Information Acquisition and Time-Risk Preference

Daniel Chen*and Weijie Zhong†

April 23, 2024

Abstract

An agent acquires information dynamically until her belief about a binary state reaches an upper or lower threshold. She can choose any signal process subject to a constraint on the rate of entropy reduction. Strategies are ordered by "time risk"—the dispersion of the distribution of threshold-hitting times. We construct a strategy maximizing time risk (*Greedy Exploitation*) and one minimizing it (*Pure Accumulation*). Under either strategy, beliefs follow a compensated Poisson process. In the former, beliefs jump to the threshold that is closer in Bregman divergence. In the latter, beliefs jump to the point with the same entropy as the current belief.

*dtchen@princeton.edu, Department of Economics, Princeton University
†wjzhong@stanford.edu, Graduate School of Business, Stanford University

1 Introduction

In this paper, we study information acquisition by a Bayesian agent about an unknown binary state $\omega \in \{0,1\}$. The agent wants to be reasonably certain about ω and is satisfied once her posterior belief that ω is 1 reaches either an upper or a lower threshold. She earns a unit payoff the first time that this happens. She has great flexibility in how she can learn but has a resource constraint that limits her rate of learning. That is, she can choose any cádlág martingale posterior belief process subject to a constraint on the rate of entropy reduction.[1]

Our simple model captures three important features of many economic settings: flexible learning, limited resources, and threshold decision rules. These features often appear in the contexts of research and development, clinical trials, digital marketing, user-experience testing, and others. In settings like these, a researcher often has a fixed budget of resources that she can spend on a variety of different experiments with the goal of testing a hypothesis at some minimal level of statistical power.

Our main contribution is to show how, in such settings, the agent's optimal learning strategy depends on her *time-risk preferences*. That is, we allow for a rich set of preferences over threshold-hitting times beyond the standard case of exponential discounting. We say that the agent is time-risk loving (averse) whenever her utility over threshold-hitting times is concave (convex).[2] We derive a learning strategy that is optimal whenever the agent is time-risk loving and a strategy that is optimal whenever she is time-risk averse. Critically, the optimality of these strategies does not depend on the shape of the agent's utility function beyond its convexity or concavity.

In reality, there are many reasons why individuals may have time preferences that differ from the predominantly-studied case of exponential discounting. For example, this may be due to external factors such as exogenous decision deadlines, explicit discounting, and flow costs associated with foregone opportunities while learning. It may also be due to internal factors such as present-bias resulting from hyperbolic discounting. We provide a simple framework that allows for these factors when studying optimal learning and derive strategies that are uniformly optimal up to the convexity

[1]Our measure of entropy is a generalization of Shannon's entropy. Let $H : [0,1] \to \mathbb{R}$ be an arbitrary strictly convex C^2 function. Given H, we define the entropy at belief $\mu \in [0,1]$ to be $-H(\mu)$. Shannon's entropy is the case when $H = \mu \ln \mu + (1-\mu)\ln(1-\mu)$.

[2]In this paper, we model the agent as an expected-utility maximizer. However, it is easy to see that all of our results will go through as long as the agent has a preference relation over stopping times that is monotonic in the mean-preserving spread order.

or concavity of the utility function over threshold-hitting times.

We now briefly describe the two learning strategies and provide intuition for them. When the agent is time-risk loving, a *Greedy Exploitation* strategy is optimal. Under this strategy, the agent myopically maximizes the instantaneous rate that her beliefs jump to a threshold. She acquires a rare but decisive signal that, upon arrival, induces her belief to jump to the threshold that is closest in the Bregman divergence.[3] By targeting the closer threshold, she can jump at a faster rate without violating her constraint on the rate of entropy reduction. In the absence of a signal arrival, her belief experiences compensating drift in the direction of the farther threshold. Eventually her belief reaches a point that is equidistant in the Bregman divergence to the two thresholds. At this point, she acquires signals such that her beliefs may jump to either threshold but at rates set so that there is no net compensating drift so that her belief is stationary in the absence of a jump.

Intuitively, Greedy Exploitation is optimal because it produces a very risky distribution of threshold hitting times. Because the strategy is greedy, it yields a high probability of an early hitting time. However, in the absence of a jump, since beliefs drift towards the farther threshold, the jump rate decreases so that the expected amount of time remaining until a threshold is reached *increases*. In this sense, the agent makes no "progress" in the absence of a jump. Thus there is a high probability of late threshold hitting times as well. We in fact show, that among all strategies that exhaust the agent's resources (in that the constraint on the rate of entropy reduction is binding at all points in time), Greedy Exploitation yields a distribution of hitting times that is maximal in the mean-preserving spread order. In this sense, it maximizes time risk.

When the agent is time-risk averse, she instead seeks to minimize time risk. In this case, an optimal strategy is *Pure Accumulation*. Under this strategy, her beliefs reach a threshold at a *deterministic* time. Her beliefs follow a compensated Poisson process that jumps in the direction of the threshold that is farther away but to an interior belief that has the same entropy as her current belief. In the absence of a jump, her belief experiences compensating drift towards the closer threshold. Pure accumulation is a continuous-time analog of the "suspense-maximal" policy in Ely et al. (2015).[4]

[3]Given the entropy function H (see footnote 1), the Bregman divergence between any two beliefs μ and μ' is $d_H(\mu', \mu) = H(\mu') - H(\mu) - H'(\mu)(\mu' - \mu)$.

[4]However, the reason why Pure Accumulation is optimal in our paper is completely different from that of Ely et al. (2015). In Ely et al. (2015), the objective is to maximize aggregate conditional

The strategy is, in effect, the "opposite" of Greedy Exploitation. Because jumps are always towards beliefs with the same entropy, in the event of a jump, there is no progress: a jump does not reduce the expected amount of time until a threshold is hit. Instead, all progress is made through drift which is why the threshold hitting time is deterministic. Thus, Pure Accumulation entails *no* time risk. It therefore produces a distribution of hitting times that is minimal in the mean-preserving spread order among all strategies that exhaust the agent's resources (in that the constraint on the rate of entropy reduction is binding at all points in time).

Our analysis of optimal learning through the lens of time-risk preferences has implications for both information acquisition in practice and economic modeling. When it is possible to acquire information flexibly, an agent who is time-risk loving should use Greedy Exploitation, whereas an agent who is time-risk averse should opt for Pure Accumulation. Moreover, learning via Brownian signals is generally suboptimal. When writing models where agents acquire information with parameterized signal structures, economists should be mindful of whether these signal structures are consistent with the time-risk preferences of the agents they seek to model.

2 Related Literature

Our paper contributes to a large literature on information acquisition. As in Wald (1947) and Arrow et al. (1949) we study a sequential sampling problem but allow the agent to flexibly design the signal process as in Zhong (2022), Hébert & Woodford (2017), Hébert & Woodford (2023), Steiner et al. (2017), and Georgiadis-Harris (2023). Whereas most of these papers restrict attention the standard case of exponential discounting or a linear delay cost, we allow for more general time preferences. For example, Zhong (2022) assumes exponential discounting which implies time-risk loving preferences whereas Hébert & Woodford (2017) and Hébert & Woodford (2023) assume a linear delay cost which implies time-risk neutral preferences. Our results suggest that the assumed time-risk preferences dictate the qualitative features of the optimal strategies identified in these papers.[5] In contrast with most of the papers

variance holding fixed the stopping time.

