# Supplemental Material:
## "Optimal Dynamic Information Acquisition"

Weijie Zhong

**Contents**



## S1 Proofs in Section 5

This section contains formal proofs for theorems and lemmas in Section 5.

### *S1.1 Useful lemmas*

I first establish a useful Lemma S.1. Lemma S.1 is an analog to three key theorems on mutual information proved in Cover and Thomas, 2012, generalizing the log-sum structure in mutual information to any function while keeping the key posterior separabiltiy.

#### S1.1.1 Information theory resutls

**Lemma S.1.** *Information measure $I(\mathcal{S};\mathcal{X}|\mu)$ satisfies the following properties:*

1. *Markov property: If $\mathcal{X}\to\mathcal{S}\to\mathcal{T}$, then $I(\mathcal{T};\mathcal{X}|\mathcal{S})=0$.*
2. *Linear additivity: $I(\mathcal{S},\mathcal{T};\mathcal{X}|\mu)=I(\mathcal{S};\mathcal{X}|\mu)+E[I(\mathcal{T};\mathcal{X}|\mathcal{S},\mu)]$.*
3. *Information processing inequality: If $\mathcal{X}\to\mathcal{S}\to\mathcal{T}$, then $I(\mathcal{T};\mathcal{X}|\mu)\leqslant I(\mathcal{S};\mathcal{X}|\mu)$.*

**Proof.**

1. *Markov property*: Suppose the signal realization of $\mathcal{S},\mathcal{T}$ are denoted by $s,t$. Then:
$$\begin{aligned}I(\mathcal{T};\mathcal{X}|\mathcal{S})&=E_s[H(\mu(x|s))-E_t[H(\mu(x|t,s))|s]]\\&=E_s[H(\mu(x|s))-E_t[H(\mu(x|s))|s]]\\&=0\end{aligned}$$
First equality is by definition of $I$. Second equality is by $\mathcal{T}\perp\mathcal{X}|\mathcal{S}$, then conditional on $s$, $t$ will not shift belief of $\mathcal{X}$ at all.

2. *Chain rule*: Suppose the signal realization of $\mathcal{S},\mathcal{T}$ are denoted by $s,t$. Then:
$$\begin{aligned}I(\mathcal{S},\mathcal{T};\mathcal{X}|\mu)&=E_{s,t}[H(\mu)-H(\mu(x|s,t))]\\&=E_{s,t}[H(\mu)-H(\mu(x|s))+(H(\mu(x|s))-H(\mu(x|s,t)))]\\&=(H(\mu)-E_s[H(\mu(x|s))])+(E_s[H(\mu(x|s))-E_t[H(\mu(x|s,t))|s]])\\&=I(\mathcal{S};\mathcal{X}|\mu)+E[I(\mathcal{T};\mathcal{X}|\mathcal{S},\mu)]\end{aligned}$$
First equality is by definition. Second equality is trivial. Third equality is by chain rule of conditional expectation.

3. *Information processing inequality*:
$$\begin{aligned}I(\mathcal{S};\mathcal{X}|\mu)&=I(\mathcal{S},\mathcal{T};\mathcal{X}|\mu)-I(\mathcal{T};\mathcal{X}|\mathcal{S},\mu)\\&=I(\mathcal{S},\mathcal{T};\mathcal{X}|\mu)\\&=I(\mathcal{T};\mathcal{X}|\mu)+I(\mathcal{S};\mathcal{X}|\mathcal{T},\mu)\\&\geqslant I(\mathcal{T};\mathcal{X}|\mu)\end{aligned}$$
First and third equalities are from chain rule. Second equality is from Markov property.

