Optimal Dynamic Information Acquisition*

(Job Market Paper)

Weijie Zhong†

January 20, 2019

Abstract. I study a dynamic model in which a decision maker (DM) acquires information about the payoffs of different alternatives prior to making her decision. The key feature of the model is the flexibility of information: the DM can choose any dynamic signal process as an information source, subject to a flow cost that depends on the informativeness of the signal. Under the optimal policy, the DM looks for a signal that arrives according to a *Poisson process*. The optimal Poisson signal confirms the DM's prior belief and is sufficiently accurate to warrant an immediate action. Over time, absent the arrival of a Poisson signal, the DM continues seeking an increasingly more precise but less frequent Poisson signal.

Keywords: dynamic information acquisition, rational inattention, stochastic control, Poisson-bandits

JEL classification: D11, D81, D83

1 Introduction

When individuals make decisions, they often have imperfect information about the payoffs of different alternatives. Therefore, the decision maker (DM) would like to acquire information to learn about the payoffs prior to making a decision. For example, when comparing new technologies, a firm may not know the profitability of alternative technologies. The firm often spends a considerable amount of money and time on R&D to identify the best technology to adopt. One practically important feature of the information acquisition process is that the choice of "what to learn" often involves considering a rich set of salient aspects. In the previous example, when designing the R&D process, a firm may choose which technology to test, how much data to collect and analyze, how intensive the testing should be, etc. Other examples include investors designing algorithms to learn about the returns of different assets, scientists conducting research to investigate the validity of different hypotheses, etc.

To capture such richness, in this paper, I consider a DM who can choose "what to learn" in terms of *all* possible aspects, as well as "when to stop learning". The main goal is to

*I am indebted to Yeon-Koo Che for guidance and support. For helpful comments and discussions, I am grateful to Sylvain Chassang, Mark Dean, Marina Halac, Johannes Hörner, Navin Kartik, Qingmin Liu, Konrad Mierendorff, Xiaosheng Mu, Pietro Ortoleva, Andrea Prat, Jakub Steiner, Philipp Strack, Tomasz Strzalecki, Andrzej Skrzypacz, Ming Yang as well as participants at conferences and seminars.

First version: June 15, 2016, Latest version: https://goo.gl/xCg6R1, Supplemental material: https://goo.gl/hzrzac.

†*Department of Economics, Columbia University*, Email: wz2269@columbia.edu.

obtain insight into dynamic information acquisition without restriction on what type of information can be acquired. In contrast to my approach, the classic approach is to focus on one aspect while leaving all other aspects exogenously fixed. The seminal works by Wald (1947) and Arrow, Blackwell, and Girshick (1949) study the choice of "when to stop" in a stopping problem with all aspects of the learning process being exogenous. Building upon the Wald framework, Moscarini and Smith (2001) endogenize one aspect of learning, the *precision*, by allowing the DM to control a precision parameter of a Gaussian signal process. Che and Mierendorff (2016) endogenize another aspect of learning, the *direction*, by allowing the DM to allocate limited attention to different news sources, each biased in a different direction. Here, by allowing all learning aspects to be endogenous, the current paper contributes by studying which learning aspect(s) is(are) endogenously relevant for the DM and how the optimal strategy is characterized in terms of these aspects.

In the model, the DM is to choose from a set of actions, whose payoffs depend on a state unknown to the DM. The state is initially selected by nature and remains fixed over time. At any instant of time, the DM chooses whether to stop learning and select an action or to continue learning by *nonparametrically* choosing the evolution of the belief process. The choice of a nonparametric belief process models the choice of a dynamic information acquisition strategy with no restriction on any aspect. I introduce two main economic assumptions. (i) The DM discounts delayed payoffs. (ii) Learning incurs a flow cost, which depends convexly on how fast the uncertainty about the unknown state is decreasing. The main model is formulated as a stochastic control-stopping problem in continuous time.

The main result shows that the optimal strategy is contained in a simple family characterized by a few endogenously relevant aspects (Theorem 1) and fully solves for the optimal strategy in these aspects (Theorems 2 and 3). Specifically, the first result states that although the model is nonparametric and allows for fully flexible strategies, the belief process can be restricted to a simple *jump-diffusion process* without loss. In other words, a combination of a *Poisson signal*—a rare and substantial breakthrough that causes a jump in belief—and a *Gaussian signal*—frequent and coarse evidence that drives belief diffusion—is endogenously optimal. A jump-diffusion belief process is characterized by four parameters: the *direction*, *size* and *arrival rate* of the jump, and the flow variance of the diffusion. The four parameters represent four key aspects of learning: the *direction*, *precision* and *frequency* of the Poisson signal, and the precision of the Gaussian signal. The first result suggests that the DM need consider only the trade-offs among these aspects; any other aspect is irrelevant for information acquisition.

The second result fully characterizes the parameters of the optimal belief process. I find that the Poisson signal strictly dominates the Gaussian signal almost surely, i.e. no resources should ever be invested in acquiring the Gaussian signal. The optimal Poisson signal satisfies the following qualitative properties in terms of the three aspects and the stopping time:

- ***Direction***: The optimal direction of learning is *confirmatory*– the arrival of a Poisson signal induces the belief to jump toward the state that the DM currently finds to be most likely. As an implication of Bayes rule, the absence of a signal causes the belief to drift gradually towards the opposite direction, namely, the DM gradually becomes less certain about the state.

- ***Precision***: The optimal signal precision is *negatively related* to the continuation value. Therefore, when the DM is less certain about the state, the corresponding continuation value is lower, which leads the DM to seek a more precise Poisson signal.
- ***Frequency***: The optimal signal frequency is *positively related* to the continuation value. In contrast to precision, the optimal signal frequency decreases when the DM is less certain.
- ***Stopping time***: The optimal time to stop learning is immediately after the arrival of the Poisson signal. Therefore, the breakthrough happens only once at the optimum. Then, the DM stops learning and chooses an optimal action based on the acquired information.

The optimal strategy is very heuristic and easy to implement. In the previous example, the firm can choose the technology to test, as well as the test precision and frequency. As a result, the optimal strategy is implementable. The optimal R&D process involves testing the most promising technology. The optimal test is designed to be difficult to pass, so good news comes infrequently, as in a Poisson process. A successful test confirms the firm's prior conjecture that the technology is indeed good and the firm immediately adopts the technology. Otherwise, the firm continues the R&D process. No good news is bad news, so the firm becomes more pessimistic about the technology and revises the choice of the most promising technology accordingly. The future tests involve higher passing thresholds and lower testing frequency. As illustrated by the example, although this paper studies a benchmark with fully flexible information acquisition, the optimal strategy applies to more general settings where information acquisition is *not* fully flexible, but involves these salient aspects.

The main intuition behind the optimal strategy is a novel *precision-frequency trade-off*. Consider a thought experiment of choosing an optimal Poisson signal with fixed direction and cost level. The remaining two parameters—precision and frequency—are pinned down by the marginal rate of substitution between them. Importantly, the trade-off depends on the continuation value. Due to discounting, when the continuation value is higher, the DM loses more from delaying the decision. Therefore, the DM finds it optimal to acquire a signal more frequently at the cost of lowering the precision to avoid costly delay. In other words, the marginal rate of substitution of frequency for precision is increasing in the continuation value. As a result, frequency (precision) is positively (negatively) related to the continuation value.

In addition to precision and frequency, this intuition also explains other aspects. First, the Gaussian signal is equivalent to a special Poisson signal with close to zero precision and infinite frequency. The previous intuition implies that infinite frequency is generally suboptimal except when the continuation value is so high that the DM would like to sacrifice almost all signal precision. As a result, the Gaussian signal is strictly suboptimal except for the non-generic stopping boundaries. Second, for any fixed learning direction, Bayes rule implies that the absence of a signal pushes belief away from the target direction; to ensure the same level of decision quality the signal precision should increase over time to offset the belief change. By acquiring a confirmatory signal, the DM becomes more pessimistic and, consequently, more patient over time. Therefore she can reconcile both incentives through reducing the signal frequency and increasing the signal precision. By contrast, if the DM acquires a contradictory signal, she becomes more impatient over time and prefers the frequency to be increasing. The two incentives become incongruent, thus, learning in a confirmatory way is optimal.

This intuition suggests that the crucial assumption for the optimal strategy is discounting — discounting drives the key precision-frequency trade-off. This observation highlights the deep connection between dynamic information acquisition and the DM's attitude toward time-risk. Discounting implies that the DM is risk loving toward payoffs with uncertain resolution time, as the exponential discounting function is convex. Intuitively, the riskiest information acquisition strategy is a "greedy strategy" that front-loads the probability of success as much as possible, at the cost of a high probability of long delays. The confirmatory Poisson learning strategy in this paper exactly resembles a greedy strategy. The key property of the strategy is that all resources are used in verifying the conjectured state directly and no intermediate step occurs before a breakthrough. By contrast, alternative strategies, such as Gaussian learning and contradictory Poisson learning, involve accumulating substantial intermediate evidence to conclude a success. The intermediate evidence in fact hedges the time risk: the DM sacrifices the possibility of immediate success to accelerate future learning.

Extensions of the main model further illustrate the role played by each key assumption. The first extension replaces discounting with a fixed flow delay cost. In this special case, all dynamic learning strategies are equally optimal, as the crucial precision-frequency tradeoff becomes value independent. This extension also illustrates that all learning strategies in the model are equally "fast" on average and differ only in "riskiness". This result further illustrates that the preference for time risk pins down the optimal strategy. Second, I consider general cost structures and find that the (strict) optimality of a Poisson signal over a Gaussian signal is surprisingly robust: it requires a minimal *continuity* assumption. Third, I study an extension where the flow cost depends linearly on the uncertain reduction speed. In this special case, learning has a constant return to signal frequency. As a result, the optimal strategy is to learn infinitely fast, that is, acquire all information at period zero.

This paper provides rich implications by allowing learning to be flexible in all aspects.

First, the main results highlight the optimality of the Poisson signal compared to the widely adopted diffusion models. Specifically, the diffusion models are shown to be justified only under the lack of discounting. Second, the characterization of the optimal strategy unifies and clarifies insights from some existing results. In these results, although the DM is limited in her learning strategy, she actually implements the flexible optimum whenever feasible and approximates the flexible optimum when infeasible. Moscarini and Smith (2001)'s insight that the "intensity" of experimentation increases in continuation value carries over to my analysis. I further unpack the design of experiment and show that higher "intensity" contributes to faster signal arrival but lower signal precision. Che and Mierendorff (2016) make same prediction about the learning direction as that of my analysis when the DM is uncertain about the state. But they predict the opposite when the DM is more certain about the state– the DM looks for a signal contradicting the prior belief. I clarify that the contradictory signal is an approximation of a high-frequency confirmatory signal when the DM is constrained in increasing the signal frequency.

The rest of this paper is structured as follows. The related literature is reviewed in Section 2. The main continuous-time model and illustrative examples are introduced in Section 3. The dynamic programming principle and the corresponding Hamilton-Jacobi-Bellman (HJB) equation are introduced in Section 4. I analyze an auxiliary discrete-time problem and verify the HJB equation in Section 5. Section 6 fully characterizes the optimal strategy and illustrates the intuition behind the result. In Section 7 I discuss the key assumptions used in the model. Section 8 explores the implications of the main model on response time in stochastic choice and on a firm's innovation. Further discussions of other assumptions are presented in Appendix A, and key proofs are provided in Appendix B. All the remaining proofs are relegated to the Supplemental material.

2 Related literature

2.1 *Dynamic information acquisition*

My paper is closely related to the literature about acquiring information in a dynamic way to facilitate decision making. The earliest works focus on the duration of learning. Wald (1947) and Arrow, Blackwell, and Girshick (1949) analyze a *stopping problem* where the DM controls the decision time and action choice given exogenous information. Moscarini and Smith (2001) extend the Wald model by allowing the DM to control the precision of a Gaussian signal. A similar Gaussian learning framework is used as the learning-theoretic foundation for the drift-diffusion model (DDM) by Fudenberg, Strack, and Strzalecki (2018). Following a different route, Che and Mierendorff (2016), Mayskaya (2016) and Liang, Mu, and Syrgkanis (2017) study the sequential choice of information sources, each of which is prescribed exogenously.

Other frameworks of dynamic information acquisition include sequential search models (Weitzman (1979), Callander (2011), Klabjan, Olszewski, and Wolinsky (2014), Ke and Villas-Boas (2016) and Doval (2018)) and multi-arm bandit models (Gittins (1974), Weber et al. (1992), Bergemann and Välimäki (1996) and Bolton and Harris (1999)). These frameworks are quite different from my information acquisition model. However, the forms of information in these models are also exogenously prescribed, and the DM has control over only whether to reveal each option.

Compared to the canonical approaches, the key new feature of my framework is that the DM can design the information generating process nonparametrically. In a similar vein to this paper, two concurrent papers Steiner, Stewart, and Matějka (2017) and Hébert and Woodford (2016) model dynamic information acquisition nonparametrically; however they focus on other implications of learning by abstracting from sequentially smoothing learning. In Steiner, Stewart, and Matějka (2017) the linear flow cost assumption makes it optimal to learn instantaneously, whereas in Hébert and Woodford (2016), the no-discounting assumption makes all dynamic learning strategies essentially equivalent.[1] By contrast, the main focus of this paper is on characterizing the optimal way to smooth learning. I analyze the setups of these two papers as special cases in Sections 7.1 and 7.3.

A main result of my paper is the endogenous optimality of Poisson signals. Section 7.2 shows a more general result: a Poisson signal dominates a Gaussian signal for generic cost functions that are continuous in the signal structure. This result justifies Poisson learning models, which are used in a wide range of problems, e.g., Keller, Rady, and Cripps (2005), Keller and Rady (2010), Che and Mierendorff (2016), and Mayskaya (2016); see also a survey by Hörner and Skrzypacz (2016).

2.2 *Rational inattention*

This paper is a dynamic extension of the static rational inattention (RI) models, which consider the flexible choice of information. The entropy-based RI framework is first introduced in Sims (2003). Matějka and McKay (2014) study the flexible information acquisition problem using an entropy-based informativeness measure and justify a generalized logit decision rule. Caplin and Dean (2015) take an axiomatization approach and characterize decision rules that can be rationalized by an RI model. On the other hand, this paper also serves as a foundation for RI models, as it characterizes, in detail, how the reduced-form decision rule is supported by acquiring information dynamically. In several limiting cases, my model completely reduces to a standard RI model.

The RI framework is widely used in models with strategic interactions (Matějka and McKay (2012), Yang (2015a), Yang (2015b), Matějka (2015), Denti (2015), etc). My paper is different from these works as no strategic interaction is considered and the focus is on repeated learning. Despite the strategic component, Ravid (2018) also studies a dynamic model with repeated learning. In Ravid (2018), an RI buyer learns sequentially about the offers from a seller and the value of the object being traded. Similar to the DM in my model,

[1] Steiner, Stewart, and Matějka (2017) assume the decision problem to be history dependent. Therefore, non-trivial dynamics remain in the optimal signal process. However, the dynamics are a results of the history dependence of the decision problem rather than the incentive to smooth information. In the dynamic learning foundation of Hébert and Woodford (2016), all signal processes are equally optimal because of a key no-discount assumption. They select a Gaussian process exogenously to justify a neighbourhood-based static information cost structure.

the buyer systematically delays trading in equilibrium, and the stochastic delay resembles the arrival of a Poisson process.[2] However, in Ravid (2018), the delay is an equilibrium property that ensures the buyer's strategy is responsive to off-path offers. By contrast, the stochastic delay in my paper is a property of an optimally smoothed learning process.

I use the reduction speed of uncertainty as a measure of the amount of information acquired per unit time. This measure captures the *posterior separability* from Caplin and Dean (2013). The posterior separable measure nests *mutual information* (introduced in Shannon (1948)) as a special case and is widely used in Gentzkow and Kamenica (2014), Clark (2016), Matyskova (2018), Rappoport and Somma (2017), etc. I provide an axiomatization for posterior separability based on the chain rule in Appendix A.4.1. Caplin, Dean, and Leahy (2017) axiomatize (uniform) posterior separability based on behavior data. Morris and Strack (2017) provide a dynamic foundation for posterior separability based on implementing an information structure with Gaussian learning. In addition to axiomatizing posterior separability, Frankel and Kamenica (2018) relates to my paper in another interesting way. The *valid measure of information* defined in their paper coincides with the uncertainty reduction speed per unit arrival rate of a Poisson signal derived in this paper.

2.3 Information design

In this paper, I use a belief-based approach to model the choice of information. This approach is widely used for studying Bayesian persuasion models (Kamenica and Gentzkow (2011), Ely (2017), Mathevet, Perego, and Taneva (2017), etc.). An important methodology in this literature is the concavification method developed in Aumann, Maschler, and Stearns (1995) (based on Carathéodory's theorem). An alternative approach to model information is the direct signal approach [3] used in both information design problems, such as Bergemann and Morris (2017), and rational inattention problems. However, neither of the two methods applies to my dynamic information acquisition problem. I take the belief-based approach as in Bayesian persuasion models, but utilize the generalized concavification method developed in Zhong (2018a).

2.4 Stochastic control

Methodologically, this paper is closely related to the theory of continuous-time stochastic control. The early theories study control processes measurable to the natural filtration of Brownian motion (see Fleming (1969) for a survey). The application of Bellman (1957)'s dynamic programming principle leads to the HJB equation characterization of the value function. On the contrary, the main stochastic control problem of this paper has general martingale control process, which is a variant of the (semi)martingale models of stochastic control studied in Davis (1979), Boel and Kohlmann (1980), Striebel (1984), etc. However, none of the existing theories are sufficiently general to nest the stochastic control problem studied in this paper. I introduce an indirect method that proves a verification theorem for

[2] Precisely speaking, in the analysis of Proposition 2, Ravid (2018) shows that when quality is deterministic, the delay time distribution is exponential, which is the same as the stopping time induced by a Poisson signal process.

[3] This approach applies to settings where without loss of generality we can restrict the problem to considering only signals that are direct recommendations of actions.

a tractable HJB equation.

3 Model setup

The main model is a continuous-time stochastic control problem. A DM chooses an irreversible action at an endogenous decision time. The DM can control the information received before the decision time in a flexible manner, bearing a cost on information.

Decision problem: Time $t \in [0, +\infty)$. The DM discounts the delayed utility with rate $\rho > 0$. The DM is a vNM expected utility maximizer with Bernoulli utility associated with action-state pair $(a, x) \in A \times X$ at time t being $e^{-\rho t} u(a, x)$. Both the action space A and the state space X are finite. The DM holds a prior belief $\mu \in \Delta(X)$ about the state. Define $F(\nu) \triangleq \max_{a \in A} E_\nu[u(a, x)]$ given belief $\nu \in \Delta(X)$.

Information: I model information using a belief-based approach. A distribution of posterior beliefs is induced by an information structure according to Bayes rule *iff* the expectation of posterior beliefs is equal to the prior. Hence, in a static environment the choice of information can be equivalently formulated as the choice of a distribution of posterior beliefs (see Kamenica and Gentzkow (2011) for example). Extending this formulation to the dynamic environment in the current paper, I assume that the DM chooses the entire posterior belief process $\langle \mu_t \rangle$ in a nonparametric way. Now Bayes' rule should be satisfied at every instant of time—$\forall s > t$, the expectation of μ_s is μ_t. Thus, I restrict $\langle \mu_t \rangle$ to be a martingale, with $\langle \mathcal{F}_t \rangle$ as its natural filtration. A formal justification that choosing a belief martingale is equivalent to choosing a dynamic information structure is provided in Appendix A.4.

It is useful to define the following operator $\mathcal{L}_t$ for any $\langle \mu_t \rangle$ and $f : \Delta(X) \to \mathbb{R}$:

$$\mathcal{L}_t f(\mu_t) = E\left[\frac{\mathrm{d}f(\mu_t)}{\mathrm{d}t}\middle|\mathcal{F}_t\right] \triangleq \lim_{t' \to t^+} E\left[\frac{f(\mu_{t'}) - f(\mu_t)}{t' - t}\middle|\mathcal{F}_t\right]$$

By definition, $\mathcal{L}_t f$ captures the expected *speed* at which $f(\mu_t)$ increases. Let $\mathscr{D}(f)$ be the domain of $\langle \mu_t \rangle$ on which $\mathcal{L}_t f(\mu_t)$ is well defined.[4] For well-behaved Markov process $\langle \mu_t \rangle$ and $C^{(2)}$ smooth f, $\mathcal{L} f$ is the standard *infinitesimal generator* (subscript t omitted).

Cost of information: I assume that the flow cost of information depends on how fast the information reduces uncertainty. The flow cost of information is $C(I_t)$, where:

Assumption 1. *$I_t = -\mathcal{L}_t H(\mu_t)$, where $H : \Delta(X) \to \mathbb{R}$ is concave and continuous.*

I call H an *uncertainty measure*—because H is concave *iff* $E[H(\mu)]$ captures the Blackwell order on the belief distribution. By Assumption 1, I_t is the *speed* at which uncertainty falls when the belief updates. I call I_t the *(flow) informativeness measure*. One example of H is the entropy function $H(\mu) = -\sum \mu_x \log(\mu_x)$. Revelation of information reduces entropy; hence, the entropy reduction speed is a natural measure of the amount information.

[4] Formally, $\langle \mu_t \rangle \in \mathscr{D}(f)$ if the uniform limit (w.r.t t) exists *almost surely*. Let $\mathscr{D} = \bigcap_{f \in C(\Delta X)} \mathscr{D}(f)$. $\mathscr{D}$ contains all Feller processes, whose transition kernels are *stochastically continuous* w.r.t. t and *continuous* w.r.t. state μ. However, $\mathscr{D}$ is much more general than Feller processes as it allows the transition kernel to be discontinuous in state μ.

Assumption 1 is the main technical assumption in my analysis. I generalize this assumption in Section 7.2. For further discussions, see Appendix A.4, where I show that it is the continuous-time analog of "posterior separability" and provide an axiom for posterior separability.

Stochastic control: The DM solves the following stochastic control problem:

$$V(\mu) = \sup_{\langle \mu_t \rangle \in \mathbb{M}, \tau} E\left[e^{-\rho\tau} F(\mu_\tau) - \int_0^\tau e^{-\rho t} C(I_t) \mathrm{d}t \right] \tag{1}$$

where $\mathbb{M}$ is the set of all martingales $\langle \mu_t \rangle$ in $\mathscr{D}(H)$ with cadlag[5] path and satisfying $\mu_0 = \mu$, and τ is a $\langle \mathcal{F}_t \rangle$-measurable stopping time.[6]

The objective function in Equation (1) is fairly standard in canonical information acquisition problems. The DM acquires information that affects $\langle \mu_t \rangle$ and chooses stopping time τ to maximize the expected stopping payoff $E\left[e^{-\rho\tau}F(\mu_\tau)\right]$ less the total information cost $E\left[\int_0^\tau e^{-\rho t} C(I_t)\mathrm{d}t\right]$. The novel feature is that the DM is allowed to fully control $\langle \mu_t \rangle$, in contrast to canonical models, where the DM controls only a few parameters determining $\langle \mu_t \rangle$. The nonparametric control of the belief process exactly captures the flexible design of information by the DM.

I make the following assumption on the cost function $C(I)$ to generate incentive to smooth learning over time.

Assumption 2. $C : \mathbb{R}^+ \to \overline{\mathbb{R}}^+$ *is weakly increasing, convex and continuous.* $\lim\limits_{I\to\infty} C'(I) = \infty$.

The increasing and continuous cost function assumption is standard. The convexity of $C(I)$ and the condition $\lim C'(I) = \infty$ give the DM strict incentive to smooth the acquisition of information. Given Assumption 2, if the DM acquires all information immediately then uncertainty falls at infinite speed and the marginal cost $C'(I)$ is infinite, hence suboptimal.[7] I solve a special case violating Assumption 2 in Section 7.3, where I assume C to be linear. In this case the optimal strategy is to acquire all information at $t = 0$ (a static strategy).

In Example 1, I present a few examples of canonical Wald-type sequential learning models, each of which is a variant of Equation (1) with additional constraints on the set of admissible belief processes. Example 1 first illustrates how different learning technologies can be systematically compared under the same framework with an entropy-based cost function. The comparison also illustrates why a fully flexible learning framework is useful.

Example 1. Let the state be binary $X = \{l, r\}$. The prior belief of state $x = r$ is $\mu \in (0,1)$. $A = \{L, R\}$. The DM wants to choose an action that matches the state: $u(L,l) = u(R,r) = 1$;

---

[5] cadlag: $\mu_t : t \mapsto \Delta(X)$ is right continuous with left limits. Note that assuming martingale $\langle \mu_t \rangle$ being cadlag can be weakened to assuming $\langle \mathcal{F}_t \rangle$ being right continuous (see the martingale modification theorem in Lowther (2009)).

[6] I postpone the formal definition of integrability in Equation (1) to Section 5.1. For now, assume that the integral is well defined for all admissible strategies. Further discussions in Remark B.2 provide a formal justification that ignoring the integrability is innocuous.

[7] A weaker sufficient condition can guarantee information smoothing: $\sup_I \bar{\lambda} I - C(I) > \rho \sup F$, where $\bar{\lambda} = \underline{\lim}_{I\to\infty} \frac{C(I)}{I}$. This condition explicitly states that when I is sufficiently large, C is sufficiently convex that the utility gain from smoothing information dominates the loss from waiting longer. All the following theorems in this paper are proved under this weaker condition.

$u(L,r) = u(R,l) = -1$. The discount rate $\rho = 1$, H is the standard entropy function: $H(\mu) = -\mu\log(\mu) - (1-\mu)\log(1-\mu)$, and the information cost $C(I) = \frac{1}{2}I^2$.

I consider three simple heuristic learning technologies: Gaussian learning, perfectly revealing breakthroughs and partially revealing evidence. A DM who uses a specific learning technology is modeled by restricting the admissible control set $\mathbb{M}$ to include only the corresponding family of processes. In each case, the DM controls a parameter that represents one aspect of learning.

1. *Gaussian learning*: The signal follows a Brownian motion whose drift is the true state, and whose variance is controlled by the DM. Therefore, the posterior belief follows a diffusion process (Bolton and Harris (1999)), so the set of admissible controls are:

$$\mathbb{M}_D = \{\langle \mu_t \rangle | \mathrm{d}\mu_t = \sigma_t \mathrm{d}W_t\}$$

The DM controls the signal precision $\langle \sigma_t \rangle$. According to Ito's lemma, $I_t = -\frac{1}{2}\sigma_t^2 H''(\mu_t) = \frac{\sigma_t^2}{2\mu_t(1-\mu_t)}$. This problem is studied in Moscarini and Smith (2001)[8], where the value function is characterized by HJB:

$$\rho V_D(\mu) = \sup_{\sigma>0} \frac{1}{2}\sigma^2 V_D''(\mu) - \frac{1}{2}\left(\frac{\sigma^2}{2\mu(1-\mu)}\right)^2$$

The solution $V_D(\mu)$ is plotted as the blue curve in Figure 1. The shaded region is the experimentation region and the non-shaded region is the stopping region.

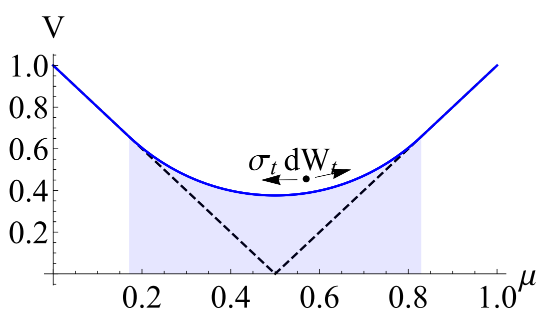


Figure 1: Incremental information

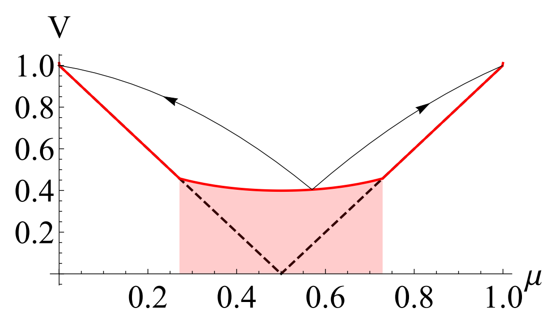


Figure 2: Breakthroughs

2. *Breakthroughs*: The DM observes breakthroughs that perfectly reveal the true state with arrival rate λ_t. Then, belief follows a Poisson process that jumps to 1 if the state is r and to 0 if the state is l. The set of admissible control is:

$$\mathbb{M}_B = \left\{\langle \mu_t \rangle | \mathrm{d}\mu_t = (1-\mu_t)\mathrm{d}J_t^1(\lambda_t\mu_t) + (0-\mu_t)\mathrm{d}J_t^0(\lambda_t(1-\mu_t))\right\}$$

$\langle J_t^i(\cdot) \rangle$ are independent Poisson counting processes with Poisson rate $(\cdot)$. The DM controls the signal frequency $\langle \lambda_t \rangle$. The Entropy reduction speed is $\lambda_t H(\mu)$. The HJB equation is as follows:

$$\rho V_B(\mu) = \sup_{\lambda>0} \lambda(\mu F(1) + (1-\mu)F(0) - V_B(\mu)) - \frac{1}{2}(\lambda H(\mu))^2$$

[8] With "belief elasticity" defined as $\mathscr{E}(\mu) = \mu(1-\mu)$ in my model.

The solution V_B is plotted as the red curve in Figure 2. The two arrows show the belief jumps induced by breakthroughs at μ.

3. *Partially revealing evidence*: The DM allocates one unit of total attention to two news sources, each revealing one state with arrival rate $\gamma = 2$. Then belief follows a compensated Poisson process, and the set of admissible belief processes is:

$$\mathbb{M}_P = \left\{ \langle \mu_t \rangle \middle| \begin{aligned} \mathrm{d}\mu_t =&(1-\mu_t)(\mathrm{d}J_t^1(\alpha_t\gamma\mu_t) - \alpha_t\gamma\mu_t\mathrm{d}t) \\ &+(0-\mu_t)(\mathrm{d}J_t^0((1-\alpha_t)\gamma(1-\mu_t)) - (1-\alpha_t)\gamma(1-\mu_t)\mathrm{d}t) \end{aligned} \right\}$$

$\langle J_t^i(\cdot)\rangle$ are independent Poisson counting processes with Poisson rate $(\cdot)$. The DM controls $\langle \alpha_t \rangle$, the attention allocated to the signal revealing state r. This control process is identical to that in Che and Mierendorff (2016). Applying their analysis, optimal α_t is a bang-bang solution, and the HJB equation is:

$$\begin{aligned}\rho V_P(\mu) = \max\Big\{&\gamma\mu\big(F(1)-V_P(\mu)-V_P'(\mu)(1-\mu)\big) - \frac{1}{2}\big(\gamma\mu(H(\mu)+H'(\mu)(1-\mu))\big)^2, \\ &\gamma(1-\mu)\big(F(0)-V_P(\mu)-V_P'(\mu)(0-\mu)\big) - \frac{1}{2}\big(\gamma(1-\mu)(H(\mu)+H'(\mu)(0-\mu))\big)^2\Big\}\end{aligned}$$

The solution V_P is plotted as the black curve in Figure 3. The optimal strategy is qualitatively the same as in Che and Mierendorff (2016). In the deep gray region, optimal learning direction is *confirmatory*: the arrival of news reveals the a priori more likely state (represented by solid arrows). In the light gray region, optimal learning direction is *contradictory*: the arrival of news reveals the a priori less likely state (dashed arrows).

