

# A Continuous-time Tractable Model for Present-biased Agents

Yasunori Akagi<sup>1</sup>, Hideaki Kim<sup>1,2</sup>, Takeshi Kurashima<sup>1</sup>

<sup>1</sup>NTT Human Informatics Laboratories, NTT Corporation,

<sup>2</sup>NTT Communication Science Laboratories, NTT Corporation  
{yasunori.akagi, kim.hideaki, takeshi.kurashima}@ntt.com

## Abstract

Present bias, the tendency to overvalue immediate rewards while undervaluing future ones, is a well-known barrier to achieving long-term goals. As artificial intelligence and behavioral economics increasingly focus on this phenomenon, the need for robust mathematical models to predict behavior and guide effective interventions has become crucial. However, existing models are constrained by their reliance on the discreteness of time and limited discount functions. This study introduces a novel continuous-time mathematical model for agents influenced by present bias. Using the variational principle, we model human behavior, where individuals repeatedly act according to a sequence of states that minimize their perceived cost. Our model not only retains analytical tractability but also accommodates various discount functions. Using this model, we consider intervention optimization problems under exponential and hyperbolic discounting and theoretically derive optimal intervention strategies, offering new insights into managing present-biased behavior.

## 1 Introduction

Humans often fail to achieve their goals despite having long-term plans, even when no unforeseen events occur. Consider a person who plans to diet for a month. When making the plan, they expect to achieve their target weight by strictly managing their diet, believing they can resist temptations like a weekend invitation to a barbecue. However, as the weekend approaches, they succumb to temptation and go to the barbecue, failing their diet. Such behavior is referred to as *time inconsistency* and is a significant topic in behavioral economics.

Time inconsistency is attributed to how humans discount future values, mainly due to *present bias*. Present bias refers to the tendency to overvalue immediate costs and rewards while undervaluing those in the distant future. The mechanism by which present bias leads to time-inconsistent behavior can be explained using the example of a person planning a diet: during the planning stage, the individual does not place a high value on the temptation of a weekend barbecue, which is still some time away, and believes they can resist it. However, as the weekend approaches, the value of the barbecue increases significantly due to present bias, making it

much harder to resist the temptation.

Constructing behavior models incorporating present bias has become an important research topic in artificial intelligence and behavioral economics. It helps us predict human behavior and determine appropriate interventions for goal achievement. In particular, Kleinberg and Oren (2014) proposed a graph-theoretic model to capture behavior under present bias. In this model, the graph's vertices represent the agent's states, while the edges represent the agent's actions. The agent moves toward the goal by repeatedly taking actions along a path that minimizes the perceived cost, accounting for the effects of present bias. This model successfully replicates time-inconsistent behaviors studied extensively in behavioral economics, such as procrastination and task abandonment, while being applicable to various real-world tasks. Building on Kleinberg and Oren's model, Akagi, Marumo, and Kurashima (2024) proposed a more feasible model restricting the tasks to *progress-based tasks*. A progress-based task is one in which an agent accumulates a real-valued measure called *progress* over a period and receives a reward if the progress reaches a predetermined goal. Common examples of such tasks in daily life include completing 20 hours of exercise within a week or finishing a graduation thesis within six months. This model allows for the closed-form description of the agent's behavior and the efficient optimization of intervention, which were challenging in Kleinberg and Oren's model due to the high computational complexity (Tang et al. 2017; Albers and Kraft 2019).

These models are crucial for examining the impact of present bias on human behavior, but they have two practical challenges when applied to real-world scenarios. One is that they are defined on a discretized time axis and demand the problematic predefinition of an appropriate step size of time (e.g., daily and weekly). If the time step is too large, the agent's behavior can only be described in coarse granularity, while a too-small time step significantly increases the computational burden. The other is that the previous models only allow for a specific discount function called quasi-hyperbolic discounting (Laibson 1997) with particular parameters to represent present bias. In behavioral economics, not only quasi-hyperbolic discounting but also other discount functions, such as exponential discounting (Samuelson 1937) and hyperbolic discounting (Ainslie 1975), are frequently employed to model human time preferences. Pre-

vious models' inability to accommodate crucial discounting functions, such as exponential or hyperbolic discounting, poses significant constraints when analyzing real-world scenarios.

This paper introduces a new behavioral model for agents engaged in progress-based tasks under present bias in a continuous time setting. Unlike previous models defined on a discretized time axis, the proposed model in a continuous time setting faces a non-trivial mathematical problem for describing and analyzing the agent behavior; the agent repeatedly selects the trajectory of their future state that minimizes the perceived cost at each point in time, makes small progress during *an infinitesimal time interval*, and accumulates *the infinite number of small progress* before reaching the goal progress. To deal with this problem properly, we formalize the agent's behavior under time discounting in terms of the variational principle, where the cost-minimizing trajectory of the agent solves a set of ordinary differential equations. The variational principle is a mathematical tool to treat the minimization/maximization problem of objective functionals, primarily utilized in physics (Berdichevsky and Berdichevsky 2009) and machine learning (Kim 2021; Zappala et al. 2023). Although the dynamics of an agent described above appear highly complex, we can obtain the agent's trajectory in a very concise form under mild assumptions about discount functions. Notably, representative discount functions like exponential and hyperbolic discounting satisfy the assumptions, allowing the agent's trajectory to be described in closed form or using easy-to-handle integrals.

Furthermore, we analyzed two optimization problems related to interventions using the derived closed-form trajectory equations. The first problem is the optimal goal-setting problem. This problem involves determining the progress goal that maximizes an individual's final progress when the period and reward are given. The problem formulation changes depending on whether permitting *exploitative rewards*, which are decoy rewards that influence the agent's behavior. We analytically derived the optimal solutions for this problem when the discount function is either exponential or hyperbolic, considering both scenarios where exploitative rewards are allowed and not. Interestingly, we found that for both discount functions, the optimal solution remains unchanged regardless of whether we use exploitative rewards, indicating that exploitative rewards have no effect. This result contrasts with previous research in discrete-time settings with quasi-hyperbolic discounting, where exploitative rewards have been shown to have a significant impact (Akagi, Marumo, and Kurashima 2024), suggesting a substantial difference between these discounting functions. Investigating which type of discount function the individual uses carefully is essential for better intervention; otherwise, there is a risk of applying a suboptimal intervention.

The second intervention optimization problem is the optimal reward scheduling problem. In this problem, given a total period and total reward, the goal is to divide the total period and decide rewards and progress targets for each sub-period to maximize the final progress achieved by the agent. We successfully derived the optimal solutions for this problem analytically in the cases of both exponential and hyper-

bolic discounting. Our results show that when the number of divisions is fixed, it is optimal to divide the period into equal intervals, regardless of the parameters of the discount function. Moreover, the greater the number of divisions, the more significant the final progress can be achieved. Additionally, when the period is divided infinitely finely, the final progress converges to a limit independent of the shape of the discount function. This result suggests that dividing the period and rewards can effectively counteract the effects of present bias. These findings are also in clear contrast to those from discrete-time models with quasi-hyperbolic discounting; in such models, depending on the parameter values, it can sometimes be optimal not to divide the period into smaller intervals.