*Q.E.D.*

#### S1.1.2 Concavification theorem

**Theorem S.1** (Concavification)**.** *Let $X$ be a finite state space, $V\in C(\Delta X)$, $\mu\in\Delta X$. $H\in C(\Delta X)$ is non-negative. $f:\mathbb{R}^+\mapsto\overline{\mathbb{R}}^+$ continuous, increasing and convex. Then there exists $P$ s.t. $|\text{supp}(P)|\leqslant 2|X|$ solving:*
$$\begin{aligned}&\sup_{P\in\Delta^2X}E_P[V(\nu)]-f(H(\mu)-E_P[H(\nu)])\\&\text{s.t. }E_P[\nu]=\mu\end{aligned}\tag{S.1}$$
*Let $I^*=H(\mu)-E_P[H(\nu)]$, there exists $\lambda\in\mathrm{d}f(I^*)$ such that:*
$$\text{co}(V+\lambda H)(\mu)=E_P[(V+\lambda H)(\mu)]$$

**Proof.** Theorem S.1 is a corollary of Lemma 1 and Theorem 4 of Zhong, 2017.

*Support size*: since objective function is monototic in $(E_P[V],E_P[H])$, optimal solution must be on the boudary of set $\{(E_P[V],E_P[H])|E_P[\nu]=\mu\}$. Lemma 1 of Zhong, 2017 implies that there exists $P$ solving Equation (S.1) and $|\text{supp}(P)|\leqslant 2|X|$.

*Concavification*: Suppose $f(I)=\infty\iff I>\bar{I}$. Since $v-f(H(\mu)-h)$ is a concave function in $(v,h)$, and $H(\mu)-h\leqslant\bar{I}$ is a linear constraint, we can apply Theorem 4 of Zhong, 2017: let $V^*$ be maximum of Equation (S.1), there exists $\lambda$ s.t.
$$\begin{cases}P\in\arg\max\limits_{\substack{P\in\Delta^2(x)\\E_P[\nu]=\mu}}E_P[\lambda_1V-\lambda_2H]\\(E_P[V],I^*)\in\arg\min\limits_{\substack{I\leqslant\bar{I}\\v-f(I)>V^*}}\lambda_1v-\lambda_2I\end{cases}$$

Then by Kuhn-Tucker condition (generalized to subgradients), there exists $\eta,\gamma\geqslant 0$ such that:

$$\begin{cases}\lambda_1=\eta\\ \lambda_2\in -\eta\partial f(I^*)-\gamma\end{cases}$$

If $\eta=0$, then $\gamma>0$ and $P$ maximizes $E_P[\gamma H]$, then optimal $P$ is uninformative and $I^*=0$, contradiction. So $\eta>0$. If $\gamma=0$, then we can normalize $(\lambda_1,-\lambda_2)$ to $(1,\lambda)$ and $\lambda\in\partial f(I^*)$. If $\gamma>0$, the complementary slackness condition implies that $I^*=\bar{I}$ and $\lambda^2/\eta\in\partial f(I^*)$. So we can also normalize $(\lambda_1,-\lambda_2)$ to $(1,\lambda)$ and $\lambda\in\partial f(I^*)$. *Q.E.D.*

### *S1.1.3 Decomposition of information*

In this section, I prove two important lemmas. Lemmas S.2 and S.3 shows that any static information structure can be decomposed into a continuous time belief process on unit time interval such that the flow reduction of informativeness is constant.

**Lemma S.2.** *$\forall\mu\in\Delta(X)$, $\forall\pi\in\Delta^2(X)$, $\int\pi(\nu)\mathrm{d}\nu=\mu$ and $|\pi|$ is finite. Then there exists probability space $(\Omega,\mathcal{F},\mathbb{P})$ and stochastic process $\langle\mu_t\rangle$ s.t.:*

1. *$\langle\mu_t\rangle$ is a Markovian martingale.*
2. *$\mu_0=\mu$, $\mu_1\sim\pi$.*
3. *$\forall t_1,t_2\in[0,1]$ and $t_1<t_2$, $E[H(\mu_{t_1})-H(\mu_{t_2})|\mathcal{F}_{t_1}]=(t_2-t_1)E[H(\mu_0)-H(\mu_1)]$.*

