Optimal Stopping under Present-Biased Preferences

Qing Li, Javad Nasiry

School of Business and Management, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong SAR. imqli, nasiry@ust.hk

Peiwen Yu

School of Management, Fudan University, Shanghai, PR China, ypw@fudan.edu.cn

Abstract

We study the optimal stopping problem in which decision makers (agents) have time-inconsistent, present-biased preferences. The agents may be naive and unaware of the bias or sophisticated and aware, and the stopping problem may involve immediate rewards or costs. We investigate whether agents follow a threshold stopping policy as well as differences in agent behavior. Prior work shows that an agent with standard time-consistent preferences follows a threshold stopping policy if the marginal payoff is monotone in state. We show that, under present-biased preferences, similar conditions are sufficient for naifs and—when rewards are immediate—for sophisticates to follow a threshold stopping policy. Yet if costs are immediate, then sophisticates might not follow a threshold policy because of preemptive stopping to avoid future self-control problems. We also compare the sets of states under which different agents stop. Sophisticates are always more likely to stop than naifs. When rewards are immediate, naifs are more likely to stop than time-consistent agents; however, the converse holds when costs are immediate. These findings extend the literature to settings where the states of agents are important. We discuss implications of our results in two stopping-problem examples, one in project management and the other in health care.

Keywords: optimal stopping, present-biased preferences, quasi-hyperbolic discounting, dynamic programming

1 Introduction

In an optimal stopping problem, a decision maker (agent) maximizes her expected utility by making a binary choice, in each time period, to stop or to continue. The decision-making environment evolves stochastically, and the problem terminates when the decision maker stops (Ross 1983).

Optimal stopping problems appear frequently in operations, finance, marketing, and economics (for a review, see Oh and Özer 2016) and are often studied under exponential discounting, which assumes that the discount rate between any two time periods is the same irrespective of when the rate is evaluated. However, this assumption is not "psychologically or normatively plausible" (Frederick et al. 2002). Decision makers are biased toward immediate gratification, and they prefer to enjoy rewards now and to postpone costs. Writing a review, for example, involves immediate costs and possible future benefits; in contrast, smoking involves immediate rewards and delayed consequences. Hence the reviewer is discouraged from writing the review and the smoker is encouraged to indulge. Such preferences are known as *present-biased* preferences (O'Donoghue and Rabin 1999, 2001). Our goal in this paper is to study optimal stopping problems under such preferences when either rewards or costs are immediate.

Exponential time discounting implies that the agent exhibits time-consistent behavior; in other words, the choice between two alternatives is independent of when the comparison is made. Yet present-biased preferences lead to time inconsistency because preferences may change over time. For instance, suppose that spending an hour on some unpleasant task a month from now is preferable to spending two hours on the same task two months from now. After a month has elapsed, however, most people postpone the task instead of doing it as originally planned. Thus, agents exhibit self-control problems. The decision maker may be naive or sophisticated—that is, she may not foresee her future self-control problems and plan just as if she were time consistent, or she may be aware of the bias and plan accordingly.

Following the extant literature, we model present-biased preferences by assuming that the decision maker has quasi-hyperbolic time preferences; this is often referred to as the (β, δ) model (see, e.g., Laibson 1997). In this framework, the one-period discount factor for the immediate future is $\beta\delta$, where a lower β corresponds to a stronger bias for immediate payoffs. However, the one-period discount factor for all future payoffs is δ. Because naifs are unaware of their future self-control problems, when making decisions in the current period they mistakenly suppose that their long-run utility is the same as that of a time-consistent (TC) agent. We can therefore define their decision-making problem by way of a dynamic programming recursion. Unlike naifs, sophisticates are fully aware of their future self-control problems and can be viewed as a collection of distinct selves, each making decisions based on her own preferences prevailing at the time. To identify the choices made by each self, we solve the multi-player game by backward induction and find the subgame-perfect Nash equilibrium.

In order to establish the problem's structural properties, we focus on the *marginal payoff*, or the difference in payoff between stopping now and stopping one period later. Prior work shows that if the marginal payoff is monotone in state, then a TC agent follows a threshold stopping policy (see, e.g., Oh and Özer 2016). We find that similar conditions are sufficient for naifs to follow a threshold stopping policy under both immediate costs and immediate rewards. For sophisticates, however, these conditions guarantee a threshold structure only when rewards are immediate. When costs are immediate, sophisticates might not generally follow a threshold policy because they might preemptively stop to avoid future self-control problems.

We compare the sets of states under which different decision makers stop. When rewards are immediate, the set of "stopping states" for TC agents (TCs, hereafter) is a subset of that for naifs, which itself is a subset of that for sophisticates. This means that sophisticates are more likely to stop and collect the immediate rewards than are naive or time-consistent agents. The reason is that naifs, because they are unaware of their future self-control problems, value the option to continue more than do sophisticates. In addition, if we assume that time-consistent behavior is ideal, then our results show that sophistication induces premature stopping decisions and hence does not benefit the agent.

When costs are immediate, the set of stopping states for naifs is a subset of that for time-consistent agents and a subset of that for sophisticates. In other words, naifs are the least likely to stop (the most likely to postpone). However, the comparison between time-consistent agents and sophisticates is in general indeterminate owing to the opposing effects of present-biased preferences and sophistication. In this case, we provide sufficient conditions for the present-bias effect to dominate the sophistication effect; under these conditions, sophisticates value the option to continue more than do TCs and hence are less likely than the latter to stop. Thus, sophistication mitigates procrastination and benefits the agent. Our results are general in the sense that they do not depend on whether the stopping decision follows a threshold structure.

The operations management literature on present-biased preferences is sparse. Su (2009) develops a model of "consumer inertia" in which consumers delay purchases even when it is optimal to purchase immediately. He argues that such inertial behavior is consistent with hyperbolic time preferences of naive agents in a purchase decision setup that is characterized as an immediate cost problem (the payment precedes the consumption). Plambeck and Wang (2013) study the pricing and scheduling of a service with immediate costs when customers have quasi-hyperbolic preferences. They show that charging for subscription is optimal for the service provider, especially when customers are naive. In a project management context characterized by costly immediate efforts, Wu et al. (2014) study the optimal contract design and team composition for achieving project goals when the workers have hyperbolic time preferences. Gao et al. (2014) study the dynamic pricing problem of a monopolist selling to strategic consumers with quasi-hyperbolic preferences and show that, from the seller's perspective, a policy of "cream skimming" (reducing prices) is generally optimal. Unlike ours, these studies do not consider present-biased preferences in the framework of stopping problems. The optimal stopping framework can be used to model important operational problems such as those in project management and in deciding when to undertake medical examinations. These problems have been traditionally studied under exponential discounting. Our results allow us to understand when and why an agent with present-biased preference behaves differently, which enriches the operations management literature.

Present-biased preferences have been studied extensively in economics under various setups; see Frederick et al. (2002) for a review. In the particular area of stopping problems, O'Donoghue and Rabin (1999) are closest to our research. In their model, there is no uncertainty and the payoffs are not dependent on states. Hence, structural properties are irrelevant, and the comparative results focus on the time at which different agents stop—not on the sets of states (i.e., stopping regions) that we are interested in.