The proofs are deferred to Section 9.

## 2 Related Work

Present bias and time-inconsistent behavior have long been significant research themes in behavioral economics (Frederick, Loewenstein, and O'Donoghue 2002; Camerer and Loewenstein 2004; Wilkinson and Klaes 2017). Present bias is modeled by time discount functions. Classical economic research has used exponential discounting (Samuelson 1937) as the time discount function. Later, hyperbolic discounting (Ainslie 1975) was introduced to explain human time-inconsistent behavior, which exponential discounting could not explain alone. Additionally, quasi-hyperbolic discounting (Laibson 1997; Phelps and Pollak 1968) emerged as a discrete-time approximation of hyperbolic discounting. Other discounting functions have also been proposed, such as generalized hyperbolic functions (Loewenstein and Prelec 1992) and generalized Weibull functions (Takeuchi 2011). In this study, we focus on exponential and hyperbolic discounting to explore the properties of the proposed model. As noted above, exponential discounting is not a model of present bias (i.e., it can not explain time-inconsistent behavior) because it has a constant discount rate. However, we include exponential discounting in this paper's analysis because it is the most fundamental model of time discounting, and comparing it with other models with present bias is insightful.

There has been extensive research on present bias, goal achievement, and incentives (O'Donoghue and Rabin 2006; O'Donoghue and Rabin 2008; Koch and Nafziger 2011; Hsiaw 2013). Recent highlighted research on the mathematical modeling of goal achievement behavior under present bias is the graph-based model by Kleinberg and Oren (2014). They model a task as a directed acyclic graph (DAG) and human behavior as agents moving through the DAG. The agent repeatedly calculates the cost for all paths to the goal, influenced by quasi-hyperbolic discounting, and proceeds along the path with the minimum cost. Their model is flexible enough to handle various tasks, and they use it to examine the relationship between present bias and time inconsistency. Despite the novelty and superiority of their model, it also faced the challenge that solving the optimization problems for various interventions is often NP-Hard. (Albers and Kraft 2019, 2021; Tang et al. 2017).

To address this issue, Akagi, Marumo, and Kurashima (2024) proposed a new model based on the work of Kleinberg and Oren (2014). They demonstrated that the agent’s behavior could be expressed in closed form by restricting their model to tasks where progress is accumulated in discrete time. Based on this formulation, they derived fast algorithms and closed-form optimal solutions for intervention optimization problems, such as optimal goal setting and reward scheduling, and analyzed their behavior. Their work dramatically inspires our research. Our contributions to their research are twofold. Firstly, we demonstrated that even when extending their model to continuous time, the agent’s trajectory can be expressed in a tractable form, allowing for a broader range of discount functions, thereby enhancing the model’s applicability. Secondly, similar to their study, we analyzed the conditions for task abandonment and the intervention optimization problem to clarify how the agent’s behavior and optimal intervention strategies change due to differences between continuous and discrete time and differences in discount functions.

Our research is closely related to reinforcement learning (Kaelbling, Littman, and Moore 1996). While our model focuses on explaining agents’ behavior and determining optimal interventions under the assumption that agents maximize discounted rewards, reinforcement learning aims to teach agents how to act in unknown situations to achieve the same goal. Thus, our research and reinforcement learning explore the same issue from different angles. Additionally, research in reinforcement learning has extended time discount functions beyond exponential discounting (Schultheis, Rothkopf, and Koepl 2022; Fedus et al. 2019), suggesting that future insights from these studies may intersect with our findings.

Furthermore, continuous-time optimal control (Lewis, Vrabie, and Syrmos 2012) is highly relevant to our study. Our study and optimal control share the goal of obtaining interventions to optimize objective functions in dynamical systems. However, there are two significant differences: (i) Description of system dynamics. Optimal control typically models system dynamics by differential/difference equations or Markov models. In contrast, the system dynamics in our study are governed by continuous decision-making processes considering temporal discounting, specifically the solutions to sequential variational problems. As a result, the system dynamics in our study are considerably more complex than those in optimal control. (ii) Focus of research interest. While optimal control primarily focuses on optimal control strategies and the algorithms for obtaining them, our study places additional emphasis on exploring the qualitative relationships between optimal interventions and the shape of the discount function (the strength of present bias). Despite these differences, using tools from optimal control to analyze the problems in this study is a promising direction for future work.

### 3 Formulation of the Proposed Model

**Progress-based Task.** Similar to the previous study (Akagi, Marumo, and Kurashima 2024), this study deals with a

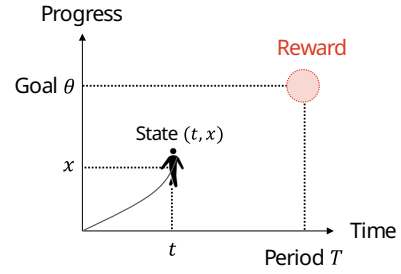


Figure 1: An illustrative example of a progress-based task.

task that aims to achieve a goal within a time limit by growing a real value called progress. We assume that progress never decreases. Such types of tasks frequently occur in daily life. For example, consider a student who aims to complete an assignment within 8 hours on a given day. If the student finishes the assignment, they will earn school credit for the course. In this case, the period corresponds to the 8 hours, progress corresponds to the percentage of the assignment completed, and the reward corresponds to the value of earning the course credit. Another example could be a person participating in a health program that involves exercising for 30 hours within a month. In this scenario, the period is one month, and progress is measured by the hours spent exercising. The reward is the incentive provided upon completing the health program.

The agent is rewarded with  $R$  if it achieves the goal progress  $\theta$  within the period  $T$ , where  $R, \theta, T \in \mathbb{R}_{\geq 0}$ . We denote the agent’s state by a tuple  $(t, x)$ , where  $t \in [0, T]$  is the time and  $x \in \mathbb{R}_{\geq 0}$  is the progress. Note that time  $t$  is not restricted to discrete values but takes continuous values, a significant difference from the previous study (Akagi, Marumo, and Kurashima 2024). Figure 1 describes an illustrative example of a progress-based task.

**Perceived Cost.** We will denote by  $\mathcal{Y}_{t,x}$  the set of absolutely continuous functions defined in  $[t, T]$  which satisfies  $y(t) = x$ . A function  $y(s) \in \mathcal{Y}_{t,x}$  represents the progress trajectory that an agent in state  $(t, x)$  will take at future time  $s \in [t, T]$ . For progress trajectory  $y(s)$ , we define *perceived cost*  $\mathcal{P}[y]$  by

$$\mathcal{P}[y] := \int_t^T L_t(s, y'(s)) ds - \gamma_t(T) R \cdot \mathbf{1}[y(T) \geq \theta], \quad (1)$$

$$L_t(s, v) := \gamma_t(s) c(v), \quad (2)$$

where  $y'(s)$  denotes the derivative function of  $y(s)$ . Note that  $\mathcal{P}[\cdot]$  is a *functional* that takes a function as an argument and returns a real value. The function  $\gamma_t : \mathbb{R} \rightarrow (0, 1]$  is the *discount function*, which represents how the agent discounts the future benefit or cost; an agent at time  $t$  calculates the benefits and costs at a future time  $s$  by multiplying the discount factor  $\gamma_t(s)$ . We assume that  $\gamma_t$  is continuously differentiable in  $[t, T]$ . The function  $c : \mathbb{R} \rightarrow \mathbb{R} \cup \{+\infty\}$  is the *cost function*; to increase progress at a rate of  $v$ , a cost of  $c(v)$  is incurred. Also,  $\mathbf{1}[\cdot]$  is an indicator function;  $\mathbf{1}[y_T \geq \theta] = 1$  if  $y_T \geq \theta$ , otherwise  $\mathbf{1}[y_T \geq \theta] = 0$ .

