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 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 so accurate as to warrant an immediate action from her. Over time, absent arrival of a Poisson signal, the DM continues seeking a Poisson signal that is increasingly more precise but arrives less frequently.

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. For example, a manufacturer is choosing from different technologies to be put into production, whose profitabilities are unknown. An asset manager is designing a portfolio of different assets, whose returns are unknown. In these situations, the decision maker (DM) can acquire information prior to making the decision. The manufacturer invests in R&D and the asset manager analyzes market data, both to facilitate decision making. Acquiring more information enables better decision making, but also incurs higher cost or longer delay. Therefore, a rational DM must decide in a sequential way what to learn and when to stop learning contingent on what is already learned, to best balance between the value of information and the cost of learning activity. The optimal decision of “when to stop” is well-studied in seminal works by Wald (1947) and Arrow et al. (1949). A growing literature extends Wald’s model and analyzes “what to learn” in different aspects.

In many situations, the available options for “what to learn” is quite rich, and there are multiple salient aspects to be considered. Again, take the manufacturer for example, one important aspect is the direction(s) of R&D — which features of the product are tested.
Meanwhile, the DM also controls the precision of information — how many data are collected from each test, the frequency of information — how intensively each test is run, etc. In order to acquire information in a most efficient way, the DM fine-tunes numerous parameters in all salient aspects jointly. Contrary to the practice, theoretical models are much more restricted in that most of them allows optimizing only one aspect when analyzing “what to learn”. For example, Moscarini and L. Smith (2001) studies the optimal precision of information while assuming other aspects of information to be exogenously fixed. Che and Mierendorff (2016) and Liang et al. (2017) focus on the optimal direction of learning, where each direction is represented by an exogenous information source with fixed precision and frequency.

The present paper focuses on dynamic information acquisition problems with a rich set of feasible learning strategies. I consider a DM who acquires information about the payoff relevant state before making a one-time choice of action. I introduce three main economic assumptions. (i) the DM controls her belief process about the state non-parametrically in continuous-time, which means that she can flexibly choose the dynamic information acquisition strategy in all possible aspects. This is the most important assumption, and it distinguishes the paper from the existing studies. (ii) the flow cost of information depends on how fast the information reduces uncertainty about the unknown state. (iii) the DM discounts future payoffs and the flow cost of information is strictly convex, which drive non-trivial dynamic learning behavior. The main model is formulated as a stochastic control-stopping problem in continuous-time with non-parametric control process.

Within this framework, I obtain two main results. First, I simplify the non-parametric stochastic control problem to a parametric problem: it is without loss of optimality to restrict the belief process to be a jump-diffusion process. In other words, it is endogenously optimal to acquire a combination of Poisson signal (which drives the jump of belief) and Gaussian signal (which drives the diffusion of belief). The Poisson signal captures seeking discrete breakthroughs in learning, and the Gaussian signal captures the acquisition of incremental information. The optimal strategy is characterized by a tractable Hamilton-Jacobi-Bellman (HJB) equation. The main methodology I develop to obtain the simplification is to analyze an auxiliary discrete-time problem, and then characterize the continuous-time HJB equation indirectly using discrete-time Bellman equations.

Second, I solve the HJB equation derived in the first result. The first finding is that Poisson signal strictly dominates Gaussian signal almost surely. The optimal Poisson signal is pinned down by three key parameters: direction, size and arrival rate of the belief jump. The three parameters exactly represent the three aforementioned salient aspects of learning: direction, precision and frequency. The second finding is the optimal strategy characterized in the three aspects, as well as the optimal stopping time:

- **Direction:** the optimal direction of learning is confirmatory — namely that arrival of the Poisson signal induces belief to jump toward the state that the DM currently finds to be more likely. As an implication of Bayes rule, absence of the signal induces belief to gradually drift towards the opposite direction. When using confirmatory learning strategy, the DM is effectively seeking a breakthrough that supports her prior conjecture. Successfully reaching such a breakthrough makes the DM discontinuously more
certain about her prior conjecture, while waiting for the breakthrough makes her less confident gradually.

- **Precision**: the optimal precision of learning is *negatively related* to the continuation value. Since optimal learning strategy is confirmatory, before a breakthrough happens the DM’s belief is becoming less and less extreme, and the corresponding continuation value is decreasing. As a result, the DM is seeking an increasingly precise Poisson signal.

- **Frequency**: the optimal frequency of learning is *positively related* to the continuation value. Therefore, before a breakthrough happens, the DM is conducting the experiment less and less frequently. Besides, it is optimal to invest decreasing amount of cost in learning.

- **Stopping time**: the optimal time to stop is immediately after the arrival of the Poisson signal. So at the optimum the breakthrough happens only once. Then the DM stops learning and chooses an optimal action based on acquired information.

The main intuition behind the optimal strategy is mostly explained by a novel *precision-frequency trade-off*. Such a trade-off determines the optimal choice of a Poisson signal. The cost of a Poisson signal is increasing in both precision and frequency. So the optimal level of precision and frequency are pinned down by the marginal gain from each parameter. Importantly, this trade-off is closely related to the level of continuation value: when continuation value is larger, the impatient DM loses more from discounting. Hence, she prefers frequency more than precision, because higher frequency enables her to decide sooner and avoid costly waiting. As a result, frequency is positively related to the continuation value, while precision is negatively related to the continuation value.

Besides precision and frequency, this intuition also explains the other aspects: (i) Gaussian signal is equivalent to a special Poisson signal with close to zero precision and infinite frequency. As a result, Gaussian signal is strictly suboptimal except when value is very high. In fact, Gaussian signal is optimal only when the DM is on the edge of stopping learning. (ii) Compare confirmatory learning and contradictory learning. In both cases, when continuing, belief is moving against the optimal posterior — precision of signal is increasing. If learning is confirmatory, belief becomes more uncertain over time. Hence, the DM prefers frequency less and precision more over time. Therefore, confirmatory learning is consistent with the precision-frequency trade-off. On the other hand, if learning is contradictory, belief becomes more certain over time and the DM prefers precision less, which is inconsistent with the trade-off.

To fully understand the role each key assumption plays, I extend the baseline model in different dimensions. First, I find that the (strict) optimality of Poisson signal over Gaussian signal is surprisingly robust to alternative cost structure: it only requires *continuity*, namely if a Poisson signal approximates Gaussian signal then the corresponding cost also converges. Second, I study an extension with no discounting but fixed waiting cost. In this special case, the crucial precision-frequency trade-off diminishes, as a result all dynamic learning strategies become equally optimal. Third, I study an extension where cost depends linearly on uncertain reduction speed. In this special case, learning has constant
return to signal frequency. As a result optimal strategy is to learn infinitely fast — acquire all information at period 0.

The rest of the paper is structured as follows. The related literature is reviewed in Section 2. The main continuous-time model and some illustrative examples are introduced in Section 3. The dynamic programming principle and the corresponding 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 my model. Section 8 explores the implications of the main model on response time in stochastic choice, and on firm’s innovation. Further discussion of other assumptions are in Appendix A and key proofs are in Appendix B. All remaining proofs are relegated to the Supplemental Material.

2 Related literature
2.1 Dynamic information acquisition

My paper is closely related to a literature about designing information acquisition strategy in a dynamic way to facilitate decision making. The canonical approach is to model information flow as a family of stochastic processes. The DM controls parameters determining the information flow, and chooses when to stop learning and make decision. The earliest works focus on the duration of learning. Wald (1947) assumes information to be exogenous, and the DM has control over decision time and action choice. The problem can be formulated as an optimal stopping problem. Moscarini and L. Smith (2001) endogenizes the experimentation intensity by allowing the DM to control precision of a Gaussian signal. A similar learning framework is used as the learning-theoretic foundation for the drift-diffusion model (DDM) by Fudenberg et al. (2015). Following a different route, Che and Mierendorf (2016), Mayskaya (2016) and Liang et al. (2017) studies the sequential choice of information sources, each of which is prescribed exogenously.

Papers more broadly related to sequential learning include Weitzman (1979), Callander (2011), Klabjan et al. (2014), Ke and Villas-Boas (2016) and Doval (2018) on sequential search, Gittins (1974), Weber et al. (1992), Bergemann and Välimäki (1996) and Bolton and Harris (1999) on multi-armed bandits, etc. These papers use frameworks quite different from the information acquisition models. However, in these models, the forms of information are also exogenously prescribed and the DM only has control over whether to reveal each option.

In contrast to the canonical approaches, the key new feature of my framework is that the DM can design the information generating process non-parametrically. In a similar vein to this paper, two concurrent papers Steiner et al. (2016) and Hébert and Woodford (2016) also model dynamic information acquisition non-parametrically. However they focus on other implications of learning by abstracting from sequentially smoothing learning: in Steiner et al. (2016) the assumption of linear flow cost makes it optimal to learn instantaneously, while in Hébert and Woodford (2016) the assumption of zero discount rate makes all dynamic learning strategies essentially equivalent.\(^1\) In contrast, the main

\(^1\)Steiner et al. (2016) assumes the decision problem to be history dependent. So all dynamics in the chosen signal process comes from the history dependence of decision problem, rather than incentive to smooth information. In the dynamic learning foundation of Hébert and Woodford (2016), almost all signal processes are equally optimal because of a key no-discount assumption. They pick
focus of this paper is exactly on characterizing the optimal way to smooth learning. I analyze these two papers’ setups as special cases in Sections 7.2 and 7.3.

A major result of my framework is the endogenous optimality of Poisson signals. A stronger result is established in Section 7.1 where I compare Poisson signal and Gaussian signal. Poisson signal dominates Gaussian signal for generic cost functions which 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), Mayskaya (2016), see also a survey by Hörner and Skrzypacz (2016). I show that the optimal Poisson signal confirms the a priori more likely state, which is consistent with the finding of Che and Mierendorff (2016) when the DM is uncertain of the state. Che and Mierendorff (2016) predicts contradictory learning when DM is more certain of the state, mainly because in their model signal frequency is exogenously fixed.

2.2 Rational inattention

The intra-period information acquisition in my paper is modeled based on the model of rational inattention. A common approach in this literature is to model information as general state contingent signal distributions (Blackwell information structures). The DM can choose from the set of all information structures subject to a cost or a constraint on information. An Entropy based rational inattention framework is first introduced in Sims (2003). Matejka and McKay (2014) studied the flexible information acquisition problem using an Entropy based informativeness measure and derived a generalized Logit decision rule. Caplin and Dean (2015) takes an axiomatization approach and characterized decision rules that can be produced by an information acquisition problem.

Rational inattentive decision rule is widely used in models with strategic interactions in Yang (2015a), Yang (2015b), Matejka (2015), Matejka and McKay (2012), Denti (2015), etc. My paper is quite different from these as mine considers no strategic interaction but repeated learning. Despite the presence of strategic component, Ravid (2018) also studies a dynamic model with repeated learning. In Ravid (2018), a buyer sequentially learns about the offers from a seller and the value of the object being traded in a manner of rational inattention. Seemingly similar to the DM in my model, the rational inattentive buyer systematically delays trading in equilibrium, and the stochastic delay resembles the arrival of a Poisson signal in continuous time. However, in Ravid (2018), the delay is an equilibrium property that ensures the buyer’s strategy being responsive to off-path offers. As a result delay persists even when there is no uncertainty on path. In contrast, the stochastic delay in my paper is an property of optimally smoothed learning process.

I use the reduction speed of uncertainty as a measure of amount of information acquired in unit time. This measure is closely related to the posterior separable information measure from Caplin and Dean (2013). Posterior separable measure of information generalizes the mutual information measure introduced in Shannon (1948), and is widely used to model the cost of information in Gentzkow and Kamenica (2014), Clark (2016), Matyskova (2018), Rappoport and Somma (2017), etc. I provide an axiomatization for

---

Gaussian process exogenously to justify a neighbourhood based static information cost structure.

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

Frankel and Kamenica (2018) does not only provide an alternative axiomatization for posterior separability, but is also related to my paper in an interesting way. Frankel and Kamenica (2018) defines a valid measure of information, which coincides with the uncertainty reduction speed per unit arrival rate of a Poisson signal process 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 in information design (Kamenica and Gentzkow (2009), Ely (2017), Mathevet et al. (2017), etc.). An important simplification method in this literature is a concavification method developed in Aumann et al. (1995) (based on the Carathéodory’s theorem). An alternative method to simplify is the direct signal approach used in both information design problems like Bergemann and Morris (2017), and rational inattention problems. However, the dynamic optimization of learning in my model involves indirect signals before choosing an action, and the objective function is not a simple expectation of state contingent payoffs due to information cost. As a result neither of the two methods is applicable. Instead, I take the belief based approach as in Bayesian persuasion models, but utilize a 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 a Brownian motion (see Fleming (1969) for a survey). The application of Bellman (1957)'s dynamic programming principle leads to the Hamilton-Jacobi-Bellman (HJB) equation characterization of value function. On the contrary, the main stochastic control problem of this paper has general martingale control process. It is a variant of the (semi)martingale models of stochastic control, studied in Davis (1979), Boel and Kohlmann (1980), Striebel (1984), etc. The existing theory provides abstract characterization and existence results, without practically tractable solving methodology provided (not even computationally tractable). This paper introduce an indirect method that establishes a tractable HJB equation. I prove that the HJB equation is identical to that of a jump-diffusion control model (see Hanson (2007)).

3 Model setup

The main model is a continuous-time stochastic control problem. A DM chooses a one-time action at endogenous decision time. The DM controls all the information received before the decision time, bearing a cost on information.

Decision problem: Time $t \in [0, +\infty)$. The DM discounts future utility with rate $\rho > 0$. Both the action space $A$ and the state space $X$ are finite. The DM is a vNM expected

---

3These models focus on settings where without loss of generality we can restrict to consider only signals that are direct recommendation of actions.
utility maximizer, with the Bernoulli utility associated with action-state pair \((a, x) \in A \times X\) at time \(t\) being \(e^{-\rho t}u(a, x)\). The DM holds a prior belief \(\mu \in \Delta(X)\) about the state. Define \(F(\nu) = \max_{a \in A} E_{\nu}[u(a, x)]\) as the expected utility from choosing optimal action given belief \(\nu \in \Delta(X)\).

**Information:** I model information using a belief based approach. It is well known that a distribution of posterior beliefs is induced by an information structure according to Bayes rule iff the expectation of posterior beliefs equals 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 (2009) for example). Extending this formulation to the dynamic environment in current paper, I assume that the DM chooses the whole posterior belief process \(\langle \mu_t \rangle\) in a non-parametric way. Now Bayes rule should be satisfied at every instant of time — \(\forall s > t\), the expectation of \(\mu_s\) is \(\mu_t\). Therefore I restrict \(\langle \mu_t \rangle\) to be a martingale, with \(\langle \mathcal{F}_t \rangle\) being its natural filtration. A formal justification that choosing a belief martingale is equivalent to choosing a dynamic information structure is provided in Section 5.

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

\[
\mathcal{L}_t f(\mu_t) = E\left[\frac{df(\mu_t)}{dt} \bigg| \mathcal{F}_t\right] = \lim_{t' \rightarrow t^+} E\left[\frac{f(\mu_{t'}) - f(\mu_t)}{t' - t} \bigg| \mathcal{F}_t\right]
\]

By definition, \(\mathcal{L}_t f\) captures the expected speed at which \(f(\mu_t)\) is increasing. Let \(\mathcal{D}(f)\) be the domain of \(\langle \mu_t \rangle\) that \(\mathcal{L}_t f(\mu_t)\) is well defined.\(^4\) When \(\langle \mu_t \rangle\) is a well-behaved Markov process and \(f\) is \(C^2\) smooth, \(\mathcal{L} f\) is the standard infinitesimal generator (subscript \(t\) omitted).\(^5\)

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

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

I call \(H\) an uncertainty measure, because \(E[H(\mu)]\) is larger whenever \(\mu\) is Blackwell less informative about the state.\(^6\) By Assumption 1, \(I_t\) is the speed at which uncertainty decreases when belief updates. I call \(I_t\) the (flow) informativeness measure. One example for \(H\) function is the Entropy function: \(H(\mu) = -\sum \mu_x \log(\mu_x)\). It is well known that receiving information reduces Entropy, and \(I_t\) is exactly the speed of Entropy reduction. Assumption 1 is the main technical assumption for my analysis. In Section 5, I show that it is the continuous-time analog of “posterior separability”. For additional discussions, see Section 7.1, where I show an axiom characterization for posterior separability and generalize Assumption 1.

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

\[
V(\mu) = \sup_{\langle \mu_t \rangle \in \mathcal{M}, \tau} E\left[e^{-\rho \tau} F(\mu_\tau) - \int_0^\tau e^{-\rho t} h(I_t) dt\right]
\]

\(^4\)Formally, \(\langle \mu_t \rangle \in \mathcal{D}(f)\) if the uniform limit (w.r.t \(t\)) exists almost surely. Define \(\mathcal{D} = \bigcap_{f \in C(\Delta X)} \mathcal{D}(f)\).

\(^5\)All Feller processes are contained in \(\mathcal{D}\). For example \(\mu_t = at + \sigma W_t\), where \(\langle W_t \rangle\) is standard Brownian motion, is Fellerian, and \(\mathcal{L} f(\mu) = a f'(\mu) + \frac{1}{2} \sigma^2 f''(\mu)\) for \(f \in C^2(\Delta X)\). Like in the definition of Feller process, \(\langle \mu_t \rangle\) should be stochastically continuous w.r.t \(t\) to make \(\mathcal{L} f\) well-defined. However, \(\mathcal{D}\) is much more general than Feller processes as it allows the transition kernel to be discontinuous in the state \(\mu\).

\(^6\)\(H\) is concave if and only if \(\forall \mu\) Blackwell more informative than \(\nu\), \(E[H(\mu)] < E[H(\nu)]\). Concave function is continuous on interior. So the only loss of generality is continuity of \(H\) on \(\partial \Delta(X)\).
where $\mathcal{M}$ is the set of admissible processes. $\mathcal{M}$ contains all martingale $\langle \mu_t \rangle$ in $\mathcal{D}(H)$ with cadlag\footnote{cadlag: $\mu_t : t \mapsto \Delta(X)$ is right continuous with left limits. Notice that assuming martingale $\langle \mu_t \rangle$ being cadlag can be weaken to assuming $\langle \mathcal{F}_t \rangle$ being right continuous (see a martingale modification theorem in Lowther (2009)).} path and satisfying $\mu_0 = \mu$. $\tau$ is a $\langle \mathcal{F}_t \rangle$-measurable stopping time.\footnote{I postpone the discussion of integrability in Equation (1) to Section 5.2. For now, assume that the integral is well defined for all admissible $\langle \mu_t \rangle$ and stopping time $\tau$. In Remark 1, it will be clear that doing so is WLOG.}

The objective function in Equation (1) is fairly standard in a canonical information acquisition problem. The DM acquires information that affects the acquisition of information. Given $\lim_{\tau \to \infty} h(\tau) = \infty$, hence suboptimal.

The comparison also illustrates why a fully flexible learning framework is useful. Systematically compared under the same framework with Entropy based cost function.

Example 1.

Assumption 2. $h : \mathbb{R}^+ \to \mathbb{R}^+$ is weakly increasing, convex and continuous. $\lim_{I \to \infty} h'(I) = \infty$.