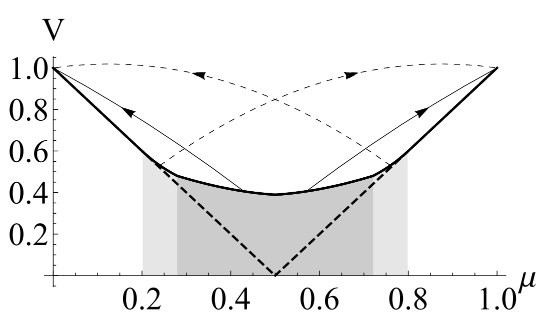


Figure 3: Partially revealing evidence

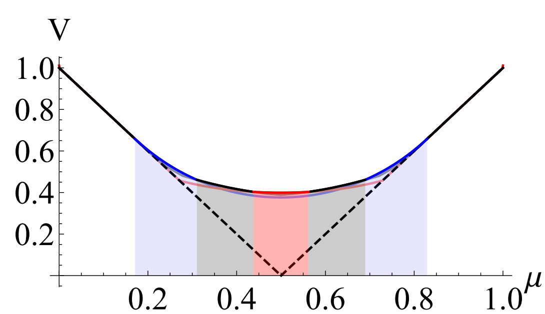


Figure 4: Comparison

In this example, the three learning technologies are analyzed for the same underlying decision problem and the same entropy cost function. Therefore, the utilities are directly comparable. I plot all three value functions in Figure 4 and use differently colored regions to illustrate the order of utility. Each color corresponds to a learning strategy being optimal: blue—Gaussian learning, red—breakthroughs, and gray—confirmatory evidence.[9] As shown in Figure 4, allowing the DM to use a rich set of strategies improves the decision-making quality.

[9] In this example, whenever contradictory learning dominates confirmatory learning, contradictory learning is dominated by Gaussian learning, thus, contradictory learning is not optimal in any region.

More interestingly, there appears to be a pattern when optimizing in different aspects. When the prior belief is highly uncertain, a fully revealing Poisson signal that can bring the DM directly to a conclusion is optimal. When the prior belief is quite uncertain but asymmetrically in favor of one state, allocating attention to the more promising direction becomes optimal. When the prior belief is very certain, an imprecise but frequent Gaussian signal becomes optimal. The formal analysis for fully flexible information acquisition in Section 6 illustrates that this pattern is systematic: the optimal direction, precision and frequency of learning are exactly the relevant aspects and are closely related to the location of the prior belief.

3.1 *Motivation for a flexible model*

Example 1 implies that single-aspect models are insufficient for modeling a dynamic information acquisition problem with a rich strategy set. For instance, the model considering only partially revealing evidence predicts that seeking contradictory evidence is generally optimal when the belief is uncertain. However, further analysis shows that this prediction is misleading when Gaussian signals are also feasible. Studying a model where information acquisition is flexible in *all* aspects enables us to obtain insights about information acquisition without interference from any ad hoc restriction. Such insights include which aspect(s) is(are) endogenously salient for information acquisition and how each of these aspects is determined by the DM's incentives.

Although the results are derived in a fully flexible model, they apply to much more general settings where information acquisition is *not* flexible in all aspects. First, all results directly apply to all settings where information acquisition is flexible in those endogenously salient aspects, as all other aspects are redundant for implementing the unconstrained optimum. Second, even for settings where some of the relevant aspects are constrained, the intuitions from the flexible model identify the DM's most important incentive and how the hypothetically ideal strategy might be approximated by adjusting other aspects. In fact, the analysis of the flexible model in Sections 4 and 6 shows that the set of endogenously salient aspects is quite small, and the optimal strategy satisfies very simple qualitative properties in these aspects. Therefore, the findings of this paper are useful in a very wide range of settings.

4 Dynamic programming and HJB equation

Solving Equation (1) is not an easy task due to the abstract strategy space. To the best of my knowledge, no general theory applicable to this stochastic control problem exists. The most closely related problems are studied in a set of remarkable papers on the martingale method in stochastic control (Davis (1979),Boel and Kohlmann (1980),Striebel (1984)). These papers introduce abstract formulations of stochastic control problems with general (semi)martingale control processes. The problems have finite horizon and specific objective functions; hence, they do not nest Equation (1).

Nevertheless, it is useful to introduce the general dynamic programming principle and

HJB characterization. On the basis of the intuition of dynamic programming, the conjecture that $V(\mu_t)$ satisfies the following HJB is reasonable:

$$\max\Big\{\underbrace{F(\mu_t)-V(\mu_t)}_{\text{stopping value}},\ \underbrace{-\rho V(\mu_t)}_{\text{discount}}+\sup_{\mathrm{d}\mu_t}\Big\{\underbrace{\mathcal{L}_t V(\mu_t)}_{\text{continuation value}}-\underbrace{C(-\mathcal{L}_t H(\mu_t))}_{\text{control cost}}\Big\}\Big\}=0 \tag{2}$$

HJB Equation (2) is conceptually the same as the standard HJB equation. Recall the definition for operator $\mathcal{L}_t$, $\mathcal{L}_t V(\mu_t)$ is the flow utility gain from continuing. The exact form of $\mathcal{L}_t V$ and $\mathcal{L}_t H$ depends on the probability space, the filtration and the control process in the neighbourhood of t (which are summarized by the symbol $\mathrm{d}\mu_t$). Therefore, Equation (2) essentially states the dynamic programming principle: at any instance when the control is chosen optimally, either stopping is optimal (the first term is 0) or continuing is optimal and the net continuation gain equals the loss from discounting (the second term is 0).

For a simple example, let $\mathbb{M}$ be a family of Markov jump-diffusion belief processes, characterized by the following SDE:

$$\mathrm{d}\mu_t=\underbrace{(\nu(\mu_t)-\mu_t)(\mathrm{d}J_t(p(\mu_t))-p(\mu_t)\mathrm{d}t)}_{\text{compensated Poisson part}}+\underbrace{\sigma(\mu_t)\mathrm{d}W_t}_{\text{Gaussian diffusion}} \tag{3}$$

where $(p,\nu,\sigma):\mu_t\mapsto\mathbb{R}^+\otimes\Delta(\mathrm{Supp}(\mu))\otimes\mathbb{R}^{|\mathrm{Supp}(\mu)|-1}$ are control parameters, $J_t(\cdot)$ is a Poisson counting process with Poisson rate $(\cdot)$, and W_t is a standard one-dimensional Wiener process. Note that this example also nests all three families of strategies in Example 1 as special cases[10]. Itô's lemma implies an explicit form for the infinitesimal generator:

$$\mathcal{L}V(\mu)=\underbrace{p(V(\nu)-V(\mu)-\nabla V(\mu)(\nu-\mu))}_{\text{flow value of Poisson jump \& drift}}+\underbrace{\frac{1}{2}\sigma^T\mathrm{H}V(\mu)\sigma}_{\text{flow value of diffusion}}$$

where ∇ and H are the gradient and Hessian operators, respectively. By replacing $\mathcal{L}$ in Equation (2) with its explicit expression, we obtain a parametrized HJB Equation (4):

$$\begin{aligned}\rho V(\mu)=\max\Big\{\rho F(\mu),\ \sup_{p,\nu,\sigma}\ & p(V(\nu)-V(\mu)-\nabla V(\mu)(\nu-\mu))+\frac{1}{2}\sigma^T\mathrm{H}V(\mu)\sigma \\ & -C\Big(p(H(\mu)-H(\nu)+\nabla H(\mu)(\nu-\mu))-\frac{1}{2}\sigma^T\mathrm{H}H(\mu)\sigma\Big)\Big\}\end{aligned} \tag{4}$$

On the other hand, when $\mathbb{M}$ is the jump-diffusion family, the jump-diffusion control theory (see textbooks, e.g., Hanson (2007)) provides a *verification theorem* that proves that the value function for Equation (1) is exactly characterized by HJB Equation (4).

This simple example illustrates how a specific stochastic control problem relates to an HJB equation. Now, consider the general problem Equation (2) without any restriction on the admissible belief process. First, we require a verification theorem stating that the HJB Equation (2) characterizes the solution of Equation (1). Second, a representation theorem

[10]The admissible control sets in the second and third cases in Example 1 are not exactly nested in Equation (3). However, they can be viewed as mixed strategies of pure Poisson-jump processes defined by Equation (3).

for the abstract operator $\mathcal{L}_t$ is also necessary to make Equation (2) practically tractable. The existing theories on martingale methods have little power for both tasks.[11] In Theorem 1, I achieve both goals by showing that the solution of Equation (1) is characterized by a simple parametric HJB equation:

Theorem 1. *Assume H is strictly concave and $C^{(2)}$ smooth on interior beliefs in $\Delta(X)$, Assumptions 1 and 2 are satisfied. Let $V(\mu) \in C^{(1)}\Delta(X)$ be a solution[12] to HJB Equation (4); then $V(\mu)$ solves Equation (1).*

Theorem 1 first states that $V(\mu)$ is characterized by a HJB equation. More surprisingly, Theorem 1 also states that the HJB is exactly Equation (4). As a direct corollary, Equation (1) can be solved by considering only the family of Markov jump-diffusion processes characterized by SDE (3). The compensated Poisson jump part and Gaussian diffusion part in SDE (3) each represents a simple learning strategy.

- ***Poisson learning***: The DM uses *Poisson learning* or acquires a *Poisson signal* when a compensated Poisson part exists in the belief process. A Poisson jump in the belief process can be induced by observing non-conclusive news whose arrival follows a Poisson process. The compensating belief drift is induced by observing no news arriving. The control variables for Poisson learning are (p, ν), which represent three endogenously relevant aspects of Poisson learning. The arrival rate p represents the *frequency* of learning. The direction of belief jump represents the *direction* of learning. The magnitude of belief jump represents the *precision* of learning.

- ***Gaussian learning***: The DM uses *Gaussian learning* or acquires a *Gaussian signal* when a diffusion part exists in the belief process. Gaussian diffusion in the belief process can be induced by observing the realization of a Gaussian process, with state x being the unobservable drift. The flow variance σ represents the signal precision.

Equation (4) suggests that to determine the optimal strategy in all relevant aspects, the DM considers four types of trade-offs : (i) the standard continuing-stopping trade-off in optimal stopping problems, captured by the outer-layer maximization; (ii) the information cost-utility gain trade-off, which determines the total cost spent on learning; (iii) the Poisson-Gaussian trade-off, which determines the proportion of cost allocated to the Poisson signal (p, ν) and the Gaussian signal σ; (iv) the precision-frequency trade-off, which determines the marginal rate of substitution of signal frequency for precision. These trade-offs, especially the precision-frequency trade-off, will be discussed in detail to characterize the solution to Equation (4) in Section 6.

[11] First, the existing martingale methods verify the HJB equation for different sets of problems that do not cover this specific problem. Moreover, the martingale method only states the existence of such $\mathcal{L}_t V$ (for example theorem 4.3.1 of Boel and Kohlmann (1980)) and does not provide an explicit representation. This issue is considered to be the main drawback of the martingale method (see discussions in Davis (1979)).

[12] The $C^{(1)}$ solution to the second-order ODE is not well defined. To be precise, V is a viscosity solution (see Crandall, Ishii, and Lions (1992)). In the viscosity solution, $\sigma^T HV(\mu)\sigma$ is replaced by $D^2V(\mu,\sigma)||\sigma||^2$, where $D^2V(\mu,\sigma) = \overline{\lim_{\delta\to 0}}\, 2\frac{V(\mu+\delta\sigma)-V(\mu)-\nabla V(\mu)\delta\sigma}{\delta\|\sigma\|^2}$.

The proof of Theorem 1 uses an indirect method. I characterize Equation (1) as the limit of a series of auxiliary discrete-time problems. The discrete-time analyses are presented in Section 5. Readers interested in the solution of HJB Equation (4) can jump to Section 6.

5 The auxiliary discrete-time problem

In this section, I introduce the steps for proving Theorem 1 using an auxiliary discrete-time problem. First, in Section 5.1 I introduce a discrete-time stochastic control problem that converges to the continuous-time problem. Then I characterize the Bellman equation for the discrete-time problem in Section 5.2. In Section 5.3, I introduce a key lemma that links all the discrete-time analyses and proves Theorem 1.

5.1 *Discrete-time problem*

I consider a stochastic control problem that is a discrete-time analog of Equation (1). Then I illustrate the discretization of the original problem. The discretization serves as a useful intermediary showing that the discrete-time problem converges to the continuous-time problem.

Decision problem: The primitives (A, X, u, μ, ρ) are the same as those in Section 3. Time is discrete $t \in \mathbb{N}$, and the period length $dt > 0$. The payoff delayed by t periods is discounted by $e^{-\rho dt \cdot t}$.

Information: The DM chooses the posterior belief process $\langle \widehat{\mu}_t \rangle$ in a nonparametric way. $\langle \widehat{\mu}_t \rangle$ is restricted to be a martingale. Let $\langle \widehat{\mathcal{F}}_t \rangle$ be the natural filtration of $\langle \widehat{\mu}_t \rangle$.

Cost of information: Define $C_{dt}(I) \triangleq C\left(\frac{I}{dt}\right)dt$. The per-period cost of information is assumed to be $C_{dt}(E[H(\widehat{\mu}_t) - H(\widehat{\mu}_{t+1})|\widehat{\mathcal{F}}_t])$. Note that this is exactly the finite-difference analog of the flow cost $C(-\mathcal{L}_t H(\mu_t))$ in the continuous-time problem.

Optimization problem: The DM solves the following stochastic control problem:

$$V_{dt}(\mu) = \sup_{\langle \widehat{\mu}_t \rangle \in \widehat{\mathbb{M}}, \widehat{\tau}} E\left[e^{-\rho dt \cdot \widehat{\tau}} F(\widehat{\mu}_{\widehat{\tau}}) - \sum_{t=0}^{\widehat{\tau}-1} e^{-\rho dt \cdot t} C_{dt}\left(E\left[H(\widehat{\mu}_t) - H(\widehat{\mu}_{t+1}) | \widehat{\mathcal{F}}_t \right] \right) \right] \tag{5}$$

where $\widehat{\mathbb{M}}$ is the set of discrete-time martingales satisfying $\widehat{\mu}_0 = \mu$, and τ is a $\langle \widehat{\mathcal{F}}_t \rangle -$measurable stopping time. Note that in this section, all discrete-time stochastic processes and random variables are labeled with "hat" to differentiate them from continuous-time processes.

The purpose of analyzing the discrete-time problem is to characterize the continuous-time value function $V(\mu)$. Therefore, the first step is to show that $V_{dt}(\mu)$ approximates $V(\mu)$. To study the relation between $V_{dt}(\mu)$ and $V(\mu)$, let us discretize the objective function in Equation (1). For any admissible strategy $(\langle \mu_t \rangle, \tau)$, consider the Riemann sum:

$$W_{dt}(\mu_t, \tau) = \sum_{i=1}^{\infty} \text{Prob}(\tau \in [(i-1)dt, idt]) E\left[e^{-i\rho dt} F(\mu_{idt}) - \sum_{j=0}^{i-1} e^{-j\rho dt} C\left(I_{jdt}\right) dt \right]$$

where $I_{jdt} = E\left[\frac{H(\mu_{jdt}) - H(\mu_{(j+1)dt})}{dt} | \mathcal{F}_{jdt} \right]$. The objective function in Equation (1) is defined in the notion of the Riemann-Stieltjes integral as $\lim_{dt \to 0} W_{dt}(\mu_t, \tau)$. I call the martingale $\langle \mu_t \rangle$

integrable if the limit $\lim_{dt\to 0} W_{dt}(\mu_t, \tau)$ exists.[13] Unless otherwise stated, $\mathbb{M}$ is restricted to contain integrable processes, an innocuous restriction that enables me to avoid technical discussions of integrability.[14] Then it follows that $V(\mu) = \sup_{\langle \mu_t \rangle \in \mathbb{M}, \tau} \lim_{dt\to 0} W_{dt}(\mu_t, \tau)$.

Now, consider the relation between W_{dt} and V_{dt}. I argue that the objective function in Equation (5) is equivalent to $W_{dt}(\mu_t, \tau)$. This result can ben verified by noting that if $(\langle \mu_t \rangle, \tau)$ and $(\langle \widehat{\mu}_t \rangle, \widehat{\tau})$ jointly satisfy $\widehat{\mu}_t = \mu_{t \cdot dt}$ and $\widehat{\tau} = \lceil \tau / dt \rceil$, then:

$$W_{dt}(\mu_t, \tau) = E\left[e^{-\rho dt \widehat{\tau}} F(\widehat{\mu}_{\widehat{\tau}}) - \sum_{t=0}^{\widehat{\tau}-1} e^{-\rho dt \cdot t} C_{dt}\Big(E\Big[H(\widehat{\mu}_t) - H(\widehat{\mu}_{t+1}) | \widehat{\mathcal{F}}_t \Big] \Big) \right]$$

Given feasible strategy $(\langle \mu_t \rangle, \tau)$, such $(\langle \widehat{\mu}_t \rangle, \widehat{\tau})$ can be constructed by simply discretizing the continuous-time strategy. Given feasible strategy $(\langle \widehat{\mu}_t \rangle, \widehat{\tau})$, such $(\langle \mu_t \rangle, \tau)$ can be constructed by the Kolmogorov extension theorem. Therefore, it follows that $V_{dt}(\mu) = \sup_{\langle \mu_t \rangle \in \mathbb{M}, \tau} W_{dt}(\mu_t, \tau)$. Now that both V and V_{dt} are characterized using W_{dt}, W_{dt} can be used as an intermediary to link V and V_{dt}:

$$\begin{cases} V(\mu) = \sup\limits_{\langle \mu_t \rangle, \tau} \lim\limits_{dt\to 0} W_{dt}(\mu_t, \tau) \\ \lim\limits_{dt\to 0} V_{dt}(\mu) = \lim\limits_{dt\to 0} \sup\limits_{\langle \mu_t \rangle, \tau} W_{dt}(\mu_t, \tau) \end{cases}$$

Clearly, V and $\lim V_{dt}$ are obtained by taking the limit of W_{dt} in different orders. Therefore, V_{dt} approximates V when the two limits are interchangeable, which is indeed true as proved in Lemma 1:

Lemma 1. *Given Assumption 1, $\forall \mu \in \Delta(X)$, $\lim_{dt\to 0} V_{dt}(\mu) = V(\mu)$.*

5.2 *Discrete-time Bellman equation*

Equation (5) is a discrete-time sequential optimization problem with bounded payoffs and exponential discounting. Therefore, standard dynamic programming theory applies and provides the Bellman equation that characterizes V_{dt}.

Lemma 2 (Discrete-time Bellman)**.** *V_{dt} is the unique solution in $C(\Delta X)$ of the following functional equation:*

$$V_{dt}(\mu) = \max\left\{ F(\mu), \max_{p_i, \nu_i} e^{-\rho dt} \sum_{i=1}^{N} p_i V_{dt}(\nu_i) - C_{dt}\Big(H(\mu) - \sum p_i H(\nu_i) \Big) \right\} \tag{6}$$
$$s.t. \sum p_i \nu_i = \mu$$

where $N = 2|X|$, $p \in \Delta(N)$, $\nu_i \in \Delta(X)$.

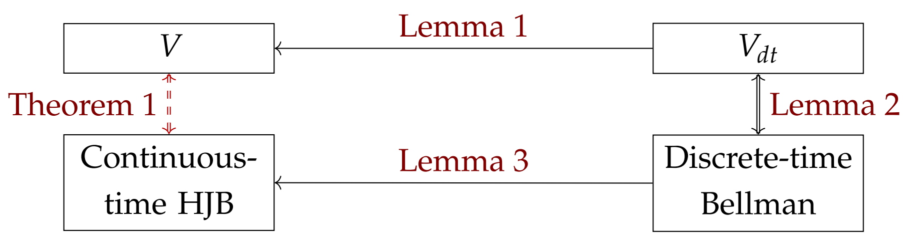
[13] The standard definition for integrability also requires the limit to exist uniformly for all alternative nonuniform discretizations of the time horizon and all alternative measurable stopping times. Here I use the weaker integrability requirement for notational simplicity. The optimal strategy actually satisfies the stronger integrability requirements, so the current definition can be used without loss. The discretization of $\langle I_t \rangle$ is WLOG given the uniform convergence in the definition of $\mathscr{D}(H)$.

[14] The detailed discussion of why restricting belief to be integrable is innocuous is in Remark B.2.

Equation (6) is a standard Bellman equation, except that it covers a restricted space of strategies. The choice of signal structure is restricted to have support size no larger than $2|X|$, while the original space contains signal structures with an arbitrary number of realizations. This simplification is based the generalized concavification methodology developed in Theorem 2 of Zhong (2018a). The standard concavification methodology is an application of the Carathéodory theorem to the graph of the objective function in the belief space.[15] Equation (6) involves an additional term $C_{dt}(H(\mu) - \sum p_i H(\nu_i))$, which makes the standard method inapplicable. The general method suggests that the maximum is characterized by concavifying a linear combination of V_{dt} and H.

5.3 *Convergence and verification theorem*

The following figure illustrates the roadmap for proving Theorem 1.



Theorem 1 is represented by the red dashed arrow on the left. The discrete-time problem's value function V_{dt} is the solution of the Bellman equation Equation (6) (the double arrow on the right, proved in Lemma 2). I have shown that V_{dt} converges to the continuous-time optimal control value V (the arrow on the top, proved in Lemma 1). In the next lemma, I show that solution of HJB Equation (4) is the limit of solution of Equation (6) (the arrow on the bottom, to be proved in Lemma 3). Therefore, the function solving HJB Equation (4) is the value function of the continuous-time stochastic control problem Equation (1).

Lemma 3. *Assume H is strictly concave and $C^{(2)}$ on interior beliefs, Assumption 2 is satisfied. Suppose $V(\mu) \in C^{(1)}$ is a solution to Equation (4). Then $V_{dt} \xrightarrow[dt \to 0]{L_\infty} V$.*

Lemma 3 proves that whenever Equation (4) has a solution, the solution is unique and coincides with the limit of solution to discrete-time problem Equation (6). Verification theorem Theorem 1 is a direct corollary of Lemmas 1, 2 and 3.

6 Optimal information acquisition

In this section I prove the existence of the solution to the continuous-time HJB Equation (4) and fully characterize the value and policy functions, assuming binary states and two forms of flow cost function: a hard cap and a smooth convex function. In both cases, the optimal strategies share the same set of qualitative properties. Then in Section 6.2, I discuss the key trade-offs in the optimization problem and provide the intuition for the optimal strategy. First, I introduce the assumptions for tractability:

[15] See Aumann, Maschler, and Stearns (1995) and Kamenica and Gentzkow (2011))

Assumption 3.

1. *(Binary states):* $|X| = 2$.
2. *(Positive payoff):* $\forall \mu \in [0,1],\ F(\mu) > 0$.
3. *(Uncertainty measure):* $H''(\mu) < 0$ *and locally Lipschitz on* $(0,1)$, $\lim_{\mu\to 0,1} |H'(\mu)| = \infty$.

Assumption 3 comprises three parts. First, I restrict the state space to be binary. Therefore, the belief space is one dimensional, and I can use ODE theory to construct a candidate solution. Although the existence of the solution technically relies on the binary state assumption, the characterization generalizes to general state spaces, as discussed in Appendix A.3. Second, I assume that the utility from decision making is strictly positive so that "delay forever" is strictly suboptimal. This restriction is made without loss of generality in the sense that we can always add a dummy "outside action" that gives ε payoff. Third, I assume that H is sufficiently smooth, strictly convex (which rules out free information) and satisfies an Inada condition (which guarantees a non-degenerate stopping region).

6.1 Main characterization theorem

Theorem 1 states that to characterize $V(\mu)$, it is sufficient to find a smooth solution to HJB Equation (4). I prove the existence of a solution and characterize the optimal strategy under Assumption 2-a or Assumption 2-b, two slightly stronger variants of Assumption 2.

Assumption 2-a (Capacity constraint)**.** *There exists* c *s.t.* $C(I) = \begin{cases} 0 & \text{when } I \leqslant c \\ +\infty & \text{when } I > c \end{cases}$

Assumption 2-a restricts the cost function C to be a hard cap: information is free when its measure is below capacity c and infinitely costly when it exceeds this capacity.[16] This condition forces the DM to smooth the information acquisition process over time.

Theorem 2. *Given Assumptions 1, 2-a and 3, there exists a quasi-convex value function* $V \in C^{(1)}(0,1)$ *solving Equation (4). Let* $E = \{\mu \in [0,1] | V(\mu) > F(\mu)\}$ *be the experimentation region. There exists policy function* $\nu : E \to [0,1]$ *satisfying:*

$$\rho V(\mu) = -c\frac{F(\nu(\mu)) - V(\mu) - V'(\mu)(\nu(\mu) - \mu)}{H(\nu(\mu)) - H(\mu) - H'(\mu)(\nu(\mu) - \mu)}$$

where $\nu(\mu)$ *is unique a.e. and satisfies the following properties.* $\exists \mu^* \in \arg\min V$ *s.t.*

1. *Poisson learning:* $\rho V(\mu) > -c\frac{V''(\mu)}{H''(\mu)}$ $\forall \mu \in E \backslash \mu^*$.
2. *Direction:* $\mu > \mu^* \implies \nu(\mu) > \mu$ *and* $\mu < \mu^* \implies \nu(\mu) < \mu$.
3. *Precision:* $|\nu(\mu) - \mu^*|$ *is decreasing in* $|\mu - \mu^*|$ *on each interval of* E.
4. *Stopping time:* $\nu(\mu) \in E^C$ *(a successful experiment lands in the stopping region).*

[16] $\lim_{I\to\infty} C'(I)$ is not well defined with Assumption 2-a. However, it is not hard to see that Assumption 2-a still satisfies the weaker formulation discussed in Footnote 7. As a result, Theorem 1 applies with Assumption 2-a.

Theorem 2 proves the existence of a solution to Equation (4) and characterizes the optimal policy function. The theorem first states that the optimal value function is implemented by a *Poisson signal*, i.e., seeking a breakthrough that causes the belief to jump to $\nu(\mu)$. Moreover, property 1 states that the Gaussian signal is strictly dominated, except for at most one critical belief. Therefore, as discussed in Section 4, the optimal strategy is Poisson learning, which can be characterized by three aspects of learning and the stopping time.

Direction: Property 2 states that the optimal direction is *confirmatory*: when $\mu > \mu^*$, the DM holds a high prior belief for state 1 and acquires a signal whose arrival induces an even higher posterior belief $\nu(\mu)$ and vice versa for $\mu < \mu^*$.

Precision: Property 3 states that the optimal precision measured by $|\nu(\mu) - \mu^*|$ is *negatively related* to how certain the belief is (measured by $|\mu - \mu^*|$). Since $\mu^* \in \arg\max V$, the property equivalently states that precision is negatively related to the continuation value.

Frequency: With Assumption 2-a, frequency is automatically determined given the precision, according to $p(\mu) = -\frac{c}{H(\nu(\mu)) - H(\mu) - H'(\mu)(\nu(\mu)-\mu)}$. Thus, the optimal frequency is *positively related* to the continuation value.

Stopping time: Property 4 states that the image of ν is always in the stopping region. In other words, the optimal stopping time is exactly the signal arrival time.

By combinging these properties, we can qualitatively determine the optimal learning dynamics. The DM seeks a signal that arrives according to a Poisson process. The arrival of the signal confirms the DM's prior belief and is sufficiently accurate to warrant an immediate action. Absent the arrival of a Poisson signal, the DM becomes less certain about the state, following Bayes' rule. The DM's continuation value decreases correspondingly; hence, she continues seeking a Poisson signal with higher frequency and lower precision.

Assumption 2-b (Convex cost)**.** $C \in C^{(2)}\mathbb{R}^+$, $C(0) = 0$, $C'(I) \geqslant 0$, $C''(I) > 0$, $\lim_{I\to\infty} C'(I) = \infty$.

Assumption 2-b restricts the cost function C to be $C^{(2)}$ smooth and strictly convex: acquiring an additional unit of information is of strictly increasing marginal cost. The condition on $\lim C'(I)$ in Assumption 2 is retained. If we replace Assumption 2 with Assumption 2-b, we obtain the following characterization theorem:

Theorem 3. *Given Assumptions 1, 2-b and 3, there exists a quasi-convex value function* $V \in C^{(1)}(0,1)$ *solving Equation (4). Let* $E = \{\mu \in [0,1] | V(\mu) > F(\mu)\}$ *be the experimentation region. There* $\exists$ *policy functions* $\nu : E \to [0,1]$ *and* $I \in C^{(1)}(E)$[17] *satisfying:*

$$\rho V(\mu) = -I(\mu) \cdot \frac{F(\nu(\mu)) - V(\mu) - V'(\mu)(\nu(\mu) - \mu)}{H(\nu(\mu)) - H(\mu) - H'(\mu)(\nu(\mu) - \mu)} - C(I(\mu))$$

where ν *and* I *are unique a.e. and satisfy the following properties.* $\exists \mu^* \in \arg\min V$ *s.t.*

1. *Poisson learning:* $\rho V(\mu) > \max_{\sigma} \frac{1}{2}\sigma^2 V''(\mu) - C(-\frac{1}{2}\sigma^2 H''(\mu))\ \forall \mu \in E \backslash \mu^*$.

[17] Note that given ν, selecting I or p is equivalent. They uniquely pin down each other according to equation $I(\mu) = p(\mu)(-H(\nu(\mu)) + H(\mu) + H'(\mu)(\nu(\mu) - \mu))$.

2. *Direction:* $\mu > \mu^* \implies \nu(\mu) > \mu$ *and* $\mu < \mu^* \implies \nu(\mu) < \mu$.
3. *Precision:* $|\nu(\mu) - \mu^*|$ *is decreasing in* $|\mu - \mu^*|$ *on each interval of* E.
4. *Stopping time:* $\nu(\mu) \in E^C$.
5. *Intensity:* $I(\mu)$ *is increasing in* $V(\mu)$.

With the exception of property 5, the optimal strategy has the same set of properties as Theorem 2. Property 5 states that the informativeness measure I of the optimal signal is higher when the continuation value is higher. Since the belief process drifts downward the value function conditional on continuation, the DM invests less in information acquisition as time passes.

The intuition for property 5 is discussed in Moscarini and Smith (2001). The marginal gain from experimentation is proportional to the continuation value while marginal cost is increasing in I. Therefore, the optimal cost is increasing in the value function. This property is called "value-level monotonicity" in Moscarini and Smith (2001), where the level (flow variance of the diffusion process) is a parameter for both the cost and precision of a Gaussian signal. My analysis identifies this intuition separately from another important trade-off between signal precision and frequency. I refer to property 5 as "value-intensity monotonicity" in this paper. Here I rename parameter I the *intensity* of learning, which is more intuitive and concise than "informativeness measure".

Examples

In this section, I first provide a minimal working example that illustrates Theorem 3 in Example 2. Then I provide supplementary examples to illustrate a rich set of implications of my model, including multiple phases of learning in Example 3 and learning from a one-sided search in Example 4.

Example 2. Consider the problem studied in Example 1. $F(\mu) = \max\{2\mu - 1, 1 - 2\mu\}$, $H(\mu) = -\mu \log(\mu) - (1 - \mu) \log(1 - \mu)$, $\rho = 1$, and $C(I) = \frac{1}{2}I^2$. No parametric assumption is placed on the set of admissible belief process.

The solution is presented in Figures 5 and 6. In Figure 5-(a), dashed lines depict $F(\mu)$, the blue curve depicts $V(\mu)$, and the blue shaded region is experimentation region E. Figure 5-(b) shows the optimal posterior $\nu(\mu)$ as a function of the prior. As stated in Theorem 3, the policy function is piecewise smooth and decreasing. The three arrows in Figure 5-(a) depict the optimal strategies prescribed at three different priors. The arrows start at the priors and point to the optimal posteriors. The blue curve in Figure 5-(c) shows the optimal intensity $I(\mu)$ as a function of the prior. Clearly, $I(\mu)$ is isomorphic to $V(\mu)$ in the experimentation region.