There are alternative methods in economics for modeling self-control problems and their consequences in special cases of optimal stopping time problems. Miao (2008) uses the "temptation proneness" model of Gul and Pesendorfer (2001) to show that temptation can result in either procrastination or preproperation (excessive haste) in an optimal option exercise problem. Fudenberg and Levine (2006) obtain similar results with a "dual-self" model in which the decision-making process is modeled as a game between a short-run impulsive self and a long-run patient self. We apply the quasi-hyperbolic discounting framework originally developed by Phelps and Pollak (1968) in a more general stopping problem.

The rest of the paper is organized as follows. Section 2 outlines the basic components of our model for different agents. In Section 3, we establish the structural properties of the optimal stopping problem under present-biased preferences. Section 4 compares the stopping behavior of different types of agents. In Section 5, we show how our model applies in operational settings characterized by either immediate rewards or immediate costs. Section 6 concludes.

2 Model

We assume that time is discrete and use s_t to denote as the state in period t. In each period $t \geq 1$, an agent must decide whether to stop (decision 1) or to continue (decision 0). If she stops, the process ends and she receives a payoff $r_t^1(s_t)$ immediately as well as a payoff $r_t^f(s_t)$ in the future. If she continues, she receives a payoff $r_t^0(s_t)$ immediately and the state will become $S_{t-1}(s_t)$ in the next period $t-1$. Here $S_{t-1}(s_t)$ is a random variable dependent on s_t. We assume that if the agent does not stop in all periods $t \geq 1$ then she must stop at period 0. All payoffs can be either positive or negative.

In this paper, we adopt the following simple framework to model present-biased preferences; it was developed by Phelps and Pollak (1968) and has since been extensively used in the literature.

Definition 1 *Let c_i be an agent's instantaneous utility in period i. She has (β, δ)-preferences if her intertemporal utility in period t can be represented by*

$$U_t(c_t, c_{t-1}, \ldots, c_1) = c_t + \beta \sum_{i=1}^{t-1} \delta^{t-i} c_i,$$

where $0 < \beta, \delta \leq 1$.

Without loss of generality, we let $\delta = 1$ throughout the paper. When $\beta = 1$, the preferences are time consistent.

Let $v_t(s_t)$ denote the maximum total expected payoff from period t to the end of the horizon when the state is s_t. For TCs, the dynamic programming recursion can be written as

$$v_t(s_t) = \max\{r_t^1(s_t) + r_t^f(s_t), r_t^0(s_t) + \mathsf{E}v_{t-1}(S_{t-1}(s_t))\},$$

where $v_0(s_0) = r_0^1(s_0) + r_0^f(s_0)$. Again without loss of generality, we assume that if stopping and continuing yield the same expected payoff, the agent chooses to continue.

Agents may have different beliefs about their present-biased preferences. A naive agent is unaware that she is time inconsistent; she believes that her future behavior is the same as that of a time-consistent agent. Therefore, naifs will stop if and only if

$$r_t^1(s_t) + \beta r_t^f(s_t) > r_t^0(s_t) + \beta \mathsf{E}v_{t-1}(S_{t-1}(s_t)).$$

In contrast, a sophisticated agent can foresee her self-control problems and correctly anticipates her future behavior. We let $u_{t-1}(s_{t-1})$ denote a sophisticated agent's expected payoff from period $t-1$ to the end of the horizon. Then a sophisticate will stop if and only if

$$r_t^1(s_t) + \beta r_t^f(s_t) > r_t^0(s_t) + \beta \mathsf{E}u_{t-1}(S_{t-1}(s_t)).$$

Here $u_{t-1}(s_{t-1})$ is the perceived value function from period-t self's perspective, in which all future payoffs are weighted equally. This function reflects the *long-run utility* of sophisticates (O'Donoghue and Rabin 1999). In general, $u_t(s_t)$ is not equal to the maximum of the two terms $r_t^1(s_t) + \beta r_t^f(s_t)$ and $r_t^0(s_t)+\beta \mathsf{E}u_{t-1}(S_{t-1}(s_t))$. Instead, a sophisticated agent compares these two terms to determine her optimal action at period t and then—based on the optimal action—updates $u_t(s_t)$ according to the following recursive relationship:

$$u_t(s_t) \quad = \quad \begin{cases} r_t^0(s_t) + \mathsf{E}u_{t-1}(S_{t-1}(s_t)) & \text{if} \quad \textit{it is optimal to continue at } s_t; \\ r_t^1(s_t) + r_t^f(s_t) & \text{if} \quad \textit{it is optimal to stop at } s_t. \end{cases}$$

We also assume that $u_0(s_0) = r_0^1(s_0) + r_0^f(s_0)$.

We consider two cases. If $r_t^1(s_t) \geq r_t^0(s_t)$ then there is an immediate benefit from stopping relative to continuing; thus the agent receives an "immediate reward" from stopping. In contrast, if $r_t^1(s_t) < r_t^0(s_t)$ then there is an immediate cost from stopping relative to continuing; so in this case, the agent incurs an "immediate cost" from stopping. The distinction is immaterial for TCs but is critical for naifs and sophisticates.

3 Threshold Stopping Structure

A question often asked in stopping problems is whether—and under what conditions—a threshold policy is optimal. Under a threshold policy, stopping is the optimal action if and only if the state is greater (or smaller) than a threshold. Threshold structure reveals complementarity (or substitutability) between states and actions; it also enables easy computation and implementation of the optimal policy.

3.1 Time-consistent Agents

The conditions under which a threshold policy is optimal for TCs are well established in the literature (see, e.g., Oh and Özer 2016). We summarize the results here. First we define the marginal payoff of continuing for one more period:

$$M_t^{TC}(s_t) = r_t^0(s_t) + \mathsf{E}[r_{t-1}^1(S_{t-1}(s_t)) + r_{t-1}^f(S_{t-1}(s_t))] - r_t^1(s_t) - r_t^f(s_t).$$

The marginal payoff represents the difference in payoff between (a) continuing for one more period and then stopping and (b) stopping immediately in the current period. If the marginal payoff is decreasing (resp., increasing) in state and the state transition is stochastically increasing, then it is optimal to stop if and only if the state is above (resp., below) a certain threshold. The concept of stochastic increasing is defined next.

Definition 2 *A set of random variables* $\{X(\theta) \mid \theta \in \mathbb{R}\}$ *is stochastically increasing in* θ *if* $\mathsf{E}g(X(\theta))$ *is increasing in* θ *for all increasing functions* g*.*

Throughout we shall make the following assumption on state transitions.

Assumption 1 *The state transition* $S_{t-1}(s_t)$ *is stochastically increasing in* s_t*.*

According to this assumption, a larger state s_t in the current period leads to a stochastically larger state S_{t-1} in the next period. For instance, Assumption 1 is met when the state transition takes the following multiplicative and additive form: $S_{t-1}(s_t) = D_{t,1}s_t + D_{t,2}$, where $D_{t,1}$ and $D_{t,2}$ are random variables and $D_{t,1}$ is positive. The implications of this assumption are discussed (within different problem contexts) in Section 5.