I make the following assumption on the cost function $h(I)$ to rule out some trivial cases where learning everything immediately at $t = 0$ is optimal.

Example 1.

Let 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 match the state: $u(L, l) = u(R, r) = 1$; $u(L, r) = u(R, l) = -1$. Discount rate $\rho = 1$. $H$ is the standard Entropy function: $H(\mu) = -\mu \log(\mu) - (1 - \mu) \log(1 - \mu)$. Information cost $h(I) = \frac{1}{8} I^2$.

I consider three simple heuristic learning technologies: Gaussian learning, perfectly revealing breakthroughs and partially revealing evidences. A DM who uses a specific learning technology is modeled by restricting the admissible control set $\mathcal{M}$ 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. It is well known that the posterior belief $\mu_t$ is increasing in $L$ and decreasing in $R$. The interpretation of convexity and the Inada condition $\lim_{\tau \to \infty} h(\tau) = \infty$ is that the DM has strict incentive to smooth the acquisition of information. Given Assumption 2, if the DM acquires all information immediately then uncertainty reduces at infinite speed and the marginal cost $h'(I)$ is infinite, hence suboptimal.\footnote{There is a weaker sufficient condition guaranteeing information smoothing: $\sup_I \mathcal{M} - h(I) > \rho \sup_I F$, where $I = \lim_{I \to \infty} \frac{h(I)}{\rho}$. This condition explicitly states that when $I$ is sufficiently large, $h(I)$ is sufficiently convex that utility gain from smoothing information dominates less from longer waiting. All following theorems in this paper are proved under this weaker condition.}
follows a diffusion process \((\text{Bolton and Harris (1999)})\). So the set of admissible control is:

\[
\mathbb{M}_D = \{\langle \mu_t \rangle | d\mu_t = \sigma_t dW_t \}
\]

The DM controls signal precision \(\langle \sigma_t \rangle\). By 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 L. Smith (2001)\(^{10}\), where optimal information acquisition 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.

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

\[
\mathbb{M}_B = \left\{\langle \mu_t \rangle | d\mu_t = (1 - \mu_t) dJ_t^1(\lambda_t \mu_t) + (0 - \mu_t) dJ_t^0(\lambda_t(1 - \mu_t)) \right\}
\]

\(\langle J_t^1(\cdot) \rangle\) are independent Poisson counting processes with Poisson rate \((\cdot)\). The DM controls signal frequency \(\langle \lambda_t \rangle\). The Entropy reduction speed is \(\lambda_t H(\mu)\). HJB equation:

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

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

3. **Partially revealing evidence**: the DM allocates 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 control is:

\[
\mathbb{M}_P = \left\{\langle \mu_t \rangle | d\mu_t = (1 - \mu_t) dJ_t^1(\alpha_t \gamma \mu_t) - \alpha_t \gamma \mu_t dt + (0 - \mu_t) dJ_t^0((1 - \alpha_t) \gamma (1 - \mu_t)) - (1 - \alpha_t) \gamma (1 - \mu_t) dt \right\}
\]

\(^{10}\)With “belief elasticity” defined as \(\varepsilon(\mu) = \mu(1-\mu)\) in my model.
\[ \langle J_i(\cdot) \rangle \text{ are independent Poisson counting processes with Poisson rate } \langle \cdot \rangle. \] The DM controls \( \langle a_i \rangle \), the attention allocated to the signal confirming state \( r \). This problem is similar to that in Che and Mierendorff (2016). Applying their analysis, optimal \( a_i \) is a bang-bang solution, and the HJB equation is:

\[
\rho V_p(\mu) = \max \left\{ \gamma \mu \left( F(1) - V_p(\mu) - V'_p(\mu)(1-\mu) \right) - \frac{1}{2} \left( \gamma \mu (H(\mu) + H'(\mu)(1-\mu)) \right)^2, \right. \\
\left. \gamma (1-\mu) \left( F(0) - V_p(\mu) - V'_p(\mu)(0-\mu) \right) - \frac{1}{2} \left( \gamma (1-\mu) (H(\mu) - H'(\mu)(0-\mu)) \right)^2 \right\}
\]

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 grey region, optimal learning direction is confirmatory: arrival of news reveals the a priori more likely state (represented by solid arrows). In the light grey region, optimal learning direction is contradictory: arrival of news reveals the a priori less likely state (represented by dashed arrows).

In this example, the three learning technologies are analyzed with 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 show the rank of utility. Each color corresponds to a learning strategy being optimal: blue — incremental information, red — breakthroughs, and grey — confirmatory evidences. As is shown in Figure 4, allowing the DM to use a rich set of strategies improves the quality of decision making.

More interestingly, there seems to be a pattern in optimizing different aspects. When prior belief is far from the stopping region, high precision Poisson signal that can bring the DM directly to conclusion is valuable. When the prior belief is more sure about one state, flexible allocation of attention to the more promising direction becomes valuable. When prior belief is very sure, imprecise Gaussian signal becomes optimal. The formal analysis of general problem Equation (1) will show that this pattern is systematic: optimal direction, precision and frequency of learning are closely tied to the location of prior belief.

A key message from Example 1 is that the optimal learning dynamics is sensitive to the restrictions imposed on feasible strategy set. It directly implies that existing single-

\[ ^{11} \text{In this example, whenever contradictory learning dominates confirmatory learning, it is dominated by Gaussian learning. So there is no region where contradictory learning is optimal.} \]
aspect models are insufficient for modeling a dynamic information acquisition problem with rich strategy set. In Example 1, each of the three models ignores important trade-offs and generates inaccurate predictions for learning dynamics. Less obviously, an indirect implication is that even richer models that allow controlling multiple salient aspects are unsatisfactory, for the following two reasons: (i) to make an accurate prediction, one must identify very precisely the set of feasible strategies when studying a specific learning problem, which can be difficult in practice. (ii) the sets of feasible strategies are unlikely to be the same across different problems, which significantly restricts the genericity of the solution to any specify model.

3.1 Motivation for a flexible model

Based on the discussions about Example 1, I can now provide the motivation for studying a model with fully flexible information acquisition strategy. On the one hand, Equation (1) is itself an interesting theoretical benchmark of fully flexible learning. More importantly, it provides general predictions for a large set of dynamic information acquisition problems with multiple salient aspects, even when these problems are far from being fully flexible. This claim is based on a very elementary argument — any feasible unconstrained optimum is a constrained optimum. Given a dynamic information acquisition problem, one just need to verify whether the flexible optimal strategy is feasible in the problem. If feasible, then the constrained problem is automatically solved. Therefore, studying a flexible model overcomes the two difficulties mentioned in the previous discussion. First, verifying whether one particular strategy is feasible is much easier than exhausting all alternatives and identifying the whole feasible strategy set. Second, predictions from the flexible model hold generally across all models where the strategy is feasible.

In fact, the analysis of the flexible model in Sections 4 and 6 will show that the optimal strategy is surprisingly simple and quite heuristic. As a result, it is practically easy to verify the feasibility of the strategy, and the strategy is likely to be feasible in various problems.

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, there is no existing general theory applicable to this stochastic control problem. The closest problems are studied in a set of remarkable papers on 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 process. These papers study problem with finite horizon and specific objective functions, hence they do not cover Equation (1).

Nevertheless, it is useful to introduce the general dynamic programming principle and HJB characterization. Based on the intuition of dynamic programming, it is a reasonable conjecture that $V(\mu_t)$ satisfies the following HJB:

$$\max \{ F(\mu_t) - V(\mu_t), -\rho V(\mu_t) + \sup_{d\mu_t} \{ L_t V(\mu_t) - h(-L_t H(\mu_t)) \} \} = 0$$

HJB Equation (2) is conceptually the same as the standard HJB equation. Recall the def-
inition for infinitesimal generator, \( L_t V(\mu_t) \) is the flow utility gain from continuing. The exact form of \( L_t V \) and \( L_t H \) depends on the underlying probability space, the filtration and the control process in a neighbourhood of \( t \) (which are summarized by the symbol \( d\mu_t \)). So 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 loss from discounting (the second term is 0).

For a simple example, let \( \mathbb{M} \) be a family of Markov jump-diffusion belief processes:

\[
d\mu_t = \left( v(\mu_t) - \mu_t \right) dJ_t(p(\mu_t)) - p(\mu_t)dt + \sigma(\mu_t)dW_t
\]

(3)

where \( (p, v, \sigma) : \mu_t \mapsto \mathbb{R}^+ \otimes \Delta(\text{Supp}(\mu)) \otimes \mathbb{R}^{\text{Supp}(\mu) - 1} \) are control parameters. \( J_t(\cdot) \) is Poisson counting process with Poisson rate \( \cdot \), \( W_t \) is a standard one-dimensional Wiener process. Itô’s lemma implies an explicit form for the infinitesimal generator:

\[
L V(\mu) = p(V(v) - V(\mu) - \nabla V(\mu)(v - \mu)) + \frac{1}{2} \sigma^T \mathcal{H}V(\mu)\sigma
\]

where \( \nabla, \mathcal{H} \) is gradient and Hessian matrix operator. The corresponding HJB becomes:

\[
\rho V(\mu) = \max \left\{ \rho F(\mu), \sup_{p, \mu, \sigma} p(V(v) - V(\mu) - \nabla V(\mu)(v - \mu)) + \frac{1}{2} \sigma^T \mathcal{H}V(\mu)\sigma \right. \\
\left. - h \left( p(H(\mu) - H(v) + \nabla H(\mu)(v - \mu)) - \frac{1}{2} \sigma^T \mathcal{H}H(\mu)\sigma \right) \right\}
\]

(4)

Now consider the general martingale \( \left\langle \mu_t \right\rangle \). Of course the first gap to be filled is a verification theorem that establishes HJB Equation (2). Second, to make Equation (2) practically tractable, a representation theorem for the abstract operator \( L_t \) is necessary. Existing martingale methods have little power on both tasks.\(^{12}\) I bypass the difficulties by using an indirect method and (i) formally establish the HJB Equation (2), (ii) simplify the operator \( L_t \) to a tractable form:

**Theorem 1.** Assume \( H \) is strictly concave and \( C^{(2)} \) on interior beliefs, Assumptions 1 and 2 are satisfied. Let \( V(\mu) \in C^{(1)} \) be a solution\(^{13}\) to 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). It is known in jump-diffusion control theory that the solution of Equation (4) is the solution of Equation (1) with admissible controls restricted to solutions of SDE (3) (see textbooks, e.g. Hanson (2007)). Therefore, Equation (1) can be solved by considering only the family of Markov jump-diffusion processes, where the control variables are arrival rate \( (p) \) and size \( (v - \mu) \) of a Poisson jump, and the flow variance \( (\sigma) \) of Gaussian diffusion.

\(^{12}\) Existing martingale methods verifies HJB equation for different sets of problems not covering this one. The martingale method only states the existence of such \( L_t V \) (for example theorem 4.3.1 of Boel and Kohlmann (1980)), instead of an explicit representation. This is considered as the main drawback of the martingale method (see discussions in Davis (1979)).

\(^{13}\) A solution to second order ODE is not well defined. To be precise, \( V \) is viscosity solution (see Crandall et al. (1992)) to Equation (4). In the viscosity solution, \( \sigma^T \mathcal{H}V(\mu)\sigma \) is replaced by \( D^2 V(\mu, \sigma)|c|^2 \), where \( D^2 V(\mu, \sigma) = \lim_{\delta \to 0} \frac{1}{2} \mathbb{E}[\langle \mu | V(\mu; \sigma + c \delta) - V(\mu; c \delta) \rangle^2] \).
The compensated Poisson jump part and Gaussian diffusion part in SDE (3) each represents a simple learning strategy. The Poisson jump in belief process can be induced by observing a non-conclusive news whose arrival follows a Poisson process. The compensating belief drift is induced by observing no news arriving. I say that the DM is using Poisson learning or is acquiring Poisson signal when there is a compensated Poisson part in belief process. The Gaussian diffusion in belief process can be induced by observing the realization of a Gaussian process, with state \( x \) being its unobservable drift. I say that the DM is using Gaussian learning or is acquiring Gaussian signal when there is a diffusion part in belief process.

The control variables in Equation (4) represent four kinds of trade-offs in a one-step decision problem: (i) the standard continuing v.s. stopping trade-off in dynamic programming problems, captured by the outer layer maximization. (ii) the information cost v.s. utility gain trade-off. More informative signal leads to higher utility gain, it also incurs higher cost due to increasing \( h \). (iii) the Poisson v.s. Gaussian trade-off. Total amount of informativeness is allocated to the Poisson signal \( (p, v) \) and the Gaussian signal \( \sigma \). (iv) the precision v.s. frequency trade-off. Fixing the informativeness of a Poisson signal, when DM increases the precision of a signal (chooses \( v \) further away from prior \( \mu \)), she has to sacrifice the arriving frequency of the signal (lower arrival rate \( p \)). These trade-offs, especially the precision-frequency trade-off, will be discussed in details to characterize the solution to Equation (4) in Section 6.

The indirect method I use to prove Theorem 1 is to characterize Equation (1) as the limit of a series of auxiliary discrete-time problems. Moreover, the discrete-time problem also serves as foundation for Equation (1). Modeling assumptions like using martingale \( \langle \mu_t \rangle \) as control, and using \(-{\mathcal{L}}_t \mathcal{H}\) as measure of informativeness are justified by more fundamental assumptions. The discrete-time analyses are presented in Section 5. Readers more interested in the solution of HJB Equation (4) can jump to Section 6.

5 Discrete-time foundation

In this section, I first propose a discrete-time problem in Section 5.1 that explicitly models flexible design of signal process. Then I show in Section 5.2 that it equivalently characterizes the discretization of the continuous-time stochastic control problem Equation (1). In Section 5.3 I introduce a key lemma that links all the discrete-time analysis and proves Theorem 1.

5.1 General discrete-time problem

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

**Strategy:** a strategy is a triplet \((S^t, \tau, A^t)\). \( S^t \) is a random process correlated with the state, called an information structure. The realization of \( S^t \) is called a signal history. The signal history up to period \( t \) is denoted by \( S^t \). Each \( S^t \) specifies the signal structure acquired in period \( t \) conditional on all histories up to period \( t \).\(^{14}\) \( \tau \) is a random variable whose realization is in \( \mathbb{N} \). \( \tau \) specifies a random decision time. The action choice \( A^t \) is a random process whose realization is in \( A \). Each \( A^t \) specifies the joint distribution of action

\(^{14}\)\( S^{-1} \) is defined as a degenerate random variable that induces belief same as prior belief \( \mu \) for notation simplicity.
choice and state conditional on making decision in period $t$. Let the marginal distribution of the state be denoted by random variable $X$.

**Cost of information:** The cost of information is assumed to be a discretization of the flow cost in Assumption 1. Define $h_{dt}(I) = h(I_{dt})dt$, where:

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

The per-period cost of information is $h_{dt}(I(S^t; X|S^{t-1}, 1_{\tau \leq t}))$. It is not difficult to see that $I(S^t; X|S^{t-1}, 1_{\tau \leq t})$ is exactly the 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 $S^t$ and unknown state $X$ (conditional on history).

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

$$V_{dt}(\mu) = \sup_{S^t, \tau, A^t} \mathbb{E} \left[ e^{-\rho dt \cdot \tau} u(A^\tau, X) - \sum_{t=0}^{\infty} e^{-\rho dt \cdot t} h_{dt}(I(S^t; X|S^{t-1}, 1_{\tau \leq t})) \right]$$

s.t. \begin{align*}
    &X \to S^{t-1} \to 1_{\tau \leq t} \\
    &X \to S^{t-1} \to A^t \text{ conditional on } \tau = t
\end{align*}

The two constraints in Equation (5) are called information processing constraints. 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 $t$. They are extensions to the standard measurability requirement, allowing randomness unrelated to unknown state to be added.

5.2 Discretization of continuous-time problem

Take $dt > 0$, consider the discretized Riemann sum of Equation (1):

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

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 Riemann-Stieltjes integral as $\lim_{dt \to 0} W_{dt}(\mu_t, \tau)$. I call the martingale $\langle \mu_t \rangle$ time-measurable if the limit $\lim_{dt \to 0} W_{dt}(\mu_t, \tau)$ exists.\footnote{\textsuperscript{17}}

**Remark 1.** Given the time-measurability condition, there are two ways to interpret Equation (1): 1) the admissible control set $\mathcal{M}$ is further restricted to time-measurable processes. Despite the loss of technical generality, this restriction does not exclude any meaningful

\textsuperscript{15} $1_{\tau \leq t}$ is an indicator whether learning is already stopped up to current period, which is known to the DM. So $\langle S^{t-1}, 1_{\tau \leq t} \rangle$ summarizes all knowledge of the DM.

\textsuperscript{16} Noticing 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.

\textsuperscript{17} The standard definition for integrability also requires limit to exist uniformly for all alternative nonuniform discretization of time horizon, and all alternative measurable stopping time. Here I am using the weaker integrability requirement for simplicity of notation. The optimal strategy actually satisfies the stronger integrability requirements so it is without loss to use the current definition. The discretization of $\langle \mu_t \rangle$ is WLOG given uniform convergence in the definition of $\mathcal{D}(H)$.
processes. In particular, this restriction w.r.t time so the discontinuity of transition kernel (highlighted in Footnote 5) is still allowed.

2) a bolder interpretation is to define the integral as sup limit. The corresponding supremum \( V \) becomes an essential upper bound of utility. However, Lemma 1 can be strengthened (in Remark B.2) to show that this relaxed upper bound can be approached by time-measurable processes. If one takes the second interpretation, then excluding non-time-measurable processes from \( \mathcal{M} \) is WLOG.

Since \( \langle \mu_t \rangle \) is a martingale and \( \tau \) is measurable to \( \mathcal{F}_t \), the discretized \( \langle \mu_{t}\rangle \) is a discrete-time martingale, and discretized \( \tau \) is a stopping time measurable to \( \langle \mu_{t}\rangle \)'s natural filtration. So any admissible strategy of Equation (1) can be naturally converted to a feasible strategy of the following discrete-time optimization problem:

\[
W_{dt}^*(\mu) = \sup_{\langle \mu_t \rangle, \tau} E \left[ e^{-\rho dt} F(\mu_{\tau}) - \sum_{t=0}^{\tau} e^{-\rho (t+1)} h_{dt}(E[H(\mu_t) - H(\mu_{t+1})]|\mathcal{F}_t]) \right]
\]  (6)

where \( \langle \mu_t \rangle \) is a discrete-time martingale and \( \tau \in \mathbb{N} \) is measurable to \( \mu_t \)'s natural filtration \( \mathcal{F}_t \). It is not hard to verify that \( W_{dt}^*(\mu) = \sup_{\langle \mu_t \rangle, \tau} W_{dt}(\mu_t, \tau) \).

On the other hand, given Assumption A, Equation (5) can be modified by removing redundant information and rewriting based on beliefs to become exactly Equation (6). So \( V_{dt}(\mu) = W_{dt}^*(\mu) \) (the formal proofs of claims made in this subsection are in Appendix B.2.1).