Figure 6 illustrates the dynamics of the optimal policy. Figure 6-(a) depicts the optimal belief process. Conditional on no signal arrival, the posterior belief drifts towards the critical belief level $\mu^* = 0.5$. In this example, two *phases* of learning occur (represented by different colors of shaded regions in Figure 6-(a)). In the first phase (blue region), the DM seeks a Poisson signal to confirm the most likely state. As time passes, the signal precision

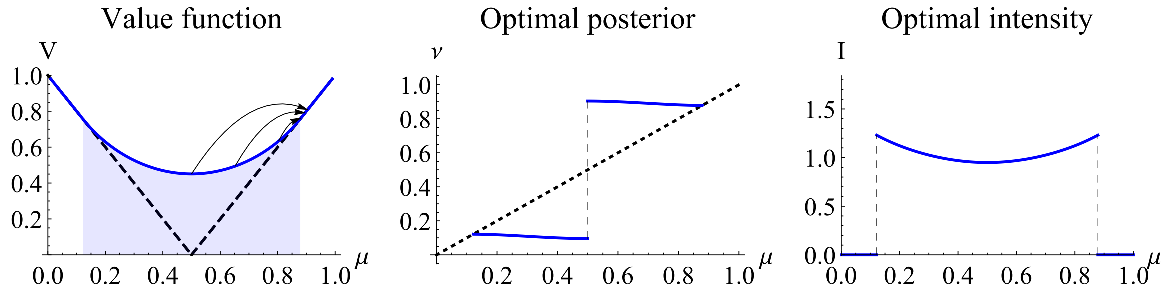


Figure 5: Value and policy functions

increases while signal frequency and learning intensity decreases (as in Figure 6-(b)&(c)). Eventually, the DM believes that the two states are equally likely and switches to the second phase (gray region). In the second phase, she seeks two signals that confirm each state in a balanced way such that before any signal arrives her posterior belief is stationary.

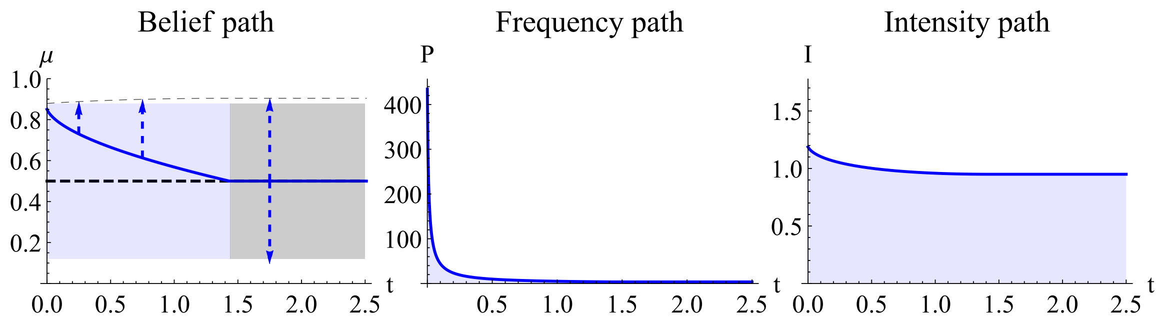


Figure 6: Dynamics of optimal policy

Recall the three learning technologies in Example 1. They approximate the full solution in Example 2. In general, the optimal signal is a confirmatory Poisson signal with varying precision and frequency. However, in Example 1, the precision and frequency of the confirmatory Poisson signal are exogenously fixed. Therefore, for very certain prior beliefs, the ideal high-frequency Poisson signal is approximated by a Gaussian signal. For very uncertain prior beliefs, the ideal signal is approximated by acquiring perfectly revealing breakthroughs with low frequency.

Example 3 (Multiple phases)**.** Figure 7 depicts an example with four actions, whose expected payoffs are represented by the four dashed lines in Figure 7-(a). The two blue dashed lines are called riskier actions, and the two red dashed lines are called safer actions. The upper envelope of the four lines is $F(\mu)$. The experimentation region contains three disjoint intervals. For the middle interval, in the red regions, the DM has a more extreme belief and searches for a signal that confirms a safer action (red arrow). In the blue region, the DM has a more ambiguous belief and searches for a riskier action (blue arrow). Figure 7-(c) depicts the optimal belief process with a prior belief in the red region. The experimentation follows three phases, the DM searches for a safer action in phase 1, searches for a riskier action in phase 2 and searches in a balanced way in phase 3.

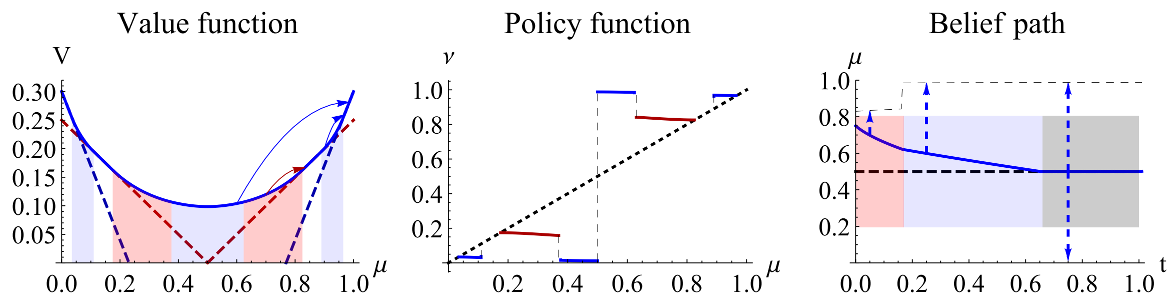


Figure 7: Example with four alternatives

Example 4 (One-sided search)**.** Figure 8 depicts an example where the optimal strategy includes only one-sided search. A safe action with deterministic payoff and a risky action whose payoff is higher than that of the safe action in state 1 exists. As illustrated in Figure 8-(a), both $F(\mu)$ and $V(\mu)$ are monotonically increasing. According to property 1, $\mu > \mu^*$ in the entire experimentation region E. Figure 8-(b) shows that the optimal strategy is always to search for a Poisson signal that induces a posterior belief higher than the prior. Figure 8-(c) shows that in this example, only one phase occurs. If no signal arrives before the belief reaches to the critical belief, the optimal solution is for the DM to stop learning and choose the safe action.

This example illustrates more precisely the definition of confirmatory evidence: the optimal belief jump is in the direction of a more *profitable* state. The profitability of a state depends jointly on its likelihood and the corresponding payoff of the actions. In this example, consider a prior belief less than 0.5. Although state 0 is more likely, since it is dominated by state 1 for any action, state 1 is unambiguously more profitable to learn about. Therefore, the optimal confirmatory evidence is always revealing state 1.

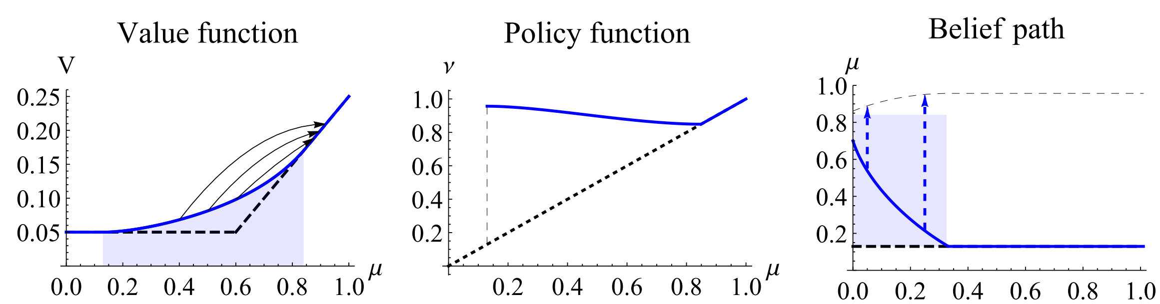


Figure 8: Example with one-sided search

6.2 *Proof methodology and key intuitions*

In Section 3, I introduce four types of trade-offs. Now, I discuss the trade-offs in detail and illustrate how they determine the optimal strategy in each salient aspect. I first derive a geometric characterization of the optimal policy in Section 6.2.1. Then, I discuss how the key trade-offs are represented by the geometric characterization and provide intuitions for the optimal policy. In Section 6.2.2, I present the sketch of a proof for Theorem 2.

6.2.1 *Geometric representation and key trade-offs*

A though experiment is useful to gain intuition. Fix the value function V and consider a simplified optimization problem:

$$\sup_{p\geqslant 0,\nu} p\big(V(\nu)-V(\mu)-V'(\mu)(\nu-\mu)\big)-C\big(p\big(H(\mu)-H(\nu)+H'(\mu)(\nu-\mu)\big)\big) \tag{7}$$

Equation (7) is more restrictive than Equation (4). I assume that the DM acquires only a Poisson signal. Let us temporarily ignore the Gaussian signal. Define:

$$\begin{cases} U(\mu,\nu)=V(\nu)-V(\mu)-V'(\mu)(\nu-\mu) \\ J(\mu,\nu)=H(\mu)-H(\nu)+H'(\mu)(\nu-\mu) \end{cases}$$

The interpretation of $U(\mu,\nu)$ is the flow value per unit arrival rate from a Poisson signal with posterior ν. Similarly, $J(\mu,\nu)$ is the flow uncertainty reduction per unit arrival rate from the Poisson signal. Then Equation (7) can be rewritten as:

$$\begin{aligned} &\sup_{p\geqslant 0,\nu} p\cdot U(\mu,\nu)-C(p\cdot J(\mu,\nu)) \\ \xLeftrightarrow{I\triangleq p\cdot J(\mu,\nu)}\ &\sup_{I\geqslant 0,\nu}\left(\frac{U(\mu,\nu)}{J(\mu,\nu)}\right)\cdot I-C(I) \end{aligned}$$

The problem is separable in choosing I and ν. The solution (ν^*,I^*) is characterized by:

$$\begin{cases} \nu^*\in\arg\max_{\nu}\frac{U(\mu,\nu)}{J(\mu,\nu)} \\ C'(I^*)=\max_{\nu}\frac{U(\mu,\nu)}{J(\mu,\nu)} \end{cases}$$

The optimal posterior ν^* maximizes $\frac{U(\mu,\nu)}{J(\mu,\nu)}$—the value to uncertainty reduction ratio. Let $\lambda=C'(I^*)=\max\frac{U(\mu,\nu)}{J(\mu,\nu)}$; then, $U(\mu,\nu)\leqslant\lambda J(\mu,\nu)$ and the equality holds at ν^*.[18] Define $G(\mu)=V(\mu)+\lambda H(\mu)$. I call $G(\mu)$ the *gross value function*. Then, the definition of U and V implies $U(\mu,\nu)-\lambda J(\mu,\nu)=G(\nu)-G(\mu)-G'(\mu)(\nu-\mu)$. Hence, $U(\mu,\nu)\leqslant\lambda J(\mu,\nu)$ implies that the gross value function has the following property:

$$\begin{cases} G(\nu)\leqslant G(\mu)+G'(\mu)(\nu-\mu) & \forall\nu\in[0,1] \\ G(\nu^*)=G(\mu)+G'(\mu)(\nu^*-\mu) \end{cases} \tag{8}$$

Equation (8) states that $G(\nu)$ is everywhere (weakly) below the tangent line of G at μ, except $G(\mu)$ and $G(\nu^*)$ touch the tangent line. The tangent line is linear (hence concave) and thus weakly dominates G's upper concave hull $\text{co}(G)$. Therefore, $G(\mu)=\text{co}(G)(\mu)$ and $G(\nu^*)=\text{co}(G)(\nu^*)$. See Figure 9 for a graphical illustration.

[18] With Assumption 2-a, $I^*=c$ and $\lambda=\max\frac{U(\mu,\nu)}{J(\mu,\nu)}$ is the Lagrangian multiplier for constraint $I\leqslant c$.

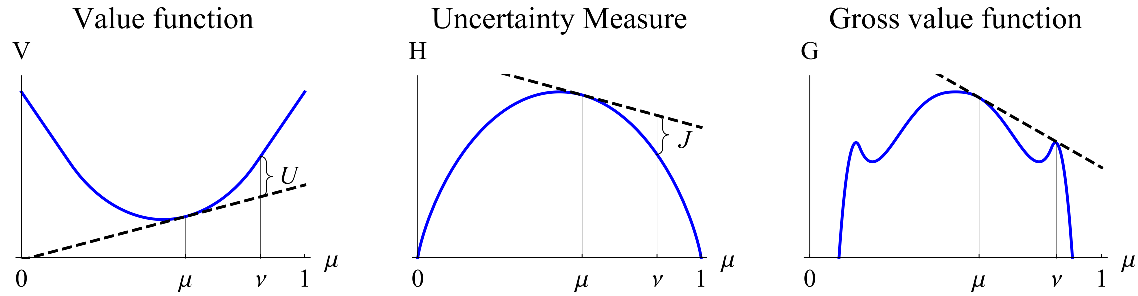


Figure 9: Concavification of the gross value function

Figure 9-(a) and Figure 9-(b) depict the value function V and the uncertainty measure H, respectively. Figure 9-(c) depicts the gross value function $G = V + \lambda H$, where λ is calculated for the prior μ. As discussed, G touches the upper concave hull at both μ and ν^*. When ν^* is unique, μ and ν^* are the two boundary points of the *concavified region* (the interval (μ, ν) on which $G < \text{co}(G)$).

Equation (8) is called a *concavification* characterization as it is an analog to the concavification method in Bayesian persuasion problems. The difference is that in a Bayesian persuasion problem, the boundary points of a concavified region are optimal posteriors, whereas in the current problem, the prior is also on the boundary of a concavified region. This property has clear economic meaning. G is called the gross value function because it integrates value function V and uncertainty measure H using marginal cost level λ. λ is a multiplier that captures the marginal effect of reducing uncertainty on flow cost. Therefore, solving:

$$\sup_{p \geqslant 0, \nu} p(G(\nu) - G(\mu) - G'(\nu)(\nu - \mu)) \tag{9}$$

is equivalent to solving Equation (7). Whether Equation (9) yields a positive payoff depends on whether $G(\mu) < \text{co}(G)(\mu)$. Suppose $G(\mu) < \text{co}(G)(\mu)$. Then, there is a strictly positive gain from information and Equation (9) is strictly positive. However, Equation (9) is linear in the signal arrival rate p. As a result the DM has incentive to increase p, which drives up marginal cost $C'(\cdot)$. Thus, when the optimum is reached, $C'(\cdot)$ (or λ) must be such that solving Equation (9) yields exactly zero utility: $G(\mu) = \text{co}(G)(\mu)$. This characterization illustrates that in the continuous time limit, information is smoothed such that uncertainty is reduced by only an infinitesimal amount at every instant of time.

Now, suppose that the HJB is satisfied, i.e., Equation (7) equals the flow discounting loss $\rho V(\mu)$. Then applying $I^* = p^* \cdot J(\mu, \nu^*)$ and $C'(I^*) = \frac{U(\mu,\nu^*)}{J(\mu,\nu^*)}$ to the HJB implies:

$$\begin{aligned} \rho V(\mu) &= p^* \cdot U(\mu, \nu^*) - C(p^* \cdot J(\mu, \nu^*)) \\ \implies \rho V(\mu) &= I^* C'(I^*) - C(I^*) \end{aligned} \tag{10}$$

Combining Equation (8) and Equation (10) dentifies the value function V and corresponding strategies p, ν.[19] Now, I analyze key trade-offs in the dynamic information acquisition problem by studying Equations (8) and (10).

[19] With Assumption 2-a, $C(I^*) = 0$ and $I^* = c$. Thereofre, $\rho V(\mu) = \lambda c$.

1. Utility gain vs. information cost

Equation (10) illustrates the utility gain vs. information cost trade-off. Since C is a convex function, $IC'(I) - C(I)$ is increasing in I[20], that is, the optimal flow informativeness measure I is isomorphic in continuation value $V(\mu)$. This property is exactly the "value-intensity monotonicity" I introduced in Section 6.1.

The intuition for this property is simple. The marginal cost of increasing the intensity of the signal *proportionately* is $IC'(I)$. The marginal gain is obtained from increasing the arrival rate proportionately (keeping the signal precision fixed, as in the envelope theorem). Increasing the arrival rate by a unit proportion reduces the waiting time by the same proportion, so the marginal gain from increasing I by a unit proportion is discount ρV plus cost $C(I)$. At the optimum, the maginal cost equals the margina gain; therefore, we obtain Equation (10) and the flow informativeness is monotonic in value function.

If we consider the case with Assumption 2-a, then λ in Equation (8) is replaced by the shadow cost of increasing informativeness (see Footnotes 18 and 19). Equation (10) can be written as $\rho V(\mu) = c\lambda$. Although the intensity is fixed, in this case, a monotonicity between the shadow cost and value function remains.

In summary, by studying the utility gain vs. information cost trade-off, I established a monotonicity between the shadow/marginal cost λ and the continuation value $V(\mu)$. (I refer to both as the "value-intensity monotonicity" for notational simplicity.) Now that I characterized λ, we can proceed to Equation (8).

2. Precision vs. frequency

A novel trade-off characterized by Equation (8) is the precision vs. frequency trade-off. The value-intensity monotonicity determines I from the value function. Now, the DM allocates total intensity I to precision (parametrized by the size of belief jumps) and frequency (parametrized by the arrival rate of jumps). Equation (8) suggests that the optimal signal precision can be solved by concavifying the gross value function $G(\mu)$. In this section, I illustrate how this trade-off changes for different priors and explain the intuition.

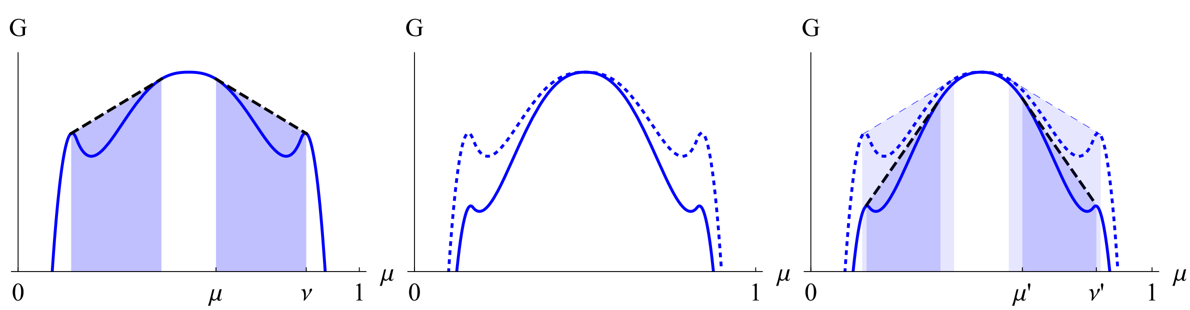


Figure 10: Precision-frequency trade-off

Figure 10 shows how varying λ affects the optimal jump size. In Figure 10-(a) the blue curve is $G(\mu)$, and the dashed curve is $\text{co}(G)$. I call the blue region, where $G(\mu) < \text{co}(G)(\mu)$, the *concavified region* and the white region, where $G(\mu) = \text{co}(G)(\mu)$, the *globally concave*

[20] $\frac{\mathrm{d}}{\mathrm{d}I}(IC'(I) - C(I)) = IC''(I) \geqslant 0$

region. The prior μ and optimal posterior ν are on the boundary of a concavified region. Consider $G_1 = V + \lambda_1 H$, where $\lambda_1 > \lambda$. Figure 10-(b) depicts both G (the dashed curve) and G_1 (the blue curve). Since G_1 is G plus a strictly concave function, any belief in the globally concave region of G is still in the globally concave region of G_1. As a result, as λ increases, the white region is expands and the blue region contracts (see Figure 10-(c)). Thus, the prior and optimal posterior move closer together. Recall that λ is monotonic in V, which means the DM is more willing to choose a signal that induces shorter belief jump when the continuation value is higher.

The intuition for this property is as follows. When the DM is more certain about the state, the continuation value is higher; hence, the utility loss from discounting is higher. The DM wants to receive a signal more frequently to benefit from the high value sooner. In other words, the marginal rate of substitution of frequency for precision is increasing in the continuation value. In this analysis, the continuation value is isomorphic to λ, which controls the shape of G. The marginal rate of substitution of frequency for precision is exactly captured by the global concavity of the gross value function; thus, the analysis presented by Figure 10 exactly illustrates the intuition.

Confirming vs. contradicting: The analysis above determines the magnitude of the optimal belief jump. The optimal jump direction remains to be determined to pin down the optimal posterior. Now, I show that the precision-frequency trade-off also implies the optimality of confirmatory learning.

Let us hypothetically consider a belief μ at which jumping toward the right is optimal (weakly). In both panels of Figure 11, μ is the prior and ν_L, ν_R are optimal posteriors on each side of μ. Jumping to ν_R (the black arrow) is better than jumping to ν_L (the dashed black arrow). Let V be increasing around μ. Now consider the DM's incentive at μ_1 slightly larger than μ (in Figure 11-(a)). Although the corresponding optimal posteriors could also move, keeping them fixed at ν_L and ν_R has only a second-order effect on utility. We can compare ν_L and ν_R to pin down the optimal posterior for μ_1. Since $\mu_1 > \mu$, ν_R is closer to prior, and ν_L is farther from prior. Moreover, $V(\mu_1) > V(\mu)$ implies that the DM has a stronger preference for frequency to precision with belief μ_1. Since $V' > 0$, the effect is first order. Therefore, ν_R is strictly preferred to ν_L at μ_1. Consider μ_2 slightly smaller than μ (in Figure 11-(b)). A similar analysis shows that now size of jump to ν_R is larger, and the DM has a stronger preference for precision with belief μ_2. Thus, ν_R is also strictly optimal for μ_2.

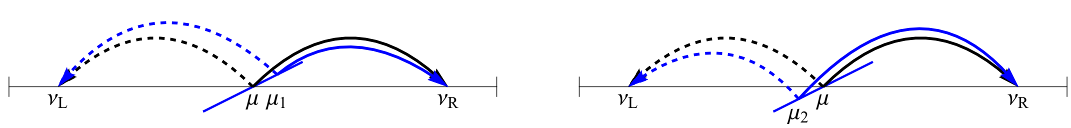


Figure 11: Confirmatory v.s. contradictory

In this analysis, jumping in the direction of increasing value function means the signal is confirmatory. When value function is quasi-convex, this property is equivalent to property 2 of Theorems 2 and 3. Therefore, the precision-frequency trade-off implies that the incentive for confirmatory learning is self-enforcing.

3. Poisson vs. Gaussian

Thus far, I have ignored the possibility of Gaussian signals. In fact, Gaussian signals are implicitly modeled in Equation (8). Consider the optimization w.r.t. Gaussian signals:

$$\begin{aligned} &\sup_{\sigma} \sigma^2 V''(\mu) - C(-\sigma^2 H''(\mu)) \\ \Longrightarrow\ &\text{FOC}: V''(\mu) + \lambda H''(\mu) = 0 \\ \Longleftrightarrow\ &G''(\mu) = 0 \end{aligned} \tag{11}$$

where $\lambda = C'(-\sigma^2 H''(\mu))$ with Assumption 2-a or $\lambda = \frac{\rho}{c} V(\mu)$ with Assumption 2-b. Comparison of Equations (8) and (11) shows that Equation (11) is exactly the limit of Equation (8) when optimal posterior ν converges to prior μ. This result is intuitive since a Gaussian signal can be approximated as a Poisson signal with very low precision and high arrival rate.

The comparison of Gaussian and Poisson signals is effectively the comparison of a special imprecise Poisson signal and other Poisson signals. Therefore, this trade-off is a special case of the precision-frequency trade-off. Selecting a Gaussian signal is a corner solution when the DM wants to sacrifice almost all precision for frequency—a slightly less patient DM is willing to avoid any waiting and stop immediately, while a slightly more patient DM is willing to wait for a more precise Poisson signal. Therefore, the Gaussian signal is optimal only on the boundaries of the experimentation regions. Given this intuition, one could imagine that the Gaussian signal is generically suboptimal except for special cases where the precision-frequency trade-off is invariant. Since the preference between precision and frequency depends on the loss from delaying, the trade-off is invariant only when the DM does not discount future payoffs. This intuition is confirmed in a no-discounting special case in Section 7.1, as well as in the model of Hébert and Woodford (2016).

4. Continuing vs. stopping

Consider the optimal stopping time. Theorems 2 and 3 states that repeated jumps are suboptimal. I prove by showing that repeated jumps can be improved by a direct jump. Let ν be the optimal posterior for prior μ (see Figure 12). Then, Equation (8) implies that $\frac{U_0}{J_0} = \frac{U_0'}{J_0'} = \lambda(\mu)$.

Hypothetically, imagine that at ν, it is optimal to continue, and the optimal posterior is ν'. Then, $\frac{U_1}{J_1} = \lambda(\nu)$, and $\lambda(\nu) > \lambda(\mu)$ by the confirmatory evidence property and value-intensity monotonicity. I want to show that this result implies $\frac{U(\mu,\nu')}{J(\mu,\nu')} = \frac{U_1+U_1'}{J_1+J_1'} > \lambda(\mu)$, i.e., jumping to posterior ν' directly is strictly better than a two-step jump. By elementary geometry, there exists α s.t $U_1' = \alpha U_0$ and $J_1' = \alpha J_0$.[21] Therefore, the value to uncertain reduction ratio $\frac{U(\mu,\nu')}{J(\mu,\nu')} = \frac{U_1+\alpha U_0}{J_1+\alpha J_0}$ is a weighted average of $\frac{U_0}{J_0}$ and $\frac{U_1}{J_1}$, which is larger than $\lambda(\mu)$.

[21]See Figure 12. $\frac{U_0}{J_0} = \frac{U_0'}{J_0'} = \lambda(\mu)$ implies $\frac{U_1'}{J_1'} = \lambda(\mu)$, hence, $\frac{U_1'}{U_0} = \frac{J_1'}{J_0}$. I assume the ratio to be α.

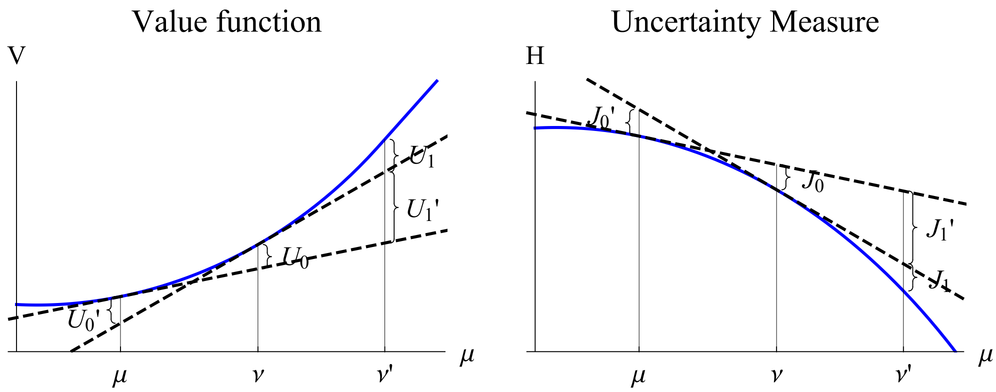


Figure 12: Continuing vs. stopping

The intuition for the stopping rule is now clear. If we combine a two-step jump into a direct jump, the flow utility gain is a weighted sum of that of the two jumps. The flow uncertainty reduction is exactly the same weighted sum of that of the two jumps. Therefore, the net value from a direct jump is a weighted average of the net values from each jump. As a result, sequentially jumping to higher values is dominated by directly jumping to the highest value.

Remark 1.

The intuition behind the value-intensity monotonicity is driven purely by convexity of cost function h and is clearly independent of the formulation of the information measure. The intuition behind the optimality of a Poisson signal over a Gaussian signal is the use of the precision-frequency trade-off to compare a generic Poisson signal with an extremely imprecise Poisson signal. The result does not depend on the exact form of I. I generalize the optimality of a Poisson signal to the generic cost of information in Theorem 5, Section 7.2. I also discuss confirmatory evidence and immediate stopping properties with generic cost functions in Section 7.2.

The precision-frequency trade-off also does not depend on the size of the state space. I confirm this result via a general characterization of optimal strategy with more states (Theorem 9) in Appendix A.3. However, the binary states assumption is crucial for proving the existence of the solution to the HJB equation. A constructive proof of the binary state case based on ODE theory is introduced in Section 6.2.2.

Our discussion thus far does not rely on the exact form of λ. The qualitative properties of all these trade-offs depend only on the monotonicity of λ in continuation value, which is true with both Assumptions 2-a and 2-b. Therefore, when I introduce the sketch of the proof, I discuss only Theorem 2, and the proof extends to Theorem 3.

6.2.2 *Sketched proof of Theorem 2*

I prove Theorem 2 by construction and verification. I conjecture that the optimal policy for Equation (4) takes the form of Theorem 2: a single confirmatory signal associated with an immediate action. I first construct $V(\mu)$ and $\nu(\mu)$ via three steps:

- *Step 1*. Determine μ^*. Since $\mu^* \in \arg\min V$, except for the special case where V is strictly monotonic, μ^* is essentially the unique belief at which $V'(\mu^*) = 0$, and searching for posteriors on either side of μ^* is equally good. The HJB equation implies:

$$\sup_{\nu \leqslant \mu^*} \frac{F(\nu)}{1 + \frac{\rho}{c} J(\mu^*, \nu)} = \sup_{\nu \geqslant \mu^*} \frac{F(\nu)}{1 + \frac{\rho}{c} J(\mu^*, \nu)}$$

$V(\mu^*)$ and $\nu(\mu^*)$ are pinned down correspondingly. The special case occurs when F is strictly monotonic. Take $F' > 0$ for example. μ^* is the smallest belief that $\rho F(\mu) \leqslant \sup_{\nu \geqslant \mu} -c\frac{F(\nu)-F(\mu)-F'(\mu)(\nu-\mu)}{J(\mu,\nu)}$, and vice versa for $F' < 0$.

- *Step 2*. Solve for the value function while holding the action fixed. Let a be the optimal action for optimal posterior ν solved in step 1. Let $F_a(\mu) = E_\mu[u(a,x)]$. Now, solve for the value function given payoff $F_a(\nu)$:

$$\rho V(\mu) = \max_{\nu \geqslant \mu} -c\frac{F_a(\nu) - V(\mu) - V'(\mu)(\nu - \mu)}{H(\nu) - H(\mu) - H'(\mu)(\nu - \mu)}$$

The primitives in the objective function are all sufficiently smooth in ν. Then, the first-order condition w.r.t. ν yields a well-behaved ODE characterizing $\nu(\mu)$ with initial condition $\nu(\mu^*)$. Therefore, we can solve for the optimal policy ν and calculate value $V(\mu)$ accordingly for $\mu \geqslant \mu^*$. $V(\mu)$ and $\nu(\mu)$ for all $\mu \leqslant \mu^*$ are solved by a symmetric process.

- *Step 3*. Update the value function w.r.t. all alternative actions and smoothly paste the solved value function piece by piece. This step begins with solving the ODE defined in step 2 at μ^*. Then, I extend the value function towards $\mu = \{0, 1\}$. Whenever I reach a belief at which two actions yield the same payoff, I setup a new ODE with the new action. This process continues until the calculated value function $V(\mu)$ smoothly pastes to $F(\mu)$. This procedure generates a quasi-convex value function (minimized at μ^*).

Solving the ODE characterizing $\nu(\mu)$ directly implies monotonicity of $\nu(\mu)$ in each connected experimentation region. Now, I need to verify the optimality of the constructed strategy. The verification takes three steps, which rule out repeated jumps, contradictory evidence and Gaussian signals. The intuition for the suboptimality of these three alternative strategies is explained in Section 6.2. The formal proof is relegated to Appendix B.3.

7 Discussion

In this section, I discuss, in detail, the assumptions I make in the baseline model, which can be categorized into three classes.

