

# Appendix

Zijian Zark Wang

## A. Proofs about mutual information

(under editing)

$$V = \sum_{t=0}^T w_t v_t$$

Calculate the first-order derivative of  $V$  with respect to  $x_t$

$$\frac{\partial V}{\partial x_t} = v'_t w_t (1 + v_t - V)$$

## B. Proofs about common difference effect

Suppose the instantaneous utilities of LL and SS are  $v_l$  and  $v_s$ , and the delays for LL and SS are  $t_l$  and  $t_s$ . Under ADU, the common difference effect implies that, if

$$\frac{1 + G(t_s)e^{-v_s}}{v_s} = \frac{1 + G(t_l)e^{-v_l}}{v_l} \quad (\text{B.1})$$

then for any  $\Delta t \geq 0$ ,

$$\frac{1 + G(t_s + \Delta t)e^{-v_s}}{v_s} > \frac{1 + G(t_l + \Delta t)e^{-v_l}}{v_l} \quad (\text{B.2})$$

If  $G(T) = T$ , we have  $G(t + \Delta t) = G(t) + \Delta t$ . Thus, after combining with Equation (B.1), Equation (B.2) holds true if and only if

$$\frac{\Delta t e^{-v_s}}{v_s} > \frac{\Delta t e^{-v_l}}{v_l} \quad (\text{B.3})$$

Given that function  $\psi(v) = e^{-v}/v$  is decreasing in  $v$  so long as  $v > 0$ , Equation (B.3) must be valid.

If  $G(T) = \frac{1}{1-\delta}(\delta^{-T} - 1)$ , we have

$$1 + G(t + \Delta t)e^{-v} = \delta^{-\Delta t}[1 + G(t)e^{-v}] + (\delta^{-\Delta t} - 1)\left(\frac{e^{-v}}{1-\delta} - 1\right)$$

Thus, after combining Equation (B.1), Equation (B.2) holds true if and only if

$$(\delta^{-\Delta t} - 1)\frac{\frac{e^{-v_s}}{1-\delta} - 1}{v_s} > (\delta^{-\Delta t} - 1)\frac{\frac{e^{-v_l}}{1-\delta} - 1}{v_l} \quad (\text{B.4})$$

Given that  $0 < \delta < 1$ , we have  $\delta^{-\Delta t} > 1$ , thus Equation (B.4) is valid if and only if

$$\frac{1}{v_s} - \frac{1}{v_l} < \frac{1}{1-\delta}\left(\frac{e^{-v_s}}{v_s} - \frac{e^{-v_l}}{v_l}\right) \quad (\text{B.5})$$

From Equation (B.1) we know that

$$\frac{1}{v_s} - \frac{1}{v_l} = \frac{1}{1-\delta} \left[ \frac{(\delta^{-t_l} - 1)e^{-v_l}}{v_l} - \frac{(\delta^{-t_s} - 1)e^{-v_s}}{v_s} \right] \quad (\text{B.6})$$

Combine Equation (B.5) and (B.6), we have

$$\delta^{-t_l} \frac{e^{-v_l}}{v_l} < \delta^{-t_s} \frac{e^{-v_s}}{v_s} \implies v_l - v_s + \ln\left(\frac{v_l}{v_s}\right) > -(t_l - t_s) \ln \delta$$

## C. Proofs about magnitude effect

Define  $V(x, t) = w_t(x)v(x)$ , which denotes the overall utility of a reward of level  $x$  delivered in period  $t$ . The reward level of LL and SS are  $x_l$  and  $x_s$ , and the delays for LL and SS are  $t_l$  and  $t_s$ . The magnitude effect implies, for any  $x_s, x_l, t_s, t_l$  such that  $V(x_s, t_l) = V(x_l, t_l)$ , we have

$$\frac{\partial}{\partial x_s}\left(\frac{x_l}{x_s}\right) < 0 \implies \frac{\partial x_l}{\partial V} \frac{\partial V}{\partial x_s} x_s - x_l < 0 \implies \frac{\partial V}{\partial x_s} x_s < \frac{\partial V}{\partial x_l} x_l \quad (\text{C.1})$$

Equation (C.1) is obtained by applying the chain rule of differentiation to decompose the expression on the left-hand side. Before proceeding with further derivation, we first prove that for any  $x_s$  and  $x_l$ , we can always find a pair of  $t_s$  and  $t_l$  such that the decision maker is indifferent between SS and LL (see Lemma 1). This allows us to focus merely on the relation between  $x_s$  and  $x_l$ .