Now consider the relation between \( V \) and \( V_{dt} \):

\[
\begin{cases}
V(\mu) = \sup_{\langle \mu_t \rangle, \tau} \lim_{dt \to 0} W_{dt}(\mu_t, \tau) \\
\lim_{dt \to 0} V_{dt}(\mu) = \lim_{dt \to 0} \sup_{\langle \mu_t \rangle, \tau} W_{dt}(\mu_t, \tau)
\end{cases}
\]

The following lemma shows that the lim and sup operators are interchangeable:

**Lemma 1.** Given Assumption 1 and Assumption A, \( \lim_{dt \to 0} V_{dt}(\mu) = V(\mu) \).

Equation (6) is a discrete-time optimization problem, with exponential discounting and bounded utility functions. Therefore, standard dynamic programming theory guarantees the Bellman equation characterization of \( V_{dt} \):

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

\[
V_{dt}(\mu) = \max_{p, \nu_i} \left\{ F(\mu), \max_{p, \nu_i} e^{-\rho dt} \sum_{i=1}^{N} p_i V_{dt}(\nu_i) - h_{dt}(H(\mu) - \sum_{i=1}^{N} p_i H(\nu_i)) \right\}
\]  (7)

s.t. \( \sum_{i=1}^{N} p_i \nu_i = \mu \)

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

Equation (7) 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 arbitrary number of
realizations. This simplification is based on a generalized concavification methodology developed in Theorem 2 of Zhong (2018a). The original concavification methodology is an application of Carathéodory theorem on graph of objective function on belief space.\textsuperscript{19} Equation (7) involves an additional term \( h_{dt}(H(\mu) - \sum p_i H(v_i)) \), which is not linear in expectation. The general method suggests that 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:

![Diagram illustrating the roadmap for proving Theorem 1](image)

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 (7) (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 (7) (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 A and Assumption 2 are satisfied. Suppose \( V(\mu) \in C^{(1)} \) is a solution to Equation (4). Then \( V_{dt} \xrightarrow{L^\infty} V \).

Lemma 3 proves that whenever Equation (4) has a solution, this solution is unique and coincides with the limit of solution to discrete-time problem Equation (7). Verification theorem Theorem 1 is a direct corollary of Lemmas 1, 2 and 3.

6 Optimal information acquisition

In this section I prove existence of 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 or a smooth convex function. In both cases, the optimal strategy is confirmatory Poisson learning, namely signal arrival induces Poisson jumps of belief towards more likely state. The optimal stopping time is immediately after signal arrival, and absence of signal is followed by searching for more precise signals. Optimal information cost decreases overtime when the cost is flexible. Then in Section 6.2 I discuss the key trade-offs in the optimization problem and provide intuition for the optimal strategy.

First of all, I introduce the assumptions for mathematically tractability:

**Assumption 3.**

1. (Binary states): \( |X| = 2 \).

\textsuperscript{19}see Aumann et al. (1995) and Kamenica and Gentzkow (2009)
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 contains three parts. First, I restrict the state space to be binary. Therefore belief space is one dimensional and I can apply tools from differential equations to construct a candidate solution. Although existence of solution technically relies on the binary state assumption, most of the characterization results generalize to general state spaces as is discussed in Appendix A.3. Second, I assume that utility from decision making is strictly positive so that “delay forever” is strictly suboptimal. In fact property 2 is without loss of generality in the sense that we can always add a dummy “outside action” that gives utility close to zero. Third, I assume that \( H \) is sufficiently smooth, strictly convex (which rules out free information) and satisfies an Inada condition (which guarantees 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 \( \text{HJB Equation (4)} \). I prove the existence of such a solution and provide characterization under either Assumption 2-a or Assumption 2-b, two slightly stronger variants of Assumption 2.

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

Assumption 2-a restricts the cost function \( h \) to be a hard cap: information is essentially free when its measure is below flow capacity \( c \) and infinitely costly when it exceeds this capacity. This condition forces the DM to smooth his information acquisition process over time.

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

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

where \( \nu(\mu) \) is unique a.e. and satisfies following properties:
1. \( \exists \mu^* \in \arg\min V \text{ s.t. } \mu > \mu^* \implies \nu(\mu) > \mu \text{ and } \mu < \mu^* \implies \nu(\mu) < \mu. \)
2. \( \nu(\mu) \in E^C. \) (successful experiment lands in stopping region)
3. \( \nu(\mu) \) is strictly decreasing on each interval of \( E \cap [0, \mu^*) \) and \( E \cap (\mu^*, 1] \).
4. \( \rho V(\mu) > -c \frac{\nu''(\mu)}{H''(\mu)} \forall \mu \in E \setminus \mu^*. \)

Theorem 2 proves existence of 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 drives belief to jump to \( \nu(\mu) \). Property 1 says that the optimal signal is confirmatory: when \( \mu > \mu^* \), the DM holds high prior belief for state 1 and she acquires information that induces even higher posterior belief. Vice versa for \( \mu < \mu^* \). Conditional on receiving no signal, the DM’s belief drifts towards
\( \mu^* \). Property 2 says that the image of \( \nu \) is always in the stopping region. In other words, the optimal stopping time is exactly the signal arrival time. Property 3 says that when the prior belief is drifting towards \( \mu^* \), the optimal posterior belief induced by signal is moving against \( \mu^* \), i.e. of increasing precision. Since Assumption 2-a means that the total informativeness of signal is bounded, a signal of increasing precision is achieved at the expense of decreasing frequency of observing a signal. Finally, the optimal policy is essentially unique: Gaussian signal is almost never optimal (Property 4) and optimal Poisson signal is almost everywhere unique.

**Assumption 2-b (Convex flow cost).** \( h : \mathbb{R}^+ \to \mathbb{R}^+ \) is \( C^2(\mathbb{R}^+) \) smooth, \( h(0) = 0 \), \( h''(c) > 0 \), \( \lim_{c \to \infty} h'(c) = \infty \).

Assumption 2-b restricts the cost function \( h \) to be \( C^2 \) smooth and strictly convex: acquiring additional unit of information is of strictly increasing marginal cost. The Inada condition in Assumption 2 is still retained. If we replace Assumption 2-a with Assumption 2-b, we have the following characterization theorem:

**Theorem 3.** Given Assumptions 1, 2-b and 3, there exists quasi-convex value function \( V \in C^1(0,1) \) solving Equation (4). Let \( E = \{ \mu \in [0,1] \mid V(\mu) > F(\mu) \} \) be the experimentation region, there \( \exists \) policy function \( \nu : E \to [0,1], I \in C^1(\mathcal{E}) \)\(^{20} \) s.t.

\[
\rho V(\mu) = -I(\mu) \cdot \frac{F(\nu(\mu)) - V(\mu)}{\frac{1}{2} \sigma^2 V''(\mu)} - h(I(\mu))
\]

where \( \nu \) and \( I \) are unique a.e. and satisfy the following properties:

1. \( \exists \mu^* \in \arg\min V \text{ s.t. } \mu > \mu^* \implies \nu(\mu) > \mu \text{ and } \mu < \mu^* \implies \nu(\mu) < \mu \).
2. \( \nu(\mu) \in E^C \).
3. \( \nu(\mu) \) is strictly decreasing on each interval of \( E \cap [0,\mu^*) \) and \( E \cap (\mu^*, 1] \).
4. \( \rho V(\mu) > \max \frac{1}{2} \sigma^2 V''(\mu) - h(-\frac{1}{2} \sigma^2 H''(\mu)) \quad \forall \mu \in \mathcal{E} \setminus \mu^* \).
5. \( I(\mu) \) is isomorphic to \( V(\mu) \).

Other than property 5, the optimal strategy shares the same set of properties as in Theorem 2. The optimal value function can be achieved through Poisson signals. Optimal stopping time is arrival time of signals. The unique optimal signal takes a form of confirmatory evidence that arrives at increasing precision and decreasing frequency conditional on continuation. Property 5 states that the optimal informativeness measure \( I \) of acquired information is higher when continuation value is higher. Since the belief process drifts downward value function conditional on continuation, this means that the DM invests less in information acquisition when time goes.

The intuition for property 5 is actually discussed in Moscarini and L. Smith (2001).

---

\(^{20}\)Noticing that given \( \nu \), choosing \( I \) and \( \rho \) are equivalent. They uniquely pin down each other according to equation \( I(\mu) = p(\mu)(-H(\nu(\mu)) + H(\mu) + H'(\mu)(\nu(\mu) - \mu)) \).
precision of experimentation. My analysis identifies this intuition separately from another important trade-off between signal precision and frequency. I refer to property 5 as “value-intensity monotonocity” in this paper. Here I rename parameter $I$ the intensity of learning, for the reason that it 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 some supplementary examples showing a rich set of results predicted in my framework, including multiple phases of learning in Example 3, and learning from one-sided search in Example 4.

Example 2. Consider the problem studied in Example 1. $F(\mu) = \max\{2\mu - 1, 1 - 2\mu\}$, $H(\mu) = -\mu \log(\mu) - (1 - \mu) \log(1 - \mu)$, $\rho = 1$, $h(I) = \frac{1}{2}I^2$. Now 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 depicts $F(\mu)$, blue curve depicts $V(\mu)$. The blue shaded region is experimentation region $E$. Figure 5-(b) shows the optimal posterior $v(\mu)$ as a function of prior. As stated in Theorem 2, the policy function is piecewise smooth and decreasing. The three arrows in Figure 5-(a) depict optimal strategies prescribed at three different priors. The arrows start at priors and point to optimal posteriors. The blue curve in Figure 5-(c) shows optimal intensity $I(\mu)$ as a function of prior. It is easy to see that $I(\mu)$ is isomorphic to $V(\mu)$ in the experimentation region.

![Figure 5: Value and policy functions](image)

Figure 5 illustrates the dynamics of the optimal policy. Figure 6-(a) depicts the optimal belief process. Conditional on no signal arrival, posterior belief drifts towards critical belief level $\mu^*$. It is clear that in this example, there are two phases of learning (represented by different colors of shaded regions in Figure 6). In the first phase (blue region), the DM seeks a Poisson signal to confirm one state. As time goes, the signal precision is increasing while signal frequency and learning intensity are decreasing (as in Figure 6-(b)&(c)). Eventually, the DM switches to the second phase (grey region). In the second phase, she seeks two signals confirming each state in a balanced way such that before any signal arrives her posterior belief is stationary.

Now it is clear why the three simple learning strategies studied in Example 1 are optimal in the corresponding regions. It is an approximation to the full solution in Example 2. The confirmatory signal in Example 1 is non-flexible, so it is optimal when the belief is not extreme. When belief is close to the boundary, optimal Poisson signal should be short
jump with high frequency — approximated by Gaussian learning. When belief is close to \( \mu^* \), optimal Poisson signal should be low cost, and jumping to each side does not differ very much — approximated by a perfect revealing signal.

**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 now contains three disjoint intervals. Looking the middle interval, in red regions, the DM has more extreme belief and searches for a safer action (red arrow). In blue region, the DM has more ambiguous belief and searches for a riskier action (blue arrow). Figure 7-(c) depicts optimal belief process with prior belief in the red region. Experimentation now has 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.

**Example 4 (One-sided search).** Figure 8 depicts an example where optimal strategy includes only one-sided search. In this example, state 1 dominates state 0 for both actions. By property 1, \( \mu > \mu^* \) in the whole experimentation region \( E \). Figure 8-(b) shows that optimal strategy is always to search for a Poisson signal inducing a posterior belief higher than the prior. Figure 8-(c) shows that in this example, there is only one phase. If no signal arrives before belief drops to the critical belief, it is optimal for the DM to stop learning and choose the safe action.

This example illustrates more precisely the definition for confirmatory evidence: optimal belief jump is in the direction of a more profitable state. Profitability of a state depends on not only its likelihood, but also the corresponding utility the DM can get in the state.
this example, consider prior beliefs 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. As a result 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 kinds of trade-offs. Now I discuss them in details and illustrate how these trade-offs leads to the properties of optimal strategy in Theorems 2 and 3. I first derive a geometric characterization for optimal policy in Section 6.2.1. Then I discuss how key tradeoffs are represented by the geometric characterization and provide intuitions for the optimal policy. In Section 6.2.2 I show a sketched proof for Theorem 2.

6.2.1 Geometric representation and key trade-offs

In order to gain intuition it is useful to conduct a thought experiment. Fix a value function \( V \) and consider a simplified optimization problem:

\[
\sup_{p \geq 0, \nu} p(V(\nu) - V(\mu) - V'(\mu)(\nu - \mu)) - h(p(H(\mu) - H(\nu) + H'(\mu)(\nu - \mu)))
\]  

Equation (8) is more restrictive than Equation (4). I assume that the DM acquires only a Poisson signal. Let us ignore Gaussian signal for a moment. Define:

\[
\begin{align*}
U(\mu, \nu) &= V(\nu) - V(\mu) - V'(\mu)(\nu - \mu) \\
J(\mu, \nu) &= H(\mu) - H(\nu) + H'(\mu)(\nu - \mu)
\end{align*}
\]

The interpretation of \( U(\mu, \nu) \) is the flow value per unit arrival rate from a Poisson signal with posterior \( \nu \). Similarly, \( J(\mu, \nu) \) is the flow uncertainty reduction per unit arrival rate from the Poisson signal. Then Equation (8) can be rewritten as:

\[
\sup_{p \geq 0, \nu} p \cdot U(\mu, \nu) - h(p \cdot J(\mu, \nu))
\]

\[
\iff \sup_{I \geq 0, \nu} \left( \frac{U(\mu, \nu)}{J(\mu, \nu)} \right) \cdot I - h(I)
\]

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

\[
\begin{align*}
\nu^* &\in \arg \max_{\nu} \frac{U(\nu, \mu)}{J(\nu, \mu)} \\
h'(I^*) &= \max_{\nu} \frac{U(\nu, \mu)}{J(\nu, \mu)}
\end{align*}
\]
The optimal posterior \( \nu^* \) maximizes \( \frac{U(\mu, \nu)}{I(\mu, \nu)} \) — the value to uncertainty reduction ratio. Let \( \lambda = h'(I^*) = \max \frac{U(\mu, \nu)}{I(\mu, \nu)} \), then \( U(\mu, \nu) \leq \lambda J(\mu, \nu) \) and equality holds at \( \nu^* \).\(^{21}\) Define \( G(\mu) = V(\mu) + \lambda H(\mu) \). I call \( G(\mu) \) the gross value function. Then definition of \( U \) and \( V \) implies: \( U(\mu, \nu) - \lambda J(\mu, \nu) = G(\nu) - G(\mu) - G'(\mu)(\nu - \mu) \). Hence, gross value function has the following property:

\[
\begin{align*}
G(\nu) &\leq G(\mu) + G'(\mu)(\nu - \mu) & \forall \nu \in [0, 1] \\
G(\nu^*) &= G(\mu) + G'(\mu)(\nu^* - \mu)
\end{align*}
\]  

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

![Figure 9: Concavification of gross value function](image)

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

**Equation (9)** 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, while in 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 \( \lambda \). \( \lambda \) is the multiplier capturing the marginal effect of reducing uncertainty on flow cost. So solving:

\[
\sup_{p \geq 0, \nu} p(G(\nu) - G(\mu) - G'(\mu)(\nu - \mu))
\]

is equivalent to solving **Equation (8)**. Whether **Equation (10)** 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 (10)** is strictly positive. However, **Equation (10)** is linear in the signal arrival rate \( p \). As a result the DM has incentive to increase \( p \). Increasing \( p \) drives up marginal cost \( h'(\cdot) \). So when the optimum is reached, \( h'(\cdot) \) (or

\(^{21}\)With Assumption 2-a, \( I^* = c \) and \( \lambda = \max \frac{U(\mu, \nu)}{I(\mu, \nu)} \) is the Lagrangian multiplier for constraint \( I \leq c \).
\[ \lambda \) must be such that solving Equation (10) yields exactly zero utility: \( G(\mu) = \text{co}(G)(\mu) \). This characterization illustrates that in the continuous time limit, information is smoothed such that only infinitesimal amount of uncertainty is reduced at every instant of time.

Now suppose that the HJB is satisfied, i.e. Equation (8) equals the flow discounting loss \( \rho V(\mu) \). Then applying \( I^* = p^* \cdot J(\mu, v^*) \) and \( h'(I^*) = \frac{U(\mu, v^*)}{J(\mu, v^*)} \) to the HJB implies:

\[
\begin{align*}
\rho V(\mu) &= p^* \cdot U(\mu, v^*) - h(p^* \cdot J(\mu, v^*)) \\
\Rightarrow \rho V(\mu) &= I^* h'(I^*) - h(I^*)
\end{align*}
\]

Combining Equation (9) and Equation (11) pins down the value function \( V \) and corresponding strategies \( p, v \). Now I can analyze key trade-offs in dynamic information acquisition problem by studying Equations (9) and (11).

1. Utility gain v.s. information cost

Equation (11) illustrates the utility gain v.s. information cost trade-off. Since \( h \) is a convex function, \( I h'(I) - h(I) \) is increasing in \( I \). That is to say, 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 quite simple. The marginal cost of increasing informativeness of the signal proportionately is \( I h'(I) \). The marginal gain is from increasing arrival rates proportionately (keeping the signal precision fixed as in envelope theorem). Increasing arrival rate by unit proportion reduces waiting time by same proportion. So marginal gain from increasing \( I \) by unit proportion is discount \( \rho V \) plus cost \( h(I) \). At the optimum, marginal cost equates marginal gain, therefore we get Equation (11) and flow informativeness is monotonic in value function.

If we consider the case with Assumption 2-a, then \( \lambda \) in Equation (9) is replaced by shadow cost of increasing informativeness (see Footnotes 21 and 22). And Equation (11) can be written as \( \rho V(\mu) = c \lambda \). Although in this case intensity is fixed, there is a monotonicity between shadow cost and value function.

To sum up, by studying the utility gain v.s. information cost trade-off, I established a monotonicity between shadow/marginal cost \( \lambda \) and continuation value \( V(\mu) \). (I refer to both of them as “value-intensity monotonicity” for notational simplicity). Now that I characterized \( \lambda \), we can move on and focus on Equation (9).

2. Precision v.s. frequency

A novel trade-off characterized by Equation (9) is precision v.s. frequency trade-off. The value-intensity monotonicity determines \( I \) from value function. Now the DM allocates total informativeness \( I \) into precision (parametrized by size of belief jumps) and frequency (parametrized by arrival rate of jumps). Equation (9) suggests that optimal signal precision can be solved by concavifying gross value function \( G(\mu) \). In this section, I illustrate how this trade-off changes at different prior and explain the intuition behind it.

Figure 10 shows how varying \( \lambda \) affects the optimal size of jump. 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

---

22 With Assumption 2-a, \( h(I^*) = 0 \) and \( I^* = c \). So \( \rho V(\mu) = \lambda c \).

23 \[ \frac{d}{dt} (I h'(I) - h(I)) = I h''(I) \geq 0 \]
Figure 10: Precision-frequency trade-off

concave region. The prior $\mu$ and optimal posterior $\nu$ are on the boundary of a concavified region. Now 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 with a larger $\lambda$ the white region is expanding and the blue region is shrinking (see Figure 10-(c)). So prior and optimal posterior are closer to each other. Recall that $\lambda$ is monotonic $V$. This means with higher continuation value, the DM is more willing to choose signal inducing shorter belief jump.