The following result on threshold policies is adopted from Oh and Özer (2016).

Lemma 1 *The following statements hold in both the immediate reward and the immediate cost cases:*

(i) if $M_t^{TC}(s_t)$ *is decreasing in* s_t*, then there exists a threshold* s_t^{TC} *such that TCs stop if and only if* $s_t \geq s_t^{TC}$*;*

(ii) if $M_t^{TC}(s_t)$ *is increasing in* s_t*, then there exists a threshold* s_t^{TC} *such that TCs stop if and only if* $s_t \leq s_t^{TC}$*.*

3.2 Naive Agents

Much as we did for time-consistent agents, for naifs we define the marginal payoff of continuing for one more period:

$$M_t^n(s_t) = r_t^0(s_t) + \beta\mathsf{E}[r_{t-1}^1(S_{t-1}(s_t)) + r_{t-1}^f(S_{t-1}(s_t))] - r_t^1(s_t) - \beta r_t^f(s_t).$$

This definition reflects the present-bias effect. Unlike the time-consistent case, here the monotonicity of M_t^n is not sufficient to guarantee a threshold stopping rule. A naive agent's period-t self regards $M_t^n(s_t)$ as the marginal payoff of continuing for one more period and then stopping at period $t-1$. However, she is unaware of her future self-control problems and believes that her future selves would behave like TCs. She will therefore likewise view $M_i^{TC}(s_i)$ as her marginal payoffs of continuation in future periods ($i \leq t-1$). A threshold policy is not guaranteed unless both M_t^{TC} and M_t^n are monotone. The results are presented formally in the following theorem.

Theorem 1 *The following statements hold in both the immediate reward and the immediate cost cases:*

(i) if both $M_t^{TC}(s_t)$ and $M_t^n(s_t)$ are decreasing in s_t, then there exists a threshold s_t^n such that naifs stop if and only if $s_t \geq s_t^n$;

(ii) if both $M_t^{TC}(s_t)$ and $M_t^n(s_t)$ are increasing in s_t, then there exists a threshold s_t^n such that naifs stop if and only if $s_t \leq s_t^n$;

Under some conditions, the monotonicity of $M_t^n(s_t)$ implies the monotonicity of $M_t^{TC}(s_t)$; under other conditions, the opposite is true. The relevant conditions are given in the following lemma. In applications, if these conditions are satisfied then we need only check the monotonicity of either $M_t^{TC}(s_t)$ or $M_t^n(s_t)$; we do not need to check both.

Lemma 2

(i) Suppose $r_t^1(s_t) - r_t^0(s_t)$ is decreasing in s_t. Then $M_t^{TC}(s_t)$ is decreasing in s_t if $M_t^n(s_t)$ is decreasing in s_t.

(ii) Suppose $r_t^1(s_t) - r_t^0(s_t)$ is increasing in s_t. Then $M_t^n(s_t)$ is decreasing in s_t if $M_t^{TC}(s_t)$ is decreasing in s_t.

3.3 Sophisticated Agents

Unlike naifs, sophisticates correctly anticipate their future behavior. Although $M_t^n(s_t)$ captures the present-bias effect of the period-t self, it does not capture future selves' present-bias effect, of which the period-t self is fully aware and accounts for when pondering whether to stop or continue. As a result, the sophisticate's stopping behavior may not follow a threshold policy even when both $M_t^n(s_t)$ and $M_t^{TC}(s_t)$ are monotone.

Consider the following example with a two-period planning horizon and immediate costs. There are two possible states in each period: $s_t \in \{0, 1\}$. Assume the state remains unchanged over time; that is, let $S_{t-1}(s_t) = s_t$. Then we clearly have a stochastically increasing state transition. Let $\beta = 0.5$, and suppose $r_t^0(s_t) = r_0^0(s_0) = r_0^1(s_0) = r_0^f(s_0) = 0$. The reward and cost schedules are

$$r_2^1(s_2) = -2, \quad r_2^f(s_2) = 5 + 4s_2;$$
$$r_1^1(s_1) = -5, \quad r_1^f(s_1) = 9 + 3s_1.$$

By definition of $M_t^n(s_t)$, it is easy to see that

$$M_1^n(0) = 0.5, \qquad M_1^n(1) = -1;$$
$$M_2^n(0) = 1.5, \qquad M_2^n(1) = 1.$$

Thus $M_t^n(s_t)$ is decreasing in s_t and $M_t^{TC}(s_t)$ is also decreasing in s_t by Lemma 2. The actions taken by different agents and the payoffs from those actions can be computed using backward induction. Table 1 summarizes the payoffs from different actions for all types of agents.

	Period 2				Period 1			
	State 0		State 1		State 0		State 1	
	Stop	Continue	Stop	Continue	Stop	Continue	Stop	Continue
TCs	3	4	7	7	4	0	7	0
Naifs	0.5	2	2.5	3.5	−0.5	0	1	0
Sophisticates	0.5	0	2.5	3.5	−0.5	0	1	0

Table 1: Payoffs of different actions for different types of agents

Observe from Table 1 that TCs always stop in period 1 and continue in period 2, irrespective of the states. Further, naifs only stop at state 1 in period 1. Both are threshold policies in line with Lemma 1 and Theorem 1. However, sophisticates stop at state 0 but not at state 1 in period 2 yet stop at state 1 but not at state 0 in period 1. Such behavior is obviously not reflective of a threshold policy.

When the state is 1, for sophisticates the period-2 self's discounted payoff is $-2 + 9/2 = 2.5$ when stopping. If the sophisticate continues, she correctly anticipates that she will end up stopping in period 1 and so receive a discounted payoff of $(-5 + 12)/2 = 3.5$, which is greater than 3. In this case, the period-2 self continues. When the state is 0, if the period-2 self stops then her discounted payoff is $-2+5/2 = 0.5$. If the sophisticate continues, she also correctly anticipates that her period-1 self will likewise continue and hence will end up with a discounted payoff of 0. So in this case, the period-2 self chooses to stop. Just like TCs, sophisticates weight all future payoffs equally, which future selves disagree. In this example, if the state is 0 then, despite the period-2 self's wishes to continue and let the period-1 self stop in period 1, she knows that the period-1 self would not stop. The period-2 self is thus better off by stopping immediately.

The non-threshold stopping behavior occurs when the costs are immediate. When rewards are immediate, however, the monotonicity of $M_t^n(s_t)$ and $M_t^{TC}(s_t)$ guarantees a threshold policy for sophisticated agents. In this case, if the present self plans to continue for one period and then stop, the next-period self would never object because the immediate reward from stopping is then even more irresistible. This point will become clearer in Section 4, where we compare the behavior of TCs, naifs, and sophisticates. Results concerning threshold policies are presented in our next theorem.