[5]Hébert & Woodford (2023) allows both discounting and linear delay cost. However, they consider the time-risk neutral limit for the majority of their analysis. Their main objective is to study how different costs/constraints of information dictate the optimal learning pattern, which is orthogonal to the objective of this paper. Zhong (2022) assumes a belief-dependent payoff function that does

above as well as our own paper, Georgiadis-Harris (2023) studies the case where the decision maker faces an *exogenous* random decision time and finds that a strategy qualitatively similar to Pure Accumulation is optimal. The Pure Accumulation strategy is also closely related to the suspense-maximal strategy in Ely et al. (2015).

In our analysis, the key summary statistic that determines the payoff from a strategy is the distribution of the time that the agent's belief first reaches a threshold. This statistic defines a *time lottery*, which is an object studied in an emerging literature on time-risk preferences. Chesson & Viscusi (2003) and Chen (2013) show that the expected discounted utility framework implies preferences that are *risk seeking over time lotteries* (RSTL). DeJarnette et al. (2020) show that within a broad class of models RSTL can not be violated if there is stochastic impatience. However, experimental evidence suggests that subjects are often *risk averse over time lotteries* (RATL) (Chesson & Viscusi (2003); Onay & Öncüler (2007)). Our model accommodates both RSTL and RATL and shows that optimal information acquisition can differ dramatically under different time-risk preferences.

The optimal learning strategies that we identify are qualitatively similar to learning strategies that have been assumed in reduced form by many papers in the literature. For example, Che & Mierendorff (2019), Mayskaya (2019), and Nikandrova & Pancs (2018) adopt a framework that restricts attention to Poisson signal processes in order to study optimal stopping with endogenous information. Poisson signals are also often assumed in the literature on strategic experimentation (see a survey by Hörner & Skrzypacz (2017)). We show that Poisson learning has an optimization foundation under time-risk loving preferences. The Pure Accumulation strategy is also related to classic models on the timing of innovation introduced by Dasgupta & Stiglitz (1980) and Lee & Wilde (1980) (see a survey by Reinganum (1989)) which involve a deterministic time of innovation. The models in these papers assume a reduced-form learning process and are non-Bayesian. However, we show that the learning strategies in these papers can emerge endogenously in a Bayesian information acquisition framework when agents have time risk-averse preferences.

Our model also allows for Gaussian learning strategies. Gaussian signal processes are often assumed in reduced-form learning models (see for example Moscarini & Smith (2001); Ke & Villas-Boas (2019); Liang et al. (2019); Morris & Strack (2019)).

not have a threshold structure. Nevertheless, the optimal learning strategy is similar to Greedy Exploitation.

Also, *drift-diffusion models* (DDM) of binary choice problems appear in Ratcliff & Rouder (1998) and Fudenberg et al. (2018). However, our results imply that Gaussian learning can not be justified by optimality except in the knife-edge case when agents have time-risk neutral preferences provided information can be acquired flexibly.

Our result that a greedy strategy is optimal for a time-risk loving agent is also the main result in Liang et al. (2019). However, the mechanisms in the two papers are very different. Liang et al. (2019)'s result crucially depends on the linear-Gaussian setup with exogenously given Gaussian information sources and holds for any time preferences. Our result allows for a flexible and endogenous choice of information sources, but crucially depends on time preferences.

We model limits on the agent's learning resources via a constraint on the rate of entropy reduction. That is, the rate of resource depletion is determined by a *uniformly posterior separable* (UPS) function. The rational inattention literature also typically models information costs or constraints using a UPS function (Sims (2003); Matějka & McKay (2014); Steiner et al. (2017); Caplin et al. (2017)). Microfoundations for the UPS formulation can be found in Frankel & Kamenica (2019); Caplin et al. (2017); Zhong & Bloedel (2021); Morris & Strack (2019). In our paper, the UPS information constraint ensures that the expected threshold hitting time is equalized for all exhaustive strategies, which allows us to isolate the role of time *risk* in information acquisition. By Theorem 3 in Zhong (2022), a UPS information constraint is both necessary and sufficient for the expected learning time to be equalized for all exhaustive strategies.

3 Model

This section presents a simple model of an agent who wants to learn over time about an unknown state. The unknown state ω takes values in $\{0, 1\}$. At $t = 0$, the agent believes that ω is 1 with probability $\mu \in (0, 1)$. She receives a unit payoff when her posterior belief μ_t that ω is 1 reaches either an upper threshold $\overline{\mu} \in (\mu, 1)$ or a lower threshold $\underline{\mu} \in (0, \mu)$. However, she is impatient and her utility is a decreasing function $\rho : \mathbb{R}_+ \to \mathbb{R}$ of the threshold-hitting time. When ρ is convex, we say that the agent is *time-risk loving*. When ρ is concave we say that she is *time-risk averse*.

The agent has great flexibility in how she can learn about ω but has limited resources and cannot learn infinitely fast. Let $\mathcal{M}$ denote the set of processes $\boldsymbol{\mu} =$

$\{\mu_t; t \geq 0\}$ that satisfy a stochastic differential equation (allowing for jumps) of the form

$$d\mu_t = \sum_{i=1}^{N} (\nu^i(t,\mu_t) - \mu_t) \left[dJ_t^i(\lambda(t,\mu_t)) - \lambda^i(t,\mu_t)\, dt \right] + \sum_{j=1}^{M} \sigma^j(t,\mu_t)\, dZ_t^j, \ \ \mu_0 = \mu$$

for some positive integers N and M and functions $\{\nu^i\}_{i=1}^N$, $\{\lambda^i\}_{i=1}^N$, and $\{\sigma^j\}_{i=1}^M$. Above, each Z_t^j is a standard Brownian Motion and each J_t^i is a Poisson point process that ticks at rate $\lambda^i(t,\mu_t^i)$. If J_t^i ticks at time t, the belief μ_t jumps to the point $\nu^i(t,\mu_t)$. The number of distinct points that the belief can jump to at time t is the integer N and the number of distinct Brownian Motions is the integer M.