The intuition for this property is as follows. When the DM is more sure about the state, continuation value is higher, hence utility loss from discounting is higher. The DM wants to receive a signal more frequently to enjoy the high value sooner. Therefore, the value of frequency outweighs the value of precision and the optimal strategy is to give up some precision to get higher arrival frequency. In this analysis, continuation value is isomorphic to $\lambda$, which controls the shape of $G$. Relative value from signal precision (to value from frequency) is determined by global concavity of gross value function. So the analysis presented earlier by Figure 10 exactly illustrates the intuition.

Confirming vs. contradicting: The analysis above pins down the absolute size of optimal belief jump. To pin down the optimal posterior, it remains to be seen which direction of jump is optimal. Now I show that the precision-frequency trade-off also implies the optimality of confirmatory learning.

Let us hypothetically consider a belief $\mu$ at which jumping towards right is optimal (weakly). In both panels of Figure 11, $\mu$ is the prior and $\nu_L, \nu_R$ are optimal posteriors on each side of $\mu$ respectively. So jumping to $\nu_R$ (the black arrow) is better than jumping to $\nu_L$ (the dashed black arrow). Let $V$ be increasing around $\mu$. Now consider the DM’s incentive at $\mu_1$ slightly larger than $\mu$ (on Figure 11-(a)). By an envelop theorem argument, $\nu_L$ and $\nu_R$ are optimal posteriors also for $\mu_1$ on each side. To pin down the optimal posterior for $\mu_1$, we just need to compare $\nu_L$ and $\nu_R$. Since $\mu_1 > \mu$, $\nu_R$ is closer to prior, while $\nu_L$ is further away from prior. Meanwhile, $V(\mu_1) > V(\mu)$ implies that the DM prefers frequency to precision more with belief $\mu_1$. Therefore, $\nu_R$ is strictly more preferred at $\mu_1$. Consider $\mu_2$ slightly smaller than $\mu$ (on Figure 11-(b)). Similar analysis shows that now size of jump to $\nu_R$ is larger, and DM prefers precision more with belief $\mu_2$. So $\nu_R$ is also strictly optimal for $\mu_2$.

In this analysis, jumping towards the direction that value function is increasing means the signal is confirmatory. When value function is quasi-convex, it is equivalent to prop-
Figure 11: Confirmatory v.s. contradictory

3. Poisson v.s. Gaussian

So far I ignored the possibility of Gaussian signals. In fact, they are implicitly modeled in Equation (9). Consider the optimization w.r.t. Gaussian signals:

\[
\sup_{\sigma} \sigma^2 V''(\mu) - h(-\sigma^2 H''(\mu)) \rightarrow \text{FOC: } V''(\mu) + \lambda H''(\mu) = 0 \\
\iff G''(\mu) = 0
\]

where \( \lambda = h'(-\sigma^2 H''(\mu)) \) with Assumption 2-a or \( \lambda = \frac{\ell}{\varphi} V(\mu) \) with Assumption 2-b. Comparing Equations (9) and (12), it is not difficult to notice that Equation (12) is exactly the limit of Equation (9) when optimal posterior \( \nu \) is converging to prior \( \mu \). This is intuitive since Gaussian signal can be approximated by a Poisson signal with very low precision and high arrival rate.

The comparison between Gaussian signal and Poisson signal is effectively the comparison between a special imprecise Poisson signal and other Poisson signals. So this trade-off is a special case of the precision v.s. frequency trade-off. Choosing Gaussian signal is a corner solution when the DM wants to trade off 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 Poisson signal with positive jump. Therefore, Gaussian signal is optimal only on boundaries of experimentation regions. Given this intuition, one could imagine that Gaussian signal is generically suboptimal except for special cases where there is no precision-frequency trade-off at all. Since the preference between precision and frequency depends on the loss from delaying, the trade-off diminishes only when the DM does not discount future payoffs. This intuition is confirmed in a no-discounting special case in Section 7.3 as well as in the model of Hébert and Woodford (2016).

4. Continuing v.s. stopping

Now consider the optimal stopping rule. Theorems 2 and 3 states that the optimal posteriors are in the stopping region, i.e., it is suboptimal to do repeated jumps. Now I prove by showing that repeated jumps can be improved by a direct jump. Let \( \nu \) be the optimal posterior for prior \( \mu \) (see Figure 12). Then Equation (9) and Equation (11) jointly implies that \( \frac{U_0}{f_0} = \frac{U_\nu}{f_\nu} = \lambda(\mu) \).

Hypothetically imaging that at \( \nu \) it is optimal to continue and optimal posterior is \( \nu' \). Then \( \frac{U_1}{f_1} = \lambda(\nu) \), and \( \lambda(\nu) > \lambda(\mu) \) by the confirmatory evidence property & value-intensity monotonicity. I want to show that this implies \( \frac{U(\nu, \nu')}{f(\nu, \nu')} = \frac{U_1 + U_1'}{f_1 + f_1'} > \lambda(\mu) \), i.e.
jumping to posterior $v'$ directly is strictly better than the two-step jump. By elementary geometry there exists $\alpha$ s.t $U'_1 = \alpha U_0$ and $J'_1 = \alpha J_0$. Therefore, the value to uncertain reduction ratio $\frac{U(J_0, J_1, \nu)}{J_0^\prime(J_0, J_1, \nu)}$ is a weighted average of $\frac{U_0}{J_0}$ and $\frac{U_1}{J_1}$, which is larger than $\lambda(\mu)$.

Now the intuition for the stopping rule is 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 jumps. As a result, sequentially jumping to higher values is dominated by directly jumping to the highest value.

**Remark 2.**

The intuition behind the value-intensity monotonicity is purely driven by convexity of cost function $h$ and is obviously independent to the formulation of the information measure. The intuition behind the optimality of Poisson signal over Gaussian signal is to use the precision-frequency trade-off to compare generic Poisson signal with extremely imprecise Poisson signal. It does not depend on the exact form of $I$ either. I generalize the optimality of Poisson to generic cost of information in Theorem 5, Section 7.1. I also discuss confirmatory evidence and immediate stopping properties with generic cost functions in Section 7.1.

The precision-frequency trade-off does not depend on the size of state space either. I confirm this with a general characterization result with more states Theorem 10 in Appendix A.3. However, the existence of solution to HJB equation relies on the binary state assumption. A constructive proof for the binary state case based on ODE theory is introduced in Section 6.2.2.

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

---

24 See Figure 12. $\frac{U_0}{J_0} = \frac{U_0}{J_0} = \lambda(\mu)$ implies $\frac{U_0}{J_0} = \lambda(\mu)$. Hence, $\frac{\mu}{\mu}\frac{U_1}{J_1}$. I assume the ratio to be $\alpha$. 

---
6.2.2 Sketched proof of Theorem 2

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

- **Step 1.** Determine $\mu^*$. $\mu^*$ can be calculated as essentially the unique belief at which $V'(\mu^*) = 0$ and searching for posteriors at either side of $\mu^*$ are equally good. Noticing that if $V'(\mu^*) = 0$, HJB implies:

\[
V(\mu^*) = \frac{F(v)}{1 + \frac{\rho}{c} J(\mu, v)}
\]

$\mu^*$ is solved as the unique belief at which:

\[
\sup_{\nu \leq \mu^*} \frac{F(v)}{1 + \frac{\rho}{c} J(\mu, v)} = \sup_{\nu \geq \mu^*} \frac{F(v)}{1 + \frac{\rho}{c} J(\mu, v)}
\]

$V(\mu^*)$ and $v(\mu^*)$ are all pinned down correspondingly.

- **Step 2.** Search for optimal posterior for a fixed action. Let $a$ be the optimal action for optimal posterior $\nu \geq \mu^*$ solved from step 1. Let $F_a(\mu) = E_{\mu}[u(a, x)]$. Now solve for optimal posterior $\nu$ with payoff $F_a(\nu)$:

\[
\rho V(\mu) = \max_{\nu \geq \mu} -c \frac{F_a(\nu) - V(\mu) - V'(\mu)(\nu - \mu)}{H(\nu) - H(\mu) - H'(\mu)(\nu - \mu)}
\]

The primitives in objective function are all sufficiently smooth in $\nu$. Then, first order condition w.r.t. $\nu$ yields a well behaved first order ordinary differential equation characterizing $v(\mu)$ with initial condition $v(\mu^*)$. Therefore we can solve for optimal policy $\nu$ and calculate value $V(\mu)$ accordingly for $\mu \geq \mu^*$. $V(\mu)$ and $v(\mu)$ for all $\mu \leq \mu^*$ is 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 starts from solving the ODE defined in step 2 at $\mu^*$. Then I extend the value function towards $\mu = \{0, 1\}$. Whenever I reach a belief at which two actions yield 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 $\mu^*$).

Solving the ODE characterizing $v(\mu)$ directly implies monotonicity of $v(\mu)$ in each connected experimentation region. Now I need to verify the optimality of the constructed strategy. The verification takes three steps, each ruling out repeated jumps, contradictory evidence and Gaussian signals respectively. The intuitions for suboptimality of these three alternative strategies are explained in Section 6.2 already. The formal proof is relegated to Appendix B.3.
7 Discussions
In this section, I discuss in details the assumptions I make in the baseline model: Assumptions 1, 2 and 3. All assumptions I made can be categorized into three classes:
1. Economic assumptions:
   - Informativeness measure (Assumption 1, or equivalently Assumption A).
   - Convexity of cost function (Assumption 2).
   - Discounting (positive $\rho$).
2. Restrictive assumptions: Finite actions and binary states (Assumption 3).

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

The restrictive assumptions do restrict the generality of my model. However, relaxing them does not fundamentally alter the key intuition, and the methodology generalizes. The discussions for these assumptions are relegated to the appendix. In Appendix A.2, I relax the finite action assumption. I showed 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 for existence no longer works with general state space, I showed that the properties in Theorem 2 extend to general state space. Technical assumptions do not restrict my model in a meaningful way so I will not discuss them.

7.1 Posterior separable measure
In this section, I first provide an axiom for Assumption A and extend Assumption 1 to generic flow information measures.

7.1.1 Axiom for Assumption A
Theorem 4. $I(S; X|\mu)$ is a non-negative function of information structure and prior belief. It satisfies Assumption A if and only if the following axiom holds:

Axiom: $\forall \mu, \forall$ information structure $S_1$ and information structure $S_2|S_1$ whose distribution depends on realization of $S_1$:

$$I((S_1, S_2); X|\mu) = I(S_1; X|\mu) + E[I(S_2; X|S_1, \mu)]$$

Theorem 4 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.

28
Given Theorem 4, Assumption A is essentially a consistency requirement on the cost of compound experiments. Technically, Assumption A (Assumption 1) helps me throughout the whole analysis. First, separability of information measure establishes Lemma S.1. It helps me eliminate redundant information and inter-temporal complexity to establish equivalence between continuous time model and limit of discrete time model. Second, the methodology of concavifying "gross value function" is only possible when expected utility gain and information measure takes a consistent form. To accommodate an even larger set of possible information measures, I study a problem with more general information measure in next section.

7.1.2 General informativeness measure

I setup a continuous time HJB equation with a generic information cost structure which imposes no specific link between prior and posterior. I want to show that one key feature of the baseline model is generic — the optimality of Poisson learning. Let \( J(\mu, \nu) \) and \( \kappa(\mu, \sigma) \) be bivariate functions. Consider the following functional equation:

\[
p \nu V(\mu) = \max \left\{ pF(\mu), \sup_{\nu, \sigma^2} p(V(\nu) - V(\mu) - V'(\mu)(\nu - \mu)) + \frac{1}{2} \sigma^2 V''(\mu) \right\}
\]

s.t. \( pJ(\mu, \nu) + \kappa(\mu, \sigma) \leq c \)  

Equation (13)

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 same as before. I assume \( J(\mu, \nu) \) to be an arbitrary function which is both prior and posterior dependent. Cost of diffusion signal is \( \kappa(\mu, \sigma) \). I impose the following assumptions on \( J(\mu, \nu) \) and \( \kappa(\mu, \sigma) \):

**Assumption 4.**

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

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”.\(^{25}\) \( J''(\nu, \mu) = 0 \) and \( J''(\nu, \mu) > 0 \) are implications of “any informative Poisson signal is costly”. Within this continuous time framework, these assumptions imposed on \( J \) are without loss of generality. The crucial assumption is the third condition: \( \kappa(\mu, \sigma) = \frac{1}{2} \sigma^2 J''(\nu, \mu) \). This assumption is essentially saying that the cost functional is 'continuous' in the space of signal structures. Consider a Poisson signal \((p, \nu)\). When \( \nu \to \mu \), the utility gain from learning this signal is:

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

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

\[
pJ(\mu, \nu) = p\left( J(\mu, \mu) + J''(\mu, \mu)(\nu - \mu) + \frac{1}{2} J''(\nu, \mu)(\nu - \mu)^2 + O(|\nu - \mu|^3) \right)
\]

\(^{25}\)In this setup, \( J(\mu, \mu) = 0 \) is WLOG. If uninformative signal has strictly positive cost, we can always shift capacity constraint \( c \) to normalize \( J(\mu, \mu) \) to 0.
\[ = \frac{1}{2} p(v - \mu)^2 f''(\mu, \mu) + pO(|v - \mu|^3) \]

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

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

\[ L(\mu) = \frac{\rho}{c} f''(\mu, \mu)^2 - \frac{2f^{(3)}(\mu, \mu)^2 + f^{(4)}(\mu, \mu)}{f''(\mu, \mu)} \]

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

\[ \rho V(\mu) = c \frac{V''(\mu)}{f''(\mu, \mu)} \]

is of zero measure.

The interpretation of Theorem 5 is that Poisson signal is almost always strictly superior to 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 interval only when \( J(\mu, \nu) \) is locally solution to the PDE. Solution to a 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 diffusion signal almost nowhere.

A trivial sufficient condition for \( L(\mu) \neq 0 \) is Assumption 1. Assumption 1 implies that \( J^{(2)}(\mu, \nu) \) is invariant in \( \mu \). In this case \( L(\mu) = \frac{\rho}{c} f''(\mu, \mu)^2 > 0 \) for sure. 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: \( \| f(\cdot) \|_\delta = \sup_{x \in [\delta, 1]} \{ |f(x, x)|, \| \nabla f(x, x) \|_{L_2} \} \).

**Corollary 5.1.** Given Assumption 4, suppose \( V \in C^{(3)}(0,1) \) solves Equation (13), then for any \( \delta > 0 \), there exists \( \varepsilon \) s.t. if \( \| J^{(3)}(\mu, \mu) \|_\delta \leq \varepsilon \):

\[ \mu \in [\delta, 1-\delta] | \rho V(\mu) = c \frac{V''(\mu)}{f''(\mu, \mu)} \]

is of zero measure.

The condition in Corollary 5.1 states that \( J''(\mu, \nu) \) is approximately a constant over \( \mu \) for \( \nu \) close to \( \mu \). This result illustrates my analysis in Section 6.2.1, that the comparison of Poisson and Gaussian signal relies only on 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 \( \Delta \) s.t. if \( \frac{\rho}{c} \geq \Delta \):

\[ \mu \in [\delta, 1-\delta] | \rho V(\mu) = c \frac{V''(\mu)}{f''(\mu, \mu)} \]

is of zero measure.
In Section 6.2.1, I show that the precision-frequency trade-off is the main driving force for the comparison between Gaussian signals and Poisson signals. When $\ell$ is high, the dependence of discount loss on continuation value is magnified. Hence, in this case Poisson is more likely to be optimal even though the information cost could be distorted in an arbitrary way that is in favor of Gaussian signals.

Although Poisson learning is generic optimal, immediate action and confirmatory evidence are not. They are implications of posterior separability (or Assumption 1). Imagine a case in which signals of high precision are relatively cheap (say an extreme case where $J(\mu, \nu)$ is truncated both below and above). Then, when prior is close to boundary of experimentation region, seeking for confirmatory evidence (with low precision and high frequency) results in very high cost, while seeking for a precise contradictory signal is quite cheap. Searching for the precise contradictory signal induces belief to drift quickly towards the more likely state, which effectively enables quick confirmation. Therefore, confirmatory evidence seeking is dominated. In fact, this example shares same intuition with the findings in Che and Mierendorff (2016). In their setup, there is no cost on signal but total attention is bounded. When allocating more attention to signals revealing the state of higher prior probability, DM is effectively increasing arrival rate of signal. Of course signals revealing the state of higher prior induces shorter jumps, thus less precise and more frequent. So the precision-frequency trade-off is still valid and drives confirmatory evidence seeking in intermediate region. On boundary, since total arrival rete is bounded and signals are costless, it is as if contradictory signal is free but high frequency confirmatory signal is infinitely expensive. As a result contradictory evidence becomes optimal.

On the other hand, consider the immediate action property. Imagine a case in which signals of low precision is relatively cheap. Then, break a long jump of posterior into multiple short jumps might become profitable. Immediate action property is named single experiment property (SEP) in Che and Mierendorff (2016). In their paper, it is also documented that SEP is not a robust property in a generic Poisson learning model.

### 7.2 Linear flow cost

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

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

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

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

$$V(\mu) = \sup_{p \in \Delta^2(X)} \mathbb{E}_p[F(\nu)] - \lambda \mathbb{E}_p[H(\mu) - H(\nu)]$$  \hspace{1cm} (14)

The intuition for this result is simple. At any instant of time, suppose it is optimal to continue learning for positive amount of time. The value is discounted future value at next instant of time $(t + dt)$ less flow cost of information. Now consider moving the
learning strategy at \( t + dt \) to current period. Then both future value at \( t + dt \) and cost are discounted \( dt \) less. If the net utility gain from learning at \( t + dt \) is non-negative, then this operation increases current utility by reducing waiting time.\(^{26}\) If the net utility gain from learning at \( t + dt \) 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.\(^{27}\)

In fact, given Assumptions 1 and 2-b, Equation (1) is a variant of the main model in Steiner et al. (2016) with the unknown payoff relevant state being constant. With linear cost function \( h(I) \), there is essentially no dynamics generated from the motivation of smoothing learning behavior. The problem in Steiner et al. (2016) does not reduce to a period-by-period myopic RI problem mainly because of the assumption that action affects future utilities.

### 7.3 No discounting

As discussed in Section 6.2, discounting is one key factor driving all the dynamics. With exponential discounting, the trade-off between arrival frequency and precision of signals changes according to continuation value. One can imagine that if we replace exponential discounting with linear discounting, i.e. the DM pays a fixed per period waiting cost, time distribution of utility gain and information cost does not matter for the DM anymore. In fact this conjecture is correct. Consider the following problem:

\[
V(\mu) = \sup_{(\mu_\tau) \in \mathcal{M}, \tau} \mathbb{E} \left[ F(\mu_\tau) - m \tau - \int_0^\tau h(I_t) dt \right] \tag{15}
\]

**Theorem 7.** Given Assumptions 1 and 2, suppose \( V(\mu) \) solves Equation (15), then:

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

**Theorem 7** illustrates that solving Equation (15) is equivalent to solving a static rational inattention problem, with \( \frac{m + f(\lambda)}{\lambda} \) being the marginal cost on information measure (see Caplin and Dean (2013) and Matejka and McKay (2014)). The optimal value function can be achieved through many different learning strategies. In fact, assuming \((P^*, \lambda^*)\) to be the solution to the problem in **Theorem 7**, then all dynamics information acquisition strategy that eventually implement \(P^*\) (i.e. \(\mu_\infty \sim P^*\)) and incur flow cost \(\lambda^*\) achieve same utility level \(V(\mu)\).\(^{28}\)

**Equation (15)** is in fact the dynamic learning foundation provided in Hébert and Woodford (2016) to justify Gaussian learning.\(^{29}\) It is illustrated in this problem that no-discounting is a knife-edge case where almost all information acquisition strategies, including Gaussian learning, are equally good. So no-discounting is also a special case.