The intuitive interpretation of perceived cost (1) is the total cost of a future progress trajectory to an agent, consider-

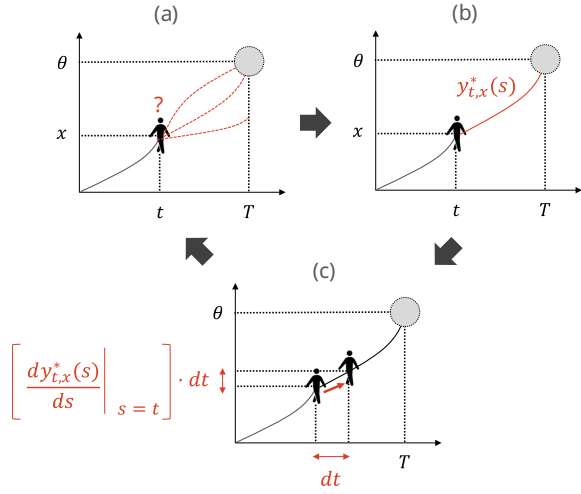


Figure 2: An illustrative example of agents behavior model. (a) The agent searches for a trajectory that minimizes the perceived cost among the candidate trajectories of future states  $\mathcal{Y}_{t,x}$ . The red dashed line is an example of a candidate trajectory in  $\mathcal{Y}_{t,x}$ . (b) Let  $y_{t,x}^*(s)$  be the found trajectory that minimizes perceived cost. (c) The agent generates progress along  $y_{t,x}^*(s)$ . Note that while the time interval  $dt$  is inherently infinitesimal, it is depicted as larger for illustrative purposes.

ing the discount function effect. The first term of (1) is the total cost perceived by the agent over the period from  $t$  to  $T$ , considering time discounting. It is calculated by integrating the agent's perceived cost  $L_t(s, y'(s))$  at time  $s$  from  $t$  to  $T$ . The second is the reward term; the agent receives the time-discounted reward  $\gamma_t(T)R$  if and only if the final progress  $y(T)$  exceeds the goal progress  $\theta$ . This term has a negative sign because rewards have the opposite effect of costs.

**Agent Behavior Model.** The agent at state  $(t, x)$  selects the trajectory  $y_{t,x}^*$  with the minimum perceived cost among the candidate trajectories  $\mathcal{Y}_{t,x}$ . Then, at time  $t$ , the agent generates progress along  $y_{t,x}^*$ ; it produces progress of  $\left( \frac{dy_{t,x}^*(s)}{ds} \right)_{s=t} \cdot dt$  during the infinitesimal time  $dt$ . Starting from the state  $(0, 0)$ , the agent repeats this procedure at each time  $t \in [0, T]$ . Figure 2 shows an illustrative example of this process.

In this agent behavior model, the trajectory  $x(t)$  actually taken by the agent is fully determined by (1) and (2) and

$$y_{t,x(t)}^* := \operatorname{argmin}_{y \in \mathcal{Y}_{t,x(t)}} \mathcal{P}[y], \quad (3)$$

$$\frac{dx(t)}{dt} = \left. \frac{dy_{t,x(t)}^*(s)}{ds} \right|_{s=t}, x(0) = 0. \quad (4)$$

This agent's behavior model is a natural extension of the discrete-time behavior models from previous research (Kleinberg and Oren 2014; Akagi, Marumo, and Kurashima 2024) to continuous time. In the previous models, the agent repeatedly chooses the path with the minimum perceived

cost among all possible future paths at each time step and acts along that path. In the model proposed in this paper, the path with the smallest perceived cost corresponds to  $y_{t,x(t)}^*$  in (3) and acting along the path corresponds to (4).

The formulation of the agent's behavior in (3) is based on the variational principle. The variational principle is a concept primarily used in physics and machine learning, stating that the state of a system is determined by a function that maximizes or minimizes a particular functional (Berdichevsky and Berdichevsky 2009; Kim 2021; Zappala et al. 2023). In our formulation, the agent's future trajectory  $y_{t,x(t)}^*$  is defined as the value that minimizes the functional  $\mathcal{P}$ , making it a clear example of the variational principle. By employing this formulation, we achieve a natural extension of existing agent behavior models to continuous time while also simplifying the analysis of agent behavior by utilizing the well-established variational methods techniques (Rockafellar 2001).

**Discount Function.** The discount function has a significant impact on the agent's behavior. In this paper, we focus on the most typical discount functions, the *exponential discount function* (Samuelson 1937)

$$\gamma_t^{\text{exp}}(s) = \exp(-k(s-t)) \quad (5)$$

and the *hyperbolic discount function* (Ainslie 1975)

$$\gamma_t^{\text{hyp}}(s) = \frac{1}{1+k(s-t)}, \quad (6)$$

where  $k > 0$  is a parameter. Exponential discounting has been classically used as a human discounting model. Hyperbolic discounting was proposed to explain human behavior that exponential discounting cannot explain, such as time inconsistency. For both discount functions, the function value decays rapidly when the parameter  $k$  is large and slowly when it is small. Thus, we can interpret the magnitude of  $k$  as the strength of the agent's present bias. Figure 3 shows the shapes of the exponential and hyperbolic discount functions.

**Cost Function.** This paper assumes that the cost function can be written as

$$c(\Delta) = \begin{cases} \Delta^\alpha & \text{if } \Delta \geq 0, \\ +\infty & \text{otherwise,} \end{cases} \quad (7)$$

where  $\alpha > 1$  is a parameter. Because this paper considers only tasks in which progress never decreases, we set  $c(\Delta) = +\infty$  when  $\Delta < 0$ . We can control the shape of the cost function by adjusting the parameter  $\alpha$ . This cost function is precisely the same as the one used in the previous study (Akagi, Marumo, and Kurashima 2024).