1. *Economic assumptions:*
   - Discounting (positive ρ).
   - Informativeness measure (Assumption 1).
   - Convexity of cost function (Assumption 2).
2. *Restrictive assumptions:* Finite actions and binary states (Assumption 3).
3. *Technical assumptions:* Smoothness and positiveness assumptions (Assumption 3).

The economic assumptions are crucial for my results and deserve an in-depth discussion. To illustrate the role of discounting, in Section 7.1, I discuss the case with no discounting but a flow waiting cost, and show that without discounting, the trade-off between precision and frequency diminishes and the dynamics of information become irrelevant. In Section 7.2, I generalize Assumption 1 to general information measures and show that a Poisson signal almost always strictly dominates a Gaussian signal. I also explain that immediate action and confirmatory learning properties are tightly tied to Assumption 1. To illustrate the role of Assumption 2, I discuss the case where the cost function is linear in Section 7.3 and show that without convexity, the optimal strategy is static.

The restrictive assumptions do limit the generality of the model. However, relaxing them does not fundamentally alter the key intuition, and the methodology generalizes. The discussion of these assumptions is relegated to the appendix. In Appendix A.2, I relax the finite action assumption and show that the problem with a continuum of actions can be approximated well by adding actions. In Appendix A.3, I relax the binary state assumption. Although the constructive proof of existence no longer works with the general state space, I show that all the properties in Theorem 2 extend. The technical assumptions do not restrict my model in a meaningful way and are therefore not discusses.

7.1 *Linear delay cost*

As is discussed in Section 6.2, discounting is the key factor driving all the dynamics. With exponential discounting, the trade-off between the arrival frequency and precision of signals changes according to the continuation value. A sensible conjecture is that if we replace exponential discounting with linear discounting, i.e., the DM pays a fixed flow cost of delay, the time distribution of the utility gain and information cost no longer matters to the DM. In fact, this conjecture is correct. Consider the following problem:

$$V(\mu) = \sup_{\langle \mu_t \rangle \in \mathbb{M}, \tau} E\left[F(\mu_\tau) - m\tau - \int_0^\tau C(I_t)\mathrm{d}t \right] \tag{12}$$

Theorem 4. *Given Assumptions 1 and 2, suppose $V(\mu)$ solves Equation (12); then:*

$$V(\mu) = \sup_{P \in \Delta^2(X), \lambda > 0} E_P[F(\nu)] - \frac{m + C(\lambda)}{\lambda} E_P[H(\mu) - H(\nu)]$$

Theorem 4 illustrates that solving Equation (12) is equivalent to solving a static rational inattention problem, with $\frac{m+C(\lambda)}{\lambda}$ being the marginal cost on the information measure (see Caplin and Dean (2013) and Matějka and McKay (2014)). The optimal value function can be obtained through various learning strategies. Assuming (P^*, λ^*) to be the solution to the problem in Theorem 4, then *all* dynamic information acquisition strategies that eventually implement P^* (i.e., $\mu_\infty \sim P^*$) and incur flow cost λ^* achieve the same utility level $V(\mu)$.[22]

Note that in Equation (12), the utility depends on the decision time only through expected delay $E[\tau]$. Therefore, the previous analysis implies that all dynamic information

[22] This result is stated and proved formally in Zhong (2018b).

acquisition strategies that eventually implement P^* and incur flow cost λ^* have the same expected delay. This result suggests that the cost structure specified by Assumptions 1 and 2 has the property that all learning strategies are equally *fast* on expectation, but they might differ in terms of *riskiness*. The linear delay cost case is a knife-edge case where the DM is risk neutral on the time dimension and, consequently, all learning strategies are equally good.

When the DM discounts delayed payoffs, as is assumed in the main model, she is risk loving on the time dimension; therefore, the DM prefers a riskier strategy. Intuitively, the riskiest information acquisition strategy is a "greedy strategy" that maximizes the probability of early decision (at the cost of a high probability of long delays as the expected delay is fixed). The confirmatory Poisson learning strategy exactly resembles such a greedy strategy. The key property of the strategy is that all resources are used in verifying the conjectured state directly, and no intermediate step exists before a breakthrough. Alternative strategies, such as Gaussian learning and contradictory Poisson learning all involve the accumulation of substantial intermediate evidence to conclude a success. The intermediate evidence accelerates future learning and hence hedges the risk of decision time. Moreover, the decision time is further dispersed by acquiring signals with decreasing frequency.

Equation (12) is the dynamic learning foundation provided in Hébert and Woodford (2016) to justify Gaussian learning.[23] The analysis of Equation (12) suggests that a linear delay cost is a knife-edge case.

7.2 *General information measure*

Technically, Assumption 1 helps throughout the entire analysis. The methodology of concavifying "the gross value function" is possible only when the expected utility gain and information measure take consistent forms. However, I want to show that one key feature of the baseline model—the optimality of Poisson learning—does not depend on this assumption. Let $J(\mu,\nu)$ and $\kappa(\mu,\sigma)$ be bivariate functions. Consider the following functional equation:

$$\rho V(\mu) = \max\left\{\rho F(\mu), \sup_{p,\nu,\sigma^2} p\big(V(\nu) - V(\mu) - V'(\mu)(\nu-\mu)\big) + \frac{1}{2}\sigma^2 V''(\mu)\right\} \tag{13}$$
$$\text{s.t. } pJ(\mu,\nu) + \kappa(\mu,\sigma) \leqslant c$$

The objective function of Equation (13) is exactly the same as that of Equation (4) with Assumption 2-a. I assume that the DM controls a jump-diffusion belief process. The gain from information is the same as before. I assume $J(\mu,\nu)$ to be an arbitrary function that is both prior and posterior dependent. The cost of the diffusion signal is $\kappa(\mu,\sigma)$. I impose the following assumptions on $J(\mu,\nu)$ and $\kappa(\mu,\sigma)$.

Assumption 4.

[23] In Hébert and Woodford (2016), informativeness measures that are more general than Assumption 1 are also considered in the Appendix.

1. $J \in C^{(4)}(0,1)^2$.
2. $\forall \mu \in (0,1)$, $J(\mu,\mu) = J'_\nu(\mu,\mu) = 0$, *and* $J''_{\nu\nu}(\mu,\mu) > 0$.
3. $\kappa(\mu,\sigma) = \frac{1}{2}\sigma^2 J''_{\nu\nu}(\mu,\mu)$.

First, J is assumed to be sufficiently smooth to eliminate technical difficulties. $J(\mu,\mu) = 0$ is the implication of "an uninformative Poisson signal is free".[24] $J'_\nu(\mu,\mu) = 0$ and $J''_{\nu\nu}(\mu,\mu) > 0$ are implications of "any informative Poisson signal is costly". Within this continuous time framework, these assumptions are imposed on J without loss of generality. The crucial assumption is the third condition: $\kappa(\mu,\sigma) = \frac{1}{2}\sigma^2 J''_{\nu\nu}(\mu,\mu)$. This assumption states that the cost functional is "continuous" in the space of the signal structures. Consider a Poisson signal (p,ν). When $\nu \to \mu$, the utility gain from learning this signal is:

$$p\big(V(\nu) - V(\mu) - V'(\mu)(\nu-\mu)\big) = p\left(\frac{1}{2}V''(\mu)(\nu-\mu)^2 + O|\nu-\mu|^3\right)$$

Therefore, (p,ν) approximates a Gaussian signal with flow variance $p(\nu-\mu)^2$. Meanwhile, the cost of this signal is:

$$\begin{aligned} pJ(\mu,\nu) =& p\left(J(\mu,\mu) + J'_\nu(\mu,\mu)(\nu-\mu) + \frac{1}{2}J''_{\nu\nu}(\mu,\mu)(\nu-\mu)^2 + O(|\nu-\mu|^3)\right) \\ =& \frac{1}{2}p(\nu-\mu)^2 J''_{\nu\nu}(\mu,\mu) + pO(|\nu-\mu|^3) \end{aligned}$$

Hence, if the cost of a Gaussian signal is consistent with the cost of imprecise Poisson signals in the limit, $\kappa(\mu,\sigma) = \frac{1}{2}\sigma^2 J''_{\nu\nu}(\mu,\mu)$.

Theorem 5. *Given Assumption 4, suppose $V \in C^{(3)}(0,1)$ solves Equation (13), and let $L(\mu)$ be defined by:*

$$L(\mu) = \frac{\rho}{c}J''_{\nu\nu}(\mu,\mu)^2 - \frac{2J^{(3)}_{\nu\nu\mu}(\mu,\mu)^2 + J^{(3)}_{\nu\nu\nu}(\mu,\mu)J^{(3)}_{\nu\nu\mu}(\mu,\mu)}{J''_{\nu\nu}(\mu,\mu)} + J^{(4)}_{\nu\nu\nu\mu}(\mu,\mu) + J^{(4)}_{\nu\nu\mu\mu}(\mu,\mu)$$

Then in the open region: $D = \left\{\mu \middle| V(\mu) > F(\mu) \text{ and } L(\mu) \neq 0\right\}$, the set of μ s.t.:

$$\rho V(\mu) = c\frac{V''(\mu)}{J''_{\nu\nu}(\mu,\mu)}$$

is of zero measure.

The interpretation of Theorem 5 is that a Poisson signal is almost always strictly superior to the diffusion signal. In the experimentation region where $L(\mu) \neq 0$, $V(\mu)$ can be achieved by a diffusion signal only at a zero measure of points. $L(\mu) = 0$ is a partial differential equation on $J(\mu,\nu)$ in the diagonal of space. Therefore, the set of points that $L(\mu) = 0$ could contain an interval only when $J(\mu,\nu)$ is a local solution to the PDE. The solution to a

[24]In this setup, $J(\mu,\mu) = 0$ is WLOG. If an uninformative signal has a strictly positive cost, we can always shift the capacity constraint c to normalize $J(\mu,\mu)$ to 0.

specific PDE is a non-generic set in the set of all functions satisfying Assumption 4. In this sense, for an arbitrary information measure $J(\mu, \nu)$, the optimal policy function contains a diffusion signal almost nowhere.

A trivial sufficient condition for $L(\mu) \neq 0$ is Assumption 1. Assumption 1 implies that $J^{(2)}_{\nu\nu}(\mu, \nu)$ is invariant in μ. In this case $L(\mu) = \frac{\rho}{c} J''_{\nu\nu}(\mu, \mu)^2 > 0$ for certain. The first corollary of Theorem 5 characterizes D when J is almost locally posterior separable. $\forall f \in C^{(1)}(0,1)^2$, define a norm $\|f(\cdot)\|_\delta = \sup_{x \in [\delta, 1-\delta]} \{|f(x,x)|, \|\nabla f(x,x)\|_{L^2}\}$.

Corollary 5.1. *Given Assumption 4, suppose $V \in C^{(3)}(0,1)$ solves Equation (13); then, for any $\delta > 0$, there exists ε s.t. if $\left\| J^{(3)}_{\nu\nu\mu} \right\|_\delta \leqslant \varepsilon$, then in the interval $[\delta, 1-\delta]$ the set of μ s.t.:*

$$\rho V(\mu) = c \frac{V''(\mu)}{J''_{\nu\nu}(\mu, \mu)}$$

is of zero measure.

The condition in Corollary 5.1 states that $J''_{\nu\nu}(\mu, \nu)$ is approximately constant over μ for ν close to μ. This result verifies my analysis in Section 6.2.1 that the comparison of Poisson and Gaussian signals relies only on the local properties of J. Another simple sufficient condition for $L(\mu) \neq 0$ is high impatience or low learning capacity.

Corollary 5.2. *Given Assumption 4, suppose $V \in C^{(3)}(0,1)$ solves Equation (13). Then, for any $\delta > 0$, there exists Δ s.t. if $\frac{\rho}{c} \geqslant \Delta$, then in the interval $[\delta, 1-\delta]$, the set of μ s.t.:*

$$\rho V(\mu) = c \frac{V''(\mu)}{J''_{\nu\nu}(\mu, \mu)}$$

is of zero measure.

Corollaries 5.1 and 5.2 complement the discussion in Section 7.1 and illustrate the complete picture of how the DM's incentives pin down the optimal learning dynamics. First, when Assumption 1 holds, Theorem 4 implies that the cost structure does not favor any learning strategy. Any positive discount rate gives the DM incentive to choose a Poisson signal. All learning strategies, including Gaussian learning, become equally optimal only when time preference is risk neutral. Second, when Assumption 1 is violated by a small amount, then even though the cost structure might favor a Gaussian signal, the incentive is dominated by discounting. Third, when the cost structure provides arbitrarily strong incentive for a Gaussian signal, sufficiently high discount rate overweights the incentive.

Although *Poisson learning* is generally optimal, *immediate action* and *confirmatory evidence* are implications of Assumption 1. Imagine a case in which high-precision signals are relatively inexpansive (e.g., $J(\mu, \nu)$ is truncated both below and above). Then, when the prior is close to the boundary of the stopping region, seeking confirmatory evidence (with low precision and high frequency) results in very high cost, whereas seeking for a precise contradictory signal is inexpensive. Searching for a contradictory signal causes the belief to drift rapidly toward the more likely state, which effectively enables quick confirmation. Therefore, the contradictory signal becomes optimal. In fact, this example has the same intuition as the findings in Che and Mierendorff (2016). In their setup, the DM allocates limited attention to two exogenous Poisson signals, each revealing a state. When the DM is more uncertain, their model predicts that the DM acquires a confirmatory signal. However, near the stopping boundary, their model predicts a contradictory signal, as the contradictory signal approximates an infeasible confirmatory signal with low precision and high frequency.

On the other hand, consider the immediate action property. Imagine a case in which low-precision signals are inexpensive. Then, breaking a long jump into multiple short jumps may be profitable. The immediate action property is called the single experiment property (SEP) in Che and Mierendorff (2016). In their paper, SEP is also shown not to be a robust property in a generic Poisson learning model.

7.3 Linear flow cost

In this subsection, I study the case where the flow cost $C(I)$ is a linear function. Assumption 2 is replaced by the following assumption:

Assumption 2′ (Linear flow cost)**.** *Function h is defined by* $C(I) = \lambda I$, $\lambda > 0$.

The convexity of $C(I)$ in Assumption 2 gives the DM incentive to smooth the acquisition of information. When $C(I)$ is a linear function, the optimal value is achieved by acquiring all the information and immediately making a decision.

Theorem 6. *Given Assumptions 1 and 2′, suppose* $V(\mu)$ *solves Equation (1), then:*

$$V(\mu) = \sup_{P \in \Delta^2(X)} E_P[F(\nu)] - \lambda E_P[H(\mu) - H(\nu)] \tag{14}$$

The intuition for this result is simple. At any instant in time, suppose that the optimal decision is to continue learning for a positive amount of time. The value is the discounted future value at the next instant of time $(t + \mathrm{d}t)$ less the flow cost of information. Now, consider moving the learning strategy at $t + \mathrm{d}t$ to the current period. Then, both the future value at $t + \mathrm{d}t$ and the cost are discounted by $\mathrm{d}t$ less. If the net utility gain from learning at $t + \mathrm{d}t$ is nonnegative, then this operation increases the current utility by reducing the waiting time.[25] If the net utility gain from learning at $t + \mathrm{d}t$ is negative, then stopping learning immediately increases current utility. This operation can always be applied recursively and strictly improves the strategy until all information is acquired at period 0.[26]

In fact, given Assumptions 1 and 2-b, Equation (1) is a variant of the more general model in Steiner, Stewart, and Matějka (2017), which considers a varying state and repeated decision making. With linear cost function $C(I)$, no motivation for smoothing the learning

[25]This step utilizes Assumption 2′, which implies that the cost of a combined signal structure is the sum of the cost of each of them.

[26]Strictly speaking, an immediate learning strategy is not admissible because its belief path is not cadlag. However, there always exists a way to implement a signal structure in an arbitrarily short period of time, and the payoff approximates the immediate learning payoff.

behavior exists. The dynamics in Steiner, Stewart, and Matějka (2017) are a result of the intertemporal dependence of decision problems.

8 Applications

8.1 Choice accuracy and response time

The two-choice sequential decision making problem has been extensively studied in the psychological and behavioral studies. One of the key objective is to explain the data on choice accuracy and response time from experiments. The drift-diffusion model (DDM) has been the most popular theoretical model for these decision problems, for the reason that DDM is very tractable and fits the accuracy/ response time data well. However, accounting for the *joint distribution* of choice accuracy and response time remains a challenge for DDM. In this section, I apply my model to predict a systematic feature in the data: the *crossover* of response time-accuracy relationship.

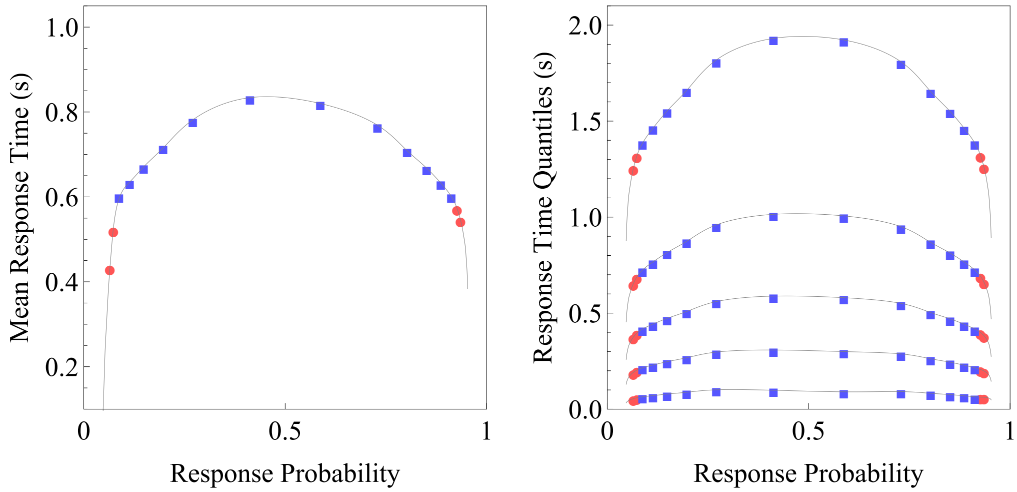
The crossover happens when the difficulty of decision problem varies: the error responses are faster than the correct responses when the task is easy; the error responses are slower than the correct responses when the task is hard (see Luce et al. (1986), Ratcliff, Van Zandt, and McKoon (1999)). First, I illustrate the crossover of time-accuracy relationship in Example 5.

Example 5. Consider the same decision problem as in Example 1. $F(\mu) = \max\{1-2\mu, 2\mu-1\}$ and $\rho = 1$. Assume prior belief $\mu_0 = 0.5$ and let $H_0(\mu)$ be the entropy function. Define uncertainty measure $H(\mu)$ as:

$$H(\mu) = \begin{cases} H_0(\mu) & \text{if } \mu \in [0.5, 0.65] \\ H_0(\mu) - |\mu - 0.5|^3 & \text{if } \mu < 0.5 \\ H_0(\mu) - 4|\mu - 0.65|^3 & \text{if } \mu > 0.65 \end{cases}$$

$H(\mu)$ is an asymmetric uncertainty measure, and $H(\mu)$ is slightly more concave than H_0 when $\mu < 0.5$ or $\mu > 0.65$. The different difficulty levels are modeled as different capacity constraints on $-\mathcal{L}H(\mu_t)$, the higher the capacity constraint is, the easier the decision problem is. I study the joint distribution of choice and decision time conditional on the true state being r ($\mu = 1$). Figure 13 depicts the *latency-probability (LP)* and *quantile-probability (QP)* plots. The horizontal coordinates of the points to the right of $p = 0.5$ shows the choice probability of the action R (the correct choice). Each such point has a corresponding point to the left of $p = 0.5$ showing the remaining probability of the action L (the error). The vertical coordinates of all points show the response time measured by mean (in LP plot) or by quantiles (in QP plot).

The crossover of time-accuracy relationship is illustrated by the differently colored points. The red points are data points where the errors happen earlier than the correct responses (measured by both mean or quantiles). They are simulated with high capacity, thus are of higher accuracy in general. On the contrary, the blue points are data points



Left panel: The latency-probability function (the thin line) and the data points simulated from 8 difficulty levels. Right panel: The quantile-probability functions (the thin lines, from bottom to top: 0.1, 0.3, 0.5, 0.7, 0.9 quantile) and the data points simulated from 8 difficulty levels. The correct responses are to the right of 0.5, the errors are to the left of 0.5. Red points: the errors have shorter response times. Blue points: the errors have longer response time.

Figure 13: LP and QP plots

where the errors happen later than the correct responses. They are simulated with low capacity, and of low accuracy in general. In fact, Figure 13 is qualitatively the same as the LP and QP plots documented in Ratcliff and Rouder (1998) and Ratcliff, Van Zandt, and McKoon (1999).

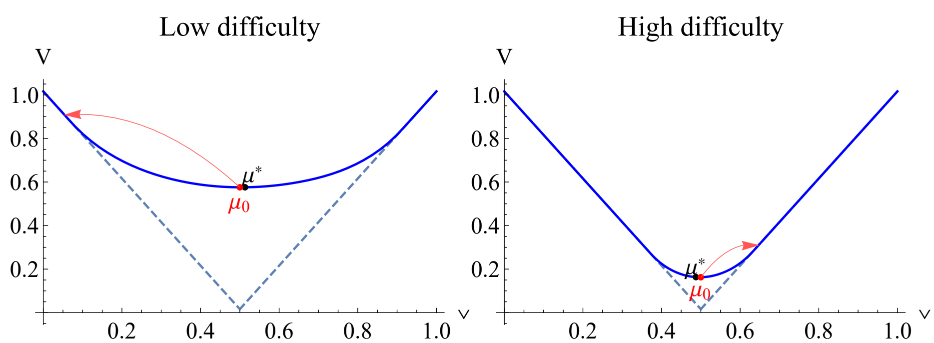


Figure 14: The critical beliefs of different difficulty levels

The main reason for the crossover is explained in Figure 14. When the capacity is low (the task difficulty is high), the optimal size of belief jump is small. By construction of $H(\mu)$, when the posterior belief is not far away from μ_0, learning the state L is more costly than learning the state R. As a result, the critical belief μ^* at which searching for both direction is indifferent is biased toward left. Since $\mu_0 > \mu^*$, the correct responses are font-loaded. Applying the same intuition, when the capacity is high, $\mu_0 < \mu^*$ and the errors are font-loaded.

Applying the idea from Example 5, creating a crossover of μ^* and μ_0 is necessary for creating a crossover of the response time-accuracy relationship.

Proposition 1. *Suppose $|A| = 2$, Assumption 2-a is satisfied. $H_0(\mu)$ and $F(\mu)$ are symmetric around $\mu_0 = 0.5$ and satisfy Assumption 3. $\forall$ partition of $\mathbb{R}^+ : \{0, c_1, \cdots, c_k, \infty\}$, there exists uncertainty measure $H(\mu)$ satisfying Assumption 3 such that:*

1. *When $c \in \{c_k\}$, $\mu^* = \mu_0$, and the optimal strategy at μ_0 is the sames as that with $H_0(\mu)$.*
2. *When c increases on $\mathbb{R}^+$, the sign of $\mu^* - \mu_0$ alternates on each partition.*

Proposition 1 states that the flexible learning model can fit an arbitrary number of crossovers of the response time-accuracy relationship at given difficulty levels. The standard DDM predicts identical decision time distribution for the correct responses and the errors (Ratcliff (1981)). To accommodate a non-trivial speed-accuracy trade-off/complementarity, DDM with varying boundary (Cisek, Puskas, and El-Murr (2009)) or DDM with random starting point and drift (Ratcliff and Rouder (1998)) are proposed, and there are a lot of debate about which variation works better. Fudenberg, Strack, and Strzalecki (2018) shows that the collapsing (expanding) boundary maps exactly to the complementarity (trade-off), and in an uncertain-difference DDM with endogenous stopping, decision boundary collapses to zero asymptotically and accuracy declines over time. These analyses suggest that DDM is able to fit the crossover, however at the cost of adding trial dependent parameters. Meanwhile, it remains to be disentangled which set of parameters in DDM are task specific and which set are subject specific. On the contrary, the flexible learning model predicts the crossovers clearly with varying only a task difficulty parameter, while keeping the task payoffs and the learning technology constant across trials.

8.2 *Radical innovation*

An important question in the study of innovation is to understand what characteristics of a firm foster innovation. The second application relates the radicality of firm's R&D and innovation to its safe option. I consider two firms: an *incumbent* (I) and an *entrant* (E). They face the identical set of risky new products. The only difference between the two firms is that the incumbent has a better existing safe product. I am interested in which firm innovates more radically in the R&D process. Intuitively, there are two competing incentives:

1. *Impatience effect*: The incumbent has an overall higher continuation value than the entrant. Therefore, by the value-precision monotonicity, the more impatient incumbent should prefer the frequency of signal to the precision of signal. So the impatience effect suggests that the entrant innovates more radically.

2. *Threshold effect*: The incumbent has a better outside option. Therefore, it has a higher threshold of belief for accepting a risky option. The relative value of a precise signal to an imprecise signal is higher for the incumbent. Therefore, the threshold effect suggests that the incumbent innovates more radically.

I model the problem using the following setup. There is one safe product P_s and K risky products $\{P_1, \cdots, P_K\}$. The state is $x \in \{G, B\}$. $x = G$ means the new technology is good, and the new products are better than the safe product: $\forall i,k\ u_i(P_k,G) > u_i(P_s,G)$. When $x = B$, the new technology fails, and $\forall i,k\ u_i(P_k,B) < u_i(P_s,B)$. $\forall x,k,\ u_I(P_k,x) = u_E(P_k,x)$ and $u_I(P_s) > u_E(P_s)$. The two firms share the same $H(\mu)$ function and capacity constraint c.[27] Let $\nu_i(\mu)$ be the two firms' optimal strategies. I define that a firm is looking for *more radical innovation* given belief μ *iff* $|\nu_i(\mu) - \mu| > |\nu_{-i}(\mu) - \mu|$, namely firm i is searching for a more precise Poisson signal.

Example 6. I calculate a simple example. There is only one risky product and $K = 1$. The incumbent's safe option pays $u_I(P_s,x) = 0.3$ and the entrant's safe option pays $u_E(P_s,x) = 0.15$. The risky option pays 1 when $x = G$ and -1 when $x = B$. H is the standard entropy function, $\rho = 1, c = 0.3$.

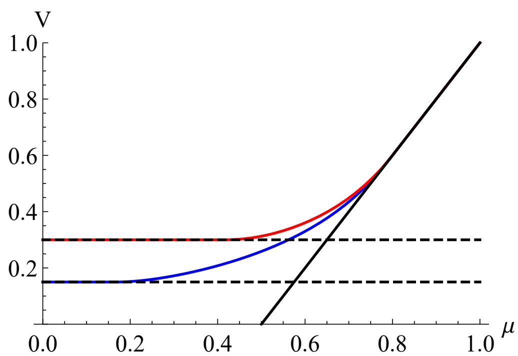


Figure 15: Value function

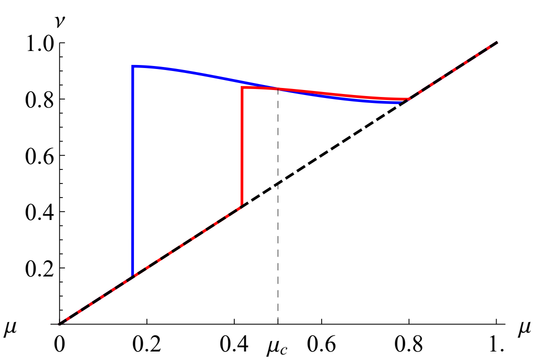


Figure 16: Policy function

Figure 15 depicts the value functions (red curve: incumbent; blue curve: entrant). The two dashed lines are the payoffs of the corresponding safe options. Figure 16 depicts the policy functions (red curve: incumbent; blue curve: entrant). There is clearly a crossover of the policy functions. In the union of the two firm's experimentation regions, when $\mu < \mu_c$ the entrant seeks more radical innovation, when $\mu > \mu_c$ the incumbent seeks more radical innovation.

The result of Example 6 can be summarized by the following proposition. Suppose $K = 1$, let E_0 be the union of the two firms' experimentation regions.

Proposition 2. *There exists μ_c s.t. $\forall \mu \in E_0$, $\mu > \mu_c \implies |\nu_I(\mu) - \mu| > |\nu_E(\mu) - \mu|$ and $\mu < \mu_c \implies |\nu_I(\mu) - \mu| < |\nu_E(\mu) - \mu|$. Moreover, $E_0 \bigcap (0,\mu_c) \neq \emptyset$ and $E_0 \bigcap (\mu_c, 1) \neq \emptyset$.*

Proposition 2 first states that there exist a threshold belief that the incumbent looks for more radical innovation if (and only if) the belief is higher than the threshold. Moreover, there exist none degenerate regions that either firm is innovating more radically than the other. Therefore, the order of radically of the two firms' innovations switches exactly once when the belief changes. Here is the intuition for the crossover. The entrant's value function is always steeper than the incumbent's, hence, the difference in the continuation

[27]It is straightforward that if the cost of R&D is flexible, the incumbent invests (strictly) more as a direct implication of the *value-intensity monotonicity*. So I fix the capacity and focus on the choice of signal precision. It is not hard to extend the results to the flexible cost case.

value is decreasing in the belief. As a result, the impatience effect is diminishing when μ increases. On the other hand, when μ is higher, it is ex ante more likely that the risky arm will be chosen. As a result, the threshold effect outweighs the impatience effect when μ increases. Therefore, when μ increases, the incumbent is increasingly favoring a more precise signal, comparing to the entrant. Thus, there is a crossover.

Proposition 2 extends to multiple risky products as well. When $K > 1$, the experimentation regions are no longer simple intervals. Instead, they are unions of open intervals. In any experimentation interval where V never touches F_s, the two firms use the identical strategy (since the outside option is never triggered). So we only consider the leftmost interval in each firm's experimentation region. Let E_0 be the union of the two firms' leftmost intervals of the experimentation region.

Proposition 3. *There exists μ_c s.t. $\forall \mu \in E_0$, $\mu > \mu_c \implies |\nu_I(\mu) - \mu| > |\nu_E(\mu) - \mu|$ and $\mu < \mu_c \implies |\nu_I(\mu) - \mu| < |\nu_E(\mu) - \mu|$. Moreover, $E_0 \bigcap (0, \mu_c) \neq \emptyset$ and $E_0 \bigcap (\mu_c, 1) \neq \emptyset$.*

9 Conclusion

This paper provides a dynamic information acquisition framework which allows fully general design of signal processes, and characterizes the optimal information acquisition strategy. My first contribution is an optimization foundation for a family of simple information generating processes: for an information acquisition problem with flexible design of information, the optimal information structure causes beliefs to follow a jump-diffusion process. Second, I characterize the optimal policy: seeking a Poisson signal whose arrival confirms the prior belief is optimal. The arrival of the signal leads to an immediate action. The absence of the signal is followed by continued learning with increasing precision and decreasing frequency.

References

Arrow, Kenneth J, David Blackwell, and Meyer A Girshick (1949). "Bayes and minimax solutions of sequential decision problems." In: *Econometrica, Journal of the Econometric Society*, pp. 213–244.

Aumann, Robert J, Michael Maschler, and Richard E Stearns (1995). *Repeated games with incomplete information*. MIT press.

Bellman, Richard (1957). *Dynamic programming*. Princeton University Press.

Bergemann, Dirk and Stephen Morris (2017). "Information design: A unified perspective."

Bergemann, Dirk and Juuso Välimäki (1996). "Learning and strategic pricing." In: *Econometrica: Journal of the Econometric Society*, pp. 1125–1149.

Boel, René and Michael Kohlmann (1980). "Semimartingale models of stochastic optimal control, with applications to double martingales." In: *SIAM Journal on Control and Optimization* 18.5, pp. 511–533.