---

\(^{26}\)This step utilizes Assumption 2', which implies that cost of a combined signal structure is sum of cost of each of them.

\(^{27}\)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 arbitrarily short period of time, and the payoff approximates the immediate learning payoff.

\(^{28}\)This result is stated and proved formally in Zhong (2018b).

\(^{29}\)In Hébert and Woodford (2016), informativeness measures that are not posterior separable (Assumption 1 and Assumption A) are also considered in the Appendix.
where no behavior prediction can be made. On the contrary, Poisson learning becomes
strictly optimal when there is discounting, no matter how low the rate $\rho$ is. Moreover, the
analysis in Section 7.1.2 suggests that for more general cost structure, Gaussian learning
is strictly dominated by Poisson learning when there is sufficiently high discount factor.
This analysis gives us a more complete picture: discounting generically generates the
precision-frequency trade-off. And optimality of Poisson learning is a result of the trade-off.
Only in a special case when there is not discounting, there is no such trade-off and all
learning dynamics are essentially equivalent.

The analysis so far illustrates how the modeling assumption on impatience affects the
optimal learning dynamics. A further generalization of time preference is provided by
Zhong (2018b), which relates optimality of Poisson to risk loving on the time dimension.
It is shown in Zhong (2018b) that within the set of all decision time distributions induced
by dynamic learning strategies implementing a same target information structure, all dis-
tributions have the same expected decision time and Poisson learning generates the most
dispersed (in MPS sense) decision time distribution.

8 Applications
8.1 Choice accuracy and response time

The two-choice sequential decision making problem has been extensively studied in
Psychology and behavioral studies. One of the key objective is to explain the data on
choice accuracy and response time from laboratories. The drift-diffusion model (DDM)
has been the most popular theoretical model for these decision problems, for the reason
that it is very tractable and fits the accuracy/ response time data well (see the survey by
Ratcliff, P. L. Smith, et al. (2016)). However, accounting for the joint distribution of choice
accuracy and response time is a challenge for DDM. In this section, I apply my model to
predict one systematic feature: the crossover of response time-accuracy relationship.

The crossover happens when the difficulty of decision problem varies: error responses
are faster than correct responses when the task is easy; error responses are slower than
correct responses when the task is hard (see Luce et al. (1986), Ratcliff, Van Zandt, et al.
(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 $\mu_0 = 0.5$. Let $H_0(\mu)$ be the Entropy function. Define un-
certainty 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 $\mu > 0.65$. Different difficulty levels are modeled as different capacity con-
straints on $-LH(\mu_i)$, 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 points to the right of $p = 0.5$ shows the choice prob-
ability of action $R$ (the correct choice). Each such point has a corresponding point to the
Figure 13: LP and QP plots

left of $p = 0.5$ showing the remaining probability of action $L$ (the error). The vertical co-
dordinates of all points show response time measured by mean (in LP plot) or by quantiles
(in QP plot).

The crossover of time-accuracy relationship is illustrated by differently colored points. The red points are data points where errors happen earlier than correct response (mea-
sured 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 where errors
happen later than 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

Figure 14: Critical beliefs of different difficulty levels

The main reason for the crossover is explained in Figure 14. When capacity is low
(difficulty is high), the optimal size of belief jump is small. By construction of $H(\mu)$, when
posterior belief is not far away from $\mu_0$, learning state $L$ is more costly than learning state
$R$. As a result, the critical belief $\mu^*$ at which searching for both direction is indifferent is
biased towards left. Since $\mu_0 > \mu^*$, correct responses are font-loaded. Applying the same
intuition, when capacity is high, $\mu_0 < \mu^*$ and errors are font-loaded.
Applying the idea from Example 5, to create a crossover of response time-accuracy relationship, one needs to create a crossover of $\mu^*$ and $\mu_0$.

**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 $\mathbb{R}^+ : \{0, c_1, \ldots, c_k, \infty\}$, 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 $\mu_0$ is the same 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 arbitrary number of crossovers of response time-accuracy relationship at given difficulty levels. It is well known that the standard DDM predicts identical decision time distribution for correct responses and errors (Ratcliff (1981)). To accommodate non-trivial speed-accuracy trade-off/complementarity, DDM with varying boundary (Cisek et al. (2009)) or DDM with random starting point and drift (Ratcliff and Rouder (1998)) are proposed, and there are a lot of debate which variation works better. Fudenberg et al. (2015) shows that collapsing (expanding) boundary maps exactly to complementarity (trade-off), and the endogenously optimal boundary shape depends on initial location. 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 task payoffs and learning technology the 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 radicality of firm’s R&D and innovation to its safe option. Consider two firms facing a decision of commercializing different new products. The two firms each facing an identical decision problem, except that one firm has a better safe option. I call the firm with better safe option the *large firm* (L) and the other the *small firm* (S).\(^{30}\) Intuitively, there are two competing incentives:

1. **Impatience effect**: The large firm has overall higher continuation value than the small firm. Therefore, by the value-precision monotonicity I developed in this paper, the more impatient large firm should prefer the frequency of signal to precision of signal. So the impatience effect suggests that smaller firm innovates more radically.

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

I model the problem using the following setup. The state is $x \in \{G, B\}$. There is one safe product $P_s$ and $K$ risky products $\{P_1, \ldots, P_K\}$. When $x = G$, the new technology

\(^{30}\)I think of the safe option as the status quo in the market. The risky product, once successfully commercialized, help the firm grab the whole market. So both firms have the identical risky option, but the small firm has a worse status quo.
The information acquisition strategy in the main model is naturally interpreted as the R&D strategy of firms. I am interested in how the two firms’ R&D strategies differ in their radically. Let $v_i(\mu)$ be the two firms’ optimal strategies. I define that a firm is looking for more radical innovation given belief $\mu$, if $|v_i(\mu) - \mu| > |v_{i-}(\mu) - \mu|$, namely firm $i$ is searching for a more precise Poisson signal. Let $E$ be the union of the two firms’ experimentation regions.

**Example 6.** I calculate a simple example. There is only one risky product and $K = 1$. The $L$ firm’s safe option pays $u_L(P_s, x) = 0.3$ and the $S$ firm’s safe option pays $u_S(P_s, x) = 0.15$. 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** depicts the value functions (red curve: large firm; blue curve: small firm). The two dashed lines are payoffs of corresponding safe options. **Figure 16** depicts the policy functions (red curve: large firm; blue curve: small firm). There is clearly a crossover of policy functions. In the union of two firm’s experimentation regions, when $\mu < \mu_c$ the small firm seeks more radical innovation, when $\mu > \mu_c$ the large firm 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 $\mu_c$ s.t. $\forall \mu \in E, \mu > \mu_c \implies |v_L(\mu) - \mu| > |v_S(\mu) - \mu|$ and $\mu < \mu_c \implies |v_L(\mu) - \mu| < |v_S(\mu) - \mu|$. Moreover, $E_0 \cap (0, \mu_c) \neq \emptyset$ and $E_0 \cap (\mu_c, 1) \neq \emptyset$.

Proposition 2 first states that there exist a threshold belief that the large firm 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 belief changes. Here is the intuition for the crossover: It is not hard to see that the small firm’s value function is always steeper than the large firm’s. So the difference in

---

31It is trivial that if the cost of R&D is flexible, the large firm 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.
continuation value is decreasing in belief. As a result, the impatience effect is diminishing when $\mu$ increase. On the other hand, when $\mu$ 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 $\mu$ increase. Therefore, with $\mu$ increases, the large firm is increasingly favoring more precise signal, comparing to the small firm. Hence 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 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 $\mu_c$ s.t. $\forall \mu \in E_0, \mu > \mu_c \implies |V_L(\mu) - \mu| > |V_S(\mu) - \mu|$ and $\mu < \mu_c \implies |V_L(\mu) - \mu| < |V_S(\mu) - \mu|$. Moreover, $E_0 \cap (0, \mu_c) \neq \emptyset$ and $E_0 \cap (\mu_c, 1) \neq \emptyset$.

9 Conclusion

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


Che, Y.-K., & Mierendorff, K. (2016). Optimal sequential decision with limited attention.


Luce, R. D. et al. (1986). *Response times: Their role in inferring elementary mental organization.* Oxford University Press on Demand.


Matyskova, L. (2018). Bayesian persuasion with costly information acquisition. unpublished manuscript.


A Further discussions

A.1 Convergence of policy

A.2 Infinite action space

A.3 General state space

B Omitted proofs

B.1 Roadmap for proofs

B.2 Proof of Theorem 1

B.3 Proof of Theorem 2

B.2.1 Proof of Lemma 1

B.2.2 Proof of Lemma 2

B.2.3 Proof of Lemma 3

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. \( \forall x \in X, u(a, x) \in C(A) \cap TB(A) \).

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, \( V_{dt}(F) \) is the unique solution of Equation (7) and \( V(F) = \lim_{dt \to 0} V_{dt}(F) \).

Lemma A.1. Given Assumption A and Assumptions 2 and 5, \( V \) is a Lipschitz continuous functional under \( L_\infty \) 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 9. 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 \( |V(F) - V(F_n)| \to 0 \) and:
1. \( V(F) \) solves Equation (4).
2. \( \forall \mu \text{ s.t. } V(\mu) > F(\mu), \text{ if each } v_n \text{ is maximizer of } V(F_n) \text{ and } v = \lim_{n \to \infty} v_n \text{ exists, then } v \text{ is the optimal posterior in Equation (4) at } \mu \).

Theorem 9 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 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 policies 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.

\[ \text{The existence of limit is guaranteed by monotonic convergence theorem.} \]
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. 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:

$$\rho V(\mu) = \max_{v, \sigma} \left\{ \rho F(\mu), \max_{v, \sigma} (V(v) - V(\mu) - \nabla V(\mu) \cdot (v - \mu)) + \sigma^T \nabla H(\mu) \sigma \right\}$$  \hspace{1cm} (16)

where $v \in \Delta(\supp(\mu))$, $p \in \Delta I$ and $\sigma \in \mathbb{R}^{[\supp(\mu)]}$. Equation (16) comes from applying Assumption 2-a and smoothness condition to Equation (4). 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 10.** 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 (16), then $\exists$ policy function $v : E \rightarrow \Delta(X)$ s.t.

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

and $v$ satisfies the following properties:

1. $D_{v(\mu)} V(\mu) \geq 0.$
2. $v(\mu) \in E^C.$
3. $D_{v(\mu)} V(\mu) \cdot \nabla H(v(\mu) \cdot (v(\mu) - \mu) \leq 0.$
4. $\rho V(\mu) \geq \sup_{\sigma} -c \sigma^T H(\mu) \sigma.$

There exists a nowhere dense set $K$ s.t. strict inequality holds on $E \backslash K$ in property 1, 3 and 4.

**Theorem 10** states that if a solution $V(\mu)$ to Equation (16) 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 2 and 4 are exactly the immediate action property and the suboptimality of Gaussian signal. Property 1 and 3 are weaker than the corresponding properties in Theorem 2. Property 1 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

---

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.

$HH(\mu)$ is defined on boundary where $V(\mu) = F(\mu)$ as continuous extension of interior Hessian’s by Kirszbraun theorem.
extension to the increasing precision property. $D_{\mu-v}(\mu)$ is the direction $v$ is moving when $\mu$ is moving against $v$. $H H(v)(v-\mu)$ is the direction $(v-\mu)$ distorted by a negative definite matrix $H H(v)$. In a special case when $H(\mu) = \|\mu-\mu_0\|^2_2$, $H H(v)(v-\mu)$ is in the same direction as $(\mu-v)$, which implies (together with property 3) that the distance between $\mu$ and $v$ is increasing when $\mu$ is drifting against $v$. In a generic case, this property does not directly predict how $\|v-\mu\|$ changes.

Figures 19 and 20 illustrate Theorem 10 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 posterior 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
Figure 21: Roadmap for proofs

B.1 Roadmap for proofs

Theorem S.1, P12

Lemma 1, P15
  Lemma 2, P15
  Lemma S.5, PS11
  Lemma S.6, PS12
  Theorem S.1, PS2

Lemma 3, P16
  Lemma S.9, PS16
  Lemma S.10, PS17
  Lemma S.7, PS13

Theorem 2, P17

Lemma B.1, P57
  Lemma B.2, P58
    Lemma S.16, PS21
    Lemma S.15, PS20
    Lemma S.17, PS23
  Lemma B.2’, PS24
    Lemma S.16’, PS24
    Lemma S.15’, PS20’
    Lemma S.17’, PS25
  Lemma S.18, PS25
  Lemma S.18’, PS27
  Lemmas S.11, S.12, S.13 and S.14, PS19

Theorem 3, P18

Lemma S.20, PS31
  Lemma S.21, PS34
    Lemma S.22, PS35
    Lemma S.21’, PS36
    Lemma S.22’, PS37
  Lemma S.23, PS37
  Lemma S.23’, PS38
    Lemmas S.11, S.12 and S.13, PS19’
  Lemma S.19, PS28
    Lemmas S.11 and S.12, PS18

Theorem 4, P28

Theorem 5, P30

Theorem 6, P31

Theorem 7, P32

Theorem 8, P41

Theorem 9, P42

Lemma A.1, P42

Lemma S.27, PS54
  Lemma S.26, PS53
    Lemma B.1, P57

Theorem 10, P43

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 (7) 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 (7). 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 5.2, it is sufficient to show that the order of limits can be switched:

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

(17)

Here $W_{dt}(\mu_t, \tau)$ is defined in Section 5.2 as the discretized payoff of continuous time strategy $\langle \mu_t \rangle, \tau$. The inner limit of LHS in Equation (17) 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 Appendix 5.5, a dependence lemma for Lemma 2). So RHS is $\lim V_{dt}$ (a technical lemma Lemma 5.8 guarantees existence of such limit).

I prove by showing inequality in two directions. The direction $V(\mu) \leq \lim V_{dt}(\mu)$ is trivial since $W_{dt}(\mu_t, \tau) \leq V_{dt}(\mu)$ for all $\langle \mu_t \rangle, \tau, dt$. The key is to prove the other direction $V(\mu) \geq \lim V_{dt}(\mu)$. 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 $\mu_t^* \text{ and } \tau_t^*$, where $\mu_{i+1}^* | \mathcal{F}_i$ has support size $N$. The goal is to construct an admissible continuous-time belief process $\langle \tilde{\mu}_t \rangle$, which satisfies two properties: 1) at each discrete time $idt$, $\mu_t$ has exactly the same distribution as $\mu_{i+1}^*$, 2) within each $dt$ period, uncertainty reduction speed of $\mu_t$ is exactly $E[H(\mu_t^*) - H(\mu_{i+1}^* ) | \mathcal{F}_i]/dt$. Such $\langle \mu_t \rangle$ with stopping time $\tau_t^*$ achieves higher payoff than $V_{dt}(\mu)$. Now this construction can be done by a technique introduced in Lemma 5.3. We will show that $\langle \tilde{\mu}_t \rangle$ is a continuous time martingale (with a corresponding probability space) satisfying: $\forall s, t \in [0,1], s > t$: $E[H(\mu_t^*) - H(\mu_{i+1}^*) | \mathcal{F}_i] = (s - t)E[H(\mu_t^*) - H(\mu_{i+1}^*) | \mathcal{F}_i]$. For $t \in [idt, (i+1)dt]$, define $\mu_t | \mathcal{F}_{idt} = \tilde{\mu}_{i}^{idt} | \mathcal{F}_i$. Therefore, $V(\mu) \geq \lim E \left[ e^{-\rho T} F(\mu_T) - \int_0^T e^{-\rho \mu(l_s)} dt \right]

\geq \lim E \left[ e^{-\rho dt} F(\mu_{dt}^*) - \sum_{t=0}^{T} \left( \frac{H(\mu_{t}^*) - \sum \sigma_t^j H(\mu_{t+1}^*)}{dt} \right) e^{-\rho dt \cdot t} \frac{1 - e^{-\rho dt}}{\rho} \right]$

46
\[
\begin{align*}
\geq & E \left[ e^{-\rho dt} F(\mu_{t^*}) - \sum_{t=0}^{T} h \left( \frac{H(\mu_{t^*}) - \sum p_i^j H(\mu_{t+1}^i)}{dt} \right) e^{-\rho dt} dt \right] \\
= & E \left[ e^{-\rho dt} F(\mu_{t^*}) - \sum_{t=0}^{T} h_{dt} \left( H(\mu_{t^*}) - \sum p_i^j H(\mu_{t+1}^i) \right) e^{-\rho dt} \right] = V_{dt}(\mu)
\end{align*}
\]

Second inequality is from \(1 - e^{-x} \leq x\). Therefore, \(V(\mu) \geq \lim V_{dt}(\mu)\). 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 5.2) is not necessary for my analysis of Theorem 1. Suppose now I extend the set of admissible belief profiles \(M\) to satisfy only the first two conditions: cadlag path, martingale property and initial value \(\mu_0 = \mu\). Then the limit of finite Riemann sum \(W_{dt}(\mu, \tau)\) might not exist (although each finite Riemann sum is always well defined). Whenever this is the case, I define the payoff of strategy \(\langle \mu_t \rangle\) as:

\[
\begin{align*}
E \left[ e^{-\rho \tau} F(\mu_{t^*}) - \int_0^\tau e^{-\rho t} h(-L^t_H(\mu_{t^*})) dt \right] = \limsup_{dt \to 0} W_{dt}(\mu, \tau)
\end{align*}
\]

Since \(W_{dt}(\mu, \tau)\) is bounded above by \(maxF\), Equation (18) is always well defined. Equation (18) 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 \(\hat{V}(\mu) = \sup \limsup_{dt \to 0} W_{dt}(\mu, \tau)\).

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 \(\hat{V}(\mu) = \lim_{dt \to 0} V_{dt}\), then Theorem 1 still holds with \(V\) replaced with \(\hat{V}\). This extension is quite trivial by observing \(\forall \langle \mu_t \rangle, \tau, dt, W_{dt}(\mu, \tau) \leq V_{dt}(\mu) \implies \limsup W_{dt}(\mu, \tau) \leq \lim V_{dt}(\mu) \implies \hat{V}(\mu) = \limsup \leq \lim V_{dt}(\mu)\).

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**

**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), \(S_t\) 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 Blackwell contract mapping condition. This is trivial as the payoff is bounded by \(maxF\) and discounted exponentially.