## 4 Properties of the Proposed Model

### 4.1 Formula of Agent's Trajectory

The trajectory of the agent  $x(t)$  is determined by (1)–(4). This dynamics involves optimizing a functional (a variational problem) and differential equations, which are expected to result in very complex behavior. However, we have

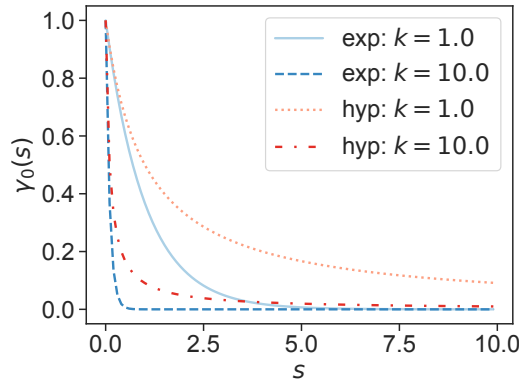


Figure 3: The plots of exponential discount function  $\gamma_0^{\text{exp}}(s)$  and hyperbolic discount function  $\gamma_0^{\text{hyp}}(s)$  for parameters  $k = 1.0, 10.0$ . “exp” means exponential discount function and “hyp” means hyperbolic discount function.

shown that under mild conditions, the agent’s trajectory  $x(t)$  can be expressed concisely. For simplicity, we introduce the following notations;

$$\Gamma_t(s) := \int \gamma_t(s)^{-\frac{1}{\alpha-1}} ds, \quad (8)$$

$$\zeta(t) := \gamma_t(T)(\Gamma_t(T) - \Gamma_t(t))^{\alpha-1}.$$

Note that (8) is an indefinite integral; we can choose the constant of integration arbitrarily because it does not affect the later results.

**Theorem 1.** Assume that  $\zeta(t)$  is non-increasing in  $[0, T]$ . The following holds:

$$x(t) = \theta \left( 1 - \exp \left( - \int_0^{\tilde{t}} \frac{ds}{\Gamma_s(T) - \Gamma_s(s)} \right) \right), \quad (9)$$

where  $\tilde{t} = \min(t, t^*)$  and  $t^*$  is the smallest  $t \in [0, T]$  such that

$$\frac{\exp \left( -\alpha \int_0^t \frac{ds}{\Gamma_s(T) - \Gamma_s(s)} \right)}{\zeta(t)} \geq \frac{R}{\theta^\alpha} \quad (10)$$

if there exists such  $t$ ; otherwise,  $t^* = T$ .

Theorem 1 shows that we can calculate the agent’s trajectory if the discount function  $\gamma_t$  satisfies the property that  $\Gamma_t(s)$  and  $\int_0^{\tilde{t}} \frac{ds}{\Gamma_s(T) - \Gamma_s(s)}$  are analytically or numerically computable. The value  $t^*$  is the *abandonment time* when the agent gives up making further progress because any additional progress would not be worth the reward. The condition that  $\zeta(t)$  is non-increasing corresponds to the scenario where once an agent abandons a task, they do not resume it (see Section 9.2). If this condition does not hold, the agent would repeatedly abandon and resume the task, resulting in more complex behavior. While this case is also interesting, we leave a detailed analysis of it for future work, and this paper focuses on the case where  $\zeta(t)$  is non-increasing.

We explain the derivation of Theorem 1 briefly. First, we use the variational method to solve for (3). Typically, the

variational method can only find local optima. However, in this case, due to the special structure of the cost function (7), the functional  $\mathcal{P}$  becomes a convex functional. This property allows us to obtain the global optimum  $y_{t,x(t)}^*$  by solving a differential equation called the Euler–Lagrange equation (Rockafellar 2001). We finally derive (9) by solving the differential equation (4) using the obtained  $y_{t,x(t)}^*$ . For details, please refer to Section 9.2.

The discount function must satisfy several conditions to apply Theorem 1. Fortunately, the two leading discount functions, exponential and hyperbolic discounting, satisfy these conditions.

**Corollary 1.** When  $\gamma_t = \gamma_t^{\text{exp}}$  defined as (5),  $\zeta(t)$  is non-increasing and

$$x(t) = \theta \left( \frac{\exp \left( \frac{k}{\alpha-1} \tilde{t} \right) - 1}{\exp \left( \frac{k}{\alpha-1} T \right) - 1} \right),$$

$$t^* = \begin{cases} 0 & \left( \frac{k}{(\alpha-1)(1 - \exp(-\frac{kT}{\alpha-1}))} \right)^{\alpha-1} \geq \frac{R}{\theta^\alpha}, \\ T & \text{otherwise.} \end{cases}$$

**Corollary 2.** When  $\gamma_t = \gamma_t^{\text{hyp}}$  defined as (6),  $\zeta(t)$  is non-increasing,

$$x(t) = \theta \left( 1 - \exp \left( - \frac{k\alpha}{\alpha-1} \eta(\tilde{t}) \right) \right),$$

$$\eta(t) := \int_0^t \frac{ds}{(1 + k(T-s))^{\frac{\alpha}{\alpha-1}} - 1},$$

and  $t^*$  is the smallest  $t \in [0, T]$  such that

$$\left( \frac{k\alpha}{(\alpha-1) \left( (1 + k(T-t))^{\frac{\alpha}{\alpha-1}} - 1 \right)} \right)^{\alpha-1} \cdot (1 + k(T-t)) \exp \left( - \frac{k\alpha^2}{\alpha-1} \eta(t) \right) \geq \frac{R}{\theta^\alpha}$$

if there exists such  $t$ ; otherwise,  $t^* = T$ .

**Corollary 3.** When  $\gamma_t = \gamma_t^{\text{hyp}}$  and  $\alpha = 2$ ,

$$x(t) = \frac{2\theta\tilde{t}}{(k(T-\tilde{t}) + 2)T}$$

and  $t^*$  is the smallest  $t \in [0, T]$  such that

$$\frac{2(kT + 2)^2 (T-t)(k(T-t) + 1)}{T^2 (k(T-t) + 2)^3} \geq \frac{R}{\theta^2} \quad (11)$$

if there exists such  $t$ ; otherwise,  $t^* = T$ .

Corollaries 1–3 indicate that we can express the agent’s trajectory in closed form for the cases of exponential discounting and hyperbolic discounting with  $\alpha = 2$ . For hyperbolic discounting with  $\alpha \neq 2$ , we cannot derive a closed-form expression of the agent’s trajectory without integration; however, we can calculate it by a simple numerical integration of a single variable.

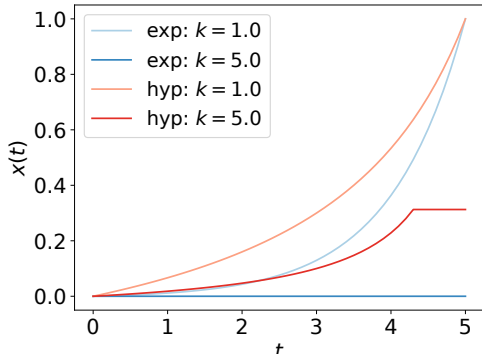


Figure 4: The agent’s trajectory plots under exponential discount function and hyperbolic discount function for  $k = 1.0, 5.0$  and  $\alpha = 2$ . We set  $R = 1.1$ ,  $T = 5$ , and  $\theta = 1$ . “exp” means exponential discount function, and “hyp” means hyperbolic discount function.

Figure 4 shows the agent’s trajectories under exponential discount function and hyperbolic discount function for  $k = 1.0, 5.0$  and  $\alpha = 2$ . When  $k = 1.0$ , the agent achieves the goal progress  $\theta = 1$  and receives the reward under both exponential and hyperbolic discounting. However, when  $k = 5.0$ , the agent gives up on reaching the target progress under both discounting functions, but the manner of giving up differs: an agent with exponential discounting abandons the goal from the outset, while an agent with hyperbolic discounting gives up midway through.