**Lemma 1:** For any  $0 < x_s < x_l < +\infty$ , there always exist  $t_s$  and  $t_l$  such that  $V(x_s, t_s) = V(x_l, t_l)$ , where  $0 \leq t_s < t_l < +\infty$ .

To prove Lemma 1, we can first fix  $t_s$ , then show that, for equation  $V(x_s, t_s) - V(x_l, t_l) = 0$ ,

there exists a solution  $t_l$  that is located in  $(t_s, +\infty)$ . This is obvious given that  $V(x, t)$  is a continuous function, and is decreasing in  $t$  and increasing in  $x$ , and converges to 0 when  $t$  approaches  $+\infty$ . When  $t_l = t_s$ , we must have  $V(x_s, t_s) - V(x_l, t_l) < 0$ ; when  $t_l \rightarrow +\infty$ , we must have  $V(x_s, t_s) - V(x_l, t_l) > 0$ , since it is easy to find a  $t_s$  such that  $V(x_s, t_s) \gg 0$ . Thus, the function  $V(x_s, t_s) - V(x_l, t_l)$  of  $t_l$  must have a zero point in  $(t_s, +\infty)$ . *QED*

Define a function  $\xi(x) = x \cdot \partial V / \partial x$ . Then Equation (C.1) is valid if and only if  $\xi(x_s) < \xi(x_l)$ . Note that

$$\frac{\partial V}{\partial x} = v'(x) \frac{1 + G(t)e^{-v(x)} + v(x) \cdot G(t)e^{-v(x)}}{[1 + G(t)e^{-v(x)}]^2} \quad (\text{C.2})$$

Meanwhile, assume  $V(x_s, t_s) = V(x_l, t_l) \equiv 1 + \alpha$ , then by the definition of  $\alpha$ , we have

$$G(t)e^{-v(x)} = \frac{v(x)}{1 + \alpha} - 1 \quad (\text{C.3})$$

Substitute the  $G(t)e^{-v(x)}$  in Equation (C.2) by Equation (C.3), we have

$$\frac{\partial V}{\partial x} = (1 + \alpha)(v(x) - \alpha) \frac{v'(x)}{v(x)} \quad (\text{C.3})$$

Let  $e_v(x)$  denote the elasticity of  $v(x)$  to  $x$ , i.e.  $e_v(x) = v'(x) \cdot x / v(x)$ . Then we can write the expression of  $\xi(x)$  as  $\xi(x) = (1 + \alpha)(v(x) - \alpha)e_v(x)$ . The condition of magnitude effect,  $\xi(x_s) < \xi(x_l)$ , thus implies that  $(v(x) - \alpha)e_v(x)$  is increasing in  $x$ . Now, we can show that this condition is weaker than the condition of magnitude effect in standard DU models (see Lemma 2).

**Lemma 2:** The magnitude effect always holds true when

$$\frac{\partial e_v(x)}{\partial x} (v(x) - \alpha) + e_v(x) v'(x) > 0 \quad (\text{C.4})$$

To prove Lemma 2, we can calculate the first-order derivative of  $(v(x) - \alpha)e_v(x)$  and set it larger than 0.

As is proposed by Loewenstein and Prelec (1992), in standard DU models, the condition to make magnitude effect appear is  $\frac{\partial e_v(x)}{\partial x} > 0$ . However, in ADU, this can be relaxed by Equation (C.4). First, note that  $v(x)$  is the instantaneous utility of reward  $x$ , and  $1 + \alpha$  is the discounted utility; thus, we must have  $v(x) > 1 + \alpha$ , which implies  $v(x) - \alpha > 1$ . Second, given  $v'(x) > 0$  and thus  $e_v(x) > 0$ , the inequality in Equation (C.4) can still hold true even when  $\frac{\partial e_v(x)}{\partial x} \leq 0$ .

Finally, by calculating the first-order derivative of  $e_v(x)$ , we can reformulate Equation (C.4)

into

$$RRA_v(x) < 1 + \frac{\alpha}{v(x) - \alpha} e_v(x) \quad (\text{C.5})$$

where  $RRA_v(x)$  is the relative risk aversion coefficient of  $v(x)$ , i.e.  $RRA_v(x) = -xv''(x)/v'(x)$ . Note that  $\alpha$  is mediated by the time of delay ( $t_s$  and  $t_l$ ). When both  $t_s$  and  $t_l$  approaches  $+\infty$ ,  $\alpha$  will converge to -1. Thus, in Equation (C.5), the  $\frac{\alpha}{v(x) - \alpha}$  will converge from above to  $-\frac{1}{v(x)+1}$ , given  $x$  is fixed. To keep Equation (C.5) valid for any feasible  $\alpha$ , we need that

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

## D. Proofs about convexity of time discounting

If  $w_t(x)$  is convex in  $t$ , we should have  $\frac{\partial^2 w_t(x)}{\partial t^2} \geq 0$ . By the definition of  $w_t(x)$ , this is equivalent to

$$2G'(t)^2 \geq (G(t) + e^{v(x)})G''(t) \quad (\text{D.1})$$

If  $G(t) = t$ , i.e. each period is initially equal-weighted, then  $G'(t) = 1$ ,  $G''(t) = 0$ . Thus, Equation (D.1) must be valid.