Theorem 2 *In the immediate reward case, the following statements hold:*

(i) if both $M_t^{TC}(s_t)$ and $M_t^n(s_t)$ are decreasing in s_t, then there exists a threshold s_t^s such that sophisticates stop if and only if $s_t \geq s_t^s$;

(ii) if both $M_t^{TC}(s_t)$ and $M_t^n(s_t)$ are increasing in s_t, then there exists a threshold s_t^s such that sophisticates stop if and only if $s_t \leq s_t^s$.

4 Comparative Statics

We now compare the stopping behavior of different types of agents. The notation R_t^i represents the *stopping region* for a type-i agent at period t for $i \in \{TC, n, s\}$. That is, for type-i agents, stopping is better than continuing if and only if $s_t \in R_t^i$. In the following theorem, we compare these regions and describe how they change as a function of β. In the examples given by O'Donoghue and Rabin (1999)— in which there is no uncertainty and states are irrelevant—the stopping regions are either empty or infinite.

Theorem 3

(i) If rewards are immediate, then $R_t^{TC} \subseteq R_t^n \subseteq R_t^s$ and both R_t^n and R_t^s are decreasing in β.

(ii) If costs are immediate, then $R_t^n \subseteq R_t^s$, $R_t^n \subseteq R_t^{TC}$ and R_t^n is increasing in β.

By this theorem, for a given state, if naifs stop then sophisticates also stop—regardless of whether it is the costs or the rewards that are immediate. This is because naive agents, being unaware of their self-control problem, assign a higher value to the option of continuing than do sophisticated agents.

When *rewards* are immediate, naifs are more likely to stop than TCs because they believe they will behave like TCs in the future but are tempted by immediate rewards now. When β increases, the stopping regions for naifs and sophisticates are moving closer to that of TCs. When *costs* are immediate, naifs are less likely to stop than TCs because naifs weight more heavily the immediate cost from stopping. When β increases, the stopping region for naifs increases and hence is closer to that of TCs.

Somewhat surprisingly, these results are very robust and do not depend on whether stopping follows a threshold policy; neither do they depend on whether stopping is optimal when the state is low or when it is high. Suppose that a threshold policy is optimal for TCs, naifs, and sophisticates and that their respective thresholds are denoted s_t^{TC}, s_t^n, and s_t^s. In the immediate reward case, for example, it follows from Theorem 3 that if stopping is optimal if and only if the state is higher than the thresholds, then $s_t^s \leq s_t^n \leq s_t^{TC}$; if stopping is optimal if and only if the state is lower than the thresholds, then $s_t^s \geq s_t^n \geq s_t^{TC}$.

Theorem 3 allows us to better understand why sophisticates may follow a non-threshold policy when *costs* are immediate. Given that naifs follow a threshold policy, any non-threshold behavior must be driven by sophistication. In other words, sophisticates may stop preemptively to avoid future self-control problems whereas naifs—unaware of the self-control problem—may continue. Take, for example, the case where both M_t^n and M_t^{TC} are decreasing. To simplify the arguments, suppose states remain unchanged over time and there are only two states: low and high. Suppose further that there is a threshold policy for periods $t-1, t-2, \ldots, 1$ (this is true at least for period 1). Now assume that, at the low state, the present self would want the period-$(t-1)$ self to stop but knows that her period-$(t-1)$ self would actually continue. As a result, she stops preemptively. When we switch to the high state, the period-$(t-1)$ self might continue but might also stop. In the latter event, the self-control problem becomes moot and there is no longer any need for preemptive stopping at period t. The consequence would be a non-threshold policy: stopping at the low state but not at the high state.

When rewards are immediate, however, self-control problems have different effects. In the low state, we suppose similarly that—for the present self's perspective—the best time to stop is at some period $t' < t-1$. Sophisticates foresee that their period-$(t-1)$ selves will not be able to resist the temptation and so will stop at period $t-1$, which for the present self is not as good as stopping at period t. Therefore, the present self stops preemptively. When we change to the high state, the period-$(t-1)$ self will also stop because of the threshold policy. Because M_t^n is decreasing, if, in the low state, stopping now is better than continuing for exactly one period followed by stopping, then the same holds in the high state.

The comparison between R_t^{TC} and R_t^s is in general indeterminate when costs are immediate. As noted by O'Donoghue and Rabin (1999), in their examples the behavior of sophisticates can be explained by two effects: the *present-bias effect* leads to procrastination while the *sophistication* *effect* leads to preproperation. These two effects work in opposite directions. The sophistication effect alleviates procrastination. However, it is possible that sophisticates stop even sooner than TCs; that is, the former might exercise preemptive overcontrol. In our setting, the sophistication effect makes the stopping region for sophisticates even larger than that for TCs. The next theorem identifies conditions under which such preemptive overcontrol would not occur.

Theorem 4 *When costs are immediate, $R_t^s \subseteq R_t^{TC}$ is true if one of the following conditions holds for all states s_t and periods t:*

(i) $r_t^f(s_t) - \mathsf{E}r_{t-1}^f(S_{t-1}(s_t)) - \mathsf{E}r_{t-1}^0(S_{t-1}(s_t)) \geq 0$*;*

(ii) $r_t^0(s_t) - r_t^1(s_t) \geq \mathsf{E}r_{t-1}^0(S_{t-1}(s_t)) - \mathsf{E}r_{t-1}^1(S_{t-1}(s_t))$*;*

(iii) $M_t^{TC}(s_t) \leq 0$.

Under these sufficient conditions, it pays to be sophisticated. That is, sophistication under these conditions can lead agents who suffer from present-bias effect to make choices that are closer to those that TCs would make—which long-run selves would appreciate. Condition (iii) in Theorem 4 is obvious since TCs always stop in this case. Condition (ii) means that the net payoffs of stopping are decreasing in time. Condition (i) is a necessary condition for $r_t^f(s_t) - u_{t-1}(s_t) \geq r_{t-1}^f(s_t) - u_{t-2}(s_t)$, which means that the net future payoffs are decreasing in time. Under these conditions, the future self-control problems are not that severe, of which sophisticates are fully aware. Hence the phenomenon of preemptive overcontrol does not arise.

Another way to understand (i) and (ii) is that, under these conditions, the long-run utility of sophisticates (u_t) is not much lower than that of TCs (v_t). In particular, we can show under condition (i) that

$$u_t(s_t) \geq \frac{1}{\beta}[v_t(s_t) - (1-\beta)(r_t^0(s_t) + r_t^f(s_t)]$$

and under condition (ii) that

$$u_t(s_t) \geq \frac{1}{\beta}[\beta v_t(s_t) - (1-\beta)(r_t^0(s_t) - r_t^1(s_t)].$$

Although the present-bias effect and sophistication effect are known in the literature, we are the first to derive conditions under which one effect dominates the other. These conditions will be further discussed in a specific context in Section 5.

5 Examples

Here we provide two examples of stopping problems. In one, rewards are immediate; in the other, costs are immediate. We discuss the conditions and results derived in previous sections in these specific contexts.

5.1 Project Management

Suppose a manager is considering when to stop a product development project. The state s_t measures the product's performance, and a higher state corresponds to a better performance. According to Assumption 1, the better the performance in the current period, the stochastically better the performance in the next period. If the manager continues the process, she incurs a development cost $r_t^0(s_t) < 0$ in the current period; if she stops the project, she incurs no cost in the current period and so $r_t^1(s_t) = 0$. The reward from the project $r_t^f(s_t)$ will come in the future. This is an immediate reward case even though the current-period payoffs are negative.