**Proof.** Define $\pi(\nu)$ as $(\nu_i,\pi_i)_{i=1}^N$. Let $M=\sum\pi_iH(\nu_i)-H(\mu)$. Consider the space $\Delta(N)$. Let $x_i=(0,\cdots,0,1,0,\cdots,0)$, which is an $N$ dimensional vector with only $i$th element being 1. Then $\{x_i\}_{i=1}^N$ is an orthogonal normal base of $\Delta(N)$. Let $x_\mu=(\pi_1,\cdots,\pi_N)$. Then it is easy to see that $x_\mu\in\Delta(N)$. $\forall\lambda_i\in[0,1]$, let $x_{i,\lambda_i}=\lambda_ix_\mu+(1-\lambda_i)x_i\in\Delta(N)$.

Define map $f:\Delta(N)\to\Delta(X)$ as $f(x)=\sum_{j=1}^N x^j\nu_j$ ($f$ is a linear map). Then consider

$$Q_i(\lambda_i)=\sum x^j_{i,\lambda_i}(H(f(x_{i,\lambda_i}))-H(\nu_j)) \tag{S.2}$$

Now consider properties of $Q_i$. First of all, since $H$ is continuous and $f$ is linear, $Q_i(\lambda_i)$ is continuous in $\lambda_i$. Second, suppose $\lambda_i'>\lambda_i$ and $\lambda_{i,\alpha}=\alpha\lambda_i+(1-\alpha)\lambda_i'$, consider:

$$\begin{aligned}&Q_i(\lambda_{i,\alpha})-\alpha Q_i(\lambda_i')-(1-\alpha)Q_i(\lambda_i)\\ =&H(f(x_{i,\lambda_{i,\alpha}}))-\alpha H(f(x_{i,\lambda_i}))-(1-\alpha)H(f(x_{i,\lambda_i'}))\\ &-\sum\left(x^j_{i,\lambda_{i,\alpha}}-\alpha x^j_{i,\lambda_i}-(1-\alpha)x^j_{i,\lambda_i'}\right)H(\nu_j)\\ =&H(f(x_{i,\lambda_{i,\alpha}}))-\alpha H(f(x_{i,\lambda_i}))-(1-\alpha)H(f(x_{i,\lambda_i'}))\\ =&H(f(\alpha x^j_{i,\lambda_i}+(1-\alpha)x^j_{i,\lambda_i'}))-\alpha H(f(x_{i,\lambda_i}))-(1-\alpha)H(f(x_{i,\lambda_i'}))\\ =&H(\alpha f(x^j_{i,\lambda_i})+(1-\alpha)f(x^j_{i,\lambda_i'}))-\alpha H(f(x_{i,\lambda_i}))-(1-\alpha)H(f(x_{i,\lambda_i'}))\\ \geqslant&0\end{aligned}$$

The first equality is by definition of $Q_i$. The second and third equalities is from linearity of $x_{i,\lambda_i}$ in $\lambda_i$. The forth equality is from linearity of $f$. The last equality is from concavity of $H$. Hence, $Q_i(\lambda_i)$ is concave in $\lambda_i$. It is easy to verify that when $\lambda_i=0$, $x_{i,\lambda_i}=x_i$ and $f(x_{i,\lambda_i})=\nu_i$ so $Q_i(0)=0$. When $\lambda_i=1$, $x_{i,\lambda_i}=x_\mu$ and $f(x_{i,\lambda_i})=\mu$ so $Q_i(1)=\sum\pi_j(H(\mu)-H(\nu_j))=M$. Since $Q_i$ is concave, the only possibility is that $Q_i$ is first increasing then decreasing. Since $Q_i$ is a continuous function, $\forall t\in[0,1]$, there exists $\lambda_i$ in increasing region of $Q_i$ s.t.:

$$Q_i(\lambda_i(t))=(1-t)M$$

Since $(1-t)M$ is strictly decreasing $M$, $\lambda_i(t)$ is strictly decreasing in $M$. When $t\in(0,1]$, $\lambda_i(t)\in[0,1)$. Let $f(x_{i,\lambda_i(t)})=\mu_i(t)$. Define:

$$\pi_i(t)=\frac{\frac{\pi_i}{1-\lambda_i(t)}}{\sum_j\frac{\pi_j}{1-\lambda_j(t)}}$$