One of the proposed model’s advantages is that the agent’s trajectory can be obtained in a tractable form for representative discount functions. In the subsequent sections, we will use these tractable formulas to investigate the properties of agent behavior with present bias in continuous time, including task abandonment and optimal interventions.

## 4.2 Abandonment Time and Time-inconsistency

Here, we will examine the task abandonment time  $t^*$  and the conditions under which the agent abandons the task, exploring the circumstances under which time-inconsistent behavior occurs.

First, consider the case of exponential discounting. According to Corollary 1, the abandonment time  $t^*$  takes either the value of 0 or  $T$ . Therefore, the agent either completes the task or does not engage with the task at all from the beginning without making any progress. In other words, an agent with exponential discounting never abandons a task once started. In this sense, an agent with exponential discounting is time-consistent.

Next, consider the case of hyperbolic discounting ( $\alpha = 2$  for simplicity).

**Proposition 1.** If  $T \leq \frac{1+\sqrt{3}}{k}$ , the left-hand side of (11) takes maximum value  $\frac{2(kT+1)}{kT+2}$  at  $t = 0$ . Otherwise, the left-hand side of (11) takes maximum value  $\frac{(kT+2)^2}{3\sqrt{3}kT^2}$  at  $t = T - \frac{1+\sqrt{3}}{k}$ .

Proposition 1 indicates that if  $T \leq \frac{1+\sqrt{3}}{k}$ ,  $t^*$  takes either the value of 0 or  $T$  and so the agent’s behavior is time-

consistent as the exponential discounting case. On the other hand, if  $T > \frac{1+\sqrt{3}}{k}$ ,  $t^*$  may take a value other than 0 or  $T$ , because the left-hand side of (11) takes maximum value at  $t = T - \frac{1+\sqrt{3}}{k}$ . This implies that an agent may initially perceive the task as beneficial and start working on it, but later, reconsidering the costs and benefits, decide that continuing the task is disadvantageous and abandon it midway. This demonstrates the possibility of time-inconsistent behavior, where the agent’s preferences change over time.

It is well-known in behavioral economics that the behavior of individuals with exponential discounting is time-consistent, while that of individuals with hyperbolic discounting is time-inconsistent (Frederick, Loewenstein, and O’Donoghue 2002). The results in this section demonstrate that our model successfully reproduces these insights in behavioral economics.

## 5 Optimal Goal-Setting Problem

In this section, we consider the problem of finding the goal  $\theta^*$  which maximizes the final progress  $x(T)$ :

$$\theta^* := \operatorname{argmax}_{\theta \geq 0} x(T), \quad (12)$$

given the period  $T$ , reward  $R$ , and parameters  $k$  and  $\alpha$ . This problem involves determining how an intervener should set goals to maximize an agent’s progress over a fixed period with a given reward.

This problem is crucial in real-world scenarios. For instance, consider a company president (intervener) who offers employees (agents) a fixed bonus for achieving a sales target within six months. How should the president set this target to maximize final sales? If the target is set too high, employees may lose motivation due to the difficulty of achieving the target, resulting in lower final sales. On the other hand, if the target is too low, employees might reach it easily, earn the bonus, and feel satisfied, resulting in reduced effort and lower sales. Considering this trade-off, setting appropriate goals, as formulated in (12), is critical.

The optimal solution to the goal-setting problem depends on whether exploitative rewards are allowed. Exploitative rewards are decoys that take advantage of an agent’s present bias. An agent with present bias may make progress even if they ultimately abandon the goal. As a result, setting high, unattainable goals can sometimes lead to more significant progress than setting lower, achievable goals, even though the agent fails to reach the final goal. This approach also benefits the intervener, because he does not need to give the reward because the agent does not achieve the goal. However, exploitative rewards raise ethical concerns, such as reducing motivation and hindering long-term effort, requiring careful consideration. Existing research has discussed the significance and propriety of exploitative rewards (Kleinberg and Oren 2014; Tang et al. 2017; Albers and Kraft 2019). A particularly important result is that using exploitative rewards for agents with quasi-hyperbolic discounting can increase overall progress significantly (Akagi, Marumo, and Kurashima 2024).

This section examines both scenarios: allowing and disallowing exploitative rewards. The optimization problem

when exploitative rewards are permitted is precisely (12), while the problem when they are not allowed adds the constraint  $x(T) = \theta$  to (12). The following theorems give the optimal solutions to the optimal goal-setting problems.

**Theorem 2.** When  $\gamma_t = \gamma_t^{\text{exp}}$ , regardless of whether exploitative rewards are permitted,

$$\theta^* = \left[ \frac{\alpha - 1}{k} \left( 1 - \exp\left(-\frac{kT}{\alpha - 1}\right) \right) \right]^{\frac{\alpha-1}{\alpha}} R^{\frac{1}{\alpha}}$$

and  $\max_{\theta \geq 0} x(T) = \theta^*$ .

**Theorem 3.** When  $\gamma_t = \gamma_t^{\text{hyp}}$  and  $\alpha = 2$ , regardless of whether exploitative rewards are permitted,

$$\theta^* = \begin{cases} \frac{T(3\sqrt{3}k)^{\frac{1}{2}}}{kT+2} R^{\frac{1}{2}} & T \geq \frac{1+\sqrt{3}}{k}, \\ \left( \frac{T(kT+2)}{2(kT+1)} \right)^{\frac{1}{2}} R^{\frac{1}{2}} & T < \frac{1+\sqrt{3}}{k} \end{cases}$$

and  $\max_{\theta \geq 0} x(T) = \theta^*$ .

From Theorem 2 and Theorem 3, we observe followings.

**Effect of  $k$ .** When  $T$  and  $R$  are fixed,  $\theta^*$  is monotonically decreasing with respect to  $k$  for both discounting functions, and  $\lim_{k \rightarrow \infty} \theta^* = 0$ . This property implies that the more rapidly the discount function decays, the smaller the achievable goal progress.

**Effect of  $T$ .** When  $k$  and  $R$  are fixed,  $\theta^*$  is monotonically increasing with respect to  $T$  for both discounting functions, and  $\lim_{T \rightarrow \infty} \theta^* = \left(\frac{\alpha-1}{k}\right)^{\frac{\alpha-1}{\alpha}} R^{\frac{1}{\alpha}}$  with exponential discounting and  $\lim_{T \rightarrow \infty} \theta^* = (3\sqrt{3}/k)^{\frac{1}{2}} R^{\frac{1}{2}}$  with hyperbolic discounting. This property implies that while longer time limits lead to higher achievable goals, there is an upper bound on the attainable goals.

**Effect of Exploitative Reward.** For both discount functions, whether we permit exploitative rewards or not does not change  $\theta^*$ . This result means that leveraging present bias cannot exploit the agents with these discount functions. This result contrasts with the significant impact of exploitative rewards under quasi-hyperbolic discounting, which approximates hyperbolic discounting (Akagi, Marumo, and Kurashima 2024). It highlights a substantial difference in the optimal interventions for hyperbolic versus quasi-hyperbolic discounting. Therefore, when designing interventions, it is crucial to determine the type of discount function the individual uses accurately; otherwise, the derived intervention may be suboptimal.