Bolton, Patrick and Christopher Harris (1999). "Strategic experimentation." In: *Econometrica* 67.2, pp. 349–374.

Callander, Steven (2011). "Searching and learning by trial and error." In: *American Economic Review* 101.6, pp. 2277–2308.

Caplin, Andrew and Mark Dean (Aug. 2013). *Behavioral Implications of Rational Inattention with Shannon Entropy*. Working Paper 19318. National Bureau of Economic Research. DOI: 10.3386/w19318. URL: http://www.nber.org/papers/w19318.

— (2015). "Revealed preference, rational inattention, and costly information acquisition." In: *The American Economic Review* 105.7, pp. 2183–2203.

Caplin, Andrew, Mark Dean, and John Leahy (Aug. 2017). *Rationally Inattentive Behavior: Characterizing and Generalizing Shannon Entropy*. Working Paper 23652. National Bureau of Economic Research. DOI: 10.3386/w23652. URL: http://www.nber.org/papers/w23652.

Che, Yeon-Koo and Konrad Mierendorff (2016). "Optimal sequential decision with limited attention."

Cisek, Paul, Geneviève Aude Puskas, and Stephany El-Murr (2009). "Decisions in changing conditions: the urgency-gating model." In: *Journal of Neuroscience* 29.37, pp. 11560–11571.

Clark, Aubrey (2016). *Contracts for information acquisition*. unpublished manuscript.

Cover, Thomas M and Joy A Thomas (2012). *Elements of information theory*. John Wiley & Sons.

Crandall, Michael G, Hitoshi Ishii, and Pierre-Louis Lions (1992). "User's guide to viscosity solutions of second order partial differential equations." In: *Bulletin of the American mathematical society* 27.1, pp. 1–67.

Davis, Mark HA (1979). "Martingale methods in stochastic control." In: *Stochastic Control Theory and Stochastic Differential Systems*. Springer, pp. 85–117.

Denti, Tommaso (2015). *Unrestricted information acquisition*. Tech. rep. Technical report. Mimeo.

Doval, Laura (2018). "Whether or not to open Pandora's box." In: *Journal of Economic Theory* 175, pp. 127–158.

Ely, Jeffrey C (2017). “Beeps.” In: *American Economic Review* 107.1, pp. 31–53.
Fleming, Wendell H (1969). “Optimal continuous-parameter stochastic control.” In: *SIAM review* 11.4, pp. 470–509.
Frankel, Alexander and Emir Kamenica (2018). “Quantifying information and uncertainty.”
Fudenberg, Drew, Philipp Strack, and Tomasz Strzalecki (2018). “Speed, accuracy, and the optimal timing of choices.” In: *American Economic Review* 108.12, pp. 3651–84.
Gentzkow, Matthew and Emir Kamenica (2014). “Costly persuasion.” In: *American Economic Review* 104.5, pp. 457–62.
Gittins, John (1974). “A dynamic allocation index for the sequential design of experiments.” In: *Progress in statistics*, pp. 241–266.
Hanson, Floyd B (2007). *Applied stochastic processes and control for Jump-diffusions: modeling, analysis, and computation*. Vol. 13. Siam.
Hébert, Benjamin and Michael Woodford (2016). “Rational Inattention with Sequential Information Sampling.”
Hörner, Johannes and Andrzej Skrzypacz (2016). “Learning, experimentation and information design.”
Kamenica, Emir and Matthew Gentzkow (2011). *Bayesian persuasion*. Tech. rep. 6, pp. 2590–2615.
Ke, T Tony and J Miguel Villas-Boas (2016). “Optimal Learning before Choice.”
Keller, Godfrey and Sven Rady (2010). “Strategic experimentation with Poisson bandits.” In: *Theoretical Economics* 5.2, pp. 275–311.
Keller, Godfrey, Sven Rady, and Martin Cripps (2005). “Strategic experimentation with exponential bandits.” In: *Econometrica* 73.1, pp. 39–68.
Klabjan, Diego, Wojciech Olszewski, and Asher Wolinsky (2014). “Attributes.” In: *Games and Economic Behavior* 88, pp. 190–206.
Liang, Annie, Xiaosheng Mu, and Vasilis Syrgkanis (2017). “Optimal Learning from Multiple Information Sources.”
Lowther, George (2009). *Cadlag Modifications*. https://almostsure.wordpress.com/2009/12/18/cadlag-modifications/. Blog.
Luce, R Duncan et al. (1986). *Response times: Their role in inferring elementary mental organization*. 8. Oxford University Press on Demand.
Matějka, Filip (2015). “Rationally inattentive seller: Sales and discrete pricing.” In: *The Review of Economic Studies* 83.3, pp. 1125–1155.
Matějka, Filip and Alisdair McKay (2012). “Simple market equilibria with rationally inattentive consumers.” In: *American Economic Review* 102.3, pp. 24–29.
— (2014). “Rational inattention to discrete choices: A new foundation for the multinomial logit model.” In: *The American Economic Review* 105.1, pp. 272–298.
Mathevet, Laurent, Jacopo Perego, and Ina Taneva (2017). “On information design in games.”
Matyskova, Ludmila (2018). *Bayesian persuasion with costly information acquisition*. unpublished manuscript.
Mayskaya, Tatiana (2016). “Dynamic Choice of Information Sources.”

Morris, Stephen and Philipp Strack (June 2017). *The Wald Problem and the Equivalence of Sequential Sampling and Static Information Costs*. Working Paper. URL: https://ssrn.com/abstract=2991567.

Moscarini, Giuseppe and Lones Smith (2001). "The optimal level of experimentation." In: *Econometrica* 69.6, pp. 1629–1644.

Rappoport, Daniel and Valentin Somma (2017). *Incentivizing Information Design*. unpublished manuscript.

Ratcliff, Roger (1981). "A theory of order relations in perceptual matching." In: *Psychological Review* 88.6, p. 552.

Ratcliff, Roger and Jeffrey N Rouder (1998). "Modeling response times for two-choice decisions." In: *Psychological Science* 9.5, pp. 347–356.

Ratcliff, Roger, Trisha Van Zandt, and Gail McKoon (1999). "Connectionist and diffusion models of reaction time." In: *Psychological review* 106.2, p. 261.

Ravid, Doron (2018). "Bargaining with rational inattention."

Shannon, Claude Elwood (1948). "A mathematical theory of communication." In: *Bell system technical journal* 27.3, pp. 379–423.

Sims, Christopher A (2003). "Implications of rational inattention." In: *Journal of monetary Economics* 50.3, pp. 665–690.

Steiner, Jakub, Colin Stewart, and Filip Matějka (2017). "Rational Inattention Dynamics: Inertia and Delay in Decision-Making." In: *Econometrica* 85.2, pp. 521–553.

Striebel, Charlotte (1984). "Martingale conditions for the optimal control of continuous time stochastic systems." In: *Stochastic processes and their applications* 18.2, pp. 329–347.

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

Weber, Richard et al. (1992). "On the Gittins index for multiarmed bandits." In: *The Annals of Applied Probability* 2.4, pp. 1024–1033.

Weitzman, Martin L (1979). "Optimal search for the best alternative." In: *Econometrica: Journal of the Econometric Society*, pp. 641–654.

Yang, Ming (2015a). "Coordination with flexible information acquisition." In: *Journal of Economic Theory* 158, pp. 721–738.

— (2015b). "Optimality of debt under flexible information acquisition."

Zhong, Weijie (Apr. 2018a). "Information design possibility set." In: *ArXiv e-prints*. arXiv: 1804.05752.

— (Sept. 2018b). "Time preference and information acquisition." In: *ArXiv e-prints*. arXiv: 1809.05120.

Appendix

A Further discussions

In Appendix A, I first discuss the convergence of discrete-time optimal policy in Appendix A.1. It is shown that the discrete-time optimal policy's support as a correspondence of prior belief converges to that of the continuous-time optimal policy. Then I complete the discussion in Section 7 by generalizing each of the restrictive assumptions. Appendix A.2 generalizes the finite actions assumption and shows that the solution of a problem with infinite actions can be approximated by solutions to a series of problems with increasing number of actions. Appendix A.3 generalizes the binary states assumption in Assumption 3 and shows that the properties of optimal policy in Theorem 2 all extend in a problem with general finite state space. The proofs of theorems stated in this section are relegated to Section S5.

A.1 Convergence of policy

By Theorems 2 and 3, the optimal policy solving Equation (4) is essentially unique in the jump-diffusion class. However, Theorem 1 does not rule out other possible optimal policies for the original stochastic control problem Equation (1). To get behavior predictions from my model, additional refinement of optimal policy of Equation (1) is necessary. In this discussion, I show that the discrete-time optimal policy of Equation (6) converges to the solutions defined in Theorems 2 and 3. I define a modified version of Lévy distance that characterizes the difference between two policy correspondences:

Definition 1 (Lévy metric)**.** *Let F,G: $[0,1]\to 2^{[0,1]}$ be two correspondences. The Lévy metric $d_{\mathscr{L}}(F,G)$ is defined as:*

$$d_{\mathscr{L}}(F,G):=\inf\left\{\varepsilon>0\,\middle|\,\inf_{|y-x|\leqslant\varepsilon} d_H(F(x),G(y))\leqslant\varepsilon,\forall x\in[0,1]\right\}$$

where d_H is the standard Hausdorff metric on $\mathbb{R}$.

$d_{\mathscr{L}}(F,G)=a$ means that $\forall\mu\in[0,1]$, $\forall y\in F(\mu)$, there exists some μ' in a-neighbourhood of μ such that y is in the a-neighbourhood of $G(\mu')$. When G is continuous at μ, and a is sufficiently small, it simply states that the images of F and G at μ are close to each other (measured by d_H). If $d_{\mathscr{L}}(F,G)=0$ then F and G are identical.

Theorem 7 (Convergence of policy)**.** *Given either Assumptions 1, 2-a and 3 or Assumptions 1, 2-b and 3, let $\nu(\mu)$ be the policy correspondence solving Equation (4). Let $N(\mu)=\{\mu\}\bigcup\nu(\mu)$. Let $N_{dt}(\mu)$ be the support of optimal posteriors solving Equation (6). Then:*

$$\lim_{dt\to 0} d_{\mathscr{L}}(N,N_{dt})=0$$

Theorem 7 states that the graph of policy function of discrete-time problem Equation (5') converges to the graph of the continuous solution defined in Theorems 2 and 3. The convergence is illustrated in Figure 17. I calculate the discrete-time policy function using parameters in Example 2. The red, blue and green lines represent the set of optimal posteriors as functions of prior when $V_{dt}>F$ with $dt=10^{-5},10^{-3}$ and 10^{-2}. As is shown in the figure, when $dt\to 0$, one of optimal posterior is converging to the prior, and the other optimal posterior is converging to the continuous time solution. The posterior converging to prior captures a drift term and the other posterior captures a Poisson jump in the limit.

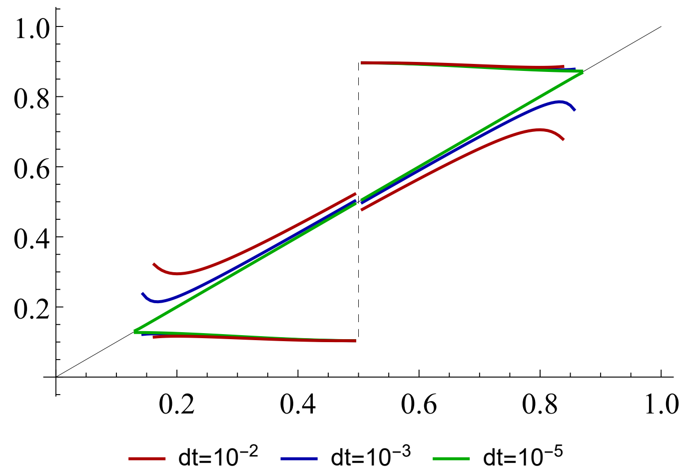


Figure 17: Convergence of policy function

A.2 *Infinite action space*

In this section, I extend my model to accommodate infinite actions (or even continuum of actions) in the underlying decision problem, i.e. $|A|=\infty$. Mathematically, the difference is that the value from immediate action $F(\mu)=\sup_{a\in A}E[u(a,x)]$ is no-longer a piecewise linear function. There are several technical problems arising from a continuum of actions. For example whether the supremum is indeed achieved and whether F has bounded subdifferentials. I impose the following assumption to rule out these technical issues:

Assumption 5. *$F(\mu)=\max_{a\in A}E[u(a,x)]$ has bounded subdifferentials.*

Assumption 5 rules out two cases. The first case is that the supremum is not achievable. The second case is that some optimal action being infinitely risky: the optimal action with belief approaching $x=0$ has utility approaching $-\infty$ at state 1 (and similar case with states swapped). A sufficient condition for Assumption 5 is:

Assumption 5′. *A is a compact set. $\forall x\in X$, $u(a,x)\in C(A)\bigcap \mathrm{TB}(A)$.*

It is useful to notice that the proof of Theorem 1 does not rely on the fact that $F(\mu)$is piecewise linear. Actually the only necessary properties of $F(\mu)$ are boundedness and continuity in Lemma 2, which prove the existence of solution to discrete time functional equation Equation (S.1). Therefore Assumption 5 guarantees that Lemma 2 and Lemma S.8 still hold when there is a continuum of actions. With Assumption 5, the problem with continuum of actions can be approximated well by a sequence of problems with discrete actions. I first define the following notation: $\forall F$ satisfying Assumption 5, $\mathcal{V}_{dt}(F)$ is the unique solution of Equation (6) and $\mathcal{V}(F)=\lim_{dt\to 0}\mathcal{V}_{dt}(F)$[28].

Lemma A.1. *Given Assumption A and Assumptions 2 and 5, $\mathcal{V}$ is a Lipschitz continuous functional under L_∞ norm.*

Lemma A.1 implies that a problem with continuum of actions can be approximated well by a sequence of problems with discrete actions in the sense of value function convergence. Next, I push the convergence criteria further to the convergence of policy function.

Theorem 8. *Given Assumptions 1, 2-a, 3 and 5, let $\{F_n\}$ be a set of piecewise linear functions on [0,1] satisfying:*

1. *$\|F_n-F\|_\infty\to 0$;*
2. *$\forall\mu\in[0,1]$, $\lim F_n'(\mu)=F'(\mu)$.*

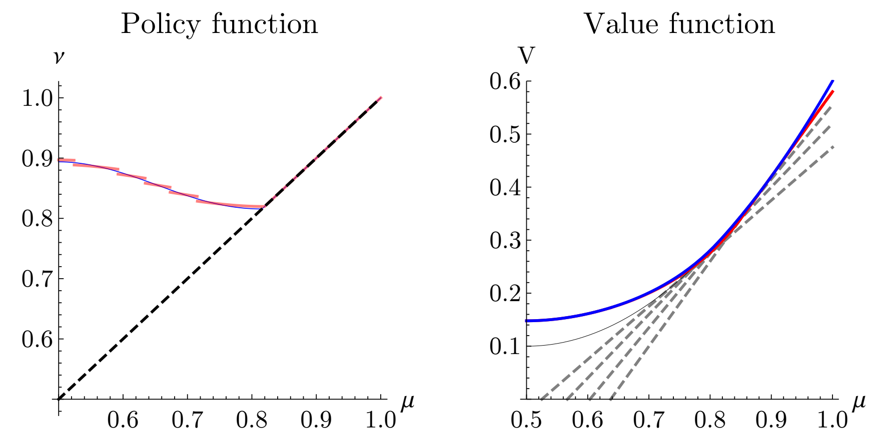
Then $|\mathcal{V}(F)-\mathcal{V}(F_n)|\to 0$ and:

1. *$\mathcal{V}(F)$ solves Equation (4).*
2. *$\forall\mu$ s.t. $V(\mu)>F(\mu)$, if each ν_n is maximizer of $\mathcal{V}(F_n)$ and $\nu=\lim_{n\to\infty}\nu_n$ exists, then ν is the optimal posterior in Equation (4) at μ.*

Theorem 8 states that to solve the problem with a continuum of actions, one can simply use both value function and policy function from problems with finite actions to approximate. As long as the immediate action values F_n converge uniformly in value and pointwise in first derivative, the optimal value functions have a uniform limit. The limit solves Equation (4) and the optimal policy function is the pointwise limit of policy functions from the finite action problems.

Figure 18 illustrates this approximation process. On both panels, only $\mu\in[0.5,1]$ is plotted (policy and value on [0,0.5] are symmetric). On the right panel, the thin black curve shows a smooth $F(\mu)$ associated with continuum of

[28]The existence of limit is guaranteed by monotonic convergence theorem.



Left panel shows the optimal policy function of discrete actions (red) and continuous actions (blue). The dashed line is $\nu=\mu$.
Right panel shows the optimal value function. The thin black line is value from immediate action $F(\mu)$, the dashed lines are discrete approximations of the continuous function F.

Figure 18: Approximation of a continuum of actions

actions. Since optimal policy only utilizes a subset of actions, I approximate the smooth function only locally as the upper envelope of dashed lines (each represents one action). The optimal value function with continuous actions is the blue curve and the discrete action approximation is the red curve. The left panel shows the approximation of policy function. The blue smooth curve is the optimal policy of the continuous action problem and the red curve with breaks is the optimal policy of the discrete action problem.

To approximate a smooth $F(\mu)$, one can simply add more and more actions to the finite action problem and use F's supporting hyper planes to approximate it. Then the optimal policy functions have more and more breaks as optimal policys involve more frequent jumps among actions. In the limit, as number of breaks grows to infinity, the size of breaks shrinks to zero and approaches a continuous policy function.

A.3 General state space

In this section, I extend the size of state space. The constructive proof for Theorems 2 and 3 relies on the ODE theory to guarantee existence of solution. With a larger state space, construction of value function relies on existence of PDE. There is no general theory ensuring existence of solution.[29] Nevertheless, the verification part still works. In fact, the discussion in Section 6.2 seems to extend to higher dimensional spaces in a natural way. I formalize a partial characterization theorem in the section.

Let $n=|X|$. Consider value function $V(\mu)$ on $\Delta(X)$. Let $V(\mu)\in C\Delta(X)$ and $C^{(2)}$ smooth when $V(\mu)>F(\mu)$. Consider the following HJB equation:

$$\rho V(\mu)=\max\left\{\rho F(\mu),\max_{\nu,p,\sigma} p(V(\nu)-V(\mu)-\nabla V(\mu)\cdot(\nu-\mu))+\sigma^T \mathrm{H}V(\mu)\sigma\right\} \tag{15}$$
$$s.t.\ -p(H(\nu)-H(\mu)-\nabla H(\mu)\cdot(\nu-\mu))-\sigma^T \mathrm{H}H(\mu)\sigma\leqslant c$$

where $\nu\in\Delta(\mathrm{supp}(\mu))$, $p\in\Delta I$ and $\sigma\in\mathbb{R}^{|\mathrm{supp}(\mu)|}$. Equation (15) comes from applying Assumption 2-a and smoothness condition to Equation (4). [30] I only discuss Assumption 2-a because the intuition is the same and similar proof methodology can be applied to Assumption 2-b to show an analog result.

Theorem 9. *Let $E=\{\mu\in\Delta(X)|V(\mu)>F(\mu)\}$ be the experimentation region. Suppose there exists $C^{(2)}$ smooth $V(\mu)$ on E solving Equation (15), then $\exists$ policy function $\nu:E\mapsto\Delta(X)$ s.t.*

$$\rho V(\mu)=-c\frac{F(\nu(\mu))-V(\mu)-\nabla V(\mu)(\nu(\mu)-\mu)}{H(\nu(\mu))-H(\mu)-\nabla H(\mu)(\nu(\mu)-\mu)}$$

and ν satisfies the following properties:

1. *Poisson learning: $\rho V(\mu)\geqslant\sup_{\sigma} -c\frac{\sigma^T\mathrm{H}V(\mu)\sigma}{\sigma^T\mathrm{H}H(\mu)\sigma}$.*
2. *Direction: $D_{\nu(\mu)-\mu}V(\mu)\geqslant 0$.*
3. *Precision: $D_{\mu-\nu(\mu)}\nu(\mu)\cdot\mathrm{H}H(\nu)(\nu-\mu)\leqslant 0$.*

[29] The maximization problem can be translated into a PDE system. What is problematic is the boundary conditions. In fact, to solve for $V(\mu)$ searching over one action, I need to use the value function at regions where DM is indifferent between two actions as a boundary condition. That boundary condition is unknown, in contrast to the one dimensional analog $V(\mu^*)$ which can be easily calculated.

[30] $\mathrm{H}H(\mu)$ is defined on boundary where $V(\mu)=F(\mu)$ as continuous extension of interior Hessian's by Kirszbraun theorem.

4. Stopping time: $\nu(\mu)\in E^C$.
There exists a nowhere dense set K s.t. strict inequality holds on $E\backslash K$ in property 1,3 and 4.

Theorem 9 states that if a solution $V(\mu)$ to Equation (15) exists, then $V(\mu)$ can be solved with only Poisson signals. The four properties are extensions to the four properties in Theorem 2 respectively. Property 1 and 4 are exactly the suboptimality of Gaussian signal and the immediate action property. Property 2 and 3 are weaker than the corresponding properties in Theorem 2. Property 2 is the extension to the confirmatory signal property. It states that optimal direction of jump is in the myopic direction that value function increases. Property 3 is the extension to the increasing precision property. $D_{\mu-\nu}\nu(\mu)$ is the direction ν is moving when μ is moving against ν. $\mathrm{H}H(\nu)(\nu-\mu)$ is the direction $(\nu-\mu)$ distorted by a negative definite matrix $\mathrm{H}H(\nu)$. In a special case when $H(\mu)=\|\mu-\mu_0\|_2^2$, $\mathrm{H}H(\nu)(\nu-\mu)$ is in the same direction as $(\mu-\nu)$, which implies (together with property 3) that the distance between μ and ν is increasing when μ is drifting against ν. In a generic case, this property does not directly predict how $\|\nu-\mu\|$ changes.

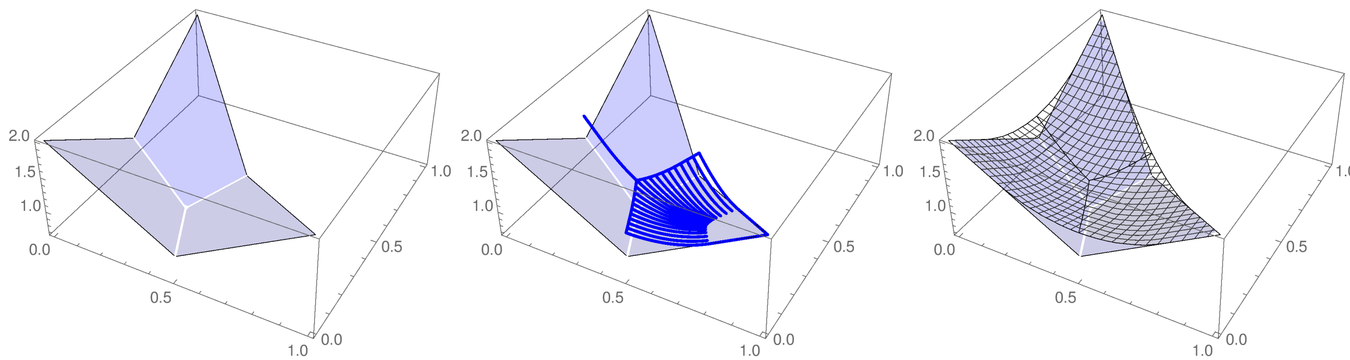
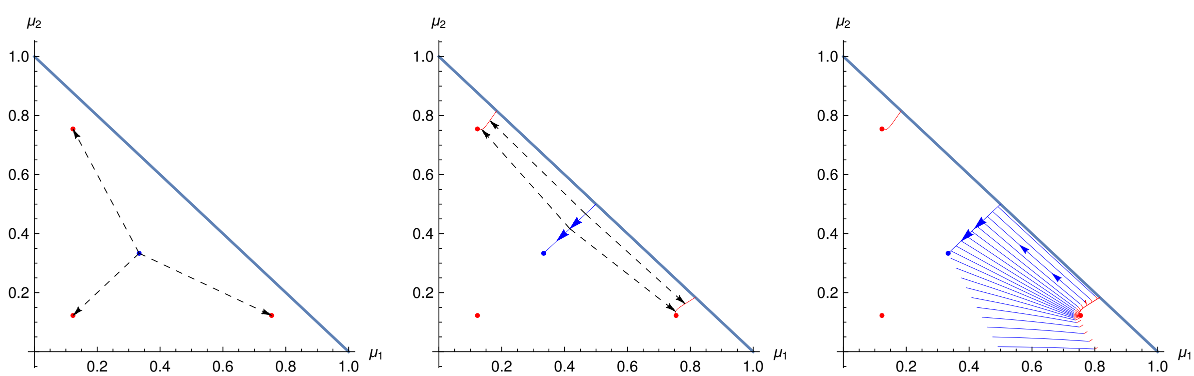


Figure 19: Value function with 3 states

Figures 19 and 20 illustrate Theorem 9 in a numerical example. There are three states and three actions. Belief space is a two-dimensional simplex. $F(\mu)$ is assumed to be a centrally symmetric function on belief space (Figure 19-(a)). Value function $V(\mu)$ is the meshed manifold in Figure 19-(c). Each blue curve in Figure 19-(b) shows a drifting path of posterior beliefs. Take a prior in lower right region. The optimal policy is to search for one posterior (red points in lower right corner of Figure 20-(c)), and posterior belief conditional on receiving no signal drifts along the curve in arrowed direction as in Figure 20-(c). Once belief reaches the boundary, optimal policy becomes searching for two posteriors in a balanced way and posterior drifts towards center of belief space (see Figure 20-(b), arrowed blue curve is belief trajectory and dashed arrows points to optimal posterior). Finally, if belief reaches center, optimal policy is to search for three posteriors in a balanced way (Figure 20-(a)).



Dashed arrows start from priors and point to optimal posteriors. Blue arrows represents drift of posrtior beliefs conditional on no signal arrival. Left panel shows a point at which a balanced search over three posteriors is optimal. MIddle panel shows a curve along which searching over two posteriors is optimal. Right panel shows curves along witch searching over one unique posterior is optimal.

Figure 20: Policy function with 3 states

A.4 Discrete-time information acquisition

In this section, I introduce a general discrete-time information acquisition problem. In the general problem, information is explicitly modeled as state-dependent signal process, and the cost of information is defined using a *posterior separable* function. I show that the discrete-time auxiliary problem Equation (5) introduced in Section 5.1 is a reduced form of the general problem. In Appendix A.4.1, I axiomatize posterior separability.

Decision problem: Time is discrete $t \in \mathbb{N}$. Period length is $dt > 0$. The other primitives (A, X, u, μ, ρ) are the same as in Section 3. The Bernoulli utility of action-state pair (a,x) in period t is $e^{-\rho dt \cdot t} u(a,x)$.

Strategy: a strategy is a triplet $(\mathcal{S}^t, \tau, \mathcal{A}^t)$. $\mathcal{S}^t$ is a random process correlated with the state, called an *information structure*. The realization of $\mathcal{S}^t$ is called a *signal history*. The signal history up to period t is denoted by $\mathcal{S}^{\mathbf{t}}$. Each $\mathcal{S}^t$ specifies the signal structure acquired in period t conditional on all histories up to period t.[31] τ is a random variable whose realization is in $\mathbb{N}$. τ specifies a random decision time. The action choice $\mathcal{A}^t$ is a random process whose realization is in A. Each $\mathcal{A}^t$ specifies the joint distribution of action choice and state conditional on making decision in period t. Let the marginal distribution of the state be denoted by random variable $\mathcal{X}$.

Cost of information: Define $C_{dt}(I) = C\left(\frac{I}{dt}\right) dt$. The per-period cost of information is $C_{dt}(I(\mathcal{S}^t; \mathcal{X} | \mathcal{S}^{\mathbf{t-1}}, \mathbf{1}_{\tau \leqslant t}))$,[32] where the measure of signal informativeness I is defined as:

Assumption A. *$I(\mathcal{S}; \mathcal{X} | \mu) = E_s[H(\mu) - H(\nu(\cdot|s))]$, where ν is the posterior belief about x according to Bayes rule.*

It is not difficult to see that $I(\mathcal{S}^t; \mathcal{X} | \mathcal{S}^{\mathbf{t-1}}, \mathbf{1}_{\tau \leqslant t})$ is exactly the finite difference formulation of $-\mathcal{L}_t H(\mu_t) dt$. Assumption A is called *(uniform) posterior separability* in the literature. If H is the standard entropy function, then I is the mutual information between signal $\mathcal{S}^t$ and unknown state $\mathcal{X}$ (conditional on history).

Dynamic optimization: The dynamic optimization problem of the DM is:

$$V_{dt}(\mu) = \sup_{\mathcal{S}^t, \tau, \mathcal{A}^t} E\left[e^{-\rho dt \cdot \tau} u(\mathcal{A}^\tau, \mathcal{X}) - \sum_{t=0}^{\infty} e^{-\rho dt \cdot t} C_{dt}\left(I\left(\mathcal{S}^t; \mathcal{X} | \mathcal{S}^{\mathbf{t-1}}, \mathbf{1}_{\tau \leqslant t} \right) \right) \right] \tag{5'}$$
$$\text{s.t.} \begin{cases} \mathcal{X} \rightarrow \mathcal{S}^{t-1} \rightarrow \mathbf{1}_{\tau \leqslant t} \\ \mathcal{X} \rightarrow \mathcal{S}^{t-1} \rightarrow \mathcal{A}^t \text{ conditional on } \tau = t \end{cases}$$

The two constraints in Equation (5′) are called the *information processing constraints*. Notation $\mathcal{X} \rightarrow \mathcal{S} \rightarrow \mathcal{T}$ means $\mathcal{X} \perp\!\!\!\perp \mathcal{T} | \mathcal{S}$. The first constraint states that signal history prior to action time is sufficient for action time. The second constraint states that signal history prior to period t is sufficient for action at time t.[33] They are extensions to the standard measurability requirement, allowing randomness unrelated to unknown state to be added.

Equation (5′) is more general than Equation (5) in that it explicitly models the fully flexible choice of information. Take any strategy in Equation (5), if we consider belief as direct signal, then it resembles a special kind of strategy which is feasible in Equation (5′). These special strategies involve no *irrelevant randomness* and *unused information*, which are permitted in Equation (5′). In fact, Equation (5′) is more general than Equation (5) *only* in permitting irrelevant randomness and unused information. It is quite intuitive that allowing those more general strategies doesn't improve utility at all. In fact, it is proved in Lemmas S.4 and S.5 that V_{dt} defined by Equation (5′) is identical to that defined by Equation (5), for which reason I do not differentiate the notation.

Given the discussion above, Equation (5′) serves as a formal justification for using a belief based approached to model dynamic information acquisition. Moreover, it also relates Assumption 1 to posterior separable function — a measure for information widely used in rational inattention problems. In addition to existing attempts to axiomatize or microfound Assumption A, I provide a different axiomitization based on sequential information decomposition in Appendix A.4.1.

A.4.1 Axiom for Assumption A

Theorem 10. *$I(\mathcal{S}; \mathcal{X} | \mu)$ is a non-negative function of information structure and prior belief. I satisfies Assumption A if and only if the following axiom holds:*
Axiom: *$\forall \mu$, $\forall$ information structure $\mathcal{S}_1$ and information structure $\mathcal{S}_2|_{S_1}$ whose distribution depends on realization of $\mathcal{S}_1$:*

$$I((\mathcal{S}_1, \mathcal{S}_2); \mathcal{X} | \mu) = I(\mathcal{S}_1; \mathcal{X} | \mu) + E[I(\mathcal{S}_2; \mathcal{X} | \mathcal{S}_1, \mu)]$$

Theorem 10 states that the *chain rule* (the name for a key property of mutual information in Cover and Thomas (2012)) is not only a necessary condition but also a sufficient condition for posterior separability. Given any experiment, we can divide it into multiple stages of "smaller" experiments. This axiom requires that the total informativeness of this sequence of small experiments is "path-independent": it always equals to the informativeness of the compound experiment.