We assume that the agent can directly choose any belief process in $\mathcal{M}$ such that[6]

$$\mathbb{E}\left[\frac{d}{dt} H(\mu_t) \Big| \mathcal{F}_t \right] \leq I \tag{1}$$

where $\{\mathcal{F}_t\}$ is the natural filtration of $\boldsymbol{\mu}$, H is a C^2 convex function defined on $[0,1]$, and $I \geq 0$ is a constant. $-H$ maps the agent's belief to its associated *entropy*. Thus, equation (1) is a constraint on the rate of entropy reduction. A special case is when $-H$ is Shannon's entropy so that (1) amounts to a constraint on the well-known mutual information rate. Without loss of generality, we normalize H so that $H(\overline{\mu}) = H(\underline{\mu}) = 0$ and set $I = 1$.[7] Thus, it follows from the optional-stopping theorem that provided (1) is binding at all times, the expected time remaining until a threshold is reached is simply the current entropy:

$$-H(\mu_t) = -\mathbb{E}\left[H(\mu_\tau) - \tau | \mathcal{F}_t \right] = \mathbb{E}\left[\tau | \mathcal{F}_t \right].$$

To state the agent's learning problem, let $\tau_{\boldsymbol{\mu}} = \inf\{t | \mu_t \in [0,\underline{\mu}] \cup [\overline{\mu},1]\}$ be the first time that her beliefs reach a threshold. Since $\tau_{\boldsymbol{\mu}}$ may be ∞ for some belief processes,

[6]Our restriction to jump-diffusion belief processes is without loss of generality within the larger class of cádlág processes such that (1) is well defined. This follows from Theorem 1 in Georgiadis-Harris (2023).

[7]Normalizing $H(\overline{\mu})$ and $H(\underline{\mu})$ to 0 is without loss because we can always redefine $H(\mu)$ to be

$$H(\mu) - \frac{H(\overline{\mu}) - H(\underline{\mu})}{\overline{\mu} - \underline{\mu}} (\mu - \underline{\mu}).$$

All the same belief processes satisfy (1) after this redefinition because beliefs are martingales. Normalizing I to 1 is without loss since we can always scale H by $1/I$.

we set $\rho(\infty) = -\infty$. She solves

$$\sup_{\boldsymbol{\mu} \in \mathcal{M}} \mathbb{E}[\rho(\tau_{\boldsymbol{\mu}})] \tag{2}$$

such that (1) holds.

Our simple model makes several assumptions in order to isolate the connection between optimal learning and time-risk preferences which is the main focus of our paper. For example, we assume that the agent experiences no costs from learning at faster speeds and that she earns a common payoff regardless of the threshold that she ultimately hits. Though these assumptions are restrictive, we believe that our model aligns well with a number of economic applications. One application, for example, is the case of a judge who wants to acquire enough evidence to decide whether to convict a suspect. The judge can only come to a decision if she is reasonably certain about the innocence or guilt of the suspect. She wants to make a decision sooner rather than later but is otherwise indifferent. She can flexibily request evidence to inform her decision. Another example is a platform that wants to learn whether a user is male or female in order to tailor its advertisements. The platform wants to be reasonably certain before it decides which ad to serve. Suppose that when it does serve an ad, it earns in expectation the same ad revenue regardless of which of the two thresholds is reached. The platform can flexibly choose how much and what kinds of users' activities to track.

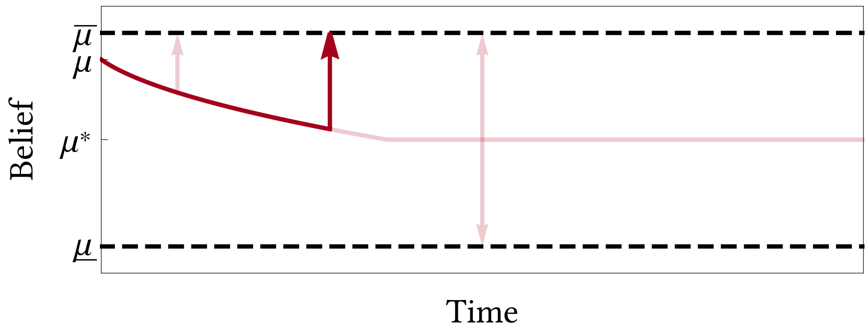
4 Optimal Learning and Time-Risk Preferences

In this section, we present our main results: a strategy that is optimal whenever the agent is time-risk loving and a strategy that is optimal whenever she is time-risk averse. All proofs can be found in the Appendix. These results illustrate the connection between optimal learning and time-risk preferences.

4.1 Time-Risk Loving

We first consider the case when the agent is time-risk loving. A formal statement of her optimal learning strategy is given below in Definition 1 but we begin by describing it informally here.

An optimal strategy for the agent is a *Greedy Exploitation* illustrated graphically below.



Notes. The dark red curve represents one possible belief path while the light red curves represent other possible belief paths.

Let

$$d_H(\tilde{\mu}, \hat{\mu}) = H(\tilde{\mu}) - H(\hat{\mu}) - H'(\hat{\mu})(\tilde{\mu} - \hat{\mu})$$

denote the Bregman divergence between any two beliefs $\hat{\mu}$ and $\tilde{\mu}$ under the function H. In the figure, μ^* represents the belief that is equidistant in the Bregman divergence to the two thresholds: $d_H(\overline{\mu}, \mu^*) = d_H(\underline{\mu}, \mu^*)$. Initially, the agent's beliefs either jump to the threshold that is closer in the Bregman divergence (in this case $\overline{\mu}$) or experiences compensating drift towards the other threshold. By jumping to the closer threshold, she greedily maximizes the "chance" that her beliefs reach a threshold in the "next instant." This is because the agent's beliefs can jump at a faster rate when she targets the closer threshold without violating her resource constraint (1). After some time, in the absence of a jump, her beliefs eventually drift to μ^*. At this point, her beliefs may jump to either threshold. The jump rates to the respective thresholds are such that there is no net compensating drift and so, in the absence of a jump, her beliefs remain stationary.

Definition 1. The *Greedy Exploitation* strategy $\boldsymbol{\mu}^{GE}$ is defined as follows. Let $\mu^* \in (0,1)$ be the unique belief such that $d_H(\overline{\mu}, \mu^*) = d_H(\underline{\mu}, \mu^*)$,

- While $\mu_t^{GE} > \mu^*$, her beliefs evolve according to

$$d\mu_t^{GE} = (\overline{\mu} - \mu_t^{GE}) \left[dJ_t^1(\lambda_t) - \lambda_t \, dt \right]$$

where $\lambda_t = I / d_H(\overline{\mu}, \mu_t^{GE})$.

- While $\mu_t^{GE} = \mu^*$, her beliefs evolve according to

$$d\mu_t^{GE} = (\overline{\mu} - \mu_t^{GE})\, dJ_t^2 \left(\frac{\mu_t^{GE} - \underline{\mu}}{\overline{\mu} - \underline{\mu}} \lambda^* \right) + (\underline{\mu} - \mu_t^{GE})\, dJ_t^3 \left(\frac{\overline{\mu} - \mu_t^{GE}}{\overline{\mu} - \underline{\mu}} \lambda^* \right)$$

  where $\lambda^* = 1/d_H(\overline{\mu}, \mu^*)$.

- While $\mu_t^{GE} < \mu^*$, her beliefs evolve according to

$$d\mu_t^{GE} = (\underline{\mu} - \mu_t^{GE}) \left[dJ_t^1(\lambda_t) - \lambda_t\, dt \right]$$

  where $\lambda_t = 1/d_H(\underline{\mu}, \mu_t^{GE})$.