## 6 Optimal Reward Scheduling Problem

This section considers an advanced intervention optimization problem, optimal reward scheduling problem. In this problem, the intervener can divide the rewards and distribute them to the agent multiple times.

The inputs to the problem are the total period  $T$ , the total reward  $R$ , and the agent parameters  $k$  and  $\alpha$ . The intervener choose some positive integer  $N$  and divide the total period  $T$  into  $N$  periods  $T_1, \dots, T_N \in \mathbb{R}_{\geq 0}$  and the total reward  $R$

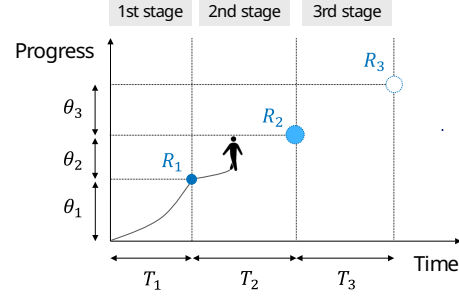


Figure 5: An example of reward schedule when  $N = 3$ .

into  $N$  rewards  $R_1, \dots, R_N \in \mathbb{R}_{\geq 0}$ . Note that  $\sum_{i=1}^N T_i = T$  and  $\sum_{i=1}^N R_i = R$  must hold. The intervener also decide the goal progress  $\theta_i \in \mathbb{R}_{\geq 0}$  for  $i \in \{1, \dots, N\}$ . We refer to the tuple  $(N, (T_i)_{i=1}^N, (R_i)_{i=1}^N, (\theta_i)_{i=1}^N)$  as reward schedule.

Once the reward schedule is determined, the agent accumulates progress in  $N$  stages. In the  $i$ -th stage, agents accumulate progress under the condition that they will receive a reward  $R_i$  if they achieve the goal progress  $\theta_i$  in the period  $T_i$ . Note that the agent makes independent progress at each stage. In other words, the agent makes progress at the  $i$ -th stage without considering the rewards or goals at stages other than the  $i$ -th. Let  $P_i$  be the increment of progress in the  $i$ -th stage. We aim to maximize the overall increment of progress  $\sum_{i=1}^N P_i$  by appropriately deciding the reward schedule. Figure 5 illustrates an example of a reward schedule in this setting.

**Theorem 4.** When  $\gamma_t = \gamma_t^{\text{exp}}$ , for fixed  $N$ , the optimal schedule is

$$T_i = T/N, R_i = R/N,$$

$$\theta_i = \left[ \frac{\alpha - 1}{k} \left( 1 - \exp\left(-\frac{kT}{(\alpha - 1)N}\right) \right) \right]^{\frac{\alpha-1}{\alpha}} \left( \frac{R}{N} \right)^{\frac{1}{\alpha}}$$

for all  $i \in \{1, \dots, N\}$ , and the objective function value is

$$f_{\alpha}^{\text{exp}}(N) := \left( \frac{N(\alpha - 1)}{k} \left( 1 - \exp\left(-\frac{kT}{(\alpha - 1)N}\right) \right) \right)^{\frac{\alpha-1}{\alpha}} R^{\frac{1}{\alpha}}. \quad (13)$$

**Theorem 5.** When  $\gamma_t = \gamma_t^{\text{hyp}}$  and  $\alpha = 2$ , for fixed  $N$ , the optimal schedule is

$$T_i = T/N, R_i = R/N,$$

$$\theta_i = \begin{cases} \frac{T(3\sqrt{3}k)^{\frac{1}{2}}}{kT+2N} \left( \frac{R}{N} \right)^{\frac{1}{2}} & \frac{T}{N} \geq \frac{1+\sqrt{3}}{k}, \\ \left( \frac{T(kT+2N)}{2N(kT+N)} \right)^{\frac{1}{2}} \left( \frac{R}{N} \right)^{\frac{1}{2}} & \frac{T}{N} < \frac{1+\sqrt{3}}{k}, \end{cases}$$

for all  $i \in \{1, \dots, N\}$ , and the objective function value is

$$f_2^{\text{hyp}}(N) := \begin{cases} \frac{T(3\sqrt{3}Nk)^{\frac{1}{2}}}{kT+2N} R^{\frac{1}{2}} & \frac{T}{N} \geq \frac{1+\sqrt{3}}{k}, \\ \left( \frac{T(kT+2N)}{2(kT+N)} \right)^{\frac{1}{2}} R^{\frac{1}{2}} & \frac{T}{N} < \frac{1+\sqrt{3}}{k}. \end{cases} \quad (14)$$

**Theorem 6.**  $f_{\alpha}^{\text{exp}}(N)$  defined by (13) is monotonically increasing with respect to  $N$ . Furthermore,  $f_{\alpha}^{\text{exp}}(N) < T^{\frac{\alpha-1}{\alpha}} R^{\frac{1}{\alpha}}$  and  $\lim_{N \rightarrow \infty} f_{\alpha}^{\text{exp}}(N) = T^{\frac{\alpha-1}{\alpha}} R^{\frac{1}{\alpha}}$  hold.

**Theorem 7.**  $f_2^{\text{hyp}}(N)$  defined by (14) is monotonically increasing with respect to  $N$ . Furthermore,  $f_2^{\text{hyp}}(N) < T^{\frac{1}{2}} R^{\frac{1}{2}}$  and  $\lim_{N \rightarrow \infty} f_2^{\text{exp}}(N) = T^{\frac{1}{2}} R^{\frac{1}{2}}$  hold.

Theorems 4 and 5 imply that when we fix  $N$ , we can write the optimal reward schedule in closed form. Furthermore, for both exponential and hyperbolic discounting, it is optimal to evenly distribute the total rewards and provide them at regular intervals. Theorems 6 and 7 indicate that the more times we divide the rewards or periods, the more significant the total progress the agent ultimately achieves. However, there is an upper bound to the total progress, and as  $N$  increases, the total progress converges to this upper bound.

An important observation is that the limit is independent of the form of the discount function. For example, the limit of  $f_{\alpha}^{\text{exp}}(N)$ ,  $T^{\frac{\alpha-1}{\alpha}} R^{\frac{1}{\alpha}}$  and the limit of  $f_2^{\text{exp}}(N)$ ,  $T^{\frac{1}{2}} R^{\frac{1}{2}}$  do not depend on the parameter  $k$  that changes the shape of the discount function. Furthermore, both  $f_2^{\text{exp}}(N)$  and  $f_2^{\text{hyp}}(N)$  converge to the same value  $T^{\frac{1}{2}} R^{\frac{1}{2}}$ . These results imply that providing intermediate rewards at sufficiently fine intervals can counteract the effects of time discounting and achieve progress independent of the individual’s discount function. In practical situations, it is impossible to divide rewards infinitely finely, and there are often constraints on the number of divisions. However, even in such cases, splitting the rewards as finely as possible can mitigate the effects of time discounting and maximize the attainable progress.