It is easy to verify that:

$$\begin{aligned}\sum\pi_i(t)\mu_i(t)=&f\left(\sum\pi_i(t)x_{i,\lambda_i(t)}\right)\\ =&f\left(\sum\pi_i(t)(\lambda_i(t)x_\mu+(1-\lambda_i(t))x_i)\right)\\ =&f\left(\frac{\sum\frac{\pi_i\lambda_ix_\mu}{1-\lambda_i(t)}+\sum\pi_ix_i}{\sum\frac{\pi_i}{1-\lambda_i}}\right)\end{aligned}$$

$$=f\left(\frac{\sum\frac{\pi_i\lambda_i}{1-\lambda_i}+1}{\sum\frac{\pi_i}{1-\lambda_i}}x_\mu\right)$$
$$=f(x_\mu)=\mu$$

Now for any $t,t'\in(0,1]$, and $t'>t$, define:

$$\pi_j(t'|\mu_i(t))=\begin{cases}\frac{\lambda_i(t)-\lambda_i(t')}{1-\lambda_i(t')}\pi_j(t') & \text{if } i\neq j\\ \frac{\lambda_i(t)-\lambda_i(t')}{1-\lambda_i(t')}\pi_i(t')+\frac{1-\lambda_i(t)}{1-\lambda_i(t')} & \text{if } i=j\end{cases}$$

It is easy to verify that:

$$\begin{aligned}\sum_j\pi_j(t'|\mu_i(t))\mu_j(t')&=f\left(\frac{1-\lambda_i(t)}{1-\lambda_i(t')}x_{i,\lambda_i(t')}+\frac{\lambda_i(t)-\lambda_i(t')}{1-\lambda_i(t')}\sum\pi_j(t')x_{j,\lambda_j(t')}\right)\\&=f\left(\frac{1-\lambda_i(t)}{1-\lambda_i(t')}x_{i,\lambda_i(t')}+\frac{\lambda_i(t)-\lambda_i(t')}{1-\lambda_i(t')}x_\mu\right)\\&=f\left(\frac{1-\lambda_i(t)}{1-\lambda_i(t')}\left(\lambda_i(t')x_\mu+(1-\lambda_i(t'))x_i\right)+\frac{\lambda_i(t)-\lambda_i(t')}{1-\lambda_i(t')}x_\mu\right)\\&=f\left(\lambda_i(t)x_\mu+(1-\lambda_i(t))x_i\right)\\&=\mu_i(t)\end{aligned}\tag{S.3}$$

and:

$$\begin{aligned}\sum_j\pi_j(t)\pi_i(t'|\mu_j(t))&=\frac{1}{\sum\frac{\pi_j}{1-\lambda_j(t)}}\left(\frac{1-\lambda_i(t)}{1-\lambda_i(t')}\frac{\pi_i}{1-\lambda_i(t)}+\sum_j\frac{\lambda_j(t)-\lambda_j(t')}{1-\lambda_j(t')}\frac{\pi_j}{1-\lambda_j(t)}\pi_i(t')\right)\\&=\frac{1}{\sum\frac{\pi_j}{1-\lambda_j(t)}}\left(\frac{\pi_i}{1-\lambda_i(t')}+\left(\sum\frac{\pi_j}{1-\lambda_j(t)}-\sum\frac{\pi_j}{1-\lambda_j(t')}\right)\pi_i(t')\right)\\&=\frac{\frac{\pi_i}{1-\lambda_i(t')}}{\sum\frac{\pi_j}{1-\lambda_j(t)}}+\pi_i(t')-\frac{\frac{\pi_i}{1-\lambda_i(t')}}{\sum\frac{\pi_j}{1-\lambda_j(t)}}\\&=\pi_i(t')\end{aligned}\tag{S.4}$$