If  $G(t) = (1 - \delta)^{-1}(\delta^{-t} - 1)$ , i.e. the initial weights decline in an exponential fashion, where  $0 < \delta < 1$ , then  $G'(t) = (1 - \delta)^{-1}(-\ln \delta)\delta^{-t}$ ,  $G''(t) = (-\ln \delta)G'(t)$ . Thus, Equation (D.1) is valid when

$$\delta^{-t} \geq (1 - \delta)e^{v(x)} - 1 \quad (\text{D.2})$$

Given  $t \geq 0$ , Equation (D.2) holds true in two cases. The first case is  $1 \geq (1 - \delta)e^{v(x)} - 1$ , which implies that  $v(x)$  is below or equal to a certain threshold  $v(\underline{x})$ , where  $v(\underline{x}) = \ln(\frac{2}{1-\delta})$ . The second case is that  $v(x)$  is above  $v(\underline{x})$  and  $t$  is above a threshold  $\underline{t}$ . In that case, we can take the logarithm of both sides of the inequality. It turns out that  $\underline{t} = \frac{\ln[(1-\delta)e^{v(x)}-1]}{-\ln \delta}$ .

## E. Proofs about violation of diminishing sensitivity

Let  $g = G(t)$ . For simplicity, hereafter we omit the parentheses and variable notations in function  $v(\cdot)$  and  $V(\cdot)$ , as well as in their derivatives. So,  $v$  denotes  $v(x)$ ,  $v'$  denotes  $v'(x)$ , and  $V$  denotes  $V(x, t)$ , etc. The first-order derivative of  $V$  with respect to  $x$  thus can be written as

$$\frac{\partial V}{\partial x} = v' \frac{e^v + V}{e^v + g} \quad (\text{E.1})$$

When  $V(x, t)$  is concave in  $x$ , we should have  $\frac{\partial^2 V}{\partial x^2} < 0$ . By Equation (E.1), we can calculate the second-order derivative of  $V$  with respect to  $x$ , and rearrange the second-order condition to

$$2\zeta(v) + \frac{1}{1 + v\zeta(v)} - 1 < \frac{-v''}{(v')^2} \equiv \frac{d}{dx} \left( \frac{1}{v'} \right) \quad (\text{E.2})$$

where  $\zeta(v) = \frac{g}{g + \exp(v)}$ . Given that  $v'' < 0$ , the RHS of Equation (E.2) is clearly positive.

To prove the first part of Proposition 5, we can show that when  $x$  is large enough, the LHS of Equation (E.2) will be non-positive. By letting the LHS be non-positive, we can derive that

$$\zeta(v) + \frac{1}{v} \leq \frac{1}{2} \quad (\text{E.3})$$

Note that  $\zeta(v)$  is decreasing in  $v$ , and  $v$  is increasing in  $x$ . Hence,  $\zeta(v) + \frac{1}{v}$  is decreasing in  $x$ . Besides, it approaches to  $+\infty$  when  $x \rightarrow 0$ , and approaches to 0 when  $x \rightarrow +\infty$ . Thus, there is a unique value of  $x$  in  $(0, +\infty)$ , say  $\bar{x}$ , making the equality in Equation (E.3) hold true. When  $x \geq \bar{x}$ , Equation (E.3) is always valid. Under this circumstance,  $V(x, t)$  is concave in  $x$ .