This result contrasts with the previous findings under quasi-hyperbolic discounting; Akagi, Marumo, and Kurashima (2024) demonstrated that under discrete time quasi-hyperbolic discounting, varying the reward intervals based on the discount-function parameter is optimal. For specific parameters, providing the rewards in a lump sum is optimal, whereas for other parameters, dividing the rewards as finely as possible is optimal. In contrast, in the continuous-time model proposed in this paper, the optimal strategy is to divide the rewards as finely as possible, regardless of the type (exponential or hyperbolic) and parameters of the discount function. The reasons for this difference may include the distinctions between discrete and continuous time and differences in the shapes of the discount functions (exponential and hyperbolic discount functions change smoothly, whereas quasi-hyperbolic discounting changes discontinuously). We will leave more detailed explanations in future work.

## 7 Discussion

### 7.1 Continuous vs. Discrete-time Model

How should we distinguish between the use of continuous-time and discrete-time models when modeling actual human behavior? One criterion is the type of task being addressed. A discrete-time model is more appropriate for tasks where agents perceive temporal chunks and make decisions based on those chunks. For example, consider a task where a pro-

fessor supervises a student over a year to complete a graduation thesis. The professor reviews the student’s progress weekly and decides on the strategy moving forward during each review. When modeling the professor as an agent, a discrete-time model is suitable because the professor’s decision-making is based on a weekly temporal chunk. Similarly, in a task where a person aims to exercise for 40 hours over a month, with exercise time tracked by a smartwatch and daily totals reported at the end of each day, the individual will likely plan based on daily exercise amounts, making a discrete model appropriate. Conversely, continuous-time modeling is more suitable for tasks with less segmented time (see examples in Section 3 “Progress-based Task”).

### 7.2 Empirical Validation of Our Model

It is important to verify whether the proposed model accurately reproduce human behavior under present bias in real-world scenarios. There are relatively few studies on the effects of goal setting and reward schedules in progress-based tasks like the ones we are examining, and even fewer studies have empirically tested these effects rather than analyzing them theoretically (see also Sections 5 and 6 of (Akagi, Marumo, and Kurashima 2024), “Comparison with the literature”). As a result, it is currently challenging to compare past experimental data with the theoretical results obtained in this study. We consider this to be a significant limitation of this paper. As a future direction, we aim to conduct experiments with human subjects to collect data and examine the consistency with our theoretical findings.

## 8 Conclusion

This paper introduces a new mathematical model for human goal achievement behavior under present bias in continuous time. The model is analytically tractable and accommodates essential discount functions, such as exponential and hyperbolic discounting. With this model, we derived analytical conditions for agents abandoning tasks and identified optimal intervention strategies, leading to valuable insights. Notably, our findings differ from those obtained using quasi-hyperbolic discounting in discrete time, emphasizing the importance of selecting the appropriate model based on the specific context.

## 9 Proofs

Due to the space limitation, we only provide proof of the main result (Theorem 1). For other proofs, please refer to the full version (Akagi, Kim, and Kurashima 2024).

### 9.1 Existing Results in Variational Methods

Here, we provide an overview of the existing results in variational methods related to our paper. We use the contents presented here in Section 9.2.

Let us denote by  $\mathcal{A}^p[a, b]$  the set of functions that are absolutely continuous on the interval  $[a, b]$  and whose derivatives are  $p$ -integrable, where  $p \in [1, \infty]$ . We consider the

following variational problem;

$$\begin{aligned} \min_{y \in \mathcal{A}^p} \quad & \int_a^b f(s, y, y'(s)) ds \\ \text{s.t.} \quad & y(a) = A, y(b) = B, \end{aligned} \quad (15)$$

where  $f(x_0, x_1, x_2)$  is continuously differentiable. Write  $f_i(x_0, x_1, x_2) = \frac{\partial}{\partial x_i} f(x_0, x_1, x_2)$  ( $i = 0, 1, 2$ ). Then, the following theorem holds.

**Theorem 8** ((Rockafellar 2001), Theorem 3). If the function  $f(x_0, x_1, x_2)$  is convex with respect to  $(x_1, x_2)$  and there exist functions  $y \in \mathcal{A}^p[a, b]$  and  $z \in \mathcal{A}^1[a, b]$  that satisfy

$$\begin{aligned} z(s) &= f_2(s, y, y'), z'(s) = f_1(s, y, y'), \\ y(a) &= A, y(b) = B, \end{aligned} \quad (16)$$

then  $y$  is an optimal solution of (15).

The equations (16) are called Euler–Lagrange equation.

## 9.2 Proof of Theorem 1

*Proof.* First, we solve (3). A function  $y$  that minimizes  $\mathcal{P}[y]$  satisfies  $y(T) = \theta$  or  $y(T) = x(T)$ . This is because  $y(T) = \theta$  holds when  $y(T) \geq \theta$  and  $y(T) = x(T)$  holds otherwise from the definitions of functional  $\mathcal{P}[y]$  and cost function  $c(\Delta)$ .

(i) When  $y(T) = x(T)$ , it is clear that the optimal choice is  $y(s) = x(s)$  for  $s \in [t, T]$ , and in this case,  $\mathcal{P}[y] = 0$ .

(ii) When  $y(T) = \theta$ , because

$$\mathcal{P}[y] = \int_t^T L_t(s, y'(s)) ds - \gamma_t(T)R$$

holds, it is sufficient to solve the problem

$$\min_{y \in \mathcal{A}^1[t, T]} \int_t^T L_t(s, y'(s)) ds \quad \text{s.t. } y(t) = x(t), y(T) = \theta. \quad (17)$$

Note that  $\mathcal{A}^1[t, T]$  consists of all absolutely continuous function in  $[t, T]$ .

Here, we consider a new variational problem

$$\min_{y \in \mathcal{A}^1[t, T]} \int_t^T \tilde{L}_t(s, y'(s)) ds \quad \text{s.t. } y(t) = x(t), y(T) = \theta, \quad (18)$$

where

$$\begin{aligned} \tilde{c}(\Delta) &:= |\Delta|^\alpha, \\ \tilde{L}_t(s, v) &:= \gamma_t(s) \tilde{c}(v). \end{aligned}$$

In this problem, if an optimal solution exists for problem (18), then an optimal solution also exists for problem (17), and the optimal solutions are identical; this is because the optimal solution to problem (18) must be non-increasing due to the form of  $\tilde{c}(\Delta)$ . Therefore, instead of solving problem (17), we consider solving problem (18). The function  $\tilde{L}_t(s, y'(s)) = \gamma_t(s)|y'(s)|^\alpha$  is continuously differentiable and convex with respect to  $(y(s), y'(s))$ . Also,

$$\begin{aligned} y(s) &= x(t) + (\theta - x(t)) \cdot \frac{\Gamma_t(s) - \Gamma_t(t)}{\Gamma_t(T) - \Gamma_t(t)}, \\ z(s) &= \alpha \left( \frac{\theta - x(t)}{\Gamma_t(T) - \Gamma_t(t)} \right)^{\alpha-1} \end{aligned} \quad (19)$$

satisfy the Euler–Lagrange equation (16) and  $y, z \in \mathcal{A}^1[a, b]$ . From Theorem 8, (18) has an optimal solution, and the solution is given by (19). Thus, (19) is also an optimal solution of (17), and in this case a simple calculation gives

$$\mathcal{P}[y] = \frac{(\theta - x(t))^\alpha - \zeta(t)R}{(\Gamma_t(T) - \Gamma_t(t))^{\alpha-1}}.$$