*Step 3.* Verify the Blackwell contract mapping condition. The contraction parameter in Equation (7) 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 (7). The final bits show that I can restrict the optimal strategy of Equation (7) to have support size \(N\). This part is proved using a generalized concavification result: Notice that the objective function in Equation (7) is not in the standard “expected valuation” form as in the literature of information design (see Kamenica and Gentzkow (2009)). Instead, there is an extra \(h_{dt}^j(\cdot)\) term. However, intuitively this problem can be handle using a Lagrange method and take the term inside \(h_{dt}^j(\cdot)\) 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 et al. (1992)). First define a space of functions on \(\Delta(X)\):

\[
\mathcal{L} = \left\{ V: \Delta(X) \to \mathbb{R}^+ \mid \forall \mu \in \Delta(X), \mu' \in \Delta(\text{supp}(\mu)), \limsup_{\mu' \to \mu} \frac{|V(\mu') - V(\mu)|}{\|\mu' - \mu\|} \in \mathbb{R} \right\}
\]
where $\|\cdot\|$ is Euclidean norm on $\Delta X$. 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 $\bar{V} \in \mathcal{L}$ which is the uniform limit of $V_{dt}$. Now I show that $\bar{V}$ coincides with the solution of the HJB equation. Consider the following HJB equation defined on $\mathcal{L}$:

$$
\rho V(\mu) = \max \left\{ \rho F(\mu), \sup_{\nu \in \Delta(\text{supp}(\mu)), p_i \in \mathbb{R}^+} \sum_{p_i} \left( V(\nu) - V(\mu) \right) - D\nu \left( \mu, \sum_{p_i} p_i v_i - \mu \right) \right\} + \frac{1}{2} \| \bar{s} \|^2 D^2 V(\mu, \bar{s}) \tag{19}
$$

where $\nabla$ and $H$ denote gradient and Hessian operator (well-defined on all interior points). Since $\bar{V}$ is not necessarily differentiable, I use operator $D$ and $D^2$ to replace the Jacobian and Hessian operators on $\bar{V}$. $D$ 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). $\forall f \in \mathcal{L}$:

$$
Df(x,y) = \liminf_{\delta \to 0} \frac{f(x+y) - f(x-y) - \delta f(x)}{\| y \|}, \\
D^2f(x,y) = \limsup_{\delta \to 0} 2 \frac{f(x+y) - f(x-y) - \delta f(x)}{\| y \|^2}.
$$

Notice that if $f \in C^1(\Delta X)$, then $Df(x,y) = \frac{\nabla f(x,y)}{\| y \|}$. If $f \in C^2(\Delta X)$ then $D^2f(x,y) = \frac{y^T Hf(x,y)}{\| y \|^2}$. It is not hard to verify that for $C^1$ smooth value function $V(\mu)$, Equation (19) is equivalent to Equation (4).

**Proof.**

Consider Lemma 3 by replacing Equation (4) with Equation (19). If the statement is proved with Equation (19), then since $\bar{V} = V$ is $C^1$ smooth, $\bar{V}$ is smooth and Equation (4) automatically holds. I prove by induction on dimensionality of $\text{supp}(\mu)$. First of all, Lemma 3 is trivially true when $\mu = \bar{x}$ since $V(\mu) = \bar{V}(\mu) = F(\mu)$ when the state is deterministic. Now it is sufficient to prove $\bar{V} = V$ on interior of $\Delta X$ conditional on $\bar{V} = V$ being true on $\partial \Delta X$ (boundary of $\Delta X$).

The proof takes three steps. Before going to the details, I introduce the steps briefly. The first step is to show that $\bar{V}$ is unimprovable in HJB Equation (19). The proof is quite standard as any continuous-time strategy that improves $\bar{V}$ can be approximated by a discrete-time strategy. The second step shows $\bar{V} \succeq V$. Proof is by a standard contradiction argument. If $\bar{V} < V$, then there exists a belief s.t. the same strategy implements strictly higher HJB with $\bar{V}$, which violates unimprovability. The last and most difficult step is to show that $V \succeq \bar{V}$.

**Unimprovability:** First I show that $\bar{V}$ is unimprovable in Equation (19). Suppose for the sake of contradiction that $\bar{V}$ is improvable at interior $\mu$, then there exists $p_i, v_i, \bar{s}, c$ such that:

$$
\rho \bar{V}(\mu) < \sum_{p_i} p_i \left( \bar{V}(v_i) - \bar{V}(\mu) \right) - D\bar{V}(\mu, \bar{s} - \sum_{p_i} p_i v_i) \left\| \sum_{p_i} p_i v_i - \mu \right\| + \sum_{p_i} D^2\bar{V}(\mu, \bar{s}) \left\| \bar{s} \right\|^2 - h(c)
$$

where $c = -\sum_{p_i} p_i \left( H(v_i) - H(\mu) - \nabla H(\mu) \cdot (v_i - \mu) \right) - \sum_{p_i} \bar{s}^T H H(\mu) \bar{s}$. Then if we compare the following two ratios:

$$
\frac{\sum_{p_i} p_i \left( \bar{V}(v_i) - \bar{V}(\mu) \right) - D\bar{V}(\mu, \bar{s} - \sum_{p_i} p_i v_i) \left\| \sum_{p_i} p_i v_i - \mu \right\|}{-\sum_{p_i} p_i \left( H(v_i) - H(\mu) - \nabla H(\mu) \cdot (v_i - \mu) \right)}; \quad \frac{D^2\bar{V}(\mu, \bar{s}) \left\| \bar{s} \right\|^2}{-\bar{s}^T H H(\mu) \bar{s}}
$$

At least one of them must be larger than $\frac{\rho \bar{V}(\mu) + h(c)}{c}$. 

**Case 1:**

$$
\sum_{p_i} p_i \left( \bar{V}(v_i) - \bar{V}(\mu) \right) - D\bar{V}(\mu, \bar{s} - \sum_{p_i} p_i v_i) \left\| \sum_{p_i} p_i v_i - \mu \right\| > \frac{\rho \bar{V}(\mu) + h(c)}{c}.
$$

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

$$
\frac{\sum_{p_i} p_i \left( \bar{V}(v_i) - \bar{V}(\mu) \right) - D\bar{V}(\mu, \bar{s} - \sum_{p_i} p_i v_i) \left\| \sum_{p_i} p_i v_i - \mu \right\|}{-\sum_{p_i} p_i \left( H(v_i) - H(\mu) - \nabla H(\mu) \cdot (v_i - \mu) \right)} \geq \frac{\rho \bar{V}(\mu) + h(c)}{c} + \epsilon \tag{20}
$$

where $\delta$ is sufficiently small that $\mu_0 = \mu - \delta(\sum_{p_i} p_i v_i - \mu) \in \Delta X^0$. Then by construction, if we assume:

$$
\begin{align*}
p_i' &= \frac{1}{1 + \delta} \\
p_i' &= \frac{1}{1 + \delta} p_i
\end{align*}
$$

48
Then \((p'_i, v'_i)\) is Bayesian plausible:

\[
\begin{cases}
\sum p'_i = 1 \\
\sum p'_i v'_i = \mu
\end{cases}
\]

where 0 is also included in indices \(i\)'s. Replacing terms in Equation (20) and let \(I(v_i | \mu) = H(\mu) - \sum p'_i H(v_i)\):

\[
\frac{\sum p'_i \nabla (v_i) - \nabla (\mu)}{-\sum p'_i H(v_i) + H(\mu)} = \frac{\rho}{c} \nabla (\mu) + \frac{h(c)}{c} + \epsilon
\]

\[
\implies \sum p'_i \nabla (v_i) - \frac{I(v_i | \mu)}{c} h(c) \geq \left(1 + \rho \frac{I(v_i | \mu)}{c}\right) \nabla (\mu) + \epsilon I(v_i | \mu)
\]

(21)

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

\[
e^{-\frac{I(v_i | \mu)}{c}} = \frac{1}{\rho} \left(1 + \rho \frac{I(v_i | \mu)}{c}\right) I(v_i | \mu) \leq \frac{\epsilon I(v_i | \mu)}{4}
\]

(22)

The equality is from Taylor expansion of exponential function. Plug Equation (22) into Equation (21):

\[
\sum p'_i \nabla (v_i) - \frac{I(v_i | \mu)}{c} h(c) \geq e^{-\frac{I(v_i | \mu)}{c}} \nabla (\mu) + \frac{\epsilon}{4} I(v_i | \mu)
\]

\[
\implies \sum p'_i \nabla (v_i) - \frac{I(v_i | \mu)}{c} h(c) \geq \nabla (\mu) + e^{-\frac{I(v_i | \mu)}{c}} I(v_i | \mu) - \left(1 - e^{-\frac{I(v_i | \mu)}{c}}\right) I(v_i | \mu) \frac{h(c)}{c}
\]

(23)

Noticing that \(1 - e^{-\frac{I(v_i | \mu)}{c}}\) \(I(v_i | \mu)\) is a second order small term. Then we can pick \(\delta\) such that Equation (23) implies:

\[
e^{-\frac{I(v_i | \mu)}{c}} \left(\sum p'_i \nabla (v_i)\right) - \frac{I(v_i | \mu)}{c} h(c) \geq \nabla (\mu) + \frac{\epsilon}{8} I(v_i | \mu)
\]

From now on, we fix \(\epsilon\) and \(\delta\). Pick \(dt = \frac{I(v_i | \mu)}{c}, dt_m = \frac{dt}{m}\). By uniform convergence, there exists \(N\) s.t. \(\forall m \geq N:\)

\[
e^{-\rho dt} \left(\sum p'_i V_{dt_m} (v_i)\right) - dt h \left(\frac{I(v_i | \mu) / m}{dt} \right) > V_{dt_m} (\mu)
\]

\[
\implies e^{-\rho dt m} \left(\sum p'_i V_{dt_m} (v_i)\right) - \sum_{\tau=0}^{m-1} e^{-\rho dt \tau} h_{dt_m} \left(\frac{I(v_i | \mu)}{m}\right) > V_{dt_m} (\mu)
\]

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 \(\mu\) for discrete time problem with \(dt_m\). Contradiction. Therefore, \(\nabla\) must be unimprovable at \(\mu\).

**Case 2:**

\[
\frac{D^2 \nabla (\mu, \sigma) \|\hat{\sigma}\|^2}{-\hat{\sigma}^2 HH(\mu, \sigma)} > \frac{\rho}{c} \nabla (\mu) + \frac{h(c)}{c}
\]

Then by the definition of operator \(D^2\) in Definition 2, there exists \(\hat{\sigma}, \delta, \epsilon > 0\) s.t.

\[
\nabla (\mu + \delta \hat{\sigma}) - \nabla (\mu) - \delta D \nabla (\mu, \hat{\sigma}) \|\hat{\sigma}\|^2 > \frac{\rho}{c} \nabla (\mu) + \frac{h(c)}{c} + 2 \epsilon
\]

Then by the definition of operator \(D\) in Definition 2, there exists \(\delta'\) s.t.

\[
\nabla (\mu + \delta \hat{\sigma}) - \nabla (\mu) - \delta' \nabla (\mu, \sigma) \|\sigma\|^2 > \frac{\rho}{c} \nabla (\mu) + \frac{h(c)}{c} + \epsilon
\]

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

\[
\sum p_i \nabla (v_i) \geq \left(1 + \rho \frac{I(v_i | \mu)}{c}\right) \nabla (\mu) + \frac{I(v_i | \mu)}{c} h(c) + \epsilon I(v_i | \mu)
\]

(24)

Noticing that Equation (24) is exactly the same as Equation (21) in Case 1. Then using same argument, This case is also ruled out.

**Equality:** I show that \(\forall\) smooth function \(V\) solving Equation (19), \(\nabla = V\). Notice that this automatically proves the uniqueness of solution of Equation (19). I prove inequality from both directions for \(\mu \in \Delta (X)^{\circ}\)
Suppose not, then consider \( U(\mu) = \mathbb{V}(\mu) - V(\mu) \). Since both \( V \) and \( \mathbb{V} \) are continuous, \( U \) is continuous. Therefore \( \text{argmin} U \) is non empty and \( \text{min} U < 0 \) according to our assumption. Choose \( \mu \in \text{argmin} U \) (\( \mu \in \Delta X^0 \) since \( V = \mathbb{V} \) on boundary). Since \( \mathbb{V}(\mu) \geq F(\mu) \), \( V(\mu) > F(\mu) \). Let \((p, v, \hat{\sigma})\) be a strategy solving \( V(\mu) \):

\[
\rho V(\mu) = \sum p_i(V(v_i) - \mathbb{V}(\mu)) - D\mathbb{V}\left(\mu, \sum p_i v_i - \mu\right) + \frac{1}{2} D^2\mathbb{V}(\mu, \hat{\sigma})\|\hat{\sigma}\|^2
\]

Now compare \( D\mathbb{V} \) and \( D\mathbb{V} \):

\[
D\mathbb{V}(\mu, \mu') - D\mathbb{V}(\mu, \mu') \left\| \mu - \mu' \right\| \leq V(\mu) - V(\mu') - V(\mu) + V(\mu') + U(\mu) - U(\mu')
\]

Therefore Equation (25) implies:

\[
\rho V(\mu) \leq \sum p_i(\mathbb{V}(v_i) - \mathbb{V}(\mu) - (U(v_i) - U(\mu))) - \frac{1}{2} D^2\mathbb{V}(\mu, \hat{\sigma})\|\hat{\sigma}\|^2 - h\left( - \sum p_i(H(v_i) - H(\mu) + \nabla H(\mu)(v_i - \mu)) - \frac{1}{2} \hat{\sigma}^T HH(\mu)\hat{\sigma} \right)
\]

The first inequality comes from replacing \( D\mathbb{V} \) and \( D^2\mathbb{V} \) with \( D\mathbb{V} \) and \( D^2\mathbb{V} \). The second inequality comes from \( U(v_i) - U(\mu) \geq 0 \) and unimprovability of \( \mathbb{V} \). Contradiction.

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

\[
\begin{align*}
V_{dt_n}(\mu^n) &= e^{-\rho dt_n} \sum p^n_i V_{dt_n}(v^n_i) - dt_n h(c_n) \\
\sum p^n_i H(\mu^n - H(v^n_i)) &= c_n dt_n \\
\sum p^n_i v^n_i &= \mu^n ; \sum p^n_i = 1
\end{align*}
\]

By definition of \( U_n(\mu) \):

\[
\begin{align*}
\sum p^n_i (V(v^n_i) - V(\mu^n)) &= \sum p^n_i (V_{dt_n}(v^n_i) - V_{dt_n}(\mu^n) - U_n(\mu^n) + U(v^n_i)) \\
& \geq \sum p^n_i (V_{dt_n}(v^n_i) - V_{dt_n}(\mu^n)) \\
& = (e^{\rho dt_n} - 1) V_{dt_n}(\mu^n) + e^{\rho dt_n} dt_n h(c_n) \\
& \geq \rho dt_n V_{dt_n}(\mu^n) + \rho dt_n h(c_n) \\
& \geq \rho dt_n \varepsilon + \rho dt_n V(\mu^n) + \rho dt_n h(c_n) \\
& \implies \rho V(\mu^n) \leq -\rho \varepsilon + \sum p^n_i (V(v^n_i) - V(\mu^n)) - e^{\rho dt_n} h(c_n)
\end{align*}
\]
The first equality is by the definition of $U_n$. The first inequality is from $\mu^n \in \text{argmin} U_n$. The second inequality is from $e^{x-1} \geq x$. The third inequality is from $U_n(\mu^n) \leq -\epsilon$. Now since the number of posteriors $v^n_i$ is no more than $2|X|$, we can take a subsequence of $n$ such that all $\lim v^n_i = v_i$. Partition $v^n_i$ into two kinds: $\lim v^n_i = v_i \neq \mu$, $\lim v^n_i = \mu$.

Since $V$ is unimprovable, $\forall c, \hat{\sigma}$ we have $D^2 V(\mu, \hat{\sigma}) ||\hat{\sigma}\|^2 \leq -\hat{\sigma}^T HH(\mu) \hat{\sigma} \left( \frac{\rho}{c} V(\mu) + \frac{h(c)}{c} \right)$. Since $V \in C^{(1)}$, $H \in C^{(2)}$, $\forall \eta$, there exists $\delta$ s.t. $\forall |\mu' - \mu| \leq \delta$:

$$
\left\{ \begin{array}{l}
|| H H(\mu) - H H(\mu') || \leq \eta \\
| V(\mu) - V(\mu') | \leq \eta
\end{array} \right.$$

$$
\implies D^2 V(\mu', \hat{\sigma}) \leq \left( \frac{\rho}{c} V(\mu') + \frac{h(c)}{c} \right) \left( - \frac{\hat{\sigma}^T HH(\mu') \hat{\sigma}}{||\hat{\sigma}\|^2} \right) + \left( \frac{\rho}{c} \sup F + \frac{h(c)}{c} \right) \eta + \frac{\rho}{c} \eta ||HH(\mu)||
$$

If we pick $\eta$ and $\delta$ properly:

$$
D^2 V(\mu', \hat{\sigma}) \leq \left( \frac{\rho}{c} V(\mu) + \frac{h(c)}{c} \right) \left( - \frac{\hat{\sigma}^T HH(\mu') \hat{\sigma}}{||\hat{\sigma}\|^2} \right) + \frac{1 + h(c)}{c} \eta
$$

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

$$
V(v^n_j) - V(\mu^n) - \nabla V(\mu^n)(v^n_j - \mu^n) = g(1) - g(0) - g'(0)
$$

$$
\leq \frac{1}{2} \sup_{\alpha \in (0, 1)} D^2 g(\alpha, 1) = \sup_{\alpha \in (0, 1)} \limsup_{d \to 0} \frac{g(\alpha + d) - g(\alpha) - g'(\alpha) d}{d^2}
$$

$$
= \sup_{\zeta \in (\mu^n, \nu^n)} \limsup_{d \to 0} \frac{V(\zeta + d(v^n_j - \mu^n)) - V(\zeta) - d \nabla V(\zeta)(v^n_j - \mu^n)}{d^2}
$$

$$
\leq \frac{1}{2} \sup_{d ||v^n_j - \mu|| \leq \delta} D^2 V(\zeta, v^n_j - \mu^n) ||v^n_j - \mu^n||^2
$$

$$
\leq \frac{1}{2} \left( \frac{\rho}{c} V(\mu) + \frac{h(c)}{c} \right) (v^n_j - \mu^n)^T HH(\mu)(v^n_j - \mu^n) + \frac{1 + h(c)}{2c} \eta ||v^n_j - \mu^n||^2
$$

(27)

Therefore, by applying Equation (27):

$$
\sum p_i^n (V(v^n_j) - V(\mu^n))
$$

$$
= \sum p_i^n (V(v^n_j) - V(\mu^n) - \nabla V(\mu^n)(v^n_j - \mu^n) + \nabla V(\mu^n)(v^n_j - \mu^n))
$$

$$
\leq \sum p_i^n (V(v^n_j) - V(\mu^n) - \nabla V(\mu^n)(v^n_j - \mu^n))
$$

$$
- \frac{1}{2} \left( \frac{\rho}{c} V(\mu) + \frac{h(c)}{c} \right) \sum p_i^n (v^n_j - \mu^n)^T HH(\mu)(v^n_j - \mu^n) + \frac{1 + h(c)}{2c} \eta \sum p_i^n ||v^n_j - \mu^n||^2
$$

(28)