[31] $\mathcal{S}^{-1}$ is defined as a degenerate random variable that induces belief same as prior belief μ for notation simplicity.

[32] $\mathbf{1}_{\tau \leqslant t}$ is an indicator whether learning is already stopped up to current period, which is known to the DM. So $(\mathcal{S}^{\mathbf{t-1}}, \mathbf{1}_{\mathcal{T} \leqslant t})$ summarizes all knowledge of the DM.

[33] Notice that in every period, the information in current period has not been acquired yet. So decision can only be taken based on the information already acquired in the past. As a result in the information processing constraints information is advanced by one period. This within period timing issue does not make a difference when going to continuous-time limit.

B Omitted proofs

B.1 Roadmap for proofs

Figure 21: Roadmap for proofs

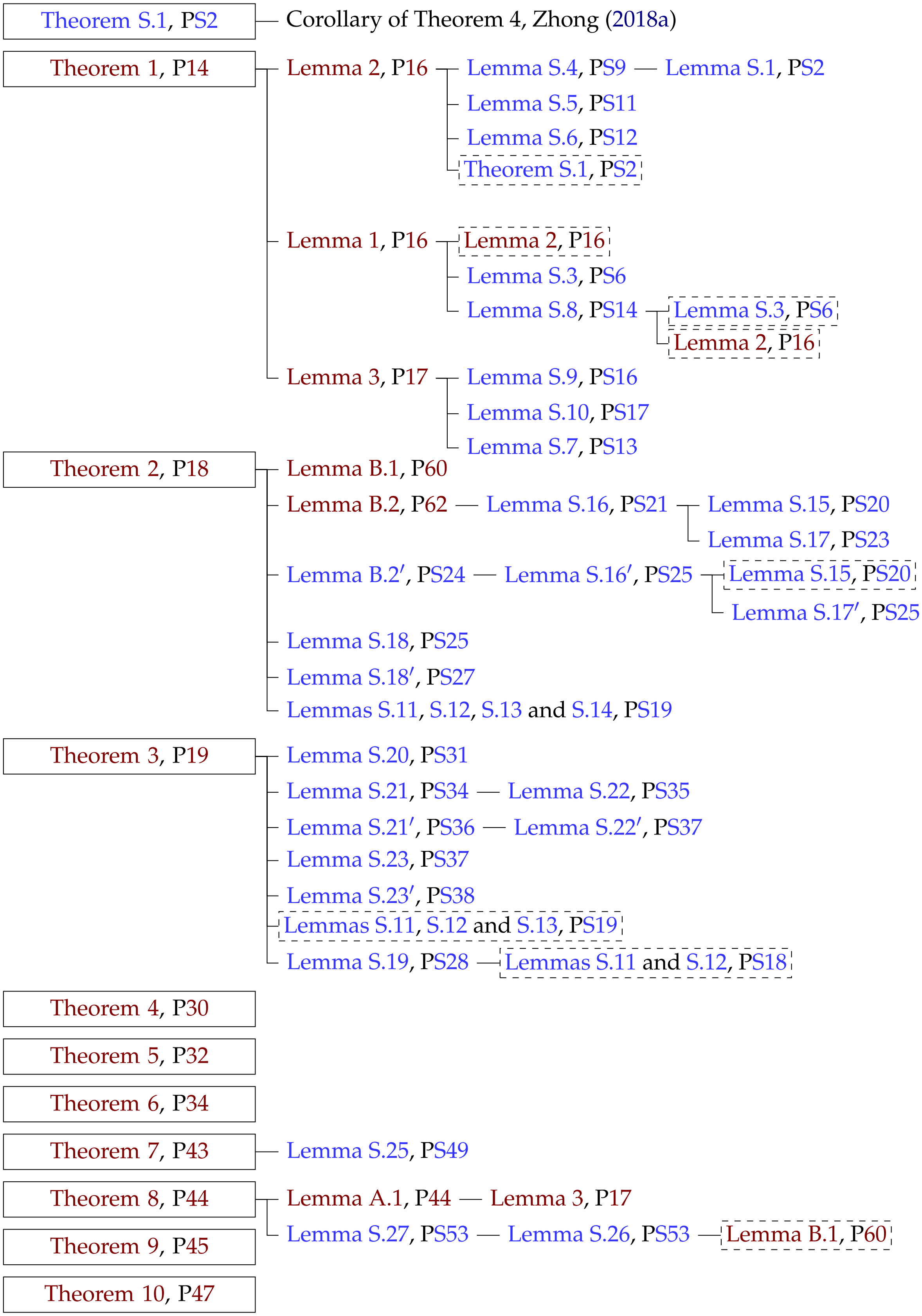


Figure 21 illustrates the roadmap for proofs in this paper. Each node in the figure displays a theorem/lemma's name and its page number. Proof of each node depends (indirectly) on all nodes linked (indirectly) to it on the right. From top to bottom, the nodes are ordered by order of proofs: each node only depends on nodes on the right of it or above it. So it is clear that there is no circular argument. Dependent nodes that have been proved earlier are boxed by dashed lines. From left to right, the nodes are ordered by importance. Lemmas in the first layer are conceptually important and are directly supporting the proof for theorems. Lemmas in the second layer or above are more technical lemmas.

B.2 *Proof of Theorem 1*

The general road map for proving Theorem 1 is introduced in Section 5.3. The proof relies on three lemmas. Lemma 1 proves that the value function V_{dt} of discrete-time optimization problem Equation (5′) converges to the value function V of continuous-time optimization problem Equation (1) as $dt \to 0$. Lemma 3 proves that the solution of discrete time Bellman Equation (6) converges to the solution of continuous time HJB Equation (4) as $dt \to 0$. Lemma 2 proves that V_{dt} is also the solution of Bellman Equation (6). Therefore, V is the solution of HJB Equation (4).

Among the three lemmas, Lemmas 1 and 2 are quite standard, and the proofs are mostly variations of standard arguments. In Appendices B.2.1 and B.2.2, I discuss only the main proof ideas and some non-standard details and relegate the standard parts and purely technical details to Section S2.1.

Lemma 3 is the key lemma for Theorem 1, as it provides an important link between discrete time Bellman and continuous time HJB. Proof of Lemma 3 is provided in details in Appendix B.2.3. The discussion also formalizes the definition of HJB Equation (4) by clarifying the notion of viscosity solution I am using.

B.2.1 *Proof of Lemma 1*

Remark B.1. The proof of Lemma 1 uses Lemma 2 for some minor technical arguments. However the main proof idea does not conceptually depend on Lemma 2. So I show the proof of Lemma 1 first.

Proof. As already stated in Section 5.1, it is sufficient to show that the order of limits can be switched:

$$\sup_{\langle\mu_t\rangle,\tau} \lim_{dt\to 0} W_{dt}(\mu_t,\tau) = \lim_{dt\to 0} \sup_{\langle\mu_t\rangle,\tau} W_{dt}(\mu_t,\tau) \tag{16}$$

Here $W_{dt}(\mu_t,\tau)$ is defined in Section 5.1 as the *discretized* payoff of continuous time strategy $\langle\mu_t\rangle,\tau$. The inner limit of LHS in Equation (16) is then by definition the payoff of strategy $\langle\mu_t\rangle,\tau$ in the continuous time problem Equation (1). So the LHS is $V(\mu)$. The inner limit of RHS is $V_{dt}(\mu)$ (as the problem optimizing W_{dt} is a discrete time problem equivalent to Equation (5′), formally shown in Lemma S.5, a dependence lemma for Lemma 2). So RHS is $\lim V_{dt}$ (a technical lemma Lemma S.8 guarantees existence of such limit).

I prove by showing inequality in two directions. The direction $V(\mu)\leqslant \lim V_{dt}(\mu)$ is trivial since $W_{dt}(\mu_t,\tau)\leqslant V_{dt}(\mu)$ for all $\langle\mu_t\rangle,\tau,dt$. The key is to prove the other direction $V(\mu)\geqslant \lim V_{dt}(\mu)$. I prove this claim by showing that $\forall dt>0$, there exists a continuous time strategy that achieves a payoff in Equation (1) no less than $V_{dt}(\mu)$.

Given time period dt, by Lemma 2 there exists discrete time optimal solution μ_t^* and τ^*, where $\mu_{t+1}^*|\mathcal{F}_t$ has support size N. The goal is to construct an admissible continuous-time belief process $\langle\mu_t\rangle$, which satisfies two properties: 1) at each discrete time idt, μ_t has exactly the same distribution as μ_i^*, 2) within each dt period, uncertainty reduction speed of μ_t is exactly $E[H(\mu_i^*)-H(\mu_{i+1}^*)|\mathcal{F}_i]/dt$. Such $\langle\mu_t\rangle$ with stopping time τ^* achieves higher payoff than $V_{dt}(\mu)$. Now this construction can be done by a technique introduced in Lemma S.3. $\forall i$ and conditional on $\mathcal{F}_i$, apply Lemma S.3 to the distribution of μ_{i+1}^* to smooth it on $[idt,(i+1)dt]$. Lemma S.3 states that there exists a continuous-time martingale $\langle\widetilde{\mu}_t\rangle$ (with a corresponding probability space) satisfying: $\forall s,t\in[0,1]$, $s>t$: $E[H(\mu_t)-H(\mu_s)|\mathcal{F}_t]=(s-t)E[H(\mu_i^*)-H(\mu_{i+1}^*)|\mathcal{F}_i]$. For $t\in[idt,(i+1)dt]$, define $\mu_t|\mathcal{F}_{idt}=\widetilde{\mu}_{\frac{t-idt}{dt}}|\mathcal{F}_i$. Therefore, $\forall t\in[idt,(i+1)dt)$:

$$\begin{aligned}
-\mathcal{L}_t H(\mu_t) &= \lim_{s\to t^+} E\left[\frac{H(\mu_t)-H(\mu_s)}{s-t}\middle|\mathcal{F}_t\right]\\
&= \lim_{s\to t^+} \frac{(s-t)E\big[H(\mu_i^*)-H(\mu_{i+1}^*)|\mathcal{F}_i\big]}{s-t}\\
&= H(\mu_i^*)-\sum p_i^j H(\mu_{i+1}^{*j})
\end{aligned}$$

Let $\tau=\tau^* dt$. It is easy to see that by construction τ is measurable to the natural filtration of μ_t. Therefore:

$$\begin{aligned}
V(\mu) &\geqslant E\left[e^{-\rho\tau}F(\mu_\tau)-\int_0^\tau e^{-\rho t}C(I_t)\mathrm{d}t\right]\\
&= E\left[e^{-\rho dt\cdot\tau^*}F(\mu_{\tau^*})-\sum_{t=0}^{\tau^*-1} C\left(\frac{H(\mu_i^*)-\sum p_i^j H(\mu_{i+1}^{*j})}{dt}\right)e^{-\rho dt\cdot t}\cdot\frac{1-e^{-\rho dt}}{\rho}\right]
\end{aligned}$$

$$\geqslant E\left[e^{-\rho dt\cdot\tau^*}F(\mu_{\tau^*})-\sum_{t=0}^{\tau^*-1}C\left(\frac{H(\mu_i^*)-\sum p_i^j H(\mu_{i+1}^{*j})}{dt}\right)e^{-\rho dt\cdot t}\cdot dt\right]$$
$$=E\left[e^{-\rho dt\cdot\tau^*}F(\mu_{\tau^*})-\sum_{t=0}^{\tau^*-1}C_{dt}\left(H(\mu_i^*)-\sum p_i^j H(\mu_{i+1}^{*j})\right)e^{-\rho dt\cdot t}\right]=V_{dt}(\mu)$$

Second inequality is from $1-e^{-x}\leqslant x$. Therefore, $V(\mu)\geqslant\lim V_{dt}(\mu)$. *Q.E.D.*

Remark B.2 (Non-integrable $\langle\mu_t\rangle$). In fact, the integrability requirement introduced in Equation (1) (defined as existence of $\lim W_{dt}$ in Section 5.1) is not necessary for my analysis of Theorem 1. Suppose now I extend the set of admissible belief profiles $\mathbb{M}$ to satisfy only the first two conditions: cadlag path, martingale property and initial value $\mu_0=\mu$. Then the limit of finite Riemann sum $W_{dt}(\mu_t,\tau)$ might not exist (although each finite Riemann sum is always well defined). Whenever this is the case, I define the payoff of strategy $\langle\mu_t\rangle,\tau$ as:

$$E\left[e^{-\rho\tau}F(\mu_\tau)-\int_0^\tau e^{-\rho t}C(-\mathcal{L}_t H(\mu_t))\mathrm{d}t\right]\triangleq\limsup_{dt\to0}W_{dt}(\mu_t,\tau) \tag{17}$$

Since $W_{dt}(\mu_t,\tau)$ is bounded above by $\max F$, Equation (17) is always well defined. Equation (17) is the essential upper-bound of payoff of an ill-behaved strategy, and when $\langle\mu_t\rangle$ is integrable it is consistent with the original definition of V. Obviously, such extension of admissible strategy set weakly increases the value of $V(\mu)$. Here I call the extended value function $\widehat{V}(\mu)=\sup_{\langle\mu_t\rangle,\tau}\limsup_{dt\to0}W_{dt}(\mu_t,\tau)$.

In the proof of Theorem 1, Lemmas 2 and 3 are not affected at all since they are about the discrete-time problem and corresponding value function V_{dt}. If Lemma 1 can be extended to $\widehat{V}(\mu)=\lim_{dt\to0}V_{dt}$, then Theorem 1 still holds with V replaced with $\widehat{V}$. This extension is quite trivial by observing $\forall\langle\mu_t\rangle,\tau,dt$, $W_{dt}(\mu_t,\tau)\leqslant V_{dt}(\mu)\implies\limsup W_{dt}(\mu_t,\tau)\leqslant\lim V_{dt}(\mu)\implies\widehat{V}(\mu)=\limsup\leqslant\lim V_{dt}(\mu)$.

To sum up, if we extend the admissible strategy set, and relax the definition of the objective function to its essential upper-bound, a solution to HJB Equation (4) still achieves the value function. Therefore, it is WLOO to eliminate all those ill-behaved strategies from the admissible control set.

B.2.2 *Proof of Lemma 2*

Remark B.3. The proof presented here is stronger than the statement of Lemma 2 in Section 5.2. It proves that the Bellman Equation (6) characterizes both Equations (5) and (5′) (while Lemma 2 only states that Equation (5) is characterized by Equation (6)). The first step of the proof shows that V_{dt} defined by Equations (5) and (5′) are identical (Lemmas S.4 and S.5), and can be rewritten as a recursive problem (Lemma S.6). To proof the Lemma 2 exactly stated in Section 5.2, one can simply skip Lemmas S.4 and S.5 and start with Lemma S.6, noticing that Equation (S.9) is simply rewriting Equation (5).

Proof. The proof of Lemma 2 is mostly the standard theory of discrete-time dynamic programming with a few tweaks. The proof involves 4 steps:

Step 1. Rewrite the sequential problem into the recursive problem. The technical details of the rewriting of problem is shown in Lemmas S.4, S.5 and S.6. The only non-standard analysis is to show that in Equation (5′), $\mathcal{S}_t$ may contain unused information/ randomness which can be discarded without loss of utility. Then the sequential problem without any redundant information can be represented in the belief space and easily written as a recursive problem.

Step 2. Verify the standard transversality condition. This is trivial as the payoff is bounded by $\max F$ and discounted exponentially.

Step 3. Verify the Blackwell contract mapping condition. The contraction parameter in Equation (6) is trivially the discount factor $e^{-\rho dt}$. The non-standard analysis is to show that the optimization operation is into the domain $C(\Delta X)$. To show this I invoke a maximum theorem in information design problems (theorem 5 of Zhong (2018a), it shows the existence of maximum as well).

Step 4. With steps 1-3, I invoke the standard contract-mapping fixed point theorem and show that value function V_{dt} is the unique solution to Equation (6). The final bits show that I can restrict the optimal strategy of Equation (6) to have support size N. This part is proved using a generalized concavification result: Notice that the objective function in Equation (6) is not in the standard "expected valuation" form as in the literature of information design (see Kamenica and Gentzkow (2011)). Instead, there is an extra $C_{dt}(\cdot)$ term. However, intuitively this problem can be handle using a Lagrange method and take the term inside $C_{dt}(\cdot)$ to combine it with $E[V]$ linearly. This intuition is formalized by Theorem S.1, which is a corollary of a more general result in Zhong (2018a). *Q.E.D.*

B.2.3 *Proof of Lemma 3*

Before going to the proof of Lemma 3, I first formally rewrite the problem to accommodate viscosity solutions (see Crandall, Ishii, and Lions (1992)). First define a space of functions on $\Delta(X)$:

$$\mathcal{L}=\left\{V:\Delta(X)\mapsto\mathbb{R}^+\middle|\forall\mu\in\Delta X,\mu'\in\Delta(\text{supp}(\mu)),\ \limsup_{\mu'\to\mu}\frac{|V(\mu')-V(\mu)|}{\|\mu'-\mu\|}\in\mathbb{R}\right\}$$

where $\|\cdot\|$ is Euclidean norm on ΔX. By definition, $\mathcal{L}$ is the set of pointwise Lipschitz functions on $\Delta(X)$. Two technical lemmas Lemmas S.8 and S.9 guarantee that $\lim V_{dt}$ is well defined, and there exists $\overline{V}\in\mathcal{L}$ which is the uniform limit of V_{dt}. Now I show that $\overline{V}$ coincides with the solution of the HJB equation. Consider the following HJB equation defined on $\mathcal{L}$:

$$\rho V(\mu)=\max\left\{\rho F(\mu),\sup_{\substack{\nu_i\in\Delta(\text{supp}(\mu)),\\ p_i\in\mathbb{R}^+,\\ \widehat{\sigma}\in\mathbb{R}^{|\text{supp}(\mu)|}}}\sum p_i(V(\nu_i)-V(\mu))-DV\left(\mu,\sum p_i\nu_i-\mu\right)\left\|\left(\sum p_i\nu_i-\mu\right)\right\|+\frac{1}{2}\|\widehat{\sigma}\|^2D^2V(\mu,\widehat{\sigma})\right.$$
$$\left.-C\left(-\sum p_i(H(\nu_i)-H(\mu)-\nabla H(\mu)\cdot(\nu_i-\mu))-\frac{1}{2}\widehat{\sigma}^T\cdot\text{H}H(\mu)\cdot\widehat{\sigma}\right)\right\}\tag{18}$$

∇ and H denote gradient and Hessian operator (well-defined on all interior points). Since $\overline{V}$ is not necessarily differentiable, I use operator D and D^2 to replace the Jacobian and Hessian operators on $\overline{V}$. D and D^2 are defined as follows. $\forall y\in B^{|\text{supp}(x)|-1}$ (Unit ball in $|\text{supp}(x)|-1$ dimensional space):

Definition 2 (General differentials)**.** *$\forall f\in\mathcal{L}$:*

$$\begin{cases}Df(x,y)=\liminf_{\delta\to0}\frac{f(x)-f(x-\delta y)}{\delta\|y\|}\\ D^2f(x,y)=\limsup_{\delta\to0}2\frac{f(x+\delta y)-f(x)-\delta\cdot Df(x,y)\cdot\|y\|}{\delta\|y\|^2}\end{cases}$$

Notice that if $f\in C^{(1)}(\Delta X)$, then $Df(x,y)=\frac{\nabla f(x)\cdot y}{\|y\|}$. If $f\in C^{(2)}(\Delta X)$ then $D^2f(x,y)=\frac{y^T\cdot\text{H}f(x)\cdot y}{\|y\|^2}$. It is not hard to verify that for $C^{(1)}$ smooth value function $V(\mu)$, Equation (18) is equivalent to Equation (4).

Proof.

Consider Lemma 3 by replacing Equation (4) with Equation (18). If the statement is proved with Equation (18), then since $\overline{V}=V$ is $C^{(1)}$ smooth, $\overline{V}$ is smooth and Equation (4) automatically holds. I prove by induction on dimensionality of supp(μ). First of all, Lemma 3 is trivially true when $\mu=\delta_x$ since $V(\mu)=\overline{V}(\mu)=F(\mu)$ when the state is deterministic. Now it is sufficient to prove $\overline{V}=V$ on interior of ΔX conditional on $\overline{V}=V$ being true on $\partial\Delta X$ (boundary of ΔX).

The proof takes three steps. Before going to the details, I introduce the steps briefly. The first step is to show that $\overline{V}$ is unimprovable in HJB Equation (18). The proof is quite standard as any continuous-time strategy that improves $\overline{V}$ can be approximated by a discrete-time strategy. The second step shows $\overline{V}\geqslant V$. Proof is by a standard contradiction argument. If $\overline{V}<V$, then there exists a belief s.t. the same strategy implements strictly higher HJB with $\overline{V}$, which violates unimprovability. The last and most difficult step is to show that $V\geqslant\overline{V}$.

Unimprovability: First I show that $\overline{V}$ is unimprovable in Equation (18). Suppose for the sake of contradiction that $\overline{V}$ is improvable at interior μ, then there exists $p_i,\nu_i,\widehat{\sigma},I$ such that:

$$\rho\overline{V}(\mu)<\sum p_i\big(\overline{V}(\nu_i)-\overline{V}(\mu)\big)-D\overline{V}(\mu,\mu-\sum p_i\nu_i)\left\|\sum p_i\nu_i-\mu\right\|+\sum D^2\overline{V}(\mu,\widehat{\sigma}_j)\|\widehat{\sigma}\|_j^2-C(I)$$
$$\text{where }I=-\sum p_i(H(\nu_i)-H(\mu)-\nabla H(\mu)\cdot(\nu_i-\mu))-\sum\widehat{\sigma}_j^T\text{H}H(\mu)\widehat{\sigma}_j$$

Then if we compare the following two ratios:

$$\frac{\sum p_i\big(\overline{V}(\nu_i)-\overline{V}(\mu)\big)-D\overline{V}(\mu,\sum p_i\nu_i-\mu)\|\sum p_i\nu_i-\mu\|}{-\sum p_i(H(\nu_i)-H(\mu)-\nabla H(\mu)\cdot(\nu_i-\mu))};\ \frac{D^2\overline{V}(\mu,\widehat{\sigma})\|\widehat{\sigma}\|^2}{-\widehat{\sigma}^T\text{H}H(\mu)\widehat{\sigma}}$$

At least one of them must be larger than $\frac{\rho\overline{V}(\mu)+C(I)}{I}$.

- *Case 1:*

$$\frac{\sum p_i\big(\overline{V}(\nu_i)-\overline{V}(\mu)\big)-D\overline{V}(\mu,\sum p_i\nu_i-\mu)\|\sum p_i\nu_i-\mu\|}{-\sum p_i(H(\nu_i)-H(\mu)-\nabla H(\mu)\cdot(\nu_i-\mu))}>\frac{\rho}{I}\overline{V}(\mu)+\frac{C(I)}{I}$$

By Definition 2, there exists $\delta,\varepsilon>0$ s.t. :

$$\frac{\sum p_i\left(\overline{V}(\nu_i)-\overline{V}(\mu)\right)-\frac{\overline{V}(\mu)-\overline{V}(\mu-\delta(\sum p_i\nu_i-\mu))}{\delta}}{\sum p_i(H(\mu)-H(\nu_i))+\frac{H(\mu)-H(\mu-\delta(\sum p_i\nu_i-\mu))}{\delta}}\geqslant\frac{\rho}{I}\overline{V}(\mu)+\frac{C(I)}{I}+\varepsilon \tag{19}$$

where δ is sufficiently small that $\mu_0=\mu-\delta(\sum p_i\nu_i-\mu)\in\Delta X^{\mathrm{o}}$. Then by construction, if we assume:

$$\begin{cases} p_0'=\frac{1}{1+\delta}\\ p_i'=\frac{\delta}{1+\delta}p_i\end{cases}$$

Then $\left(p_i',\nu_i'\right)$ is Bayesian plausible:

$$\begin{cases} \sum p_i'=1\\ \sum p_i'\nu_i=\mu\end{cases}$$

where 0 is also included in indices i's. Replacing terms in Equation (19) and let $I(\nu_i|\mu)=H(\mu)-\sum p_i'H(\nu_i)$:

$$\frac{\sum p_i'\overline{V}(\nu_i)-\overline{V}(\mu)}{-\sum p_i'H(\nu_i)+H(\mu)}\geqslant\frac{\rho}{I}\overline{V}(\mu)+\frac{C(I)}{I}+\varepsilon$$

$$\Longrightarrow\sum p_i'\overline{V}(\nu_i)-\frac{I(\nu_i|\mu)}{I}C(I)\geqslant\left(1+\rho\frac{I(\nu_i|\mu)}{I}\right)\overline{V}(\mu)+\varepsilon I(\nu_i|\mu) \tag{20}$$

It is easy to verify that $I(\nu_i|\mu)$ is continuous in δ and it is zero when $\delta=0$. So δ can be chosen sufficiently small that

$$e^{\rho\frac{I(\nu_i|\mu)}{I}}-\left(1+\rho\frac{I(\nu_i|\mu)}{I}\right)=\sum_{k=1}^{\infty}\frac{1}{(k+1)!}\left(\frac{\rho}{I}\right)^{k+1}I(\nu_i|\mu)^k\cdot I(\nu_i|\mu)\leqslant\frac{\varepsilon I(\nu_i|\mu)}{4\sup F} \tag{21}$$

The equality is from Taylor expansion of exponential function. Plug Equation (21) into Equation (20):

$$\sum p_i'\overline{V}(\nu_i)-\frac{I(\nu_i|\mu)}{I}C(I)\geqslant e^{\rho\frac{I(\nu_i|\mu)}{I}}\overline{V}(\mu)+\frac{\varepsilon}{4}I(\nu_i|\mu)$$

$$\Longrightarrow e^{-\rho\frac{I(\nu_i|\mu)}{I}}\left(\sum p_i'\overline{V}(\nu_i)\right)-\frac{I(\nu_i|\mu)}{I}C(I)\geqslant\overline{V}(\mu)+e^{-\rho\frac{I(\nu_i|\mu)}{I}}\frac{\varepsilon I(\nu_i|\mu)}{4}-\left(1-e^{-\rho\frac{I(\nu_i|\mu)}{I}}\right)\frac{I(\nu_i|\mu)}{I}C(I) \tag{22}$$

Noticing that $\left(1-e^{-\rho\frac{I(\nu_i|\mu)}{I}}\right)I(\nu_i|\mu)$ is a second order small term. Then we can pick δ such that Equation (22) implies:

$$e^{-\rho\frac{I(\nu_i|\mu)}{I}}\left(\sum p_i'\overline{V}(\nu_i)\right)-\frac{I(\nu_i|\mu)}{I}C(I)\geqslant\overline{V}(\mu)+\frac{\varepsilon}{8}I(\nu_i|\mu)$$

From now on, we fix ε and δ. Pick $dt=\frac{I(\nu_i|\mu)}{I}$, $dt_m=\frac{dt}{m}$. By uniform convergence, there exists N s.t. $\forall m\geqslant N$:

$$e^{-\rho dt}\left(\sum p_i'V_{dt_m}(\nu_i)\right)-dt\cdot C\left(\frac{I(\nu_i|\mu)/m}{dt_m}\right)>V_{dt_m}(\mu)$$

$$\Longrightarrow e^{-\rho m dt_m}\left(\sum p_i'V_{dt_m}(\nu_i)\right)-\sum_{\tau=0}^{m-1}e^{-\rho\tau dt_m}C_{dt_m}\left(\frac{I(\nu_i|\mu)}{m}\right)>V_{dt_m}(\mu)$$

That is to say we find a feasible experiment, whose cost can be spread into m periods (the split of experiment is done by applying Lemma S.3). This experiment strictly dominates the optimal experiment at μ for discrete time problem with dt_m. Contradiction. Therefore, $\overline{V}$ must be unimprovable at μ.

- *Case 2*:

$$\frac{D^2\overline{V}(\mu,\widehat{\sigma})\|\widehat{\sigma}\|^2}{-\widehat{\sigma}^T\mathrm{H}H(\mu)\widehat{\sigma}}>\frac{\rho}{I}\overline{V}(\mu)+\frac{C(I)}{I}$$

Then by the definition of operator D^2 in Definition 2, there exists $\widehat{\sigma}$, $\delta,\varepsilon>0$ s.t.:

$$\frac{\overline{V}(\mu+\delta\widehat{\sigma})-\overline{V}(\mu)-\delta D\overline{V}(\mu,\widehat{\sigma})\|\widehat{\sigma}\|}{-H(\mu+\delta\widehat{\sigma})+H(\mu)+\delta\nabla H(\mu)\cdot\widehat{\sigma}}\geqslant\frac{\rho}{I}\overline{V}(\mu)+\frac{C(I)}{I}+2\varepsilon$$

Then by the definition of operator D in Definition 2, there exists δ' s.t.:

$$\frac{\overline{V}(\mu+\delta\widehat{\sigma})-\overline{V}(\mu)-\delta\frac{\overline{V}(\mu)-\overline{V}(\mu-\delta'\widehat{\sigma})}{\delta'}}{-H(\mu+\delta\widehat{\sigma})+H(\mu)+\delta\frac{H(\mu)-H(\mu-\delta'\widehat{\sigma})}{\delta'}}\geqslant\frac{\rho}{I}\overline{V}(\mu)+\frac{C(I)}{I}+\varepsilon$$

Let $\mu_1=\mu-\delta'\hat{\delta}$ and $\mu_2=\mu+\delta\hat{\sigma}$, $p_1=\frac{\delta'}{\delta+\delta'},p_2=\frac{\delta}{\delta+\delta'}$, then:

$$\sum p_i\overline{V}(\nu_i)\geqslant\left(1+\rho\frac{I(\nu_i|\mu)}{I}\right)\overline{V}(\mu)+\frac{I(\nu_i|\mu)}{I}C(I)+\varepsilon I(\nu_i|\mu) \tag{23}$$

Noticing that Equation (23) is exactly the same as Equation (20) in Case 1. Then using same argument, This case is also ruled out.