Now we verify the dynamic consistency of $\pi_i$. $\forall r>s>t$:

$$\begin{aligned}&\sum_j\pi_j(s|\mu_i(t))\pi_k(r|\mu_j(s))\\&=\sum_j\frac{\lambda_i(t)-\lambda_i(s)}{1-\lambda_i(s)}\pi_j(s)\pi_k(r|\mu_j(s))+\frac{1-\lambda_i(t)}{1-\lambda_i(s)}\pi_k(r|\mu_i(s))\\&=\frac{\lambda_i(t)-\lambda_i(s)}{1-\lambda_i(s)}\pi_k(r)+\frac{1-\lambda_i(t)}{1-\lambda_i(s)}\frac{\lambda_i(s)-\lambda_i(r)}{1-\lambda_i(r)}\pi_k(r)+\mathbf{1}_{k=i}\frac{1-\lambda_i(t)}{1-\lambda_i(s)}\frac{1-\lambda_i(s)}{1-\lambda_i(r)}\\&=\frac{\lambda_i(t)-\lambda_i(r)}{1-\lambda_i(r)}\pi_k(r)+\mathbf{1}_{k=i}\frac{1-\lambda_i(t)}{1-\lambda_i(r)}\\&=\pi_k(r|\mu_i(t))\end{aligned}\tag{S.5}$$

The second equality is from Equation (S.4). Now define the stochastic process $\langle\mu_t\rangle$. First, I complete the definition of $\mu_i(t)$ and $\pi_i(t)$. Let $\mu(0)=\mu$, $\pi_i(t|\mu(0))=\pi_i(t)$. $\forall t>1$, define $\mu_i(t)=\nu_i$, $\pi_j(t|\mu_i(1))=\mathbf{1}_{i=j}$. Define $S_i=\{\mu_i(t)|t\in(0,1]\}$. Since $\nu_i$ are distinct, $S_i$ are disjoint sets. Since $\lambda_i(t)$ is strictly decreasing, $\mu_i(t)$ is a one-to-one map from $(0,1]$ to $S_i$. Let $S=(\bigcup S_i)\bigcup\{\mu\}$. Define: $\tau:S\to[0,1]$:

$$\tau(\nu)=\begin{cases}\mu_i(t)^{-1}(\nu) & \text{if } \mu\in S_i\\ 0 & \text{if } \nu=\mu\end{cases}$$

Now we can define the Markov transition kernel of $\langle\mu_t\rangle$: $\forall x,y\in S$, $t\in\mathbb{R}^+$,

$$P_t(x,y)=\sum_i\mathbf{1}_{y=\mu_i(\tau(x)+t)}\pi_i(\tau(x)+t|x)$$

We verify the Chapman-Kolmogorov equation: $\forall x,z\in S$, $t,s\in\mathbb{R}^+$:

- If $\tau(x)+t+s\leqslant 1$, then:

$$\begin{aligned}\int P_t(x,y)P_s(y,z)\mathrm{d}y &= \sum_{i,j}\mathbf{1}_{z=\mu_i(\tau(\mu_j(\tau(x)+t))+s)}\pi_j(\tau(x)+t|x)\pi_i\big(\tau(\mu_j(\tau(x)+t))+s|\mu_j(\tau(x)+t)\big)\\ &= \sum_{i,j}\mathbf{1}_{z=\mu_i(\tau(x)+t+s)}\pi_j(\tau(x)+t|x)\pi_i(\tau(x)+t+s|\mu_j(\tau(x)+t))\\ &= \sum_{i}\mathbf{1}_{z=\mu_i(\tau(x)+s+t)}\pi_i(\tau(x)+t+s|x)\\ &= P_{t+s}(x,z)\end{aligned}$$

  The second equality is from definition of $\tau$. The third equality is from Equation (S.5).