Suppose that $r_t^0(s_t)$ and $\mathsf{E}r_{t-1}^f(S_{t-1}(s_t)) - r_t^f(s_t)$ are both decreasing in s_t. The former means that it is more costly to improve an already high performance; the latter means that the marginal benefit of continuing the project is lower when the performance is higher. So in this scenario, TCs, naifs and sophisticates all stop the process if and only if product performance exceeds certain thresholds. Theorem 3 implies that, from the time-consistent perspective, naifs may be stopping when the performance is not sufficiently high and that sophisticates may be stopping when the performance is even less sufficient.

Suppose $r_t^0(s_t)$ and $\mathsf{E}r_{t-1}^f(S_{t-1}(s_t)) - r_t^f(s_t)$ are both increasing in s_t. This could happen if, for example, r_t^0 is a fixed cost (i.e., is independent of s_t), $S_{t-1}(s_t) = s_t + 1$, and $r_t^f(s_t) = b(s_t - a)^+$ for all t. The functional form of r_t^f means that if performance is below a then there is no reward; this kind of payoff has been much discussed in the project management literature (e.g., Huchzermeier and Loch 2001; Santiago and Vakili 2005). In this case, TCs, naifs, and sophisticates all stop the process if and only if the performance is below certain thresholds; in other words, they abandon the project if progress seems hopelessly slow. Here TCs view naifs and sophisticates as quitters and naifs think the same of sophisticates.

In summary, regardless of why the agents may stop, the comparative statics results about the impact of present-biased preferences and sophistication are robust.

5.2 Undertaking Medical Examinations

A person needs to decide whether and when to take a medical examination to detect for potential cancer. Examinations are costly and generate unpleasant feelings. The state s_t represents the individual's health status, where a higher state corresponds to being less healthy. Assumption 1 means that the more unhealthy a person is now, the more unhealthy she will be (stochastically) in the future. This is a common assumption in the healthcare literature. Under these circumstances, it is reasonable to assume that $r_t^0(s_t) = 0$ and $r_t^1(s_t) < 0$ and that $r_t^1(s_t)$ is independent of the state. The future reward $r_t^f(s_t)$ represents the health benefit after undertaking the medical examination. This is an immediate cost case.

If $\mathsf{E}r_{t-1}^f(S_{t-1}(s_t)) - r_t^f(s_t)$ is decreasing in s_t, then both $M_t^{TC}(s_t)$ and $M_t^n(s_t)$ are decreasing. This condition means that the marginal health benefit from delaying the examination is lower (or that the marginal cost is higher) when the subject is more unhealthy. By Lemma 1 and Theorem 1, both TCs and naifs choose to undertake the examination if and only if the health status is higher (worse) than certain thresholds. Sophisticates may or may not use a threshold stopping rule. According to Theorem 3(ii), there are situations where TCs decide to take the examination but naifs postpone taking it, and there are also situations where sophisticates decide to take the examination but naifs postpone taking it.

For a given health status, do sophisticates ever undergo the examination when TCs do not? If delaying medical examinations brings no health benefit—that is, if $r_t^f(s_t) \geq \mathsf{E}r_{t-1}^f(S_{t-1}(s_t))$, which is probably true in most cases—then condition (i) in Theorem 4 is satisfied. And since the cost of taking the examination is most likely fixed and independent of time, it follows that condition (ii) in Theorem 4 is also satisfied. Therefore, sophisticates would not exercise overcontrol. In this context, sophistication can alleviate the negative effects created by present-bias effect, which is appreciated by the long-run selves.

These results provide an alternative behavioral perspective from which we can better understand why (besides reasons of regret, worry or perceived risk) individuals might decline to engage in such preventive medical interventions as vaccination (Chapman and Coups 2006; Connolly and Reb 2003). If present-biased preferences are a driver of individual health decisions then the nature of remedies to improve those decisions will be affected.

6 Conclusion

In this paper, we study optimal stopping problems under general assumptions about the payoff structure and state transitions when decision makers have quasi-hyperbolic time preferences. We show that marginal payoff's monotonicity in state suffices for a threshold stopping policy to be optimal for all types of agents when the rewards are immediate. However, when costs are immediate, these conditions ensure the optimality of threshold policy only for time-consistent agents and naive agents. In this case, sophisticates might not follow a threshold policy because they are inclined to taking preemptive action that will preclude future self-control problems. We also compare the behavior of different agents and show that, when rewards are immediate, sophisticated agents are the most likely to stop prematurely. When costs are immediate, naive agents are the ones most likely to procrastinate.

Our model applies to a wide array of binary stop–continue decisions in project management, financial decisions, health-care decisions, and marketing. For example, unpleasant and costly preventive medical interventions such as vaccination may be postponed by present-biased patients, exacerbating their health status. Future research could focus on the implications of our findings in a variety of operational problems and on their possible remedies. In studying these problems, it would also be interesting to incorporate the interaction between the agents' risk attitudes and their time preferences (Baucells et al. 2017).

References

Baucells, M., N. Osadchiy, A. Ovchinnikov. 2017. Behavioral anomalies in consumer wait-or-buy decision and their implications for markdown management. *Operations Research* **65**(2) 357-378.

Chapman, G. B., E. J. Coups. 2006. Emotions and preventive health behavior: Worry, regret, and influenza vaccination. *Health Psychology* **25**(1) 82-90.

Frederick, S., G. Loewenstein, T. O'Donoghue. 2002. Time discounting and time preference: A critical review. *Journal of Economic Literature* **40**(2) 351-401.

Fudenberg, D., D. K. Levine. 2006. A dual-self model of impulse control. *American Economic Review* **96**(5) 1449-1476.

Gao, X., X. Chen, Y-J. Chen. 2014. Dynamic pricing with time-inconsistent consumers. Working paper.

Gul, F., Pesendorfer, W. 2001. Temptation and self-control. *Econometrica* **69**(6) 1403-1436.

Huchzermeier, A., C. H. Loch. 2001. Project management under risk: Using the real options approach to evaluate flexibility in R&D. *Management Science* **47**(1) 85-101.

Laibson, D. 1997. Golden eggs and hyperbolic discounting. *Quarterly Journal of Economics* **112**(2) 443-477.

Miao, J. 2008. Option exercise with temptation. *Economic Theory* **34**(3) 473-501.

O'Donoghue, T., M. Rabin. 1999. Doing it now or later. *American Economic Review* **89**(1) 103-124.

O'Donoghue, T., M. Rabin. 2001. Choice and procrastination. *Quarterly Journal of Economics* **116**(1) 121-160.

Oh, S., Ö. Özer. 2016. Characterizing the structure of optimal stopping problems. *Production and Operations Management* **25**(11) 1820-1838.

Phelps, E. S., R.A. Pollak. 1968. On second-best national saving and game-equilibrium growth. *Rev. Econom. Stud.* **35**(2) 185-199.

