5). On the other hand, by Definition 2, we know $\lim_{w_t \to 0} f_t'(w_t) = -\infty$. Given $f_t'(w_t)$ is continuous and strictly increasing, there must be a solution w_t for $f_t'(w_t) = v(s_t) - v_{\max} + v_{\min}$. Second, we show that for any $\omega \in (0,1)$, there exists some $v(s_{T+1})$ such that $\phi_{T+1}(v(s_{T+1}) - v_{\max} + v_{\min}) = \omega$. This can be achieved by setting $v(s_{T+1}) = f_{T+1}'(\omega) + v_{\max} - v_{\min}$.

As a result, for any period t in $s_{0\to T}$, by Equation (A7), we have $\ln \phi_t(v(s_t)) = \ln \phi_t(v(s_t) - \eta) + \kappa\eta$ as long as $\eta \in [0, v_{\max} - v_{\min}]$, where $\kappa > 0$. We can rewrite each $\ln \phi_t(v(s_t))$ as $\ln \phi_t(v_{\min}) + \kappa(v(s_t) - v_{\min})$. Therefore, we have

$$w_t \propto \phi_t(v_{\min}) \cdot e^{\kappa(v(s_t) - v_{\min})} \tag{A8}$$

and $\sum_{t=0}^{T} w_t = 1$. In Equation (A8), setting $\phi_t(v_{\min}) = d_t$, $\lambda = 1/\kappa$, and apply the corollary of the third step, we can conclude that $w_t \propto d_t e^{u(s_t)/\lambda}$, which is of the AAD form. *QED.*