To prove the second part of Proposition 5, firstly note that when  $x = 0$ , the LHS of Equation (E.2) will become  $\frac{2g}{g+1}$ . If  $\frac{d}{dx} \left( \frac{1}{v'(0)} \right)$  is smaller than this value, then the LHS of Equation (E.2) should be greater than its RHS at the point of  $x = 0$ . From the first part of proposition 5, we know the LHS is smaller than the RHS at the point of  $x = \bar{x}$ . Therefore, if  $\frac{d}{dx} \left( \frac{1}{v'(x)} \right)$  is continuous in  $[0, \bar{x}]$ , there must be a point within this interval, such that the LHS of Equation (E.2) equals its RHS. Let  $x^*$  denote the minimum value of  $x$  that make the equality valid. Then, for any  $x \in (0, x^*)$ , we must have the LHS of Equation (E.2) is greater than its RHS, which implies  $V(x, t)$  is convex in  $x$ . Given that  $t \geq 1$ , we have  $g \geq 1$  and thus  $\frac{2g}{g+1} \geq 1$ . Thus, when  $\frac{d}{dx} \left( \frac{1}{v'(0)} \right) < 1$ ,  $V(x, t)$  can be convex in  $x$  for any  $x \in (0, x^*)$ , regardless of  $g$ .

To prove the third part of Proposition 5, firstly note that  $v(x) = u(x)/\lambda$ . So,

$$\frac{d}{dx} \left( \frac{1}{v'} \right) = \lambda \frac{d}{dx} \left( \frac{1}{u'} \right) \quad (\text{E.4})$$

Second, we can arbitrarily take a point in interval  $(0, \bar{x})$ . For simplicity, we focus on  $x = \ln g$ . In this case, the LHS of Equation (E.2) becomes  $\frac{2}{2 + \ln g}$ . Define a function  $f(x)$ , where the value of  $f(x)$  is the LHS of Equation (E.2) minus its RHS. Note that  $f(x)$  is continuous at  $x = \ln g$ . By the definition of continuity, for any positive real number  $b$ , there must exist a positive real number  $c$  such that, when  $x \in (\ln g - c, \ln g + c)$ , we have

$$f(\ln g) - b < f(x) < f(\ln g) + b \quad (\text{E.4})$$

If  $f(\ln g) - b \geq 0$ , then  $f(x)$  will keep positive in interval  $(\ln g - c, \ln g + c)$ , which implies the LHS of Equation (E.2) is keeping greater than its RHS.

Finally, let us derive the condition for  $f(\ln g) - b \geq 0$ . Suppose when  $x = \ln g$ ,  $\frac{d}{dx} \left( \frac{1}{u'} \right) = a$  (note that  $\frac{d}{dx} \left( \frac{1}{u'} \right) < +\infty$  at this point given its continuity). Combining with Equation (E.3), we know that  $f(\ln g) - b = \frac{2}{2+\ln g} - \lambda a - b$ . By letting this value be non-negative, we have

$$\lambda \leq \frac{2}{a(2+\ln g)} - \frac{b}{a} \quad (\text{E.5})$$

Given that  $t \geq 1$ , we have  $g \geq 1$  and thus  $\frac{2}{2+\ln g}$  should be positive. Meanwhile, given that  $u' > 0$  and  $u'' < 0$ ,  $a$  is also positive. Since  $b$  can be any positive real number, Equation (E.5) holds true if and only if  $\lambda < \frac{2}{a(2+\ln g)}$ . That is, when  $\lambda$  is (positive but) smaller than a certain threshold, there must be an interval  $(\ln g - c, \ln g + c)$  such that the LHS of Equation (E.2) is greater than the RHS. Set  $x_1 = \max\{0, \ln g - c\}$ ,  $x_2 = \min\{\bar{x}, \ln g + c\}$ . When  $x \in (x_1, x_2)$ , function  $V(x, t)$  must be convex in  $x$ .

## F. Others

Suppose when having no information, the weight on period  $t$  across each potentially realized sequence is equal ( $\equiv w_t^0$ ). Let  $W$  and  $P$  be the minimal sets that contain all available function  $w$  and  $p$  respectively. We can use an optimization problem to represent the described evaluation procedure.

Meanwhile, by definition,  $D_{KL}$  is increasing in  $w(s_t|s)$  and is convex. To solve the model, one can simply set a Lagrange multiplier  $\gamma$ , then construct the FOC condition that, for any  $s \in X$  and  $t \in \{0, 1, \dots, T\}$ ,  $\frac{\partial D_{KL}}{\partial w(s_t|s)} = u(s_t) + \gamma$ . Given that  $u' > 0$  and the convexity of  $D_{KL}$  ensures its first derivative increasing in  $w(s_t|s)$ , the solution of  $w(s_t|s)$  should be increasing with  $s_t$ . This enables the decision maker to perform greater patience for a larger reward.