Plambeck, E., Q. Wang. 2013. Implications of hyperbolic discounting for optimal pricing and scheduling of unpleasant services that generate future benefits. *Management Science* **59**(8) 19270-1946.

Ross, S. M. 1983. Introduction to stochastic dynamic programming. Academic Press.

Santiago, L. P., P. Vakili. 2005. On the value of flexibility in R&D projects. *Management Science* **51**(8) 1206-1218.

Su, X. 2009. Model of consumer inertia with applications to dynamic pricing. *Production and Operations Management* **18**(4) 365-380.

Wu, Y., K. Ramachandran, V. Krishnan. 2014. Managing cost salience and procrastination in projects: Compensation and team composition. *Production and Operations Management* **23**(8) 1299-1311.

Appendix

We first introduce the following notation for (respectively) TCs, naifs, and sophisticates.

$$B_t^{TC}(s_t) = r_t^0(s_t) + \mathsf{E}v_{t-1}(S_{t-1}(s_t)) - r_t^1(s_t) - r_t^f(s_t),$$
$$B_t^n(s_t) = r_t^0(s_t) + \beta\mathsf{E}v_{t-1}(S_{t-1}(s_t)) - r_t^1(s_t) - \beta r_t^f(s_t),$$
$$B_t^s(s_t) = r_t^0(s_t) + \beta\mathsf{E}u_{t-1}(S_{t-1}(s_t)) - r_t^1(s_t) - \beta r_t^f(s_t).$$

Recall from Section 2 that a type i agent stops at state s_t in period t if and only if $B_t^i(s_t) \leq 0$.

Proof of Lemma 1. We prove only part (i) of the lemma; part (ii) can be proved similarly. First, we can show that the following relationship between $B_t^{TC}(s_t)$ and $M_t^{TC}(s_t)$ holds:

$$\begin{aligned}
B_t^{TC}(s_t) &= r_t^0(s_t) + \mathsf{E}v_{t-1}(S_{t-1}(s_t)) - r_t^1(s_t) - r_t^f(s_t) \\
&= M_t^{TC}(s_t) + \mathsf{E}(v_{t-1}(S_{t-1}(s_t)) - r_{t-1}^1(S_{t-1}(s_t)) - r_{t-1}^f(S_{t-1}(s_t))) \\
&= M_t^{TC}(s_t) + \mathsf{E}\max\{0, r_{t-1}^0(S_{t-1}(s_t)) + \mathsf{E}v_{t-2}(S_{t-2}(S_{t-1}(s_t))) - r_{t-1}^1(S_{t-1}(s_t)) - r_{t-1}^f(S_{t-1}(s_t))\} \\
&= M_t^{TC}(s_t) + \mathsf{E}\max\{0, B_{t-1}^{TC}(S_{t-1}(s_t))\}.
\end{aligned}$$

We shall identify the threshold policy by using induction to show that $B_t^{TC}(s_t)$ is decreasing in s_t. By assumption, $B_1^{TC}(s_1) = M_1^{TC}(s_1)$ is decreasing in s_1. Suppose $B_{t-1}^{TC}(s_{t-1})$ is decreasing in s_{t-1} for some $t-1 \geq 1$, then the function $\max\{0, B_{t-1}^{TC}(s_{t-1})\}$ is decreasing s_{t-1}. Since the state variable $S_{t-1}(s_t)$ is stochastically increasing in s_t, it follows that the function $\mathsf{E}\max\{0, B_{t-1}^{TC}(S_{t-1}(s_t))\}$ is decreasing in s_t. By assumption, $M_t^{TC}(s_t)$ is decreasing and so $B_t^{TC}(s_t)$ is decreasing in s_t, which completes the induction.

Proof of Theorem 1. The proof is similar to the one for Lemma 1. Note that we have the following relationship between $B_t^n(s_t)$ and $M_t^n(s_t)$:

$$B_t^n(s_t) = M_t^n(s_t) + \beta \mathsf{E}\max\{0, B_{t-1}^{TC}(S_{t-1}(s_t))\}.$$

Here, the monotonicity of $M_t^{TC}(s_t)$ is needed to ensure the monotonicity of $B_t^{TC}(s_t)$. Then, given our additional assumption on the monotonicity of $M_t^n(s_t)$, we conclude that threshold stopping rules are optimal for naifs.

Proof of Lemma 2. The results are an immediate consequence of the following two equalities:

$$M_t^{TC}(s_t) = \frac{1}{\beta} M_t^n(s_t) + (\frac{1}{\beta} - 1)(r_t^1(s_t) - r_t^0(s_t));$$

and

$$M_t^n(s_t) = \beta M_t^{TC}(s_t) - (1-\beta)(r_t^1(s_t) - r_t^0(s_t)).$$

Proof of Theorem 2. We prove only part (i), since part (ii) can be shown similarly. For sophisticates,

$$\begin{aligned} B_t^s(s_t) &= r_t^0(s_t) + \beta \mathsf{E} u_{t-1}(S_{t-1}(s_t)) - r_t^1(s_t) - \beta r_t^f(s_t) \\ &= M_t^n(s_t) + \beta \mathsf{E}[u_{t-1}(S_{t-1}(s_t)) - r_{t-1}^1(S_{t-1}(s_t)) - r_{t-1}^f(S_{t-1}(s_t))]. \end{aligned}$$

We use induction on $u_t(s_t) - r_t^1(s_t) - r_t^f(s_t)$ to show that $B_t^s(s_t)$ is decreasing. It is obvious that $u_0(s_0) - r_0^1(s_0) - r_0^f(s_0) = 0$ is decreasing. Suppose that $u_{t-1}(s_{t-1}) - r_{t-1}^1(s_{t-1}) - r_{t-1}^f(s_{t-1})$ is decreasing for some $t-1 \geq 0$. Then $B_t^s(s_t)$ is decreasing, which results in a threshold policy. To complete the induction, we must show that $u_t(s_t) - r_t^1(s_t) - r_t^f(s_t)$ is also a decreasing function. From the discussion in Section 2, we have

$$u_t(s_t) = \begin{cases} r_t^0(s_t) + \mathsf{E} u_{t-1}(S_{t-1}(s_t)) & \text{if} \quad s_t < s_t^s; \\ r_t^1(s_t) + r_t^f(s_t) & \text{if} \quad s_t \geq s_t^s. \end{cases}$$

Hence

$$u_t(s_t) - r_t^1(s_t) - r_t^f(s_t) = \begin{cases} r_t^0(s_t) + \mathsf{E} u_{t-1}(S_{t-1}(s_t)) - r_t^1(s_t) - r_t^f(s_t) & \text{if} \quad s_t < s_t^s; \\ 0 & \text{if} \quad s_t \geq s_t^s. \end{cases}$$

We know that $r_t^0(s_t) + \mathsf{E}u_{t-1}(S_{t-1}(s_t)) - r_t^1(s_t) - r_t^f(s_t)$ is decreasing because

$$r_t^0(s_t) + \mathsf{E}u_{t-1}(S_{t-1}(s_t)) - r_t^1(s_t) - r_t^f(s_t) = M_t^{TC}(s_t) + \mathsf{E}[u_{t-1}(S_{t-1}(s_t)) - r_{t-1}^1(S_{t-1}(s_t)) - r_{t-1}^f(S_{t-1}(s_t))]$$

is decreasing. In addition, we know that

$$r_t^0(s_t) + \mathsf{E}u_{t-1}(S_{t-1}(s_t)) - r_t^1(s_t) - r_t^f(s_t) \geq r_t^0(s_t) - r_t^1(s_t) + \frac{1}{\beta}(r_t^1(s_t) - r_t^0(s_t)) \geq 0$$

for $s_t \leq s_t^s$. Therefore, $u_t(s_t) - r_t^1(s_t) - r_t^f(s_t)$ is decreasing in s_t.