From (i) and (ii), we have

$$y_{t,x(t)}^*(s) = \begin{cases} x(t) & (\theta - x(t))^\alpha - \zeta(t)R \geq 0, \\ x(t) + \frac{(\theta - x(t))(\Gamma_t(s) - \Gamma_t(t))}{\Gamma_t(T) - \Gamma_t(t)} & \text{otherwise.} \end{cases} \quad (20)$$

Next, we solve (4). By differentiating (20), we get

$$\frac{dy_{t,x(t)}^*(s)}{ds} = \begin{cases} 0 & (\theta - x(t))^\alpha - \zeta(t)R \geq 0, \\ (\theta - x(t)) \cdot \frac{\gamma_t(s)^{-\frac{1}{\alpha-1}}}{\Gamma_t(T) - \Gamma_t(t)} & \text{otherwise.} \end{cases}$$

If the condition  $(\theta - x(\hat{t}))^\alpha - \zeta(\hat{t})R \geq 0$  is satisfied for a certain  $\hat{t}$ , then it follows from the form of (4) and non-increasingness of  $\zeta(t)$ , that  $x(t)$  remains constant for  $t \in [\hat{t}, T]$ . Let  $\tilde{t}$  denote the smallest such  $\hat{t}$ . Then,

$$\frac{dy_{t,x(t)}^*(s)}{ds} = (\theta - x(t)) \cdot \frac{\gamma_t(s)^{-\frac{1}{\alpha-1}}}{\Gamma_t(T) - \Gamma_t(t)} \quad (21)$$

holds for  $t \in [0, \tilde{t}]$ . In this case, solving (21) under the condition  $x(t) = 0$  yields

$$x(t) = \theta \left( 1 - \exp \left( - \int_0^t \frac{ds}{\Gamma_s(T) - \Gamma_s(s)} \right) \right).$$

Because

$$\begin{aligned} & (\theta - x(t))^\alpha - \zeta(t)R = \\ & \theta^\alpha \exp \left( -\alpha \int_0^t \frac{ds}{\Gamma_s(T) - \Gamma_s(s)} \right) - \zeta(t)R \end{aligned}$$

holds,  $\tilde{t}$  is the minimum  $t$  that satisfies (10).  $\square$

## References

- Ainslie, G. 1975. Specious reward: a behavioral theory of impulsiveness and impulse control. *Psychological Bulletin*, 82(4): 463.
- Akagi, Y.; Kim, H.; and Kurashima, T. 2024. A Continuous-time Tractable Model for Present-biased Agents. *arXiv preprint arXiv:2409.11225*.
- Akagi, Y.; Marumo, N.; and Kurashima, T. 2024. Analytically Tractable Models for Decision Making under Present Bias. In *Proceedings of the 38th AAAI Conference on Artificial Intelligence*, 9441–9450.
- Albers, S.; and Kraft, D. 2019. Motivating time-inconsistent agents: A computational approach. *Theory of Computing Systems*, 63(3): 466–487.

- Albers, S.; and Kraft, D. 2021. On the value of penalties in time-inconsistent planning. *ACM Transactions on Economics and Computation*, 9(3): 1–18.
- Berdichevsky, V.; and Berdichevsky, V. 2009. *Variational Principles*. Springer.
- Camerer, C. F.; and Loewenstein, G. 2004. Behavioral economics: Past, present, future. *Advances in Behavioral Economics*, 1: 3–51.
- Fedus, W.; Gelada, C.; Bengio, Y.; Bellemare, M. G.; and Larochelle, H. 2019. Hyperbolic discounting and learning over multiple horizons. *arXiv preprint arXiv:1902.06865*.
- Frederick, S.; Loewenstein, G.; and O'Donoghue, T. 2002. Time discounting and time preference: A critical review. *Journal of Economic Literature*, 40(2): 351–401.
- Hsiaw, A. 2013. Goal-setting and self-control. *Journal of Economic Theory*, 148(2): 601–626.
- Kaelbling, L. P.; Littman, M. L.; and Moore, A. W. 1996. Reinforcement Learning: A Survey. *Journal of Artificial Intelligence Research*, 4: 237–285.
- Kim, H. 2021. Fast Bayesian inference for Gaussian Cox processes via path integral formulation. *Advances in Neural Information Processing Systems*, 34: 26130–26142.
- Kleinberg, J.; and Oren, S. 2014. Time-inconsistent planning: a computational problem in behavioral economics. In *Proceedings of the 15th ACM Conference on Economics and Computation*, 547–564.
- Koch, A. K.; and Nafziger, J. 2011. Self-regulation through goal setting. *Scandinavian Journal of Economics*, 113(1): 212–227.
- Laibson, D. 1997. Goden eggs and hyperbolic discounting. *Quarterly Journal of Economics*, 112(2): 443–477.
- Lewis, F. L.; Vrabie, D.; and Syrmos, V. L. 2012. *Optimal control*. John Wiley & Sons.
- Loewenstein, G.; and Prelec, D. 1992. Anomalies in intertemporal choice: Evidence and an interpretation. *The Quarterly Journal of Economics*, 107(2): 573–597.
- O'Donoghue, T.; and Rabin, M. 2006. Incentives and self-control. *Econometric Society Monographs*, 42: 215.
- O'Donoghue, T.; and Rabin, M. 2008. Procrastination on long-term projects. *Journal of Economic Behavior & Organization*, 66(2): 161–175.
- Phelps, E. S.; and Pollak, R. A. 1968. On second-best national saving and game-equilibrium growth. *The Review of Economic Studies*, 35(2): 185–199.
- Rockafellar, R. 2001. Convex analysis in the calculus of variations. *Advances in Convex Analysis and Global Optimization: Honoring the Memory of C. Caratheodory (1873–1950)*, 135–151.
- Samuelson, P. A. 1937. A note on measurement of utility. *The Review of Economic Studies*, 4(2): 155–161.
- Schultheis, M.; Rothkopf, C. A.; and Koepl, H. 2022. Reinforcement learning with non-exponential discounting. *Advances in Neural Information Processing Systems*, 35: 3649–3662.
- Takeuchi, K. 2011. Non-parametric test of time consistency: Present bias and future bias. *Games and Economic Behavior*, 71(2): 456–478.
- Tang, P.; Teng, Y.; Wang, Z.; Xiao, S.; and Xu, Y. 2017. Computational issues in time-inconsistent planning. In *Proceedings of the 31st AAAI Conference on Artificial Intelligence*, 3665–3671.
- Wilkinson, N.; and Klaes, M. 2017. *An Introduction to Behavioral Economics*. Bloomsbury Publishing.
- Zappala, E.; Fonseca, A. H. d. O.; Moberly, A. H.; Higley, M. J.; Abdallah, C.; Cardin, J. A.; and van Dijk, D. 2023. Neural integro-differential equations. In *Proceedings of the 37th AAAI Conference on Artificial Intelligence*, 11104–11112.