Notice that Equations (27) and (28) are true uniform to $c$, so we can replace $c$ with $c_n$ and Equation (28) is still true. Now let $\tilde{p}_i^n = \frac{p_i^n}{d_{tn}}$, $-\hat{\sigma}^n_T HH(\mu^n) \hat{\sigma}^n d_{tn} = \sum \tilde{p}_i^n \left( H(\mu^n) - H(v^n_j) + \nabla H(\mu)(v^n_j - \mu^n) \right)$, we have:

$$
\sum \tilde{p}_i^n \left( H(\mu^n) - H(v^n_j) + H'(\mu^n)(v^n_j - \mu^n) \right) - \hat{\sigma}^n_T HH(\mu^n) \hat{\sigma}_n = c_n
$$

(29)

$(\tilde{p}_i^n, v^n_j, \hat{\sigma}_n)$ is a feasible experiment for Equation (19). Therefore, by optimality of $V$ at $\mu^n$, we have

$$
\left\{ \begin{array}{l}
\sum \tilde{p}_i^n (V(v^n_j) - V(\mu^n) - \nabla V(\mu^n)(v^n_j - \mu^n)) \leq (c_n + \hat{\sigma}^n_T HH(\mu^n) \hat{\sigma}_n) \left( \frac{\rho}{c_n} V(\mu^n) + \frac{h(c_n)}{c_n} \right)
\\
D^2 V(\mu^n, \hat{\sigma}_n) \leq -\frac{\hat{\sigma}^n T HH(\mu^n) \hat{\sigma}_n}{||\hat{\sigma}_n||^2} \left( \frac{\rho}{c_n} V(\mu^n) + \frac{h(c_n)}{c_n} \right)
\end{array} \right.
$$

(30)
Then we study term $\sum p^n_j (v^n_j - \mu^n)^2$. Apply Lemma S.10 to $g(\alpha) = H(\alpha v^n_j + (1-\alpha) \mu^n)$:

$$
\sum p^n_j \left( H(\mu^n) - H(v^n_j) + \nabla H(\mu^n)(v^n_j - \mu^n) \right)
\geq \frac{1}{2} \inf \sum p^n_j \left(- (v^n_j - \mu^n)^T HH(\varphi^n_j)(v^n_j - \mu^n) \right)
\geq -\frac{1}{2} \sum p^n_j ((v^n_j - \mu^n)^T HH(\mu)(v^n_j - \mu^n)) - \frac{1}{2} \eta \sum p^n_j \|v^n_j - \mu^n\|^2
$$

(31)

Therefore, to sum up:

$$
\sum \frac{p^n_j}{dt} \left( V(v^n_{i,j}) - V(\mu^n) \right) \leq \sum \frac{p^n_j}{dt} \left( (v^n_j - \mu^n)^T \nabla V(\mu^n)(v^n_j - \mu^n) \right)
+ \frac{1}{2} \sum \frac{p^n_j}{dt} \left(- (v^n_j - \mu^n)^T HH(\mu)(v^n_j - \mu^n) \left( \frac{\rho}{c_n} V(\mu) + \frac{h(c_n)}{c_n} \right) \right)
+ \sum \frac{p^n_j}{dt} \left( 1 + h(c_n) \right) \|v^n_j - \mu^n\|^2
\leq \left( c_n + \tilde{\sigma}^n T HH(\mu^n) \tilde{\sigma}^n \right) \left( \frac{\rho}{c_n} V(\mu^n) + \frac{h(c_n)}{c_n} \right)
+ \left( \sum \frac{p^n_j}{dt} \left( H(\mu^n) - H(v^n_j) + \nabla H(\mu^n)(v^n_j - \mu^n) \right) \right)
+ \frac{1}{2} \eta \sum p^n_j \|v^n_j - \mu^n\|^2 \left( \frac{1 + h(c_n)}{2c_n} \right) \eta
\leq \rho V(\mu^n) + h(c_n) + \frac{1}{dt} \sum p^n_j \|v^n_j - \mu^n\|^2 \left( \frac{1 + \rho V(\mu) + 2h(c_n)}{2c_n} \right) \eta + \rho \eta
$$

The first inequality is Equation (28). The second inequality comes from Equation (30) and Equation (31). The next equality comes from definition of $\tilde{\sigma}^n$. The last inequality comes from canceling out terms and $-\tilde{\sigma}^n T HH(\mu^n) \tilde{\sigma} \leq c_n$ (Notice the difference between $V(\mu)$ and $V(\mu^n)$). Then by plug into Equation (26):

$$
\rho V(\mu^n) \leq -\rho \varepsilon + \rho V(\mu^n) + \frac{1}{dt} \sum p^n_j \|v^n_j - \mu^n\|^2 \left( \frac{1 + \rho V(\mu) + 2h(c_n)}{2c_n} \right) \eta + \rho \eta
$$

Moreover:

$$
\sum p^n_j \|v^n_j - \mu^n\|^2 \inf \left| \frac{\sigma^T HH(\mu) \sigma}{\|\sigma\|^2} \right|
\leq \sum p^n_j (v^n_j - \mu^n) HH(\mu)(\mu^n - \mu^n) \leq c_n dt_n + \eta \sum p^n_j \|v^n_j - \mu^n\|^2
\Rightarrow \sum p^n_j \|v^n_j - \mu^n\|^2 \leq \frac{c_n dt_n}{\inf \left| \frac{\sigma^T HH(\mu) \sigma}{\|\sigma\|^2} \right| - \eta}
\Rightarrow \rho \varepsilon \leq \frac{1}{2} \left( 1 + \rho V(\mu) + 2h(c_n) \right) \frac{\eta}{\inf \left| \frac{\sigma^T HH(\mu) \sigma}{\|\sigma\|^2} \right| - \eta} + \rho \eta
$$

By Lemma S.7, $h(c_n)$ is uniformly bounded above. Since $H$ is strictly concave $\inf \left| \frac{\sigma^T HH(\mu) \sigma}{\|\sigma\|^2} \right|$ is positive. The
inequality holds when $\eta$ is chosen smaller than $\inf_{v} |v^THH(\mu)v|/\|v\|^2$. 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 $\mu_k$ be each kink points of $F$: $F(\mu) = F_k(\mu) \iff \mu \in \left[\frac{F_{k-1}}{F_k}\right]$. $\overline{m}$ is the smallest index s.t. $F'_m > 0$.

**Algorithm:**

In this part, I introduce the algorithm for constructing $V(\mu)$ and $v(\mu)$. I only discuss the case $\mu \geq \mu^*$. The remaining case $\mu \leq \mu^*$ follows by a symmetric method. The main steps are illustrated in **Figure 22**. The first step is to find critical the belief $\mu^*$ 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.

---

**Figure 22:** Construction of optimal value function.

* Step 1: Define:

$$\nabla^+(\mu) = \max_{\nu \geq \mu} \frac{F_m(\nu)}{1 + \frac{\nu}{\epsilon} J(\mu, \nu)}$$

$$\nabla^-(\mu) = \max_{\nu \leq \mu} \frac{F_m(\nu)}{1 + \frac{\nu}{\epsilon} J(\mu, \nu)}$$
In Lemma B.1 I analyze the technical details of \( \nabla^+ \) and \( \nabla^- \). The main property is that: \( \nabla^+ \) is increasing and \( \nabla^- \) is decreasing. There exists \( \mu^* \in [0, 1] \) s.t. \( \nabla^+(\mu) \geq \nabla^-(\mu) \) when \( \mu \geq \mu^* \) and \( \nabla^-(\mu) \leq \nabla^-(\mu) \) when \( \mu \leq \mu^* \). Define \( \nabla(\mu) = \max\{ \nabla^+(\mu), \nabla^-(\mu) \} \).

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

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

\[
\nabla(\mu^*) = \frac{F_m(\nu(\mu^*))}{1 + \frac{c}{\nabla(\mu^*)}}
\]

Initial condition \((\mu_0 = \mu^*, V_0 = \nabla(\mu^*), V_0 = 0)\) satisfies Lemma B.2, which states that there exists \( V_m(\mu) \) solving:

\[
V_m(\mu) = \max_{\nu \geq \mu} \frac{\epsilon F_m(\nu) - V(\mu) - V'(\mu)(\nu - \mu)}{J(\mu, \nu)}
\]

This refers to Figure 22-1. Define

\[
V_{\mu^*}(\mu) = \begin{cases} F(\mu) & \text{if } \mu \leq \mu^* \\ V_m(\mu) & \text{if } \mu \geq \mu^* \end{cases}
\]

Be Lemma B.2, when \( V_{\mu^*}(\mu) > F(\mu) \), \( V_{\mu^*} \) 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 \( \hat{\mu}_m \) be the smallest \( \mu \geq \mu^* \) such that:

\[
V_{\mu^*}(\hat{\mu}_m) = \max_{\nu \geq \hat{\mu}_m} \frac{\epsilon F_{m-1}(\nu) - V_{\mu^*}(\hat{\mu}_m) - V_{\mu^*}(\nu)(\nu - \hat{\mu}_m)}{J(\hat{\mu}_m, \nu)} \quad (32)
\]

At \( \hat{\mu}_m \), searching posterior on \( F_{m-1} \) first dominates searching posterior on \( F_m \). This step refers to Figure 22-2. \( \hat{\mu}_m \) is the smallest intersection point of blue curve \((V_{\mu^*}(\mu), \text{LHS of Equation (32)})\) and thin red curve \((\text{RHS of Equation (32)})\). If \( V_m(\hat{\mu}_m) > F_{m-1}(\hat{\mu}_m) \) then solve for \( V_{m-1} \) with initial condition \( \mu_0 = \hat{\mu}_m, V_0 = V_m(\hat{\mu}_m), V_0' = V_m'(\hat{\mu}_m) \) according to Lemma B.2 and redefine \( V_{\mu^*}(\mu) = V_{m-1}(\mu) \) when \( \mu \geq \hat{\mu}_m \). Otherwise skip to looking for \( \tilde{\mu}_{m-1} \). If \( m-1 \geq \tilde{m} \), continue this procedure by looking for \( \tilde{\mu}_{m-1} \) and update \( V_{\mu^*}(\tilde{\mu}_{m-1}) \) with corresponding \( V_{m-2} \) \ldots until \( m = \tilde{m} \) (No action with the slope of \( F_{\tilde{m}} \), being negative is considered). This refers to Figure 22-3. Now suppose \( V_{\tilde{m}} \) first hits \( F(\mu) \) at some point \( \mu^{**} \) (\( \mu^{**} > \mu^* \) since \( V_m(\mu^*) > F(\mu^*) \)). \( V_{\mu^*} \) is a (piecewise) smooth function on \([\mu^*, \mu^{**}]\) such that:

\[
V_{\mu^*}(\mu) = \begin{cases} F(\mu) & \text{if } \mu \leq \mu^* \text{ or } \mu \geq \mu^{**} \\ V_k(\mu) & \text{if } \mu \in [\hat{\mu}_k, \tilde{\mu}_{k-1}] \end{cases} \quad (33)
\]

By construction, optimal posterior \( V_{\mu^*}(\mu) \) is smoothly decreasing on each \( (\hat{\mu}_{k+1}, \hat{\mu}_k) \) and jumps down at each \( \hat{\mu}_k \). 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 for any \( \mu \in [\mu^*, \mu^{**}] \):

\[
V_{\mu^*}(\mu) = \max_{\nu \geq \mu, k} \frac{\epsilon F_k(\nu) - V_{\mu^*}(\nu) - V_{\mu^*}(\mu)(\nu - \mu)}{J(\mu, \nu)} \quad (33)
\]

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

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

\[
\nabla(\mu) = \max_{\nu \geq \mu, k} \frac{\epsilon F_k(\nu) - F(\mu) - F'(\nu) - F'(\nu) - F(\mu)}{J(\nu, \mu)}
\]

Define, \( \mu^{**} = \inf\{ \mu | \nabla(\mu) > F(\mu) \} \). By Assumption 3, \( \lim_{\mu \to 0} |F'(\mu)| = \infty \), then there exists \( \delta \) s.t. \( \forall \mu < \delta, \forall \nu > \mu^{**}, \sup_{J(\mu, \nu)} \leq \inf F \). Therefore \( \mu^{**} \geq \delta > 0 \). This step refers to Figure 22-4.

---

35Existence is guaranteed by smoothness of \( V_{\mu^*} \) and \( J \). Noticing that \( V_m(\mu_0) \geq F_{m-1}(\mu_0) \). Otherwise, there will be a \( \tilde{\mu}_m \) \( < \mu \) s.t. \( V_m(\tilde{\mu}_m) = F_{m-1}(\tilde{\mu}_m) \) and it is easy to verify that \( V_m \) is weakly larger than the maximum. So there is an even smaller \( \tilde{\mu}_m \), contradiction.

36Define \( \tilde{\mu}_{m+1} = \mu^* \) and \( \tilde{\mu}_m = \mu^{**} \) for consistency.

37Since \( F_{k-1} \) 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.
Finally, I prove unimprovability of each interval when

Lemmas S.11

I need to verify that for technical use (for example, the validity of using $m^1$).

Let $m$ be the index of optimal action. Solve for $V_m$ with initial condition $\mu_0=\mu^\circ, V_0=F(\mu^\circ), V'_0=F'^-(\mu^\circ)$. Then take same steps in Step 2 and solve for $\hat{\mu}_k$ and $V_{k-1}$ sequentially until $V_{m_0}$ first hits $F$. This step refers to Figure 22-45. Now suppose $V_{m_0}$ first hits $F(\mu)$ at some point $\mu^\circ$ (can potentially be $\mu$), define:

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

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

$$V_\mu(\mu)=\max_{\nu\in[\nu_1,\nu_0]} \frac{c F(\nu)-V(\mu)-V'(\nu)(\nu-\mu)}{j(\nu,\nu)}$$

Let $\Omega$ be the set of all such $\mu^\circ$s.

Step 4: Define:

$$V(\mu)=\begin{cases} V_{\mu^*}(\mu) & \text{if } \mu\in[\mu^\circ,\mu^*] \\ \sup \{ V_{\mu^\circ}(\mu) \} & \text{if } \mu>\mu^\circ \end{cases}$$

In the algorithm, I only discussed the case $\mu^*<1$ and constructed the value function on the right of $\mu^*$.

Smootheness:

I need to verify that $V_{\mu}$ that defined as Equation (35) 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 S.2.1 in Lemmas S.11, S.12, S.13 and S.14. In addition, it is shown in Section S.2.1 that there exists a set of $\mu_0$ such that on each interval when $V(\mu)>F(\mu), V(\mu)$ is defined as one $V_{\mu_0}$.

Unimprovability:

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

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

$$V(\mu)=\max_{\nu_1,\nu_0} \frac{c F(\nu)-V(\mu)-V'(\nu)(\nu-\mu)}{j(\nu,\nu)} \begin{cases} v\geq \mu^* & \text{if } \mu\geq \mu^* \\ v< \mu^* & \text{if } \mu< \mu^* \end{cases}$$

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 (33) for $\mu\in E$. So it is sufficient to prove Equation (P-C) for $\mu\in E^C$. Suppose for the sake of contradiction that there exists $\mu\geq \mu^*$ s.t. Equation (P-C) is violated. Let $F(\mu)=F_k(\mu)$. Then it is equivalently stating that:

$$U(\mu)=\max_{\nu_1,\nu_0} \frac{c F(\nu)-F_k(\nu)-F'(\nu)(\nu-\mu)}{j(\nu,\nu)} > F_k(\mu)$$

Consider $\nu_k$ (the intersection of $F_k$ and $F_{k-1}$). By Lemma S.11, there exists $I_k$ s.t. $\nu_k\in I_k$. At $b_k=\sup I_k, U(b_k)>F_k(b_k)$.

Therefore, since $U(\mu)$ is continuous, by intermediate value theorem there exists largest $\mu'$ between $\nu_k$ and $\mu$ s.t. $U(\mu')=F_k(\mu')$. Then Equation (34) is satisfied at $\mu'$ so consider $V_{\mu'}$. Sicne $V_{\mu'}(\mu)\leq V(\mu)=F_k(\mu)$, there exists $\mu''\in(\mu',\mu)$ s.t. $V_{\mu'}(\mu'')=F_k(\mu'')$ and $V_{\mu'}(\mu'')=F_k(\mu'')$. Therefore $U(\mu'')>F_k(\mu'')$ implies $V_{\mu'}(\mu'')>F_k(\mu'')$, contradiction.

Apply a symmetric argument to $\mu<\mu^*$, I prove Equation (P-C).

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

$$V(\mu)=\max_{\nu_1,\nu_0} \frac{c V(\nu)-V(\mu)-V'(\nu)(\nu-\mu)}{j(\nu,\nu)}$$

38By definition of $\mu^*\mu_0$ is bounded away from $\{0,1\}$ and Equation (34) implies conditions in Lemma B.2 are satisfied.

39Define $\hat{\mu}_{k+1}=\mu^\circ$ and $\hat{\mu}_{m_0}=\mu^\circ$ for consistency.
\[
\begin{aligned}
&v \geq \mu \text{ when } \mu \geq \mu^* \\
&v \leq \mu \text{ when } \mu \leq \mu^*
\end{aligned}
\]

**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 \( v \) is realized, the DM is allowed to continue experimentation instead of being forced to take an action.