Equality: I show that $\forall$ smooth function V solving Equation (18), $\overline{V}=V$. Notice that this automatically proves the uniqueness of solution of Equation (18). I prove inequality from both directions for $\mu\in\Delta(X)^{o}$:

- $\overline{V}(\mu)\geqslant V(\mu)$: Suppose not, then consider $U(\mu)=\overline{V}(\mu)-V(\mu)$. Since both V and $\overline{V}$ are continuous, U is continuous. Therefore $\arg\min U$ is non empty and $\min U<0$ according to our assumption. Choose $\mu\in\arg\min U$ ($\mu\in\Delta X^o$ since $V=\overline{V}$ on boundary). Since $\overline{V}(\mu)\geqslant F(\mu)$, $V(\mu)>F(\mu)$. Let $(p_i,\nu_i,\hat{\sigma})$ be a strategy solving $V(\mu)$:

$$\begin{aligned}\rho V(\mu)=&\sum p_i(V(\nu_i)-V(\mu))-DV\left(\mu,\sum p_i\nu_i-\mu\right)\left\|\sum p_i(\nu_i-\mu)\right\|+\frac{1}{2}D^2V(\mu,\hat{\sigma})\|\hat{\sigma}\|^2\\&-C\left(-\sum p_i(H(\nu_i)-H(\mu)-\nabla H(\mu)(\nu_i-\mu))-\frac{1}{2}\hat{\sigma}^T\mathrm{H}H(\mu)\hat{\sigma}\right)\end{aligned} \tag{24}$$

  Now compare $D\overline{V}$ and DV:

$$\begin{aligned}&\frac{\overline{V}(\mu)-\overline{V}(\mu')}{\|\mu-\mu'\|}=\frac{V(\mu)-V(\mu')+U(\mu)-U(\mu')}{\|\mu-\mu'\|}\leqslant\frac{V(\mu)-V(\mu')}{\|\mu-\mu'\|}\\ \Longrightarrow&\liminf\frac{\overline{V}(\mu)-\overline{V}(\mu')}{\|\mu-\mu'\|}\leqslant\lim\frac{V(\mu)-V(\mu')}{\|\mu-\mu'\|}\\ \Longrightarrow& D\overline{V}(\mu,\mu'-\mu)\left\|\mu'-\mu\right\|\leqslant\nabla V(\mu)\cdot(\mu'-\mu)\end{aligned}$$

  Compare $D^2\overline{V}$ and D^2V:

$$\begin{aligned}&\frac{\overline{V}(\mu')-\overline{V}(\mu)-D\overline{V}(\mu,\mu'-\mu)\|\mu'-\mu\|}{\|\mu'-\mu\|^2}\geqslant\frac{V(\mu')-V(\mu)-\nabla V(\mu)\cdot(\mu'-\mu)+U(\mu')-U(\mu)}{\|\mu-\mu'\|^2}\\ &\Longrightarrow D^2\overline{V}(\mu,\hat{\sigma})\geqslant D^2V(\mu,\hat{\sigma})\end{aligned}$$

  Therefore Equation (24) implies:

$$\begin{aligned}\rho V(\mu)\leqslant&\sum p_i\left(\overline{V}(\nu_i)-\overline{V}(\mu)-(U(\nu_i)-U(\mu))\right)\\&-D\overline{V}(\mu,\sum\nu_i-\mu)\left\|\sum\nu_i-\mu\right\|+\frac{1}{2}D^2\overline{V}(\mu,\hat{\sigma})\|\hat{\sigma}\|^2\\&-C\left(-\sum p_i(H(\nu_i)-H(\mu)+\nabla H(\mu)(\nu_i-\mu))-\frac{1}{2}\hat{\sigma}^T\mathrm{H}H(\mu)\hat{\sigma}\right)\\ \leqslant&\rho\overline{V}(\mu)\end{aligned}$$

  The first inequality comes from replacing DV and D^2V with $D\overline{V}$ and $D^2\overline{V}$. The second inequality comes from $U(\nu_i)-U(\mu)\geqslant 0$ and unimprovability of $\overline{V}$. Contradiction.

- $V(\mu)\geqslant\overline{V}(\mu)$: I prove by showing that $\forall dt>0$, $V\geqslant V_{dt}$. Suppose not, then there exists μ',dt s.t. $V_{dt}(\mu')>V(\mu')$. Let $dt_n=\frac{dt}{2^n}$. Since V_{dt_n} is increasing in n, there exists $\varepsilon>0$ s.t. $V_{dt_n}(\mu')-V(\mu')\geqslant\varepsilon\ \forall n\in\mathbb{N}$. Now consider $U_n=V-V_{dt_n}$. U_n is continuous by Lemma 2 and $U_n(\mu')\leqslant-\varepsilon$. Pick $\mu^n\in\arg\min U_n$. Since $\Delta(X)$ is compact, there exists a converging sequence $\lim\mu^n=\mu$. By assumption, $U_n(\mu^n)\leqslant-\varepsilon$, therefore since $U(\mu)=\lim U_n(\mu^n)\leqslant-\varepsilon$, μ must be in interior of $\Delta(X)$. So without loss, μ^n can be picked that $\mu^n\in\Delta(X)^O$. Now consider the optimal strategy of discrete time problem:

$$\begin{cases}V_{dt_n}(\mu^n)=e^{-\rho dt_n}\sum p_i^n V_{dt_n}(\nu_i^n)-dt_nC(I_n)\\ \sum p_i^n(H(\mu^n)-H(\nu_i^n))=I_ndt_n\\ \sum p_i^n\nu_i^n=\mu^n;\ \sum p_i^n=1\end{cases}$$

  By definition of $U_n(\mu)$:

$$\begin{aligned}\sum p_i^n(V(\nu_i^n)-V(\mu^n))=&\sum p_i^n\left(V_{dt_n}(\nu_i^n)-V_{dt_n}(\mu^n)-U_n(\mu^n)+U(\nu_i^n)\right)\\ \geqslant&\sum p_i^n\left(V_{dt_n}(\nu_i^n)-V_{dt_n}(\mu^n)\right)\end{aligned}$$

$$
\begin{aligned}
&=\left(e^{\rho dt_n}-1\right)V_{dt_n}(\mu^n)+e^{\rho dt_n}dt_nC(I_n)\\
&\geqslant \rho dt_nV_{dt_n}(\mu^n)+e^{\rho dt_n}dt_nC(I_n)\\
&\geqslant \rho dt_n\varepsilon+\rho dt_nV(\mu^n)+e^{\rho dt_n}dt_nC(I_n)\\
\Longrightarrow \rho V(\mu_n)&\leqslant -\rho\varepsilon+\sum\frac{p_i^n}{dt_n}(V(\nu_i^n)-V(\mu^n))-e^{\rho dt_n}C(I_n)\\
\Longrightarrow \rho V(\mu_n)&\leqslant -\rho\varepsilon+\sum\frac{p_i^n}{dt_n}(V(\nu_i^n)-V(\mu^n))-C(I_n)
\end{aligned}
\tag{25}
$$

The first equality is by the definition of U_n. The first inequality is from $\mu^n\in\arg\min U_n$. The second inequality is from $e^x-1\geqslant x$. The third inequality is from $U_n(\mu^n)\leqslant-\varepsilon$. Now since the number of posteriors ν_i^n is no more than $2|X|$, we can take a subsequence of n such that all $\lim\nu_i^n=\nu_i$. Partition ν_i^n into two kinds: $\lim\nu_i^n=\nu_i\neq\mu$, $\lim\nu_j^n=\mu$. Since V is unimprovable, $\forall c,\widehat{\sigma}$ we have $D^2V(\mu,\widehat{\sigma})\|\widehat{\sigma}\|^2\leqslant-\widehat{\sigma}^T\mathrm{H}H(\mu)\widehat{\sigma}\left(\frac{\rho}{I}V(\mu)+\frac{C(I)}{I}\right)$. Since $V\in C^{(1)}$, $H\in C^{(2)}$, $\forall\eta$, there exists δ s.t. $\forall|\mu'-\mu|\leqslant\delta$:

$$
\begin{cases}\|\mathrm{H}H(\mu)-\mathrm{H}H(\mu')\|\leqslant\eta\\ |V(\mu)-V(\mu')|\leqslant\eta\end{cases}
$$

$$
\begin{aligned}
\Longrightarrow D^2V(\mu',\widehat{\sigma})&\leqslant\left(\frac{\rho}{I}V(\mu')+\frac{C(I)}{I}\right)\left(-\frac{\widehat{\sigma}^T\mathrm{H}H(\mu')\widehat{\sigma}}{\|\widehat{\sigma}\|^2}\right)\\
&\leqslant\left(\frac{\rho}{I}V(\mu)+\frac{C(I)}{I}\right)\left(-\frac{\widehat{\sigma}^T\mathrm{H}H(\mu)\widehat{\sigma}}{\|\widehat{\sigma}\|^2}\right)+\left(\frac{\rho}{I}\sup F+\frac{C(I)}{I}\right)\eta+\frac{\rho}{I}\eta\|\mathrm{H}H(\mu)\|
\end{aligned}
$$

If we pick η and δ properly:

$$
D^2V(\mu',\widehat{\sigma})\leqslant\left(\frac{\rho}{I}V(\mu)+\frac{C(I)}{I}\right)\left(-\frac{\widehat{\sigma}^T\mathrm{H}H(\mu)\widehat{\sigma}}{\|\widehat{\sigma}\|^2}\right)+\frac{1+C(I)}{I}\eta
$$

Then there exists N s.t. $\forall n\geqslant N$, $\left|\nu_j^n-\mu\right|<\delta$, $|\mu^n-\mu|<\delta$. Now I want to do a second-order approximation of $V(\nu_j^n)-V(\mu^n)-\nabla V(\mu^n)(\nu_j^n-\mu^n)$. To apply Taylor expansion to a not necessarily twice differentiable function V, I invoke a technical Lemma S.10 to $g(\alpha)=V(\alpha\nu_j^n+(1-\alpha)\mu^n)$:

$$
\begin{aligned}
&V(\nu_j^n)-V(\mu^n)-\nabla V(\mu^n)(\nu_j^n-\mu^n)=g(1)-g(0)-g'(0)\\
\leqslant&\frac{1}{2}\sup_{\alpha\in(0,1)}D^2g(\alpha,1)=\sup_{\alpha\in(0,1)}\limsup_{d\to0}\frac{g(\alpha+d)-g(\alpha)-g'(\alpha)d}{d^2}\\
=&\sup_{\xi\in(\mu^n,\nu_j^n)}\limsup_{d\to0}\frac{V(\xi+d(\nu_j^n)-\mu^n)-V(\xi)-dJV(\xi)(\nu_j^n-\mu^n)}{d^2}\\
\leqslant&\frac{1}{2}\sup_{|\xi-\mu|\leqslant\delta}D^2V(\xi,\nu_j^n-\mu^n)\left\|\nu_j^n-\mu^n\right\|^2\\
\leqslant&-\frac{1}{2}\left(\frac{\rho}{I}V(\mu)+\frac{C(I)}{I}\right)(\nu_j^n-\mu^n)^T\mathrm{H}H(\mu)(\nu_j^n-\mu^n)+\frac{1+C(I)}{2I}\eta\left\|\nu_j^n-\mu^n\right\|^2
\end{aligned}
\tag{26}
$$

Therefore, by applying Equation (26):

$$
\begin{aligned}
&\sum p_{i,j}^n\left(V(\nu_{i,j}^n)-V(\mu^n)\right)\\
=&\sum p_i^n(V(\nu_i^n)-V(\mu^n)-\nabla V(\mu^n)(\nu_i^n-\mu^n))+\sum p_j^n\left(V(\nu_j^n)-V(\mu^n)-\nabla V(\mu^n)\left(\nu_j^n-\mu^n\right)\right)\\
\leqslant&\sum p_i^n(V(\nu_i^n)-V(\mu^n)-\nabla V(\mu^n)(\nu_i^n-\mu^n))\\
&-\frac{1}{2}\left(\frac{\rho}{I}V(\mu)+\frac{C(I)}{I}\right)\sum p_j^n\left(\nu_j^n-\mu^n)^T\mathrm{H}H(\mu)(\nu_j^n-\mu^n)\right)+\frac{1+C(I)}{2I}\eta\sum p_j^n\left\|\nu_j^n-\mu^n\right\|^2
\end{aligned}
\tag{27}
$$

Notice that Equations (26) and (27) are true uniform to I, so we can replace I with I_n and Equation (27) is still true. Now let $\bar{p}_i^n=\frac{p_i^n}{dt_n}$, $-\widehat{\sigma}_n^T\mathrm{H}H(\mu^n)\widehat{\sigma}^n dt_n=\sum p_j^n\left(H(\mu^n)-H(\nu_j^n)+\nabla H(\mu)(\nu_j^n-\mu^n)\right)$, we have:

$$\sum\bar{p}_i^n\left(H(\mu^n)-H(\nu_i^n)+H'(\mu^n)(\nu_i^n-\mu^n)\right)-\widehat{\sigma}_n^T\mathrm{H}H(\mu^n)\widehat{\sigma}_n=I_n \tag{28}$$

$\left(\bar{p}_i^n,\nu_i^n,\widehat{\sigma}_n\right)$ is a feasible experiment for Equation (18). Therefore, by optimality of V at μ^n, we have

$$\begin{cases}\sum\bar{p}_i^n(V(\nu_i^n)-V(\mu^n)-\nabla V(\mu^n)(\nu_i^n-\mu^n))\leqslant\left(I_n+\widehat{\sigma}^{nT}\mathrm{H}H(\mu^n)\widehat{\sigma}^n\right)\left(\frac{\rho}{I_n}V(\mu^n)+\frac{C(I_n)}{I_n}\right)\\ D^2V(\mu^n,\widehat{\sigma}^n)\leqslant-\frac{\widehat{\sigma}^{nT}\mathrm{H}H(\mu)\widehat{\sigma}^n}{\|\widehat{\sigma}^n\|^2}\left(\frac{\rho}{I_n}\overline{V}(\mu^n)+\frac{C(I_n)}{I_n}\right)\end{cases} \tag{29}$$

Then we study term $\sum p_j^n(\nu_j^n-\mu^n)^2$. Apply Lemma S.10 to $g(\alpha)=H(\alpha\nu_j^n+(1-\alpha)\mu^n)$:

$$\begin{aligned}&\sum p_j^n\left(H(\mu^n)-H(\nu_j^n)+\nabla H(\mu^n)(\nu_j^n-\mu^n)\right)\\ \geqslant&\frac{1}{2}\inf_{\xi_j^n\in[\mu^n,\nu_n^j]}\sum p_j^n\left(-(\nu_j^n-\mu^n)^T\mathrm{H}H(\xi_j^n)(\nu_j^n-\mu^n)\right)\\ \geqslant&-\frac{1}{2}\sum p_j^n((\nu_j^n-\mu^n)^T\mathrm{H}H(\mu)(\nu_j^n-\mu^n))-\frac{1}{2}\eta\sum p_j^n\left\|\nu_j^n-\mu^n\right\|^2\end{aligned} \tag{30}$$

Therefore, to sum up:

$$\begin{aligned}\sum\frac{p_{i,j}^n}{dt_n}\left(V(\nu_{i,j}^n)-V(\mu^n)\right)\leqslant&\sum\bar{p}_i^n(V(\nu_i^n)-V(\mu^n)-\nabla V(\mu^n)(\nu_i^n-\mu^n))\\&+\frac{1}{2}\sum\frac{p_j^n}{dt_n}\left(-(\nu_j^n-\mu^n)^T\mathrm{H}H(\mu)(\nu_j^n-\mu^n)\left(\frac{\rho}{I_n}V(\mu)+\frac{C(I_n)}{I_n}\right)\right)\\&+\sum\frac{p_j^n}{dt_n}\left(\frac{1+C(I_n)}{2I_n}\eta\left\|\nu_j^n-\mu^n\right\|^2\right)\\ \leqslant&\left(I_n+\widehat{\sigma}^{nT}\mathrm{H}H(\mu^n)\widehat{\sigma}^n\right)\left(\frac{\rho}{I_n}V(\mu^n)+\frac{C(I_n)}{I_n}\right)\\&+\left(\sum\frac{p_j^n}{dt_n}(H(\mu^n)-H(\nu_j^n)+\nabla H(\mu^n)(\nu_j^n-\mu^n))\right.\\&\left.+\frac{1}{dt_n}\frac{\eta}{2}\sum p_j^n\left\|\nu_j^n-\mu^n\right\|^2\right)\left(\frac{\rho}{I_n}V(\mu)+\frac{C(I_n)}{I_n}\right)\\&+\frac{1}{dt_n}\sum p_j^n\left\|\nu_j^n-\mu^n\right\|^2\frac{1+C(I_n)}{2I_n}\eta\\ =&\left(I_n+\widehat{\sigma}^{nT}\mathrm{H}H(\mu^n)\widehat{\sigma}^n\right)\left(\frac{\rho}{I_n}V(\mu^n)+\frac{C(I_n)}{I_n}\right)\\&+\left(-\widehat{\sigma}^{nT}\mathrm{H}H(\mu^n)\widehat{\sigma}^n+\frac{1}{dt_n}\frac{\eta}{2}\sum p_j^n\left\|\nu_j^n-\mu^n\right\|^2\right)\left(\frac{\rho}{I_n}V(\mu)+\frac{C(I_n)}{I_n}\right)\\&+\frac{1}{dt_n}\sum p_j^n\left\|\nu_j^n-\mu^n\right\|^2\frac{1+C(I_n)}{2I_n}\eta\\ \leqslant&\rho V(\mu^n)+C(I_n)+\frac{1}{dt_n}\sum p_j^n\left\|\nu_j^n-\mu^n\right\|^2\left(\frac{1+\rho V(\mu)+2C(I_n)}{2I_n}\right)\eta+\rho\eta\end{aligned}$$

The first inequality is Equation (27). The second inequality comes from Equation (29) and Equation (30). The next equality comes from definition of $\widehat{\sigma}_n^2$. The last inequality comes from canceling out terms and $-\widehat{\sigma}^{nT}\mathrm{H}H(\mu^n)\widehat{\sigma}^n\leqslant I_n$ (Notice the difference between $V(\mu)$ and $V(\mu^n)$). Then by plug into Equation (25):

$$\rho V(\mu^n)\leqslant-\rho\varepsilon+\rho V(\mu^n)+\frac{1}{dt_n}\sum p_j^n\left\|\nu_j^n-\mu^n\right\|^2\left(\frac{1+\rho V(\mu)+2C(I_n)}{2I_n}\right)\eta+\rho\eta$$

Moreover:

$$\sum p_j^n\left\|\nu_j^n-\mu^n\right\|^2\inf_{\sigma}\frac{|\sigma^T\mathrm{H}H(\mu)\sigma|}{\|\sigma\|^2}$$

$$\leqslant \sum p_j^n (\nu_j^n - \mu^n) \mathrm{H}H(\mu)(\mu^n - \mu^n) \leqslant I_n dt_n + \eta \sum p_j^n \left\| \nu_j^n - \mu^n \right\|^2$$

$$\Longrightarrow \sum p_j^n \left\| \nu_j^n - \mu^n \right\|^2 \leqslant \frac{I_n dt_n}{\inf_\sigma \frac{|\sigma^T \mathrm{H}H(\mu)\sigma|}{\|\sigma\|^2} - \eta}$$

$$\Longrightarrow \rho\varepsilon \leqslant \frac{1}{2}(1 + \rho V(\mu) + 2C(I_n)) \frac{\eta}{\inf_\sigma \frac{|\sigma^T \mathrm{H}H(\mu)\sigma|}{\|\sigma\|^2} - \eta} + \rho\eta$$

By Lemma S.7, $C(I_n)$ is uniformly bounded above. Since H is strictly concave $\inf_\sigma \frac{|\sigma^T \mathrm{H}H(\mu)\sigma|}{\|\sigma\|^2}$ is positive. The inequality holds when η is chosen smaller than $\inf_\sigma \frac{|\sigma^T \mathrm{H}H(\mu)\sigma|}{\|\sigma\|^2}$. By taking $\eta \to 0$, the LHS is eventually larger than the RHS. Contradiction. Therefore:

$$V(\mu) = \limsup_{dt \to 0} V_{dt}(\mu) = \overline{V}(\mu)$$

Q.E.D.

B.3 Proof of Theorem 2

Proof. I prove Theorem 2 by guess and verification. To simplify notation, I define a flow version of information measure:

$$J(\mu, \nu) = H(\mu) - H(\nu) + H'(\mu)(\nu - \mu)$$

Then total flow information cost is $p \cdot J(\mu,\nu)$. Let $F_m = E_\mu[u(a_m, x)]$ and reorder a_m s.t. F'_m is increasing in m. Let $\underline{\mu}_k$ be each kink points of F: $F(\mu) = F_k(\mu) \Longleftrightarrow \mu \in \left[\underline{\mu}_{k-1}, \underline{\mu}_k\right]$. $\overline{m}$ is the smallest index s.t. $F'_{\overline{m}} \geqslant 0$.

Algorithm:

In this part, I introduce the algorithm for constructing $V(\mu)$ and $\nu(\mu)$. I only discuss the case $\mu \geqslant \mu^*$. The remaining case $\mu \leqslant \mu^*$ follows by a symmetric method. The main steps are illustrated in Figure 22. The first step is to find critical the belief μ^* at which two sided stationary Poisson signal is optimal ($\mu^* = 0.5$ in a symmetric problem). Then value function is solved by searching over optimal posterior beliefs, given choosing an action (say a_m). Then the remaining actions are added one by one to consideration. And value function is updated when each additional action is added. Finally, after all actions have been considered, I complete the construction of value function.

- *Step 1*: Define:

$$\overline{V}^+(\mu) = \max_{\nu \geqslant \mu} \frac{F_m(\nu)}{1 + \frac{\rho}{c} J(\mu,\nu)}$$

$$\overline{V}^-(\mu) = \max_{\nu \leqslant \mu} \frac{F_m(\nu)}{1 + \frac{\rho}{c} J(\mu,\nu)}$$

  In Lemma B.1 I analyze the technical details of $\overline{V}^+$ and $\overline{V}^-$. The main property is that: $\overline{V}^+$ is increasing and $\overline{V}^-$ is decreasing. There exists $\mu^* \in [0,1]$ s.t. $\overline{V}^+(\mu) \geqslant \overline{V}^-(\mu)$ when $\mu \geqslant \mu^*$ and $\overline{V}^-(\mu) \leqslant \overline{V}^-(\mu)$ when $\mu \leqslant \mu^*$. Define $\overline{V}(\mu) = \max\{\overline{V}^+(\mu), \overline{V}^-(\mu)\}$.

- *Step 2:* I construct the first piece of $V(\mu)$ to the right of μ^*. There are three possible cases of μ^* to be discussed (I omitted $\mu^* = 1$ by symmetry).

  Case 1: Suppose $\mu^* \in (0,1)$ and $\overline{V}(\mu^*) > F(\mu^*)$. Then, there exists m and $\nu(\mu^*) \in (\mu^*, 1)$ s.t.

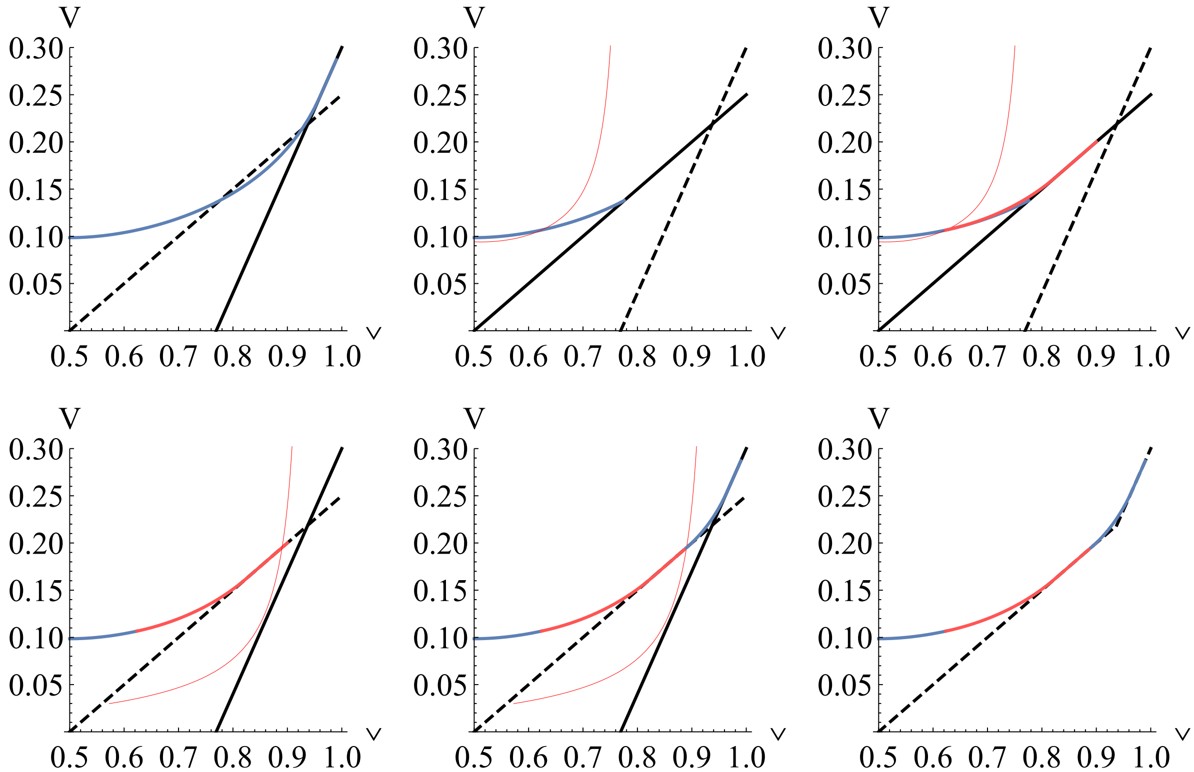
$$\overline{V}(\mu^*) = \frac{F_m(\nu(\mu^*))}{1 + \frac{\rho}{c} J(\mu^*, \nu(\mu^*))}$$

  Initial condition $\left(\mu_0 = \mu^*, V_0 = \overline{V}(\mu^*), V'_0 = 0\right)$ satisfies Lemma B.2, which states that there exists $V_m(\mu)$ solving:

$$V_m(\mu) = \max_{\nu \geqslant \mu} \frac{c}{\rho} \frac{F_m(\nu) - V(\mu) - V'(\mu)(\nu - \mu)}{J(\mu,\nu)}$$

  This refers to Figure 22-1. Define

$$V_{\mu^*}(\mu) = \begin{cases} F(\mu) \text{ if } \mu \leqslant \mu^* \\ V_m(\mu) \text{ if } \mu \geqslant \mu^* \end{cases}$$



The two black (dashed and solid) lines are $F_{m-1}(\mu), F_m(\mu)$.
The blue line is optimal value function from taking immediate action m.
The red line is optimal value function from taking immediate action $m-1$.

Figure 22: Construction of optimal value function.

Be Lemma B.2, when $V_{\mu^*}(\mu) > F(\mu)$, V_{μ^*} is smoothly increasing and optimal $\nu(\mu)$ is smoothly decreasing.

Now update $V_{\mu^*}(\mu)$ with respect to more actions (in the order of decreasing index m). First consider F_{m-1} and let $\widehat{\mu}_m$ be the smallest $\mu \geqslant \mu^*$ such that:

$$V_{\mu^*}(\widehat{\mu}_m) = \max_{\nu \geqslant \widehat{\mu}_m} \frac{c}{\rho} \frac{F_{m-1}(\nu) - V_{\mu^*}(\widehat{\mu}_m) - V'_{\mu^*}(\widehat{\mu}_m)(\nu - \widehat{\mu}_m)}{J(\widehat{\mu}_m, \nu)} \tag{31}$$

At $\widehat{\mu}_m$, searching posterior on F_{m-1} first dominates searching posterior on F_m[34]. This step refers to Figure 22-2. $\widehat{\mu}_m$ is the smallest intersection point of blue curve ($V_{\mu^*}(\mu)$, LHS of Equation (31)) and thin red curve (RHS of Equation (31)). If $V_m(\widehat{\mu}_m) > F_{m-1}(\widehat{\mu}_m)$ then solve for V_{m-1} with initial condition $\mu_0 = \widehat{\mu}_m, V_0 = V_m(\widehat{\mu}_m), V'_0 = V'_m(\widehat{\mu}_m)$ according to Lemma B.2 and redefine $V_{\mu *}(\mu) = V_{m-1}(\mu)$ when $\mu \geqslant \widehat{\mu}_m$. Otherwise skip to looking for $\widehat{\mu}_{m-1}$. If $m-1 > \overline{m}$, continue this procedure by looking for $\widehat{\mu}_{m-1}$ and update $V_{\mu^*}|_{\mu \geqslant \widehat{\mu}_{m-1}}$ with corresponding $V_{m-2} \ldots$ until $m = \overline{m}$ (No action with the slope of F'_m being negative is considered). This refers to Figure 22-3. Now suppose $V_{\overline{m}}$ first hits $F(\mu)$ at some point μ^{**} ($\mu^{**} > \mu^*$ since $V_m(\mu^*) > F(\mu^*)$). V_{μ^*} is a (piecewise) smooth function on $[\mu^*, \mu^{**}]$ such that:

$$V_{\mu^*}(\mu) = \begin{cases} F(\mu) \text{ if } \mu \leqslant \mu^* \text{ or } \mu \geqslant \mu^{**} \\ V_k(\mu) \text{ if } \mu \in [\widehat{\mu}_k, \widehat{\mu}_{k-1}]^{35} \end{cases}$$

By construction, optimal posterior $\nu_{\mu^*}(\mu)$ is smoothly decreasing on each $(\widehat{\mu}_{k+1}, \widehat{\mu}_k)$ and jumps down at each $\widehat{\mu}_k$[36]. Notice that it is not yet proved that this order of value function updating is WLOO. It is possible that optimal policy function is non-monotonic. This is taken care of by Lemma S.18, which proves the order of updating being WLOO. I relegate the proof of Lemma S.18 to supplemental materials to conserve space, but it uses exactly the techniques of the verification step 2. Now I can claim that $\forall \mu \in [\mu^*, \mu^{**}]$:

$$V_{\mu^*}(\mu) = \max_{\nu \geqslant \mu, k} \frac{c}{\rho} \frac{F_k(\nu) - V_{\mu^*}(\mu) - V'_{\mu^*}(\mu)(\nu - \mu)}{J(\mu, \nu)} \tag{32}$$

[34] Existence is guaranteed by smoothness of V_{μ^*} and J. Noticing that $V_m(\widehat{\mu}_m) \geqslant F_{m-1}(\widehat{\mu}_m)$. Otherwise, there will be a $\widehat{\mu}'_m < \widehat{\mu}_m$ s.t. $V_m(\widehat{\mu}'_m) = F_{m-1}(\widehat{\mu}'_m)$ and it is easy to verify that V_m is weakly larger than the maximum. So there is an even smaller $\widehat{\mu}_m$, contradiction.

[35] Define $\widehat{\mu}_{m+1} = \mu^*$ and $\widehat{\mu}_{\overline{m}} = \mu^{**}$ for consistency.

[36] Since F_{k-1} always crosses F_k from above, when indifference between choosing F_{k-1} and F_k, the posterior corresponding to F_{k-1} must be smaller.

Case 2: Suppose $\mu^* \in (0,1)$ but $\overline{V}(\mu^*) = F(\mu^*)$, let $\mu^{**} = \inf\{\mu \geqslant \mu^* | \overline{V}(\mu) > F(\mu)\}$.

Case 3: Suppose $\mu^* = 0$, then $F'(0) \geqslant 0$ (by Lemma B.1). Consider

$$\widetilde{V}(\mu) = \max_{\nu \geqslant \mu, k} \frac{c}{\rho} \frac{F_k(\nu) - F_1(\mu) - F_1'(\nu - \mu)}{J(\mu,\nu)}$$

Define, $\mu^{**} = \inf\left\{\mu | \widetilde{V}(\mu) > F_1(\mu)\right\}$. By Assumption 3, $\lim_{\mu \to 0} |H'(\mu)| = \infty$, then there exists δ s.t. $\forall \mu < \delta$, $\forall \nu \geqslant \underline{\mu}_2$, $\frac{\sup F}{J(\mu,\nu)} \leqslant \inf F$. Therefore $\mu^{**} \geqslant \delta > 0$. This step refers to Figure 22-4.