- If $\tau(x)+t=1$, then:

$$\begin{aligned}\int P_t(x,y)P_s(y,z)\mathrm{d}y &= \sum_{i,j}\mathbf{1}_{z=\mu_i(\tau(\mu_j(\tau(x)+t))+s)}\pi_j(\tau(x)+t|x)\pi_i\big(\tau(\mu_j(\tau(x)+t))+s|\mu_j(\tau(x)+t)\big)\\ &= \sum_{i,j}\mathbf{1}_{z=\nu_i}\pi_j(1|x)\pi_i(1+s|\nu_j)\\ &= \sum_{i}\mathbf{1}_{z=\nu_i}\pi_i(1|x)\\ &= \sum_{i}\mathbf{1}_{z=\mu_i(\tau(x)+t+s)}\pi_i(\tau(x)+t+s|x)\\ &= P_{t+s}(x,z)\end{aligned}$$

- If $\tau(x)=1$, then the C-K equation is trivially satisfied:

$$\int P_t(x,y)P_s(y,z)\mathrm{d}y=\sum_{i,j}\mathbf{1}_{z=\nu_i}\mathbf{1}_{\nu_i=\nu_j}\mathbf{1}_{\nu_j=x}=\mathbf{1}_{z=x}=P_{t+s}(x,z)$$

- Now for any general case $\tau(x)<1$ and $\tau(x)+t+s>1$, we can add 1, and apply the C-K equation in the last two cases jointly to establish the C-K equation in the general case.

Since we verified the C-K equation for Markov transition kernel $P_t(\cdot,\cdot)$, it is easy to see that $\forall t_1,\dots,t_n$, the measure given by:

$$P_{t_1,\dots,t_n}(x_1,\dots,x_n)=P_{t_1}(\mu,x_1)\prod P_{t_{i+1}-t_i}(x_i,x_{i+1})$$

satisfies the conditions in Daniell-Kolmogorov theorem (see Dellacherie and Meyer, 1979). Then, a simple corollary of Daniell-Kolmogorov theorem states that there exists a probability space $(\Omega,\mathcal{F},\mathbb{P})$ and stochastic process $\langle\mu_t\rangle$ such that any finite dimensional marginal distribution of $\langle\mu_t\rangle$ is given by $P$. Now Equation (S.3) implies $\langle\mu_t\rangle$ is a martingale and the C-K equation implies that $\langle\mu_t\rangle$ is Markovian.

What remains to be verified is the third property of Lemma S.3. $\forall t_1,t_2\in[0,1]$ and $t_1<t_2$, $\forall\mu_{t_1}\in\{\mu_i(t_1)\}$,

$$\begin{aligned}E[H(\mu_{t_1})-H(\mu_{t_2})|\mathcal{F}_{t_1}] &= H(\mu_{t_1})-E[H(\mu_{t_2})|\mu_{t_1}]\\ &= H(\mu_{t_1})-\int P_{t_2-t_1}(\mu_{t_1},\mu_{t_2})H(\mu_{t_2})\mathrm{d}y\\ &= H(\mu_{t_1})-\sum_i\pi_i(\tau(\mu_{t_1})+t_2-t_1|\mu_{t_1})H(\mu_i(\tau(\mu_{t_1})+t_2-t_1))\\ &= H(\mu_{t_1})-\sum_i\pi_i(t_2|\mu_{t_1})H(\mu_i(t_2))\\ &= H(\mu_{t_1})-\sum_i\pi_i(t_2|\mu_{t_1})\left(\sum_j\pi_j(1|\mu_i(t_2))H(\nu_j)\right)\\ &\quad -\sum_i\pi_i(t_2|\mu_{t_1})\left(H(\mu_i(t_2))-\sum_j\pi_j(1|\mu_i(t_2))H(\nu_j)\right)\\ &= \left(H(\mu_{t_1})-\sum_j\pi_j(1|\mu_{t_1})H(\nu_j)\right)-\sum_i\pi_i(t_2|\mu_{t_1})Q_i(\lambda_i(t_2))\end{aligned}$$

$$
\begin{aligned}
&=(1-t_1)M-\sum_i \pi_i(t_2|\mu_{t_1})(1-t_2)M\\
&=(t_2-t_1)(H(\mu_0)-E[H(\mu_1)])
\end{aligned}
$$