Above J_t^1, J_t^2 and J_t^3 are Poisson point processes with jump rates indicated in parentheses.

Theorem 1. *If the agent is time-risk loving, then Greedy Exploitation is optimal.*

Proof. See Appendix A. □

To prove the theorem, we need only prove that Greedy Exploitation is optimal for the special cases when the utility function is of the form $\rho_T = \max\{T - t, 0\}$ for each $T > 0$. By the results in Müller (1996), this suffices to prove optimality for all convex ρ. Our proof strategy is to write down the Hamilton-Jacobi Bellman (HJB) equation for the agent's problem and then check directly that the value function under the conjectured strategy satisfies the HJB equation. We can compute the value function analytically, since in the absence of a jump, the belief path is characterized by a separable differential equation. Given the belief path, we immediately have the jump points and jump rates.

Since Greedy Exploitation is uniformly optimal for all convex discount functions, it induces the riskiest distribution of threshold hitting times among all strategies that are exhaustive in that (1) is satisfied at all points in time. To make this precise, we first state the following definition.

Definition 2. $\mathcal{T}_{EX} = \left\{ \tau_{\boldsymbol{\mu}} \middle| \boldsymbol{\mu} \in \mathcal{M} \text{ such that (1) binds at all } t \right\}$.

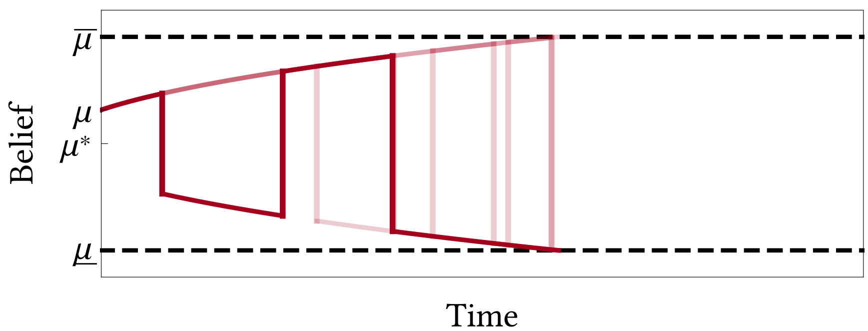
The Greedy Exploitation strategy produces a threshold-hitting time that is maximal in the mean-preserving spread order among all threshold-hitting times in this set.

Proposition 1. *It holds that* $\tau_{\boldsymbol{\mu}^{GE}} \succeq_{\text{mps}} \tau$ *for each* $\tau \in \mathcal{T}_{EX}$.

This result hinges on our assumption that the constraint on learning is of the form in (1). Because of this assumption, all exhaustive strategies have the same expected threshold hitting time which is equal to the initial entropy $-H(\mu)$.

4.2 Time-Risk Averse

When the agent is time-risk averse, her optimal learning strategy is *Pure Accumulation*, illustrated graphically below.



Notes. The dark red curve represents one possible belief path. The vertical segments represent jumps. The light red curves represent other possible belief paths.

The Pure Accumulation strategy is closely related to the *suspense-maximal strategy* from Ely et al. (2015). Under this strategy, the agent's belief either jumps in the direction of the farther threshold or experiences compensating drift. When her belief jumps, it jumps to a point with the same entropy as her current belief so that all progress is made through drift.

Definition 3. The *Pure Accumulation* strategy is defined as follows. Let $\mu^H : [0,1] \setminus \{\mu^*\} \to [0,1]$ denote the function that maps a belief $\hat{\mu}$ to the unique belief $\mu^H(\hat{\mu}) \neq \hat{\mu}$ such that $H(\mu^H(\hat{\mu})) = H(\hat{\mu})$.

$$d\mu_t^{PA} = \left[\mu^H(\mu_t^{PA}) - \mu_t^{PA}\right] dJ_t(\lambda_t) - \lambda_t \left[\mu^H(\mu_t^{PA}) - \mu_t^{PA}\right] dt$$

where J_t is a Poisson point process that ticks at rate $\lambda_t = 1/d_H(\mu^H(\mu_t^{PA}), \mu_t^{PA})$.

Then, Pure Accumulation is optimal for the agent because it entails no time-risk: she is guaranteed to hit a threshold at $t = -H(\mu)$. We immediately have the following result.

Proposition 2. *It holds that* $\tau \succeq_{\text{mps}} \tau_{\boldsymbol{\mu}^{PA}}$ *for each* $\tau \in \mathcal{T}_{EX}$.

5 Concluding Discussion

In this paper, we have studied the relationship between time-risk preferences and optimal information acquisition. We have shown that an optimal strategy for a time-risk loving agent is Greedy Exploitation. This strategy produces the riskiest distribution over threshold hitting times among all exhaustive strategies. On the other hand, an optimal strategy for a time-risk averse agent is Pure Accumulation. This strategy produces a deterministic threshold hitting time and thus entails no time risk. Both of these strategies are uniformly optimal up to the convexity or concavity of the utility function, provided the agent is impatient. Thus they are immune to dynamic inconsistency. In practice, agents may have time preferences that differ from the well-studied case of exponential discounting. Our analysis predicts how these agents may seek to acquire information and provides guidance for the kinds of signal structures that economists should use when modeling these agents.

There are two promising avenues to explore in future work. The first is to explore how our results may extend to the case when the agent is neither time-risk loving nor time-risk averse. For these more general preferences, what are the qualitative features of optimal information acquisition? A second avenue to explore is to try to embed our model of information acquisition into strategic settings where there are multiple agents in order to study the implications of flexible information acquisition in games.

References

Arrow, K. J., Blackwell, D., & Girshick, M. A. (1949). Bayes and minimax solutions of sequential decision problems. *Econometrica, Journal of the Econometric Society*, (pp. 213–244).

Caplin, A., Dean, M., & Leahy, J. (2017). *Rationally Inattentive Behavior: Characterizing and Generalizing Shannon Entropy*. Working Paper 23652, National Bureau of Economic Research.

Che, Y.-K. & Mierendorff, K. (2019). Optimal dynamic allocation of attention. *American Economic Review*, 109(8), 2993–3029.

Chen, M. K. (2013). The effect of language on economic behavior: Evidence from savings rates, health behaviors, and retirement assets. *American Economic Review*, 103(2), 690–731.

Chesson, H. W. & Viscusi, W. K. (2003). Commonalities in time and ambiguity aversion for long-term risks. *Theory and Decision*, 54(1), 57–71.

Dasgupta, P. & Stiglitz, J. (1980). Uncertainty, industrial structure, and the speed of r&d. *The Bell Journal of Economics*, (pp. 1–28).

DeJarnette, P., Dillenberger, D., Gottlieb, D., & Ortoleva, P. (2020). Time lotteries and stochastic impatience. *Econometrica*, 88(2), 619–656.

Ely, J., Frankel, A., & Kamenica, E. (2015). Suspense and surprise. *Journal of Political Economy*, 123(1), 215–260.

Frankel, A. & Kamenica, E. (2019). Quantifying information and uncertainty. *American Economic Review*, 109(10), 3650–80.