- *Step 3:* Solve for V to the right of μ^{**}. For all $\mu^\diamond \geqslant \mu^{**}$ such that:

$$F(\mu^\diamond) = \max_{\nu \geqslant \mu, k} \frac{c}{\rho} \frac{F_k(\nu) - F(\mu^\diamond) - F'^-(\mu^\diamond)(\nu - \mu^\diamond)}{J(\mu^\diamond,\nu)} \tag{33}$$

Let m be the index of optimal action. Solve for V_m with initial condition $\mu_0 = \mu^\diamond, V_0 = F(\mu^\diamond), V_0' = F'^-(\mu^\diamond)$. [37] Then take same steps in *Step 2* and solve for $\widehat{\mu}_k$ and V_{k-1} sequentially until V_{m_0} first hits F. This step refers to Figure 22-4,5. Now suppose V_{m_0} first hits $F(\mu)$ at some point $\mu^{\diamond\diamond}$ (can potentially be μ), define:

$$V_{\mu^\diamond}(\mu) = \begin{cases} F(\mu) & \text{if } \mu < \mu^\diamond \text{ or } \mu > \mu^{\diamond\diamond} \\ V_k(\mu) & \text{if } \mu \in [\widehat{\mu}_{k+1}, \widehat{\mu}_k]^{38} \end{cases}$$

By Lemma B.2, V_μ is piecewise smooth are pasted smoothly. So V_μ is a smooth function on $[\mu, \mu'']$. Optimal posterior $\nu_{\mu^\diamond}(\mu)$ is smoothly decreasing on each $(\widehat{\mu}_{k+1}, \widehat{\mu}_k)$ and jumps down at each $\widehat{\mu}_k$. By Lemma S.18 and our construction, $\forall \mu \in [\mu^\diamond, \mu^{\diamond\diamond}]$:

$$V_\mu(\mu) = \max_{\nu \geqslant \mu^\diamond, k} \frac{c}{\rho} \frac{F_k(\nu) - V_{\mu^\diamond}(\mu) - V_{\mu^\diamond}'(\mu)(\nu - \mu)}{J(\mu,\nu)} \tag{32}$$

Let Ω be the set of all such $\mu^\diamond$'s.

- *Step 4:* Define:

$$V(\mu) = \begin{cases} V_{\mu^*}(\mu) & \text{if } \mu \in [\mu^*, \mu^{**}] \\ \sup\limits_{\mu^\diamond \in \Omega} \{V_{\mu^\diamond}(\mu)\} & \text{if } \mu \geqslant \mu^{**} \end{cases} \tag{34}$$

In the algorithm, I only discussed the case $\mu^* < 1$ and constructed the value function on the right of μ^*. On the left of μ^*, V can be defined using a totally symmetric argument by referring to Lemma B.2′ and Lemma S.18′.

Smoothness:

I need to verify that $V(\mu)$ that defined as Equation (34) is a $C^{(1)}$ smooth function on $[0,1]$. This claim is purely for technical use (for example, the validity of using V' and V''). I relegate this technical proof to Section S2.1 in Lemmas S.11, S.12, S.13 and S.14. In addition, it is shown in Section S2.1 that there exists a set of μ_0 such that on each interval when $V(\mu) > F(\mu)$, $V(\mu)$ is defined as one V_{μ_0}.

Unimprovability:

Finally, I prove unimprovability of $V(\mu)$.

- *Step 1*: I first show that $V(\mu)$ solves the following problem:

$$V(\mu) = \max\left\{F(\mu), \max_{\nu, m} \frac{c}{\rho} \frac{F_m(\nu) - V(\mu) - V'(\mu)(\nu - \mu)}{J(\mu,\nu)}\right\} \tag{P-C}$$

$$\begin{cases} \nu \geqslant \mu \text{ when } \mu \geqslant \mu^* \\ \nu \leqslant \mu \text{ when } \mu \leqslant \mu^* \end{cases}$$

Equation (P-C) is the maximization problem over all confirmatory evidence seeking with immediate decision making upon arrival of signals. Equation (P-C) is implied by Equation (32) for $\mu \in E$. So it is sufficient to prove Equation (P-C) for $\mu \in E^C$. Suppose for the sake of contradictoin that there exists $\mu \geqslant \mu^*$ s.t. Equation (P-C) is violated. Let $F(\mu) = F_k(\mu)$. Then it is equivalently stating that:

$$U(\mu) = \max_{\nu, k' > k} \frac{c}{\rho} \frac{F_k'(\nu) - F_k(\mu) - F_k'(\nu - \mu)}{J(\mu,\nu)} > F_k(\mu)$$

[37] By definition of μ^{**}, μ_0 is bounded away from $\{0,1\}$ and Equation (33) implies conditions in Lemma B.2 are satisfied.

[38] Define $\widehat{\mu}_{m+1} = \mu^\diamond$ and $\widehat{\mu}_{m_0} = \mu^{\diamond\diamond}$ for consistency.

Consider $\underline{\mu}_k$ (the intercection of F_k and F_{k-1}). By Lemma S.11, there exists I_k s.t. $\underline{\mu}_k \in I_k$. At $b_k = \sup I_k$, $U(b_k) \leqslant F_k(b_k)$. Therefore, since $U(\mu)$ is continuous, by intermediate value theorem there exists largest μ' between $\underline{\mu}_k$ and μ s.t .$U(\mu') = F_k(\mu')$. Then Equation (33) is satisfied at μ' so consider $V_{\mu'}$. Sicne $V_{\mu'}(\mu) \leqslant V(\mu) = F_k(\mu)$, there exists $\mu'' \in (\mu', \mu)$ s.t. $V_{\mu'}(\mu'') \leqslant F_k(\mu'')$ and $V'_{\mu'}(\mu'') \leqslant F_k(\mu'')$. Therefore $U(\mu'') > F_k(\mu'')$ implies $V_{\mu'}(\mu'') > F_k(\mu'')$, contradiction. Apply a symmetric argument to $\mu \leqslant \mu^*$, I prove Equation (P-C).

- *Step 2*: I show that $V(\mu)$ solves the following problem:

$$V(\mu) = \max\left\{F(\mu), \max_{\nu} \frac{c}{\rho} \frac{V(\nu) - V(\mu) - V'(\mu)(\nu - \mu)}{J(\mu, \nu)}\right\} \tag{P-D}$$
$$\begin{cases} \nu \geqslant \mu \text{ when } \mu \geqslant \mu^* \\ \nu \leqslant \mu \text{ when } \mu \leqslant \mu^* \end{cases}$$

Equation (P-D) is the maximization problem over all confirmatory learning strategies. It has less constraint than Equation (P-C): when a signal arrives and posterior belief ν is realized, the DM is allowed to continue experimentation instead of being forced to take an action.

I only show the case $\mu \geqslant \mu^*$ and a totally symmetric argument applies to $\mu \leqslant \mu^*$. Suppose Equation (P-C) is violated at μ, then there exists ν' such that:

$$V(\mu) = \max_{\nu \geqslant \mu, m} \frac{c}{\rho} \frac{F_m(\nu) - V(\mu) - V'(\mu)(\nu - \mu)}{J(\mu, \nu)} < \frac{c}{\rho} \frac{V(\nu') - V(\mu) - V'(\mu)(\nu' - \mu)}{J(\mu, \nu')} \tag{35}$$

Let $\widetilde{V} = V(\mu)$. Suppose the maximizer is ν, m. Optimality implies first order conditions Equation (41) and Equation (40):

$$\begin{cases} F'_m + \frac{\rho}{c}\widetilde{V}H'(\nu) = V'(\mu) + \frac{\rho}{c}\widetilde{V}H'(\mu) \\ \left(F_m(\nu) + \frac{\rho}{c}\widetilde{V}H(\nu)\right) - \left(V(\mu) + \frac{\rho}{c}\widetilde{V}H(\mu)\right) = \left(V'(\mu) + \frac{\rho}{c}V(\mu)H'(\mu)(\nu - \mu)\right) \end{cases}$$

We define $L(V, \lambda, \mu)(\nu)$ and $G(V, \lambda)(\mu)$ as:

$$\begin{cases} L(V, \lambda, \mu)(\nu) = (V(\mu) + \lambda H(\mu)) + (V'(\mu) + \lambda H'(\mu))(\nu - \mu) \\ G(V, \lambda)(\nu) = V(\nu) + \lambda H(\nu) \end{cases} \tag{36}$$

Then L is a linear function of ν and $G(F_m, \frac{\rho}{c}\widetilde{V})(\nu)$ is a strictly concave smooth function of ν. Consider:

$$L\left(V, \frac{\rho}{c}\widetilde{V}, \mu\right)(\nu) - G\left(F_m, \frac{\rho}{c}\widetilde{V}\right)(\nu)$$

Equation (41) implies that it attains minimum 0 at ν. For any m' other than m,

$$L\left(V, \frac{\rho}{c}\widetilde{V}, \mu\right)(\nu) - G\left(F_{m'}, \frac{\rho}{c}\widetilde{V}\right)(\nu)$$

is convex and weakly larger than zero. However by Equation (35):

$$L\left(V, \frac{\rho}{c}\widetilde{V}, \mu\right)(\nu') - G\left(V, \frac{\rho}{c}\widetilde{V}\right)(\nu') = -\left(V(\nu') - V(\mu) - V'(\mu)(\nu' - \mu) - \frac{\rho}{c}\widetilde{V}J(\mu, \nu')\right) < 0$$

Therefore $L\left(V, \frac{\rho}{c}\widetilde{V}, \mu\right)(\nu) - G\left(V, \frac{\rho}{c}\widetilde{V}\right)(\nu)$ has strictly negative minimum. Suppose it's minimized at $\widetilde{\mu}$ ($\widetilde{\mu} > \mu$ since $L(V, \lambda, \mu)(\mu) \equiv G(V, \lambda)(\mu)$). Then FOC is a necessary condition:

$$V'(\mu) + \frac{\rho}{c}\widetilde{V}H'(\mu) = V'(\widetilde{\mu}) + \frac{\rho}{c}\widetilde{V}H'(\widetilde{\mu})$$

Consider:

$$\begin{aligned}
&L\left(V, \frac{\rho}{c}\widetilde{V}, \widetilde{\mu}\right)(\nu(\widetilde{\mu})) - G\left(F_m, \frac{\rho}{c}\widetilde{V}\right)(\widetilde{\nu}) \\
=&L\left(V, \frac{\rho}{c}\widetilde{V}, \mu\right)(\nu(\widetilde{\mu})) - G\left(F_m, \frac{\rho}{c}\widetilde{V}\right)(\nu(\widetilde{\mu})) \\
&+ V(\widetilde{\mu}) - V(\mu) + \frac{\rho}{c}\widetilde{V}(H(\widetilde{\mu}) - H(\mu)) - \left(V'(\mu) + \frac{\rho}{c}\widetilde{V}H'(\mu)\right)(\widetilde{\mu} - \mu) \\
\geqslant& V(\widetilde{\mu}) - V(\mu) + \frac{\rho}{c}\widetilde{V}(H(\widetilde{\mu}) - H(\mu)) - \left(V'(\mu) + \frac{\rho}{c}\widetilde{V}H'(\mu)\right)(\widetilde{\mu} - \mu) \\
=&G\left(V, \frac{\rho}{c}\widetilde{V}\right)(\widetilde{\mu}) - L\left(V, \frac{\rho}{c}\widetilde{V}, \mu\right)(\widetilde{\mu}) > 0
\end{aligned}$$

In the first equality I used Equation (41) at $\tilde{\mu}$. In first inequality I used suboptimality of $\tilde{\mu}$ at μ. However for m' and $\nu(\tilde{\mu})$ being optimizer at $\tilde{\mu}$:

$$
\begin{aligned}
0=&L\left(V,\frac{\rho}{c}V(\tilde{\mu}),\tilde{\mu}\right)(\nu(\tilde{\mu}))-G\left(F_{m'},\frac{\rho}{c}V(\tilde{\mu})\right)(\nu(\tilde{\mu}))\\
=&L\left(V,\frac{\rho}{c}\tilde{V},\tilde{\mu}\right)(\nu(\tilde{\mu}))-G\left(F_{m'},\frac{\rho}{c}\tilde{V}\right)(\nu(\tilde{\mu}))\\
&+\frac{\rho}{c}(V(\tilde{\mu})-\tilde{V})(H(\tilde{\mu})-H(\nu(\tilde{\mu}))+H'(\tilde{\mu})(\nu(\tilde{\mu})-\tilde{\mu}))\\
>&\frac{\rho}{c}(V(\tilde{\mu})-\tilde{V})J(\tilde{\mu},\nu(\tilde{\mu}))
\end{aligned}
$$

Contradiction. Therefore, I proved Equation (P-D).

- *Step 3*: I show that V satisfies Equation (18), which is less restrictive than Equation (P-D) by allowing 1) diffusion experiments. 2) evidence seeking of all possible posteriors instead of just confirmatory evidence.

  First, since V is smoothly increasing and has a piecewise differentiable optimizer ν, envelope theorem implies:

$$
\begin{aligned}
V'(\mu)=&\frac{c}{\rho}\frac{-V''(\mu)(\nu-\mu)}{J(\mu,\nu)}+V(\mu)\frac{-H''(\mu)(\nu-\mu)}{J(\mu,\nu)}\\
=&-\frac{c}{\rho}\frac{\nu-\mu}{J(\mu,\nu)}\left(V''(\mu)+\frac{\rho}{c}V(\mu)H''(\mu)\right)>0\\
\Longrightarrow&\ V''(\mu)+\frac{\rho}{c}V(\mu)H''(\mu)<0
\end{aligned}
$$

  Therefore, allocating to diffusion experiment is strictly suboptimal. Moreover, consider:

$$
\begin{aligned}
V^{-}(\mu)=&\max_{\nu\leqslant\mu}\frac{c}{\rho}\frac{V(\nu)-V(\mu)-V'(\mu)(\nu-\mu)}{J(\mu,\nu)}\\
\Longrightarrow V^{-\prime}(\mu)=&-\frac{c}{\rho}\frac{\nu-\mu}{J(\mu,\nu)}\left(V''(\mu)+\frac{\rho}{c}V^{-}H''(\mu)\right)
\end{aligned}
$$

  $V^{-}(\mu)$ is by definition the utiltiy gain from searching contradictive evidence, given value function $V(\mu)$. By definition of μ^*, $V^{-}(\mu^*)=V(\mu^*)$ and whenever $V(\mu)=V^{-}(\mu)$ $V^{-\prime}(\mu)<0$. Therefore, $V^{-}(\mu)$ can never cross $V(\mu)$ from below — $V^{-}(\mu)$ is lower than $V(\mu)$ and $V(\mu)$ is unimprovable with contraditive evidence. That is to say:

$$
\begin{aligned}
\rho V(\mu)=&\max\left\{\rho F(\mu),\max_{\nu,p,\sigma}p(V(\nu)-V(\mu)-V'(\mu)(\nu-\mu))+\frac{1}{2}V''(\mu)\sigma^2\right\}\\
\text{s.t. }&pJ(\mu,\nu)+\frac{1}{2}H''(\mu)\sigma^2\leqslant c
\end{aligned}
$$

To sum up, I construct a policy function $\nu(\mu)$ and value function $V(\mu)$ solving Equation (18). Now consider the four properties in Theorem 2. First, by my construction algorithm, in the case $\mu^*\in\{0,1\}$, I can replace μ^* with $\mu^{**}\in(0,1)$. Therefore WLOG $\mu^*\in(0,1)$. Second, $E=\{\mu\in[0,1]|V(\mu)>F(\mu)\}$ is a union of disjoint open intervals $E=\bigcup I_m$. By my construction, $V(\mu)=V_{\mu^m}(\mu)|_{\mu\in I_m}$. On each I_m, $\nu_{\mu^m}(\mu)$ is sctrictly decreasing and jumps down at finite $\hat{\mu}_k$'s. Finally, uniqueness argument in Lemma B.2 implies that ν is uniquely determined by FOC. Therefore, except for those discountinous points of ν, ν is uniquely defined. Number of such discontinuous points is countable, thus of zero measure. *Q.E.D.*

Lemma B.1. *Define $\overline{V}^{+}$ and $\overline{V}^{-}$:*

$$
\begin{aligned}
\overline{V}^{+}(\mu)=&\max_{\nu\geqslant\mu,m}\frac{F_m(\nu)}{1+\frac{\rho}{c}J(\mu,\nu)}\\
\overline{V}^{-}(\mu)=&\max_{\nu\leqslant\mu,m}\frac{F_m(\nu)}{1+\frac{\rho}{c}J(\mu,\nu)}
\end{aligned}
$$

There exists $\mu^*\in[0,1]$ s.t. $\overline{V}^{+}(\mu)\geqslant\overline{V}^{-}(\mu)$ $\forall\mu\geqslant\mu^*$; $\overline{V}^{+}(\mu)\leqslant\overline{V}^{-}(\mu)$ $\forall\mu\leqslant\mu^*$.

Proof. I define function U_m^{+} and U_m^{-} as follows:

$$
\begin{aligned}
U_m^{+}(\mu)=&\max_{\nu\geqslant\mu}\frac{F_m(\nu)}{1+\frac{\rho}{c}J(\mu,\nu)}\\
U_m^{-}(\mu)=&\max_{\nu\leqslant\mu}\frac{F_m(\nu)}{1+\frac{\rho}{c}J(\mu,\nu)}
\end{aligned}
$$

First of all, I solve U_m^+, U_m^- on interior $\mu \in (0,1)$. Since $F_m(\mu)$ is a linear function, $J(\mu,\nu) \geqslant 0$ is smooth, the objective function is a continuous function on compact domain. Therefore both maximization operators are well defined. Existence is already guaranteed, therefore I can refer to first order condition to characterize the maximizer:

$$\text{FOC: } F_m'\left(1+\frac{\rho}{c}J(\mu,\nu)\right)+F_m(\nu)\frac{\rho}{c}\left(H'(\nu)-H'(\mu)\right)=0 \tag{37}$$

$$\text{SOC: } \frac{\rho}{c}F_m'\left(H'(\nu)-H'(\mu)\right) \tag{38}$$

First discuss solving for $\nu \geqslant \mu$. Since $(1+\frac{\rho}{c}J)>0$, $H''<0$, $H'(\nu)-H'(\mu) \leqslant 0$ and inequality is strict when $\nu>\mu$. Therefore, if $F_m'<0$, FOC being held will imply SOC being strictly positive at $\nu>\mu$. So $\forall F_m'<0$, optimal ν is a corner solution. Moreover:

$$\frac{F_m(\mu)}{1+\frac{\rho}{c}J(\mu,\mu)}=F_m(\mu)>F_m(1)>\frac{F_m(1)}{1+\frac{\rho}{c}J(\mu,1)}$$

So $U_m^+(\mu)=F_m(\mu)$. If $F_m'=0$, then $\forall \nu>\mu$:

$$\frac{F_m(\mu)}{1+\frac{\rho}{c}J(\mu,\mu)}=F_m(\mu)=F_m(\nu)\geqslant\frac{F_m(\nu)}{1+\frac{\rho}{c}J(\mu,\nu)}$$

Therefore $\forall F_m' \leqslant 0$, $U_m^+(\mu)=F_m(\mu)$. Then consider the case $F_m'>0$. It can be easily verified that SOC is strictly negative when FOC holds and $\nu>\mu$. Therefore solution of FOC characterizes maximizer. Consider:

$$\lim_{\nu\to\mu} F_m'(1+\frac{\rho}{c}J(\mu,\nu))+F_m(\nu)\frac{\rho}{c}(H'(\nu)-H'(\mu))=F_m'>0$$

$$\lim_{\nu\to 1} F_m'(1+\frac{\rho}{c}J(\mu,\nu))+F_m(\nu)\frac{\rho}{c}(H'(\nu)-H'(\mu))=-\infty$$

Therefore be intermediate value theorem a unique solution $\nu \in (\mu,1)$ exists by solving FOC. Since FOC is a smooth function of μ,ν and SOC is strictly negative, implicit function theorem implies ν being a smooth function of μ. This is sufficient to apply envelope theorem:

$$\frac{d}{d\mu}U_m^+(\mu)=\frac{F_m(\nu)(-H''(\mu)(\nu-\mu))}{\left(1+\frac{\rho}{c}J(\mu,\nu)\right)^2}>0$$

Moreover, Equation (37) is strictly positive when $\nu=\mu$. This implies $U_m^+(\mu)>F_m(\mu)$ when $F_m'>0$.

New consider limit of U_m^+ when $\mu\to 0,1$. When $\mu\to 1$, $U_m^+(\mu)\leqslant \max_{\nu\geqslant\mu}F_m(\nu)=F(1)$. When $\mu\to 0$, consider FOC Equation (37):

$$\begin{aligned}
&\lim_{\mu\to 0} F_m'\left(1+\frac{\rho}{c}J(\mu,\nu)\right)+F_m(\nu)\frac{\rho}{c}\left(H'(\nu)-H'(\mu)\right)\\
=&\lim_{\mu\to 0} F_m'\left(1+\frac{\rho}{c}J(\nu,\mu)\right)+F_m(\mu)\frac{\rho}{c}\left(H'(\nu)-H'(\mu)\right)\\
=&F_m'\left(1+\frac{\rho}{c}J(\nu,0)\right)+\lim_{\mu\to 0}F_m(\mu)\frac{\rho}{c}\left(H'(\nu)-H'(\mu)\right)=-\infty
\end{aligned}$$

Therefore, when $\mu\to 0$, optimal $\nu\to 0$. Therefore $\frac{F_m(\nu)}{1+\frac{\rho}{c}J(\mu,\nu)}\leqslant F_m(\nu)\to F_m(0)$. To conclude, $U_m^+(\mu)=F_m(\mu)$ when $\mu=0,1$.

Let $\overline{m}$ be the first $F_{\overline{m}}'>0$ (not necessarily exists). Let:

$$U^+(\mu)=\max_{m\geqslant\overline{m}}U_m^+(\mu)$$

Then $U^+(\mu)$ is a strictly increasing function when $\overline{m}$ exists. Symmetrically I can define $\underline{m}$ to be last $F_m'<0$ and:

$$U^-(\mu)=\max_{m\leqslant\underline{m}}U_m^-(\mu)$$

There are three cases:

- *Case 1*: when F is not monotonic, then both U^+ and U^- exists. Moreover, $F(0)>F_{\overline{m}}(0)$ and $F(1)>F_{\underline{m}}(1)$. Therefore, $U^+(0)<U^-(0)$ and $U^+(1)>U^-(1)$. There must exists unique $\mu^*\in(0,1)$ s.t. $U^+(\mu^*)=U^-(\mu^*)$.
- *Case 2*: when $F'\geqslant 0$, then define $\mu^*=0$.
- *Case 3*: when $F'\leqslant 0$, then define $\mu^*=1$.

Finally, define $\overline{V}$:

$$\overline{V}^+(\mu)=\max\{F(\mu),U^+(\mu)\}$$

$$\overline{V}^-(\mu)=\max\{F(\mu),U^-(\mu)\}$$

$$\overline{V}(\mu)=\max\{\overline{V}^+(\mu),\overline{V}^-(\mu)\}$$

Given our construction, μ^* always exists and satisfies the conditions in Lemma B.1.

Q.E.D.

Lemma B.2. *Assume $\mu_0\geqslant\mu^*$, $F_m'\geqslant 0$, $V_0,V_0'\geqslant 0$ satisfies:*

$$\begin{cases}\overline{V}(\mu_0)\geqslant V_0\geqslant F_m(\mu_0)\\ V_0=\max\limits_{\nu\geqslant\mu_0}\frac{c}{\rho}\frac{F_m(\nu)-V_0-V_0'(\nu-\mu_0)}{J(\mu_0,\nu)}\end{cases}$$

Then there exists a $C^{(1)}$ smooth and strictly increasing $V(\mu)$ defined on $[\mu_0,1]$ satisfying

$$V(\mu)=\max_{\nu\geqslant\mu}\frac{c}{\rho}\frac{F_m(\nu)-V(\mu)-V'(\mu)(\nu-\mu)}{J(\mu,\nu)} \tag{39}$$

and initial condition $V(\mu_0)=V_0, V'(\mu_0)=V_0'$. Maximizer $\nu(\mu)$ is $C^{(1)}$ and strictly decreasing on $\{\mu|V(\mu)>F_m(\mu)\}$.

Proof. I start from deriving the FOC and SOC for Equation (39):

$$\begin{aligned}&\text{FOC: }\frac{F_m'-V'(\mu)}{J(\mu,\nu)}+\frac{F_m(\nu)-V(\mu)-V'(\mu)(\nu-\mu)}{J(\mu,\nu)^2}\big(H'(\nu)-H'(\mu)\big)=0\\ &\text{SOC: }\frac{H'(\nu)-H'(\mu)}{J(\mu,\nu)}\left(\frac{F_m'-V'(\mu)}{J(\mu,\nu)}+\frac{F_m(\nu)-V(\mu)-V'(\mu)(\nu-\mu)}{J(\mu,\nu)^2}\big(H'(\nu)-H'(\mu)\big)\right)\\ &\qquad+\frac{H''(\nu)}{J(\mu,\nu)}\big(F_m(\nu)-V(\mu)-V'(\mu)(\nu-\mu)\big)\leqslant 0\end{aligned}$$

If feasibility is imposed:

$$V(\mu)=\frac{c}{\rho}\frac{F_m(\nu)-V(\mu)-V'(\mu)(\nu-\mu)}{J(\mu,\nu)} \tag{40}$$

FOC and SOC reduces to:

$$\text{FOC: }F_m'-V'(\mu)+\frac{\rho}{c}V(\mu)(H'(\nu)-H'(\mu))=0 \tag{41}$$

$$\text{SOC: }\frac{\rho}{c}H''(\nu)V(\mu)\leqslant 0 \tag{42}$$

Let us proceed as follows. I use FOC and feasiblity to derive an ODE system with intial value defined by V_0,V_0'. Then I prove that the solution V must be strictly positive. Therefore, SOC is strict at the point where FOC is satisfied, the solution must be a local maximizer. Moreover, since $H'(\nu)-H'(\mu)<0$, when FOC is negative, SOC must be strictly negative, then FOC can cross zero only from above and hence the solution to FOC is unique. Therefore the solution I get from the ODE system is the maximizer in Equation (39).

$$\begin{cases}\text{Equation (40)}\Longrightarrow V(\mu)=\dfrac{F_m(\nu)-V'(\mu)(\nu-\mu)}{1+\frac{\rho}{c}J(\mu,\nu)}\\ \text{Equation (41)}\Longrightarrow V'(\mu)=F_m'+\dfrac{\rho}{c}V(\mu)(H'(\nu)-H'(\mu))\end{cases}$$

$$\Longrightarrow\begin{cases}V(\mu)=\dfrac{F_m(\mu)}{1-\frac{\rho}{c}J(\nu,\mu)}\\ V'(\mu)=F_m'+\dfrac{\frac{\rho}{c}F_m(\mu)(H'(\nu)-H'(\mu))}{1-\frac{\rho}{c}J(\nu,\mu)}\end{cases} \tag{43}$$

Consistency of Equation (43) implies that $\nu=\nu(\mu)$ is characterized by the following ODE:

$$\frac{\partial}{\partial\mu}\frac{F_m(\mu)}{1-\frac{\rho}{c}J(\nu,\mu)}+\frac{\partial}{\partial\nu}\frac{F_m(\mu)}{1-\frac{\rho}{c}J(\nu,\mu)}\dot{\nu}=F_m'+\frac{\frac{\rho}{c}F_m(\mu)(H'(\nu)-H'(\mu))}{1-\frac{\rho}{c}J(\nu,\mu)} \tag{44}$$

Simplifying Equation (44):

$$\begin{aligned}&\frac{F_m'}{1-\frac{\rho}{c}J(\nu,\mu)}+\frac{\frac{\rho}{c}F_m(\mu)(H'(\nu)-H'(\mu))}{\left(1-\frac{\rho}{c}J(\nu,\mu)\right)^2}+\frac{\frac{\rho}{c}F_m(\mu)H''(\nu)(\mu-\nu)}{\left(1-\frac{\rho}{c}J(\nu,\mu)\right)^2}\dot{\nu}\\ =&\frac{F_m'+\frac{\rho}{c}(-F_m'J(\nu,\mu)+F_m(\mu)(H'(\nu)-H'(\mu)))}{1-\frac{\rho}{c}J(\nu,\mu)}\end{aligned}$$

$$\implies F_m(\mu)(H'(\nu)-H'(\mu))+F_m(\mu)H''(\nu)(\mu-\nu)\dot{\nu}=(-F_m'J(\nu,\mu)+F_m(\mu)(H'(\nu)-H'(\mu)))(1-\frac{\rho}{c}J(\nu,\mu))$$

$$\implies F_m(\mu)H''(\nu)(\mu-\nu)\dot{\nu}=-F_m'J(\nu,\mu)(1-\frac{\rho}{c}J(\nu,\mu))-\frac{\rho}{c}J(\nu,\mu)F_m(\mu)(H'(\nu)-H'(\mu))$$

$$\implies \dot{\nu}=J(\nu,\mu)\frac{F_m'(1-\frac{\rho}{c}J(\nu,\mu))+\frac{\rho}{c}F_m(\mu)(H'(\nu)-H'(\mu))}{F_m(\mu)H''(\nu)(\nu-\mu)}$$

Since I want to solve for V_0 on $[\mu_0,1]$, I solve for ν_0 at μ_0 as the initial condition of ODE for ν. The technical details proving the existence of solution to the ODE is relegated to Lemma S.16, which checks standard conditions and invokes the Picard-Lindelof theorem. Lemma S.16 requires an inequality condition and I show it here:

The FOC characterizing ν is Equation (43):

$$(F_m'-V_0')\left(1-\frac{\rho}{c}J(\nu_0,\mu_0)\right)+\frac{\rho}{c}F_m(\mu_0)\left(H'(\nu_0)-H'(\mu_0)\right)=0$$

$$\iff F_m'\left(1+\frac{\rho}{c}J(\mu_0,\nu_0)\right)+\frac{\rho}{c}F_m(\nu_0)(H'(\nu_0)-H'(\mu))=V_0'\left(1-\frac{\rho}{c}J(\nu_0,\mu_0)\right)$$

$$\iff F_m(\mu_0)\left(F_m'\left(1+\frac{\rho}{c}J(\mu_0,\nu_0)\right)+\frac{\rho}{c}F_m(\nu_0)(H'(\nu_0)-H'(\mu))\right)=V_0'F_m(\mu_0)\left(1-\frac{\rho}{c}J(\nu_0,\mu_0)\right)$$

Since $V_0=\frac{F_m(\mu_0)}{1-\frac{\rho}{c}J(\nu_0,\mu_0)}\geqslant 0$, LHS is weakly positive. This satisifes the condition in Lemma S.16. Then Lemma S.16 guarantees existence of unique $\nu(\mu)$, and $\nu(\mu)$ is continuously decreasing from μ_0 until it hits $\nu=\mu$. Suppose $\nu(\mu)$ hits $\nu=\mu$ at $\bar{\mu}_m<1$, define $V(\mu)$ as following:

$$V(\mu)=\begin{cases}\dfrac{F_m(\mu)}{1-\frac{\rho}{c}J(\nu(\mu),\mu)} & \text{if } \mu\in[\mu_0,\bar{\mu}_m)\\ F_m(\mu) & \text{if } \mu\in[\bar{\mu}_m,1]\end{cases}$$

Then I prove the properies of V:

1. V is by construction smooth except for at $\bar{\mu}$. When $\mu\to\bar{\mu}_m$, $\nu(\mu)\to\mu$. Therefore $J(\nu,\mu)\to 0$. This implies $V(\mu)\to F_m(\mu)$. So V is continuous.

2. By Equation (43), when $\mu\in[\mu_0,\bar{\mu}_m)$:

$$V'(\mu)=F_m'+\frac{F_m(\mu)(H'(\nu(\mu))-H'(\mu))}{\frac{c}{\rho}-J(\nu(\mu),\mu)}$$

   When $\mu\to\bar{\mu}_m$, $H'(\nu(\mu))-H'(\mu)\to 0$, $J(\nu(\mu),\mu)\to 0$. Thus $V'(\mu)\to F_m'$. So $V'\in C[\mu_0,1] \implies V\in C^{(1)}[\mu_0,1]$.

3. Rewrite Equation (43) on $[\mu_0,1]$:

$$V'(\mu)=\frac{F_m'\left(1+\frac{\rho}{c}J(\mu,\nu)\right)+F_m(\nu)(H'(\nu)-H'(\mu))}{1-\frac{\rho}{c}J(\nu,\mu)} \tag{45}$$

   According to proof of Lemma S.16, $V'(\mu)>0\ \forall\mu\in(\mu_0,1]$. Moreover since $V_0\geqslant 0$, $\forall\mu\in(\mu_0,1]\ V(\mu)>0$.

Q.E.D.