*Q.E.D.*

**Lemma S.3.** *$\forall\mu\in\Delta(X)$, $\pi\in\Delta^2(X)$ and $\int\pi(\nu)\mathrm{d}\nu=\mu$. Then there exists probability space $(\Omega,\mathcal{F},\mathbb{P})$ and stochastic process $\langle\mu_t\rangle$ s.t.:*

1. *$\langle\mu_t\rangle$ is a martingale.*
2. *$\mu_0=\mu$, $\mu_1\sim\pi$.*
3. *$\forall t_1,t_2\in[0,1]$ and $t_1<t_2$, $E[H(\mu_{t_1})-H(\mu_{t_2})|\mathcal{F}_{t_1}]=(t_2-t_1)E[H(\mu_0)-H(\mu_1)]$.*

**Proof.** If $|\mathrm{Supp}(\pi)|$ is finite, the Lemma S.3 is identical to Lemma S.2 and the proof is done. Now I discuss the general case where $\mathrm{Supp}(\pi)$ is an infinite set. Let $M=\int\pi(\nu)(H(\mu)-H(\nu))\mathrm{d}\nu$.

*Step 1*. Discretizing $\Delta(X)$. Since $H(\mu)$ is a continuous function on $\Delta(X)$, by Heine-Cantor theorem $H(\mu)$ is uniformly continuous. Pick $\varepsilon_k=\frac{M}{2^k}$ and let $\delta_k$ be corresponding continuity parameter for $\varepsilon_k$. Discretize $\Delta(X)$ into a set of nested grids with grid size $d_k\leqslant\delta_k$. Let $D^k_\nu$ be each $d_k$-cube containing $\mu$. Then $\forall\mu\in\Delta(X)$, $\forall\pi'\in\Delta(D^k_\mu)$:

$$
\int\pi'(\nu)(H(\mu)-H(\nu))\leqslant\varepsilon_k
$$

*Step 2*. Index all $d_1$-cubes as $\left\{D^1_{i_1}\right\}_{i_1\in I_1}$. $\forall i_1\in I_1$, define:

$$
\begin{cases}
\mu^1_{i_1}=\dfrac{\int_{\nu\in D^1_{i_1}}\nu\pi(\nu)\mathrm{d}\nu}{\int_{\nu\in D^1_{i_1}}\pi(\nu)\mathrm{d}\nu}\\
\pi^1_{i_1}(\nu)=\dfrac{\mathbf{1}_{\nu\in D^1_{i_1}}\pi(\nu)}{\int_{\nu\in D^1_{i_1}}\pi(\nu)\mathrm{d}\nu}\\
q^1_{i_1}=\displaystyle\int_{\nu\in D^1_{i_1}}\pi(\nu)\mathrm{d}\nu
\end{cases}
\quad\text{and}\quad
\begin{cases}
M^1_{i_1}=\displaystyle\int\pi^1_{i_1}(\nu)(H(\mu^1_{i_1})-H(\nu))\mathrm{d}\nu\\
M^0=\displaystyle\sum_{i_1}q^1_{i_1}\left(H(\mu)-H(\mu^1_{i_1})\right)
\end{cases}
$$

It is easy to verify that:

$$
\begin{cases}
\sum q^1_{i_1}\mu^1_{i_1}=\mu\\
\int\pi^1_{i_1}\nu\mathrm{d}\nu=\mu^1_{i_1}\\
\sum q^1_{i_1}\pi^1_{i_1}(\nu)=\pi(\nu)\\
M^1_{i_1}\leqslant\varepsilon_1\\
M^1+\sum q^1_{i_1}M^1_{i_1}=M
\end{cases}
$$

Now consider distribution $(q^1_{i_1},\mu^1_{i_1})$. Let $x_{i_1}=(0,\dots,0,1,0,\dots,0)$ where only $i_1$th element is 1. Then $\{x_{i_1}\}$ is an orthogonal normal base of $\Delta(I_1)$. Let $x_\mu=(q^1_1,\dots,q^1_{I_1})$. Then it is easy to see than $x_\mu\in\Delta(N)$. $\forall\lambda\in[0,1]$, define $x_{i_1,\lambda}=\lambda x_\mu+(1-\lambda)x_{i_1}$.