Fudenberg, D., Strack, P., & Strzalecki, T. (2018). Speed, accuracy, and the optimal timing of choices. *American Economic Review*, 108(12), 3651–84.

Georgiadis-Harris, A. (2023). Information acquisition and the timing of actions.

Hébert, B. & Woodford, M. (2017). *Rational inattention and sequential information sampling*. Technical report, National Bureau of Economic Research.

Hébert, B. & Woodford, M. (2023). Rational inattention when decisions take time. *Journal of Economic Theory*, (pp. 105612).

Hörner, J. & Skrzypacz, A. (2017). Learning, experimentation, and information design. *Advances in Economics and Econometrics*, 1, 63–98.

Ke, T. T. & Villas-Boas, J. M. (2019). Optimal learning before choice. *Journal of Economic Theory*, 180, 383–437.

Lee, T. & Wilde, L. L. (1980). Market structure and innovation: A reformulation. *The Quarterly Journal of Economics*, 94(2), 429–436.

Liang, A., Mu, X., & Syrgkanis, V. (2019). Dynamically aggregating diverse information. *arXiv preprint arXiv:1910.07015*.

Matějka, F. & McKay, A. (2014). Rational inattention to discrete choices: A new foundation for the multinomial logit model. *The American Economic Review*, 105(1), 272–298.

Mayskaya, T. (2019). Dynamic choice of information sources. unpublished manuscript.

Morris, S. & Strack, P. (2019). *The Wald Problem and the Equivalence of Sequential Sampling and Static Information Costs*. Working paper.

Moscarini, G. & Smith, L. (2001). The optimal level of experimentation. *Econometrica*, 69(6), 1629–1644.

Müller, A. (1996). Orderings of risks: A comparative study via stop-loss transforms. *Insurance: Mathematics and Economics*, 17(3), 215–222.

Nikandrova, A. & Pancs, R. (2018). Dynamic project selection. *Theoretical Economics*, 13(1), 115–143.

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

Ratcliff, R. & Rouder, J. N. (1998). Modeling response times for two-choice decisions. *Psychological Science*, 9(5), 347–356.

Reinganum, J. F. (1989). The timing of innovation: Research, development, and diffusion. *Handbook of industrial organization*, 1, 849–908.

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

Steiner, J., Stewart, C., & Matějka, F. (2017). Rational inattention dynamics: Inertia and delay in decision-making. *Econometrica*, 85(2), 521–553.

Wald, A. (1947). Foundations of a general theory of sequential decision functions. *Econometrica, Journal of the Econometric Society*, (pp. 279–313).

Zhong, W. (2022). Optimal dynamic information acquisition. *Econometrica*, 90(4), 1537–1582.

Zhong, W. & Bloedel, A. (2021). The cost of optimally acquired information.

A Proofs for Theorem 1

In this Appendix we prove Theorem 1 which will follow from Lemmas 1-4 below.

We first define some useful notation. Let

$$U(\mu, t) = \int_0^t \lambda_s e^{-\int_0^s \lambda_z \, dz} \, ds$$

denote the probability that a jump arrives by time t under Greedy Exploitation. It is easy to show that $\frac{\partial}{\partial \mu} U(\mu, t) \geq 0$ for $\mu \in [\mu^*, \overline{\mu}]$ and that $\frac{\partial}{\partial \mu} U(\mu, t) \leq 0$ for $\mu \in [\underline{\mu}, \mu^*]$.

Also, let

$$V(\mu, t) = \int_0^t (t - s) \lambda_s e^{-\int_0^s \lambda_z \, dz} \, ds$$

denote the payoff under Greedy Exploitation for the discount function $\rho^t(s) = \max\{t - s, 0\}$ for $s \in [0, \infty)$. One can verify that $\partial V(\mu, t)/\partial t = U(\mu, t)$. In what follows we often write $V_t(\mu)$ and $U_t(\mu)$ in place of $V(\mu, t)$ and $U(\mu, t)$ to ease notation.

Finally, for any two beliefs ν and μ let

$$d_{V_t}(\nu, \mu) = V_t(\nu) - V_t(\mu) - V_t'(\mu)(\nu - \mu).$$

If V_t was convex then d_{V_t} would be a Bregman-divergence.

The following Lemma 1 formulates the Hamilton-Jacobi Bellman equation for the agent's problem (2) and states that V is the true value function if it satisfies this equation.

Lemma 1. *Given $T > 0$, if V satisifes*

$$\max\left\{t\mathbb{1}_{\mu\in\{\overline{\mu},\underline{\mu}\}} - V_t(\mu), \max\left\{\sup_{\nu}\frac{d_{V_t}(\nu,\mu)}{d_H(\nu,\mu)}, \frac{V_t''(\mu)}{H''(\mu)}\right\} - U_t(\mu)\right\} = 0 \tag{3}$$

at each $(t,\mu) \in [\underline{\mu},\overline{\mu}] \times [0,T]$ then $V_T(\mu)$ is equal to (2) *when $\rho = \rho^T$.*

Proof. The Hamilton-Jacobi Bellman equation for the agent's problem (2) is

$$\max\left\{ - V_t(\mu) + t\mathbb{1}_{\mu\in\{\overline{\mu},\underline{\mu}\}}, -U_t(\mu) + \sup_{(\nu^i),(\lambda^i),\sigma} \mathcal{A}^{\nu,\lambda,\sigma}V_t(\mu)\right\} = 0, \tag{4}$$

$$\text{s.t. } \mathcal{A}^{\nu,\lambda,\sigma}H(\mu_t) \leq 1 \tag{5}$$

where $\mathcal{A}^{\nu,\lambda,\sigma}$ is the operator defined for functions $f \in C^2(\underline{\mu},\overline{\mu})$ by

$$\mathcal{A}^{\nu,\lambda,\sigma}f(\mu) = \sum_i \lambda^i\left[f(\nu^i) - f(\mu) - f'(\mu)(\nu^i-\mu)\right] + \frac{1}{2}\sum_j(\sigma^j)^2 f''(\mu).$$

Here, $\mathcal{A}^{\nu,\lambda,\sigma}$ is the infinitessimal generator for a compensated jump-diffusion process with jump points (ν^i), jump rates (λ^i), and volatilities (σ^j). Because the operator is additively separable it suffices to choose either a single jump point or volatility to achieve the sup in (4). In either case, the jump rate or volatility will be set so that the constraint (5) binds. Using these facts, it follows that if V satisfies (3) then it must also satisfy (4).