Proof of Theorem 3. (i) We know that $u_t(s_t) \leq v_t(s_t)$ for any state s_t because $v_t(s_t)$ is the maximum total discounted reward under a consistent discount rate. Therefore, $B_t^s(s_t) \leq B_t^n(s_t)$ and we have $R_t^n \subseteq R_t^s$.

(ii) In what follows, we rewrite $B_t^i(s_t)$ as $B_t^i(s_t, \beta)$ in order to stress its dependence on β. For sophisticates, a key component in defining $B_t^s(s_t, \beta)$ is the function $u_t(s_t)$. We also rewrite $u_t(s_t)$ as $u_t(s_t, \beta)$ because of its dependence on β. Similarly, we also rewrite R_t^i as $R_t^i(\beta)$ for $i = \{n, s\}$.

Suppose $\beta_1 \leq \beta_2$, then we need to show that $R_t^i(\beta_2) \subseteq R_t^i(\beta_1)$ for $i = n, s$. We first look at the behavior of naifs. If $s_t \in R_t^n(\beta_2)$, then $B_t^n(s_t, \beta_2) \leq 0$ and

$$\begin{aligned} B_t^n(s_t, \beta_1) &= r_t^0(s_t) + \beta_1 \mathsf{E}v_{t-1}(S_{t-1}(s_t)) - r_t^1(s_t) - \beta_1 r_t^f(s_t) \\ &= \frac{\beta_1}{\beta_2} B_t^n(s_t, \beta_2) + (r_t^0(s_t) - r_t^1(s_t))(1 - \frac{\beta_1}{\beta_2}) \\ &\leq 0. \end{aligned}$$

Therefore, $s_t \in R_t^n(\beta_1)$ and the proof for naifs is complete.

To prove the result for sophisticates, we must perform an extra induction on the function $u_t(s_t, \beta)$. More specifically, we shall prove that $u_t(s_t, \beta)$ is increasing in β. This is true when $t = 0$ because $u_0(s_0, \beta) = r_0^1(s_0) + r_0^f(s_0)$ is independent of β. Suppose $u_{t-1}(s_{t-1}, \beta)$ is increasing in β. When $s_t \in R_t^s(\beta_2)$, we have $B_t^s(s_t, \beta_2) \leq 0$ and

$$\begin{aligned} B_t^s(s_t, \beta_1) &= r_t^0(s_t) + \beta_1 \mathsf{E}u_{t-1}(S_{t-1}(s_t), \beta_1) - r_t^1(s_t) - \beta_1 r_t^f(s_t) \\ &\leq r_t^0(s_t) + \beta_1 \mathsf{E}u_{t-1}(S_{t-1}(s_t), \beta_2) - r_t^1(s_t) - \beta_1 r_t^f(s_t) \\ &= \frac{\beta_1}{\beta_2} B_t^s(s_t, \beta_2) + (r_t^0(s_t) - r_t^1(s_t))(1 - \frac{\beta_1}{\beta_2}) \\ &\leq 0. \end{aligned}$$

Therefore, $s_t \in R_t^s(\beta_1)$ and we have $R_t^s(\beta_2) \subseteq R_t^s(\beta_1)$. To complete the induction, we still need to show that $u_t(s_t, \beta_1) \leq u_t(s_t, \beta_2)$. For $s_t \in R_t^s(\beta_2)$, the result holds since $u_t(s_t, \beta_1) = u_t(s_t, \beta_2) =$

$r_t^1(s_t) + r_t^f(s_t)$. For $s_t \in R_t^s(\beta_1)$ but $s_t \notin R_t^s(\beta_2)$,

$$\begin{aligned} u_t(s_t, \beta_2) &= r_t^0(s_t) + \mathsf{E}u_{t-1}(S_{t-1}(s_t), \beta_2) \\ &\geq r_t^1(s_t) + r_t^f(s_t) \\ &= u_t(s_t, \beta_1). \end{aligned}$$

The inequality holds because

$$r_t^0(s_t) + \mathsf{E}u_{t-1}(S_{t-1}(s_t), \beta_2) - r_t^1(s_t) - r_t^f(s_t) \geq r_t^0(s_t) - r_t^1(s_t) + \frac{1}{\beta_2}(r_t^1(s_t) - r_t^0(s_t)) \geq 0$$

for $s_t \notin R_t^s(\beta_2)$. Finally, for $s_t \notin R_t^s(\beta_1)$ we also have

$$\begin{aligned} u_t(s_t, \beta_2) &= r_t^0(s_t) + \mathsf{E}u_{t-1}(S_{t-1}(s_t), \beta_2) \\ &\geq r_t^0(s_t) + \mathsf{E}u_{t-1}(S_{t-1}(s_t), \beta_1) \\ &= u_t(s_t, \beta_1). \end{aligned}$$

This completes the induction.

(iii) Suppose $\beta_1 \leq \beta_2$, then we need to show that $R_t^n(\beta_1) \subseteq R_t^n(\beta_2)$. When $s_t \in R_t^n(\beta_1)$, we have $B_t^n(s_t, \beta_1) \leq 0$ and

$$\begin{aligned} B_t^n(s_t, \beta_2) &= r_t^0(s_t) + \beta_2 \mathsf{E}v_{t-1}(S_{t-1}(s_t)) - r_t^1(s_t) - \beta_2 r_t^f(s_t) \\ &= \frac{\beta_2}{\beta_1} B_t^n(s_t, \beta_1) + (r_t^0(s_t) - r_t^1(s_t))(1 - \frac{\beta_2}{\beta_1}) \\ &\leq 0. \end{aligned}$$

Therefore $s_t \in R_t^n(\beta_2)$, completing the proof.

Proof of Theorem 4. (i) For sophisticates,

$$\begin{aligned} B_t^s(s_t) &= \beta \mathsf{E}u_{t-1}(S_{t-1}(s_t)) + r_t^0(s_t) - r_t^1(s_t) - \beta r_t^f(s_t) \\ &= B_t^{TC}(s_t) + \mathsf{E}[\beta u_{t-1}(S_{t-1}(s_t)) - v_{t-1}(S_{t-1}(s_t)) + (1-\beta)(r_{t-1}^0(S_{t-1}(s_t)) + r_{t-1}^f(S_{t-1}(s_t)))] \\ &\quad +(1-\beta)[r_t^f(s_t) - \mathsf{E}r_{t-1}^f(S_{t-1}(s_t)) - \mathsf{E}r_{t-1}^0(S_{t-1}(s_t))]. \end{aligned}$$

The result will follow from our demonstration that if $B_t^{TC}(s_t) \geq 0$ then $B_t^s(s_t) \geq 0$. For this it suffices to show that $\beta u_t(s_t) - v_t(s_t) + (1-\beta)(r_t^0(s_t) + r_t^f(s_t)) \geq 0$ for any period t and state s_t.