While building the model, I was mainly inspired by the theories of rational inattention (Matějka and McKay 2015; Jung et al. 2019; Maćkowiak, Matějka, and Wiederholt 2023). In Matějka and McKay (2015)'s theory of rational inattention, the decision maker makes choices between discrete alternatives; she evaluates each alternative via a costly information

acquisition process, then decides the optimal choice strategy. The theory deduces the probability of each alternative being chosen should follow a logistic-like distribution. In ADU, I assume the discounting factors are generated by a similar process; hence, she subjectively weights each time period according to a logistic-like distribution as well.

The reason why I use Shannon mutual information as the cognitive cost function is twofold. First, note that  $w(s_t|s) \propto w_t^0 e^{u(s_t)/\lambda}$ . Given a certain stream  $s$  and two time periods  $t_1$  and  $t_2$  ( $t_2 > t_1$ ), the relative weight between them  $\frac{w(s_{t_1}|s)}{w(s_{t_2}|s)}$  is only relevant to  $s_{t_1}$  and  $s_{t_2}$ . Therefore, changing the reward of a third period has no impact on how the reward in  $t_2$  should be discounted relative to that in  $t_1$ . Second, under such settings, the objective function can be rewritten as

$$\sum_{s \in X} p(s)[w(s_t|s)u(s_t) - \lambda D_{KL}]$$

where  $D_{KL}$  is the KL divergence between the initial weights over time periods and the weights updated given the stream  $s$  is drawn. Clearly, the determination of  $w(s_t|s)$  in each  $s$  can be separated from each other. In other words, given two potentially realized streams  $s$  and  $s'$ , the changes in  $s'$  has no impact on the determination of discounting factors in  $s$ . This property is consistent with many forms of optimal sequential learning (Caplin, Dean, and Leahy 2022). Matějka and McKay (2015) show that the two properties are jointly satisfied if and only if the solution of  $w(s_t|s)$  follows Equation (1).

Suppose  $p(s)$  also equals to the true probability that  $s$  is realized. After decides the  $w(s_t|s)$ , the decision maker can obtain the discounted utility (DU) of each potentially realized stream  $s$ . I assume she wants to find a risky reward stream, denoted by  $p$ , that maximizes her expected discounted utility. Therefore,

$$p = \arg \max_{p \in P} \left\{ \sum_{s \in X} \sum_{t=0}^T p(s) w(s_t|s) u(s_t) \right\}$$

Given that the discounting factor  $w(s_t|s)$  is formed by an attention-adjusted evaluation procedure. I term the model alike as attention-adjusted DU (ADU). ADU suggests that the so-called discounting factors are the attention weights that decision makers assign to each time period.

The limited attention and costly attention adjustment can be characterized by intertemporal correlation aversion and magnitude-increasing patience. The rationale for using Shannon mutual information as the cognitive cost function is that, under a certain state, the discounting factor in a certain period should be independent from irrelevant periods and irrelevant states.

While building the model, I was mainly inspired by the theories of rational inattention (Matějka and McKay 2015; Jung et al. 2019; Maćkowiak, Matějka, and Wiederholt 2023). In Matějka and McKay (2015)’s theory of rational inattention, the decision maker makes choices between discrete alternatives; she evaluates each alternative via a costly information acquisition process, then decides the optimal choice strategy. The theory deduces the probability of each alternative being chosen should follow a logistic-like distribution. In ADU, I assume the discounting factors are generated by a similar process; hence, she subjectively weights each time period according to a logistic-like distribution – as Equation (1) does – as well.

- Caplin, Andrew, Mark Dean, and John Leahy. 2022. “Rationally Inattentive Behavior: Characterizing and Generalizing Shannon Entropy.” *Journal of Political Economy* 130 (6): 1676–1715. <https://doi.org/10.1086/719276>.
- Jung, Junehyuk, Jeong Ho (John) Kim, Filip Matějka, and Christopher A Sims. 2019. “Discrete Actions in Information-Constrained Decision Problems.” *The Review of Economic Studies* 86 (6): 2643–67. <https://doi.org/10.1093/restud/rdz011>.
- Loewenstein, G., and D. Prelec. 1992. “Anomalies in Intertemporal Choice: Evidence and an Interpretation.” *The Quarterly Journal of Economics* 107 (2): 573–97. <https://doi.org/10.2307/2118482>.
- Maćkowiak, Bartosz, Filip Matějka, and Mirko Wiederholt. 2023. “Rational Inattention: A Review.” *Journal of Economic Literature* 61 (1): 226–73. <https://doi.org/10.1257/jel.20211524>.
- Matějka, Filip, and Alisdair McKay. 2015. “Rational Inattention to Discrete Choices: A New Foundation for the Multinomial Logit Model.” *American Economic Review* 105 (1): 272–98. <https://doi.org/10.1257/aer.20130047>.