Suppose that V satisfies (3). Consider an arbitrary admissible strategy (ν^i), (λ^i), (σ^j) with first threshold-hitting time τ. By Itô's formula for jump diffusions,

$$\begin{aligned}
&V(\mu_{\tau\wedge T}, T-\tau\wedge T) - V(\mu,T)\\
&= \int_0^{\tau\wedge T} -U(\mu_t,T-t) + \frac{1}{2}\sum_j(\sigma_t^j)^2\frac{\partial^2 V(\mu_t,T-t)}{\partial\mu^2}\\
&\quad - \sum_i \lambda_t^i\frac{\partial V(\mu_t,T-t)}{\partial\mu}(\nu_t^i-\mu_t)\,dt + \sum_j\int_0^{\tau\wedge T}\frac{\partial V(\mu_t,T-t)}{\partial\mu}\sigma_t^j dZ_t\\
&\quad + \sum_i\int_0^{\tau\wedge T}\left[V(\nu_t^i,T-t) - V(\mu_t,T-t)\right]dJ_t^i(\lambda_t^i)
\end{aligned}$$

where we have used the fact that $\partial V(\mu,t)/\partial t = U(\mu,t)$.

Rearranging and taking an expectation of both sides yields

$$
\begin{aligned}
V(\mu,T) =&\mathbb{E}\Bigg[V(\mu_{\tau\wedge T}, T-\tau\wedge T) \\
&- \int_0^{\tau\wedge T} -U(\mu_t, T-t) + \frac{1}{2}\sum_j (\sigma_t^j)^2 \frac{\partial^2 V(\mu_t, T-t)}{\partial \mu^2} \\
&- \sum_i \lambda_t^i \frac{\partial V(\mu_t, T-t)}{\partial \mu}(\nu_t^i - \mu_t)\, dt \\
&+ \sum_i \int_0^{\tau\wedge T} \big(V(\nu_t^i, T-t) - V(\mu_t, T-t)\big)\, dJ_t^i(\lambda_t^i)\Bigg] \\
=&\mathbb{E}\left[V(\mu_{\tau\wedge T}, T-\tau\wedge T) - \int_0^{\tau\wedge T} -U(\mu_t, T-t) + \mathcal{A}^{\nu,\lambda,\sigma} V(\mu_t, T-s)\, dt\right] \\
\geq&\mathbb{E}\left[V(\mu_{\tau\wedge T}, T-\tau\wedge T)\right] \\
\geq&\mathbb{E}\left[T-\tau\wedge T\right] \\
=&\mathbb{E}\left[\rho^T(\tau)\right]
\end{aligned}
$$

where the first and second inequalities follow from the fact that V satisfies (4). Thus, no admissible strategy can achieve a higher value. Since $V(\mu,T)$ is achieved by Greedy Exploitation, the proof is complete. □

The next three lemmas verify, step by step, that V satisfies the conditions of Lemma 1.

Lemma 2. *At each $t \in [0,\infty)$ the following hold:*

1. *If $\mu \geq \mu^*$, then*

$$U_t(\mu) = \frac{d_{V_t}(\overline{\mu}, \mu)}{d_H(\overline{\mu}, \mu)}.$$

2. *If $\mu \leq \mu^*$, then*

$$U_t(\mu) = \frac{d_{V_t}(\underline{\mu}, \mu)}{d_H(\underline{\mu}, \mu)}.$$

Proof. Suppose that $\mu_s \in (\mu^*, \overline{\mu}]$. While in this region, in the absence of a jump, beliefs evolve according to

$$\dot{\mu}_s = -\frac{\overline{\mu} - \mu_s}{d_H(\overline{\mu}, \mu_s)}, \quad s \in [0,\infty). \tag{6}$$

Throughout the rest of this proof, we abuse notation and let μ_s denote the belief conditional on no jump occuring up to and including time s. We first use (6) to derive a couple of useful relationships. Let t^* denote the first time that the beliefs reach μ^* in the absence of a jump. Then we see that for $s < t^*$, (6) implies

$$\ln\left(\frac{\overline{\mu}-\mu_s}{\overline{\mu}-\mu}\right) = \int_0^s \lambda_z \, dz \tag{7}$$

and

$$\int_\mu^{\mu_s} \frac{d_H(\overline{\mu}, z)}{\overline{\mu}-z} \, dz = -s. \tag{8}$$

From (8), we find that

$$\frac{dt^*}{d\mu} = \frac{d_H(\overline{\mu}, \mu)}{\overline{\mu}-\mu} \tag{9}$$

and

$$\frac{d\mu_s}{d\mu} = \frac{d_H(\overline{\mu}, \mu)}{d_H(\overline{\mu}, \mu_s)} \frac{\overline{\mu}-\mu_s}{\overline{\mu}-\mu}. \tag{10}$$

Next, under the conjectured strategy

$$\begin{aligned} V_t(\mu) &= \int_0^t \rho_t(s)\lambda_s e^{-\int_0^s \lambda_z \, dz} \, ds \\ &= t - \int_0^{t^*\wedge t} \left(\frac{\overline{\mu}-\mu}{\overline{\mu}-\mu_s}\right) ds - \int_{t^*\wedge t}^t e^{-\lambda^*(s-t^*)} \frac{\overline{\mu}-\mu}{\overline{\mu}-\mu_*} \, ds \end{aligned} \tag{11}$$

where the second equality follows by integrating by parts and (7).

Differentiating with respect to μ yields

$$\begin{aligned} V_t'(\mu) = &\int_0^{t^*\wedge t} \left[\frac{1}{\overline{\mu}-\mu_s} - (\overline{\mu}-\mu)\left(\frac{1}{\overline{\mu}-\mu_s}\right)^2 \frac{d\mu_s}{d\mu}\right] ds \\ &+ \int_{t^*\wedge t}^t \frac{e^{-\lambda^*(s-t^*)}}{\overline{\mu}-\mu_*} \, ds - \lambda^* d_H(\overline{\mu}, \mu) \int_{t^*\wedge t}^t \frac{e^{-\lambda^*(s-t^*)}}{\overline{\mu}-\mu_*} \, ds \end{aligned} \tag{12}$$

where we have used (9). Substituting in (10) yields

$$\begin{aligned} V_t'(\mu) = &\int_0^{t^*\wedge t} \left[\frac{1}{\overline{\mu}-\mu_s} - \frac{d_H(\overline{\mu}, \mu)}{d_H(\overline{\mu}, \mu_s)} \frac{1}{\overline{\mu}-\mu_s}\right] ds \\ &+ \int_{t^*\wedge t}^t \frac{e^{-\lambda^*(s-t^*)}}{\overline{\mu}-\mu_*} \, ds - \lambda^* d_H(\overline{\mu}, \mu) \int_{t^*\wedge t}^t \frac{e^{-\lambda^*(s-t^*)}}{\overline{\mu}-\mu_*} \, ds. \end{aligned} \tag{13}$$

Multiplying by $\overline{\mu} - \mu$ yields

$$(\overline{\mu} - \mu)\, V_t'(\mu) = \int_0^{t^* \wedge t} \left[\frac{\overline{\mu} - \mu}{\overline{\mu} - \mu_s} - \frac{d_H(\overline{\mu}, \mu)}{d_H(\overline{\mu}, \mu_s)} \frac{\overline{\mu} - \mu}{\overline{\mu} - \mu_s} \right] ds + \int_{t^* \wedge t}^{t} e^{-\lambda^*(s - t^*)} \frac{\overline{\mu} - \mu}{\overline{\mu} - \mu_*}\, ds - \lambda^* d_H(\overline{\mu}, \mu) \int_{t^* \wedge t}^{t} e^{-\lambda^*(s - t^*)} \frac{\overline{\mu} - \mu}{\overline{\mu} - \mu_*}\, ds.$$

Using equation (11), we have

$$(\overline{\mu} - \mu)\, V_t'(\mu) + V_t(\mu) = V_t(\overline{\mu}) - \int_0^{t^* \wedge t} \frac{d_H(\overline{\mu}, \mu)}{d_H(\overline{\mu}, \mu_s)} \frac{\overline{\mu} - \mu}{\overline{\mu} - \mu_s}\, ds - \lambda^* d_H(\overline{\mu}, \mu) \int_{t^* \wedge t}^{t} e^{-\lambda^*(s - t^*)} \frac{\overline{\mu} - \mu}{\overline{\mu} - \mu_*}\, ds \quad (14)$$

where we have used the fact that $t = V_t(\overline{\mu})$.