The proof is by induction. When $t = 0$, we have

$$\beta u_t(s_t) - v_t(s_t) + (1-\beta)(r_t^0(s_t) + r_t^f(s_t)) = (1-\beta)(r_t^0(s_t) - r_t^1(s_t)) \geq 0.$$

Suppose $\beta u_{t-1}(s_{t-1}) - v_{t-1}(s_{t-1}) + (1-\beta)(r_{t-1}^0(s_{t-1}) + r_{t-1}^f(s_{t-1})) \geq 0$, then $R_t^s \subseteq R_t^{TC}$. We consider three cases as follows.

Case 1: $s_t \in R_t^s$. Then both TCs and sophisticates stop. We have

$$\begin{aligned} \beta u_t(s_t) - v_t(s_t) + (1-\beta)(r_t^0(s_t) + r_t^f(s_t)) &= (1-\beta)(r_t^0(s_t) - r_t^1(s_t)) \\ &> 0. \end{aligned}$$

Case 2: $s_t \in R_t^{TC}$ but $s_t \notin R_t^s$. Then TCs stop but sophisticates continue. We have

$$\begin{aligned} \beta u_t(s_t) - v_t(s_t) + (1-\beta)(r_t^0(s_t) + r_t^f(s_t)) &= \beta[\mathsf{E}u_{t-1}(S_{t-1}(s_t)) - r_t^f(s_t)] + r_t^0(s_t) - r_t^1(s_t) \\ &= B_t^s(s_t) \\ &\geq 0. \end{aligned}$$

Case 3: $s_t \notin R_t^{TC}$. Then both TCs and sophisticates continue. We have

$$\begin{aligned} \beta u_t(s_t) - v_t(s_t) + (1-\beta)(r_t^0(s_t) + r_t^f(s_t)) &= \beta \mathsf{E}u_{t-1}(S_{t-1}(s_t)) - \mathsf{E}v_{t-1}(S_{t-1}(s_t)) + (1-\beta)r_t^f(s_t) \\ &\geq (1-\beta)[r_t^f(s_t) - \mathsf{E}r_{t-1}^f(S_{t-1}(s_t)) - \mathsf{E}r_{t-1}^0(S_{t-1}(s_t))] \\ &\geq 0. \end{aligned}$$

The first inequality holds because of the induction hypothesis. Hence for any state s_t, we have

$$\beta u_t(s_t) - v_t(s_t) + (1-\beta)(r_t^0(s_t) + r_t^f(s_t)) \geq 0,$$

which completes the induction.

(ii) We can decompose $B_t^s(s_t)$ as

$$\begin{aligned} B_t^s(s_t) &= \beta \mathsf{E}u_{t-1}(S_{t-1}(s_t)) + r_t^0(s_t) - r_t^1(s_t) - \beta r_t^f(s_t) \\ &= \beta B_t^{TC}(s_t) + \mathsf{E}[\beta u_{t-1}(S_{t-1}(s_t)) - \beta v_{t-1}(S_{t-1}(s_t)) + (1-\beta)(r_{t-1}^0(S_{t-1}(s_t)) - r_{t-1}^1(S_{t-1}(s_t)))] \\ &\quad +(1-\beta)[r_t^0(s_t) - r_t^1(s_t) - (\mathsf{E}r_{t-1}^0(S_{t-1}(s_t)) - \mathsf{E}r_{t-1}^1(S_{t-1}(s_t)))]. \end{aligned}$$

The result follows from showing that if $B_t^{TC}(s_t) \geq 0$ then $B_t^s(s_t) \geq 0$. It suffices to show that, for any period t and state s_t, we have $\beta u_t(s_t) - \beta v_t(s_t) + (1-\beta)(r_t^0(s_t) - r_t^1(s_t)) \geq 0$.

The proof is by induction. When $t = 0$, we have

$$\beta u_t(s_t) - \beta v_t(s_t) + (1-\beta)(r_t^0(s_t) - r_t^1(s_t)) = (1-\beta)(r_t^0(s_t) - r_t^1(s_t)) \geq 0.$$

Suppose $\beta u_{t-1}(s_{t-1}) - \beta v_{t-1}(s_{t-1}) + (1-\beta)(r_{t-1}^0(s_{t-1}) - r_{t-1}^1(s_{t-1})) \geq 0$, then $R_t^s \subseteq R_t^{TC}$. Again we consider three cases.

Case 1: $s_t \in R_t^s$. Then both TCs and sophisticates stop. We have

$$\begin{aligned} \beta u_t(s_t) - \beta v_t(s_t) + (1-\beta)(r_t^0(s_t) - r_t^1(s_t)) &= (1-\beta)(r_t^0(s_t) - r_t^1(s_t)) \\ &> 0. \end{aligned}$$

Case 2: $s_t \in R_t^{TC}$ but $s_t \notin R_t^s$. Then TCs stop but sophisticates continue. We have

$$\begin{aligned}\beta u_t(s_t) - \beta v_t(s_t) + (1-\beta)(r_t^0(s_t) - r_t^1(s_t)) &= \beta[\mathsf{E}u_{t-1}(S_{t-1}(s_t)) - r_t^f(s_t)] + r_t^0(s_t) - r_t^1(s_t) \\ &= B_t^s(s_t) \\ &\geq 0.\end{aligned}$$

Case 3: $s_t \notin R_t^{TC}$. Then both TCs and sophisticates continue. We have

$$\begin{aligned}\beta[u_t(s_t) - v_t(s_t)] + (1-\beta)[r_t^0(s_t) - r_t^1(s_t)] &= \beta\mathsf{E}[u_{t-1}(S_{t-1}(s_t)) - v_{t-1}(S_{t-1}(s_t))] + (1-\beta)[r_t^0(s_t) - r_t^1(s_t)] \\ &\geq (1-\beta)[r_t^0(s_t) - r_t^1(s_t) - (\mathsf{E}r_{t-1}^0(S_{t-1}(s_t)) - \mathsf{E}r_{t-1}^1(S_{t-1}(s_t)))] \\ &\geq 0.\end{aligned}$$

The first inequality holds because of the induction hypothesis. Thus the induction is complete.

(iii) If $M_t^{TC}(s_t) \leq 0$ for all periods t and states s_t, then it follows from

$$B_t^{TC}(s_t) = M_t^{TC}(s_t) + \mathsf{E}\max\{0, B_{t-1}^{TC}(S_{t-1}(s_t))\}$$

that the inequality $B_t^{TC}(s_t) \leq 0$ always holds. Hence TCs always stop, and the result follows.