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

\[
V(\mu) = \max_{\nu \geq \mu, \nu \geq \rho} \frac{c F_m(\nu) - V(\mu) - V'(\mu)(\nu - \mu)}{f(\mu, \nu)} \leq \frac{c V(\nu') - V(\mu) - V'(\mu)(\nu' - \mu)}{f(\mu, \nu')}
\]  
\( (36) \)

Let \( \tilde{V} = V(\mu) \). Suppose the maximizer is \( \nu, \mu \). Optimality implies first order conditions **Equation (42)** and **Equation (41)**:

\[
\begin{aligned}
&F'_m + \frac{\rho}{c} \tilde{V} H'(v) = V'(\mu) + \frac{\rho}{c} \tilde{V} H'(\mu) \\
&(F_m(\nu) + \frac{\rho}{c} \tilde{V} H(\nu)) - (V(\mu) + \frac{\rho}{c} \tilde{V} H(\mu)) = (V'(\mu) + \frac{\rho}{c} \tilde{V} H'(\mu)(\nu - \mu))
\end{aligned}
\]

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

\[
\begin{aligned}
L(V, \lambda, \mu)(\nu) &= (V(\mu) + \lambda H(\mu)) + (V'(\mu) + \lambda H'(\mu))(\nu - \mu) \\
G(V, \lambda)(\nu) &= V(\nu) + \lambda H(\nu)
\end{aligned}
\]  
\( (37) \)

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

\[
\begin{aligned}
L(V, \frac{\rho}{c} \tilde{V}, \mu)(\nu) - G(F_m, \frac{\rho}{c} \tilde{V})(\nu)
\end{aligned}
\]

**Equation (42)** implies that it attains minimum 0 at \( \nu \). For any \( \nu' \) other than \( \mu \),

\[
\begin{aligned}
L(V, \frac{\rho}{c} \tilde{V}, \mu)(\nu) - G(F_m, \frac{\rho}{c} \tilde{V})(\nu)
\end{aligned}
\]

is convex and weakly larger than zero. However by **Equation (36)**:

\[
\begin{aligned}
L(V, \frac{\rho}{c} \tilde{V}, \mu)(\nu') - G(V, \frac{\rho}{c} \tilde{V})(\nu') = -\left(V'(\nu) - V(\mu) - V'(\mu)(\nu' - \mu) - \frac{\rho}{c} \tilde{V} H'(\mu)(\nu' - \mu)\right) < 0
\end{aligned}
\]

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

\[
V'(\mu) + \frac{\rho}{c} \tilde{V} H'(\mu) = V'(\tilde{\mu}) + \frac{\rho}{c} \tilde{V} H'(\tilde{\mu})
\]

Consider:

\[
\begin{aligned}
L(V, \frac{\rho}{c} \tilde{V}, \mu)(\nu(\tilde{\mu})) - G(F_m, \frac{\rho}{c} \tilde{V})(\nu(\tilde{\mu})) \\
= L(V, \frac{\rho}{c} \tilde{V}, \mu)(\nu(\tilde{\mu})) - G(F_m, \frac{\rho}{c} \tilde{V})(\nu(\tilde{\mu})) \\
+ V(\tilde{\mu}) - V(\mu) + \frac{\rho}{c} \tilde{V}(H(\tilde{\mu}) - H(\mu)) - \left(V'(\mu) + \frac{\rho}{c} \tilde{V} H'(\mu)(\tilde{\mu} - \mu)\right) \\
\geq V(\tilde{\mu}) - V(\mu) + \frac{\rho}{c} \tilde{V}(H(\tilde{\mu}) - H(\mu)) - \left(V'(\mu) + \frac{\rho}{c} \tilde{V} H'(\mu)(\tilde{\mu} - \mu)\right) \\
= G(V, \frac{\rho}{c} \tilde{V})(\tilde{\mu}) - L(V, \frac{\rho}{c} \tilde{V}, \mu)(\tilde{\mu}) > 0
\end{aligned}
\]

In the first equality I used **Equation (42)** at \( \tilde{\mu} \). In first inequality I used suboptimality of \( \tilde{\mu} \) at \( \mu \). However for \( \nu(\tilde{\mu}) \) being optimizer at \( \tilde{\mu} \):

\[
0 = L(V, \frac{\rho}{c} V(\tilde{\mu}), \tilde{\mu})(\nu(\tilde{\mu})) - G(F_m, \frac{\rho}{c} V(\tilde{\mu}))(\nu(\tilde{\mu})) \\
= L(V, \frac{\rho}{c} \tilde{V}, \tilde{\mu})(\nu(\tilde{\mu})) - G(F_m, \frac{\rho}{c} \tilde{V})(\nu(\tilde{\mu})) \\
+ \frac{\rho}{c} \tilde{V} (V(\tilde{\mu}) - \tilde{\mu} + H(\tilde{\mu}) - H(\nu(\tilde{\mu}))) + H'(\tilde{\mu})(\nu(\tilde{\mu}) - \tilde{\mu}) \\
\geq \frac{\rho}{c} \tilde{V} (V(\tilde{\mu}) - \tilde{\mu} + H(\tilde{\mu}) - H(\nu(\tilde{\mu}))) + H'(\tilde{\mu})(\nu(\tilde{\mu}) - \tilde{\mu})
\]

Contradiction. Therefore, I proved **Equation (P-D)**.
Step 3: I show that $V$ satisfies Equation (19), 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 $\nu$, envelope theorem implies:

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

$$= \frac{c}{\rho} \frac{v-\mu}{J(\mu, \nu)} \left( V''(\mu) + \frac{\rho}{c} V(\mu) H''(\mu) \right) > 0$$

$$\Rightarrow V''(\mu) + \frac{\rho}{c} V(\mu) H''(\mu) < 0$$

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

$$V^-(\mu) = \max_{v \leq \mu} c \frac{V(v) - V(\mu) - V'(\mu)(v-\mu)}{J(\mu, \nu)}$$

$$\Rightarrow V^{-1}(\mu) = - \frac{c}{\rho} \frac{v-\mu}{J(\mu, \nu)} \left( V''(\mu) + \frac{\rho}{c} V^{-1}(\mu) \right)$$

$V^-(\mu)$ is by definition the utility gain from searching contradictory evidence, given value function $V(\mu)$. By definition of $\mu^*$, $V^-(\mu^*) = V(\mu^*)$ and whenever $V(\mu) = V^-(\mu)$ $V^{-1}(\mu) < 0$. Therefore, $V^-(\mu)$ can never cross $V(\mu)$ from below — $V^-(\mu)$ is lower than $V(\mu)$ and $V(\mu)$ is unimprovable with contradictory evidence. That is to say:

$$\rho V(\mu) = \max \left\{ \rho F(\mu) \max_{v, p, \nu} \left( V(\nu) - V(\mu) - V'(\mu)(v-\mu) + \frac{1}{2} V''(\mu) \sigma^2 \right) \right\}$$

s.t. $p J(\mu, \nu) + \frac{1}{2} H''(\mu) \sigma^2 \leq c$

To sum up, I construct a policy function $v(\mu)$ and value function $V(\mu)$ solving Equation (19). Now consider the four properties in Theorem 2. First, by my construction algorithm, in the case $\mu^* \in \{0, 1\}$, I can replace $\mu^*$ 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, $V(\mu) = V''(\mu)|_{\mu \in I_m}$. On each $I_m$, $v_\mu(\mu)$ is strictly decreasing and jumps down at finite $\mu_k$’s. Finally, uniqueness argument in Lemma B.2 implies that $v$ is uniquely determined by FOC. Therefore, except for those discontinuous points of $v$, $v$ is uniquely defined. Number of such discontinuous points is countable, thus of zero measure. Q.E.D.

**Lemma B.1.** Define $\nabla^+$ and $\nabla^-$:

$$\nabla^+(\mu) = \max_{v \geq \mu, m} \frac{F_m(v)}{1 + \frac{\rho}{c} J(\mu, \nu)}$$

$$\nabla^-(\mu) = \max_{v \leq \mu, m} \frac{F_m(v)}{1 + \frac{\rho}{c} J(\mu, \nu)}$$

There exists $\mu^* \in [0, 1] \text{ s.t. } \nabla^+(\mu) \geq \nabla^-(\mu) \forall \mu \geq \mu^*$; $\nabla^+(\mu) \leq \nabla^-(\mu) \forall \mu \leq \mu^*$.

**Proof.** I define function $U^+_m$ and $U^-_m$ as follows:

$$U^+_m(\mu) = \max_{v \geq \mu} \frac{F_m(v)}{1 + \frac{\rho}{c} J(\mu, \nu)}$$

$$U^-_m(\mu) = \max_{v \leq \mu} \frac{F_m(v)}{1 + \frac{\rho}{c} J(\mu, \nu)}$$

First of all, I solve $U^+_m, U^-_m$ on interior $\mu \in (0, 1)$. Since $F_m(\mu)$ is a linear function, $J(\mu, \nu) \geq 0$ 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:

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

$$\text{SOC: } \frac{\rho}{c} F_m'(H'(v) - H'(\mu))$$

First discuss solving for $v \geq \mu$. Since $(1 + \frac{\rho}{c} J) > 0$, $H'' < 0$, $H'(v) - H'(\mu) \leq 0$ and inequality is strict when $v > \mu$. Therefore, if $F_m < 0$, FOC being held will imply SOC being strictly positive at $v > \mu$. So $\forall F_m < 0$, optimal $v$ 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 \(\forall v > \mu\):
\[
\frac{F_m(\mu)}{1 + \frac{\rho}{c} J(\mu, \mu)} = F_m(\mu) = F_m(v) \geq \frac{F_m(v)}{1 + \frac{\rho}{c} J(\mu, v)}
\]
Therefore \(\forall F_m' \leq 0, U_m^+(\mu) = F_m(\mu)\). Then consider the case \(F_m' > 0\). It can be easily verified that SOC is strictly negative when FOC holds and \(v > \mu\). Therefore solution of FOC characterizes maximizer. Consider:
\[
\lim_{v \to \mu} F_m'(1 + \frac{\rho}{c} J(\mu, v)) + F_m(v) \frac{\rho}{c} (H'(v) - H'(\mu)) = F_m' > 0
\]
\[
\lim_{v \to 1} F_m'(1 + \frac{\rho}{c} J(\mu, v)) + F_m(v) \frac{\rho}{c} (H'(v) - H'(\mu)) = -\infty
\]
Therefore be intermediate value theorem a unique solution \(v \in (\mu, 1)\) exists by solving FOC. Since FOC is a smooth function of \(\mu, v\) and SOC is strictly negative, implicit function theorem implies \(v\) being a smooth function of \(\mu\). This is sufficient to apply envelope theorem:
\[
\frac{d}{d\mu} U_m^+(\mu) = \frac{F_m(v)(-H''(\mu)(v - \mu))}{(1 + \frac{\rho}{c} J(\mu, v))^2} > 0
\]
Moreover, Equation (38) is strictly positive when \(v = \mu\). This implies \(U_m^+(\mu) > F_m(\mu)\) when \(F_m' > 0\).

New consider limit of \(U_m^+\) when \(\mu \to 0\). When \(\mu \to 1\), \(U_m^+(\mu) \leq \max_{v \geq \mu} F_m(v) = F(1)\). When \(\mu \to 0\), consider FOC Equation (38):
\[
\lim_{\mu \to 0} F_m'(1 + \frac{\rho}{c} J(\mu, v)) + F_m(v) \frac{\rho}{c} (H'(v) - H'(\mu)) = \lim_{\mu \to 0} F_m'(1 + \frac{\rho}{c} J(v, 0)) + \lim_{\mu \to 0} F_m(0) \frac{\rho}{c} (H'(v) - H'(\mu)) = -\infty
\]
Therefore, when \(\mu \to 0\), optimal \(v \to 0\). Therefore \(\frac{F_m(v)}{1 + \frac{\rho}{c} J(\mu, v)} \to F_m(v) \to F_m(0)\). To conclude, \(U_m^+(\mu) = F_m(\mu)\) when \(\mu = 0, 1\).

Let \(\bar{m}\) be the first \(F_m' > 0\) (not necessarily exists). Let:
\[
U_+^+(\mu) = \max_{\bar{m} \leq \mu} U_m^+(\mu)
\]
Then \(U_+^+(\mu)\) is a strictly increasing function when \(\bar{m}\) exists. Symmetrically I can define \(\bar{m}\) to be last \(F_m' < 0\) and:
\[
U_-^+(\mu) = \max_{\bar{m} \leq \mu} U_m^-(\mu)
\]
There are three cases:

- **Case 1:** when \(F\) is not monotonic, then both \(U_+^+\) and \(U_-^+\) exists. Moreover, \(F(0) > F(0)\) and \(F(1) > F(1)\). 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 \(F' \geq 0\), then define \(\mu^* = 0\).

- **Case 3:** when \(F' \leq 0\), then define \(\mu^* = 1\).

Finally, define \(V\):
\[
V^+_+(\mu) = \max \{ F(\mu), U_+^+(\mu) \}
\]
\[
V^+_-(\mu) = \max \{ F(\mu), U_-^+(\mu) \}
\]
\[
V(\mu) = \max \{ V^+_+(\mu), V^-_-(\mu) \}
\]
Given our construction, \(\mu^*\) always exists and satisfies the conditions in Lemma B.1. Q.E.D.

**Lemma B.2.** Assume \(\mu_0 \geq \mu^*\), \(F_m' = 0\), \(V_0, V_0' \geq 0\) satisfies:
\[
\begin{align*}
\nabla(\mu_0) & \geq V_0 \geq F_m(\mu_0) \\
V_0 & = \max_{v \geq \mu_0} \frac{c}{\rho} F_m(v) - V_0 - V_0'(v - \mu_0) \\
\end{align*}
\]
Then there exists a $C^1$ smooth and strictly increasing $V(\mu)$ defined on $[\mu_0,1]$ satisfying
\[
V(\mu) = \max_{v \geq \mu} \frac{c F_m(v) - V(\mu) - V'(\mu)(v - \mu)}{f(\mu, v)}
\tag{40}
\]
and initial condition $V(\mu_0) = V_0, V'(\mu_0) = V'_0$. Maximizer $v(\mu)$ is $C^1$ and strictly decreasing on $\{\mu | V(\mu) > F_m(\mu)\}$.

**Proof.** I start from deriving the FOC and SOC for Equation (40):

**FOC:**
\[
\frac{F'_m - V'(\mu)}{f(\mu, v)} + \frac{F_m(v) - V(\mu) - V'(\mu)(v - \mu)}{f(\mu, v)^2}(H'(v) - H'(\mu)) = 0
\]

**SOC:**
\[
\frac{H'(v) - H'(\mu)}{f(\mu, v)} \left( \frac{F'_m - V'(\mu)}{f(\mu, v)} + \frac{F_m(v) - V(\mu) - V'(\mu)(v - \mu)}{f(\mu, v)^2}(H'(v) - H'(\mu)) \right) + \frac{H''(v)}{f(\mu, v)}(F_m(v) - V(\mu) - V'(\mu)(v - \mu)) \leq 0
\]

If feasibility is imposed:
\[
V(\mu) = \frac{c F_m(v) - V(\mu) - V'(\mu)(v - \mu)}{f(\mu, v)}
\tag{41}
\]

FOC and SOC reduces to:

**FOC:**
\[
F'_m - V'(\mu) + \frac{\rho}{c} V(\mu)(H'(v) - H'(\mu)) = 0
\tag{42}
\]

**SOC:**
\[
\frac{\rho}{c} H''(v) V(\mu) \leq 0
\tag{43}
\]

Let us proceed as follows. I use FOC and feasibility to derive an ODE system with initial 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 $H'(v) - H'(\mu) < 0$, 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 (40).

\[
\begin{cases}
\text{Equation (41)} \implies V(\mu) = \frac{F_m(v) - V(\mu) - V'(\mu)(v - \mu)}{1 + \frac{\rho}{c} f(\mu, v)} \\
\text{Equation (42)} \implies V'(\mu) = F'_m + \frac{\rho}{c} V(\mu)(H'(v) - H'(\mu)) \\
V(\mu) = \frac{F_m(\mu)}{1 - \frac{\rho}{c} f(v, \mu)} \quad V'(\mu) = F'_m + \frac{\rho}{c} \frac{F_m(\mu)(H'(v) - H'(\mu))}{1 - \frac{\rho}{c} f(v, \mu)}
\end{cases}
\tag{44}
\]

Consistency of Equation (44) implies that $v = v(\mu)$ is characterized by the following ODE:
\[
\frac{\partial}{\partial \mu} F_m(\mu) + \frac{\partial}{\partial v} F_m(\mu) \frac{\dot{v} - F'_m + \frac{\rho}{c} F_m(\mu)(H'(v) - H'(\mu))}{1 - \frac{\rho}{c} f(v, \mu)} = 0
\tag{45}
\]

Simplifying Equation (45):
\[
\frac{F'_m + \frac{\rho}{c} F_m(\mu)(H'(v) - H'(\mu))}{1 - \frac{\rho}{c} f(v, \mu)} + \frac{\frac{\partial}{\partial v} F_m(\mu) H''(v)(v - \mu)}{(1 - \frac{\rho}{c} f(v, \mu))^2} \dot{v} = 0
\]

\[
\implies F_m(\mu)(H'(v) - H'(\mu)) + F_m(\mu) H''(v)(v - \mu) \dot{v} = -F'_m f(v, \mu) + F_m(\mu)(H'(v) - H'(\mu))(1 - \frac{\rho}{c} f(v, \mu))
\]

\[
\implies F_m(\mu) H''(v)(v - \mu) \dot{v} = -F'_m f(v, \mu) + F_m(\mu)(H'(v) - H'(\mu)) - \frac{\rho}{c} f(v, \mu) F_m(\mu)(H'(v) - H'(\mu))
\]

\[
\implies \dot{v} = \frac{F_m(\mu) H''(v)(v - \mu)}{F_m(\mu) H''(v)(v - \mu)}
\]\n
Since I want to solve for $V_0$ on $[\mu_0,1]$, I solve for $v_0$ at $\mu_0$ as the initial condition of ODE for $v$. 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:
Then I prove the properties of $V$.

\[
\begin{align*}
(F'_m - V'_0) \left(1 - \frac{\rho}{\varepsilon} J(v_0, \mu_0)\right) + \frac{\rho}{\varepsilon} F_m(\mu_0) (H'(v_0) - H'(\mu_0)) &= 0 \\
&\iff F'_m \left(1 + \frac{\rho}{\varepsilon} J(\mu_0, v_0)\right) + \frac{\rho}{\varepsilon} F_m(v_0) (H'(v_0) - H'(\mu)) = V'_0 \left(1 - \frac{\rho}{\varepsilon} J(v_0, \mu_0)\right) \\
&\iff F_m(\mu_0) \left(F'_m \left(1 + \frac{\rho}{\varepsilon} J(\mu_0, v_0)\right) + \frac{\rho}{\varepsilon} F_m(v_0) (H'(v_0) - H'(\mu))\right) = V'_0 F_m(\mu_0) \left(1 - \frac{\rho}{\varepsilon} J(v_0, \mu_0)\right)
\end{align*}
\]

Since $V_0 = \frac{F_m(\mu_0)}{1 - \frac{\rho}{\varepsilon} J(v_0, \mu_0)} \geq 0$, LHS is weakly positive. This satisfies the condition in Lemma S.16. Then Lemma S.16 guarantees existence of unique $v(\mu)$, and $v(\mu)$ is continuously decreasing from $\mu_0$ until it hits $v = \mu$. Suppose $v(\mu)$ hits $v = \mu$ at $\mu_m < 1$, define $V(\mu)$ as following:

\[
V(\mu) = \begin{cases} 
F_m(\mu) & \text{if } \mu \in [\mu_0, \mu_m) \\
\frac{F_m(\mu)}{1 - \frac{\rho}{\varepsilon} J(v(\mu), \mu)} & \text{if } \mu \in [\mu_m, 1]
\end{cases}
\]

Then I prove the properties of $V$:

1. $V$ is by construction smooth except for at $\mu_m$. When $\mu \to \mu_m$, $v(\mu) \to \mu$. Therefore $J(v(\mu), \mu) \to 0$. This implies $V(\mu) \to F_m(\mu)$. So $V$ is continuous.

2. By Equation (44), when $\mu \in [\mu_0, \mu_m)$:

\[
V'(\mu) = F'_m + \frac{F_m(\mu) (H'(v(\mu)) - H'(\mu))}{\frac{\rho}{\varepsilon} J(v(\mu), \mu)}
\]

When $\mu \to \mu_m$, $H'(v(\mu)) - H'(\mu) \to 0$, $J(v(\mu), \mu) \to 0$. Thus $V'(\mu) \to F'_m$. So $V' \in C[\mu_0, 1] \implies V \in C^1[\mu_0, 1]$.

3. Rewrite Equation (44) on $[\mu_0, 1]$:

\[
V'(\mu) = \frac{F'_m (1 + \frac{\rho}{\varepsilon} J(\mu, v) + F_m(v) (H'(v) - H'(\mu))}{1 - \frac{\rho}{\varepsilon} J(v, \mu)}
\]

(46)

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

Q.E.D.