Therefore,

$$\begin{aligned} \frac{(\overline{\mu} - \mu)\, V_t'(\mu) + V_t(\mu) - V_t(\overline{\mu})}{d_H(\overline{\mu}, \mu)} &= - \int_0^{t} \frac{1}{d_H(\overline{\mu}, \mu_s)} \frac{\overline{\mu} - \mu}{\overline{\mu} - \mu_s}\, ds \\ &\quad - \lambda^* \int_{t^* \wedge t}^{t} e^{-\lambda^*(s - t^*)} \frac{\overline{\mu} - \mu}{\overline{\mu} - \mu_*}\, ds \\ &= - \int_0^{T} \lambda_s e^{-\int_0^s \lambda_z\, dz}\, ds \\ &= -U_t(\mu) \end{aligned}$$

where we have used (7) to obtain the second equality.

□

Lemma 3. *At each* $(\mu, t) \in (\underline{\mu}, \overline{\mu}) \times [0, \infty)$ *it holds that*

$$U_t(\mu) = \sup_{\nu} \frac{d_{V_t}(\nu, \mu)}{d_H(\nu, \mu)}.$$

Proof. We will prove the case when $\mu > \mu^*$. The proof of the case when $\mu \leq \mu^*$ is analogous. We have

$$\frac{d}{d\nu} \frac{d_{V_t}(\nu, \mu)}{d_H(\nu, \mu)} = \frac{V_t'(\nu) - V_t'(\mu)}{d_H(\nu, \mu)} - \frac{d_{V_t}(\nu, \mu)}{d_H(\nu, \mu)^2} \left[H'(\nu) - H'(\mu) \right].$$

Suppose that $\nu > \mu$. Then the derivative is nonnegative if and only if

$$\frac{V_t'(\nu) - V_t'(\mu)}{H'(\nu) - H'(\mu)} \geq \frac{d_{V_t}(\nu, \mu)}{d_H(\nu, \mu)}. \tag{15}$$

Both d_{V_t} and d_H respect the law of cosines so we can rewrite this condition as

$$\frac{d_{V_t}(\overline{\mu}, \mu) - d_{V_t}(\overline{\mu}, \nu) - d_{V_t}(\nu, \mu)}{d_H(\overline{\mu}, \mu) - d_H(\overline{\mu}, \nu) - d_H(\nu, \mu)} \geq \frac{d_{V_t}(\nu, \mu)}{d_H(\nu, \mu)}$$

which rearranges to

$$\frac{d_{V_t}(\overline{\mu}, \mu) - d_{V_t}(\overline{\mu}, \nu)}{d_H(\overline{\mu}, \mu) - d_H(\overline{\mu}, \nu)} \geq \frac{d_{V_t}(\nu, \mu)}{d_H(\nu, \mu)}. \tag{16}$$

We now argue that $\overline{\mu}$ achieves the global maximum in the region $\nu \in (\mu, \overline{\mu}]$. Notice that (16) holds with equality at $\nu = \overline{\mu}$. Consider any local extrema in $(\mu, \overline{\mu}]$ so that(16) holds with equality. We can prove that all such local extrema are necessarily local maxima simply by proving that the derivative of the left-hand side of (16) is negative. This is because the derivative of the right-hand side is always 0 at a local extremum since the right-hand side expression is the objective function.

The left-hand side is decreasing in ν since $d_{V_t}(\overline{\mu}, \nu)/d_H(\overline{\mu}, \nu) = U_t(\nu)$ is increasing in ν. This can be seen since

$$\begin{aligned}\frac{d}{d\nu}\frac{d_{V_t}(\overline{\mu}, \mu) - d_{V_t}(\overline{\mu}, \nu)}{d_H(\overline{\mu}, \mu) - d_H(\overline{\mu}, \nu)} &= \frac{d}{d\nu}\frac{U_t(\mu)d_H(\overline{\mu}, \mu) - U_t(\nu)d_H(\overline{\mu}, \nu)}{d_H(\overline{\mu}, \mu) - d_H(\overline{\mu}, \nu)} \\ &< \frac{d}{d\nu}\frac{U_t(\mu)d_H(\overline{\mu}, \mu) - U_t(\mu)d_H(\overline{\mu}, \nu)}{d_H(\overline{\mu}, \mu) - d_H(\overline{\mu}, \nu)} = 0.\end{aligned}$$

Therefore, since any local extrema must be maxima in this region and since $\overline{\mu}$ is a local maximum it follows $d_{V_t}(\nu, \mu)/d_H(\nu, \mu)$ must be nondecreasing on the region. Therefore $\overline{\mu}$ must be a global maximum for $\nu \in (\mu, \overline{\mu}]$.

Now consider $\nu \in [\mu^*, \mu)$. In this region, (following the same steps as before) $d_{V_t}(\nu, \mu)/d_H(\nu, \mu) > 0$ is increasing if

$$\frac{d_{V_t}(\overline{\mu}, \mu) - d_{V_t}(\overline{\mu}, \nu)}{d_H(\overline{\mu}, \mu) - d_H(\overline{\mu}, \nu)} \leq \frac{d_{V_t}(\nu, \mu)}{d_H(\nu, \mu)}. \tag{17}$$

This is the same condition as (16) except the inequality has flipped.

As before, to determine whether a local extremum is a maximum or minimum it suffices to check how the left-hand side changes as ν increases. In this case, the left-hand side is increasing. This can be seen since

$$\frac{d}{d\nu}\frac{d_{V_t}(\overline{\mu},\mu)-d_{V_t}(\overline{\mu},\nu)}{d_H(\overline{\mu},\mu)-d_H(\overline{\mu},\nu)} = \frac{d}{d\nu}\frac{U_t(\mu)d_H(\overline{\mu},\mu)-U_t(\nu)d_H(\overline{\mu},\nu)}{d_H(\overline{\mu},\mu)-d_H(\overline{\mu},\nu)}$$
$$> \frac{d}{d\nu}\frac{U_t(\mu)d_H(\overline{\mu},\mu)-U_t(\mu)d_H(\overline{\mu},\nu)}{d_H(\overline{\mu},\mu)-d_H(\overline{\mu},\nu)} = 0$$

where we have used the fact that the denominator is negative in this case. Thus in this region, any local extremum must be a local minimum.

Therefore, to complete the proof it suffices to show that

$$\frac{d_{V_t}(\overline{\mu},\mu)}{d_H(\overline{\mu},\mu)} \geq \frac{d_{V_t}(\nu,\mu)}{d_H(\nu,\mu)} \tag{18}$$

for all $\nu \in [\underline{\mu}, \mu^*]$. Following the same steps used to derive (16), except using the law of cosines with $\underline{\mu}$ instead of $\overline{\mu}$, we derive that the derivative of the objective is positive if and only if

$$\frac{d_{V_t}(\underline{\mu},\mu)-d_{V_t}(\underline{\mu},\nu)}{d_H(\underline{\mu},\mu)-d_H(\underline{\mu},\nu)} > \frac{d_{V_t}(\nu,\mu)}{d_H(\nu,\mu)}. \tag{19}$$

We will prove that the left-hand side of (19) is bounded above by $d_{V_t}(\overline{\mu},\mu)/d_H(\overline{\mu},\mu)$. Thus there can not be a point $\nu \in [\underline{\mu}, \mu^*]$ that achieves the supremum for the objective higher than $d_{V_t}(\overline{\mu},\mu)/d_H(\overline{\mu},\mu)$, since if there was, at that point, the derivative of the objective would be negative. Thus (18) must hold.

To show this, we first observe that by the law of cosines

$$d_{V_t}(\underline{\mu},\mu) = d_{V_t}(\underline{\mu},\overline{\mu}) + d_{V_t}(\overline{\mu},\mu) - (\underline{\mu}-\overline{\mu})\left(V_t'(\mu)-V_t'(\overline{\mu})\right), \tag{20}$$

and

$$d_H(\underline{\mu},\mu) = d_H(\underline{\mu},\overline{\mu}) + d_H(\overline{\mu},\mu) - (\underline{\mu}-\overline{\mu})\left(H'(\mu)-H'(\overline{\mu})\right). \tag{21}$$

Define $f(\mu)$ and $g(\mu)$ as

$$f(\mu) = d_{V_t}(\overline{\mu},\mu) - (\underline{\mu}-\overline{\mu})\left(V_t'(\mu)-V_t'(\overline{\mu})\right) \tag{22}$$

and

$$g(\mu) = d_H(\overline{\mu},\mu) - (\underline{\mu}-\overline{\mu})\left(H'(\mu)-H'(\overline{\mu})\right). \tag{23}$$

Since (15) binds when $\nu = \overline{\mu}$, it follows that

$$\frac{f(\mu)}{g(\mu)} = \frac{d_V(\overline{\mu},\mu) - (\underline{\mu} - \overline{\mu})\,(V'(\mu) - V'(\overline{\mu}))}{d_H(\overline{\mu},\mu) - (\underline{\mu} - \overline{\mu})\,(H'(\mu) - H'(\overline{\mu}))} = \frac{d_V(\overline{\mu},\mu)}{d_H(\overline{\mu},\mu)}. \tag{24}$$

Also since $d_V(\overline{\mu},\mu^*)/d_H(\overline{\mu},\mu^*) = d_V(\underline{\mu},\mu^*)/d_H(\underline{\mu},\mu^*)$,

$$\frac{f(\mu^*)}{g(\mu^*)} = \frac{d_{V_t}(\underline{\mu},\overline{\mu}) + f(\mu^*)}{d_H(\underline{\mu},\overline{\mu}) + g(\mu^*)} \Rightarrow \frac{f(\mu^*)}{g(\mu^*)} = \frac{d_{V_t}(\underline{\mu},\overline{\mu})}{d_H(\underline{\mu},\overline{\mu})}. \tag{25}$$

Thus,

$$\begin{aligned}
\frac{d_{V_t}(\underline{\mu},\mu) - d_{V_t}(\underline{\mu},\nu)}{d_H(\underline{\mu},\mu) - d_H(\underline{\mu},\nu)} &= \frac{d_{V_t}(\underline{\mu},\overline{\mu}) + f(\mu) - d_{V_t}(\underline{\mu},\nu)}{d_H(\underline{\mu},\overline{\mu}) + g(\mu) - d_H(\underline{\mu},\nu)} \\
&= \frac{U_t(\mu^*)d_H(\underline{\mu},\overline{\mu}) + U_t(\mu)g(\mu) - U_t(\nu)d_H(\underline{\mu},\nu)}{d_H(\underline{\mu},\overline{\mu}) + g(\mu) - d_H(\underline{\mu},\nu)} \\
&\leq \frac{U_t(\mu^*)d_H(\underline{\mu},\overline{\mu}) + U_t(\mu)g(\mu) - U_t(\mu^*)d_H(\underline{\mu},\nu)}{d_H(\underline{\mu},\overline{\mu}) + g(\mu) - d_H(\underline{\mu},\nu)} \\
&\leq U_t(\mu) = \frac{d_{V_t}(\overline{\mu},\mu)}{d_H(\overline{\mu},\mu)}.
\end{aligned}$$

as desired. The first line uses (20), (21), (22), and (23). The second line uses (24) and (25) and Lemma 2. The third line uses the fact that $U_t(\nu)$ is decreasing for $\nu \in [\underline{\mu},\mu^*]$. □

Lemma 4. *At each $(\mu, t) \in (\underline{\mu},\overline{\mu}) \times [0,\infty)$, it holds that*

$$U_t(\mu) \geq \frac{V_t''(\mu)}{H''(\mu)}.$$

Proof. When $\mu \in (\mu^*,\overline{\mu})$, $U_t'(\mu) > 0$. Therefore, using Lemma 2,

$$\begin{aligned}
&\frac{d}{d\mu}\frac{t - V_t(\mu) - V_t'(\mu)(\overline{\mu} - \mu)}{H(\overline{\mu}) - H(\mu) - H'(\mu)(\overline{\mu} - \mu)} \\
&\qquad = \frac{-d_H(\overline{\mu},\mu)V_t''(\mu)(\overline{\mu} - \mu) + d_{V_t}(\overline{\mu},\mu)H''(\mu)(\overline{\mu} - \mu)}{d_H(\overline{\mu},\mu)^2} > 0
\end{aligned}$$

We therefore have

$$\frac{V_t''(\mu)}{H''(\mu)} < \frac{d_{V_t}(\overline{\mu},\mu)}{d_H(\overline{\mu},\mu)} = U_t(\mu)$$

as desired. An analogous argument applies when $\mu \in (\underline{\mu}, \mu^*)$. We can use continuity to show weak inequality when $\mu = \mu^*$.

□

We have therefore verified that the HJB in Lemma 1 is satisfied with Greedy Exploitation which must therefore be optimal for all discount functions of the form ρ^T. By the results of Müller (1996) we immediately have optimality for all decreasing convex ρ. Thus the proof of Theorem 1 is complete.