Define linear map $f:\Delta(I_1)\to\Delta(X)$ as $f(x)=\sum_{i=1}^{I_1}x^{i_1}\mu^1_{i_1}$ ($x^{i_1}$ is $i_1$th coordinate of vector $x$). Then consider:

$$
Q_{i_1}(\lambda)=\sum_{j=1}^{I_1}x^j_{i_1,\lambda}(H(f(x_{i_1,\lambda}))-H(\mu^1_j)+M^1_j)
$$

Now consider properties of $Q_{i_1}$. First of all, since $H$ is continuous and $f$ is linear, $Q_{i_1}(\lambda)$ is continuous in $\lambda$. Second, $Q_{i_1}(0)=M^1_{i_1}$ and $Q_{i_1}(1)=M$. Since $M>\varepsilon_1\geqslant M^1_{i_1}$, by intermediate value theorem there exists $\lambda_{i_1}$ s.t. $Q_{i_1}(\lambda_{i_1})=\varepsilon_1$. Now define:

$$\begin{cases}\widetilde{\mu}^1_{i_1} = f(x_{i_1,\lambda_{i_1}}) \\ \widetilde{q}^1_{i_1} = \dfrac{\frac{q^1_{i_1}}{1-\lambda_{i_1}}}{\sum_{j=1}^{I_1}\frac{\pi^1_j}{1-\lambda_j}} \\ \widetilde{\pi}^1_{i_1}(\nu) = \sum_{j=1}^{I_1} x^j_{i_1,\lambda_{i_1}}\pi^1_j(\nu)\end{cases} \quad \text{and} \quad \begin{cases}\widetilde{M}^1_{i_1} = \int \widetilde{\pi}^1_{i_1}(\nu)(H(\widetilde{\mu}^1_{i_1}) - H(\nu))\mathrm{d}\nu \\ \widetilde{M}^0 = \sum_{I_1}\widetilde{q}^1_{i_1}\left(H(\mu) - H(\widetilde{\mu}^1_{i_1})\right)\end{cases}$$

It is easy to verify that:

$$\begin{cases}\sum \widetilde{q}^1_{i_1}\widetilde{\mu}^1_{i_1} = \mu \\ \int \widetilde{\pi}^1_{i_1}\nu\mathrm{d}\nu = \widetilde{\mu}^1_{i_1} \\ \sum \widetilde{q}^1_{i_1}\widetilde{\pi}^1_{i_1}(\nu) = \pi(\nu) \\ \widetilde{M}^1_{i_1} \equiv \frac{M}{2}, \widetilde{M}^0 = \frac{M}{2}\end{cases}$$

*Step 3*. Recursively apply *step 2*. Suppose I have defined a discrete time stochastic process for $i \in \{0,\ldots,k\}$ satisfying $\mu^0 = \mu$, $\mu^k \sim \pi$ and:

$$\begin{cases}\left|\text{Supp}(\mu^i)\middle|\mathcal{F}_i\right| \leqslant I_i < \infty \ \forall i < k \\ E\left[\mu^i\middle|\mathcal{F}_j\right] = \mu^i \ \forall j < i \\ E\left[H(\mu^j) - H(\mu^i)\middle|\mathcal{F}_j\right] = \sum_{l=j}^{i-1}\frac{M}{2^{l+1}} \\ E\left[H(\mu^k) - H(\mu^i)\middle|\mathcal{F}_i\right] = \frac{M}{2^i}\end{cases} \tag{S.6}$$

Noticing that I have verified that $\left(\mu,(\widetilde{q}^i_{i_1},\widetilde{\mu}^1_{i_1}),\widetilde{\pi}^1_{i_1}\right)$ we find in step one satisfies this condition for $k = 1$. Now we prove that we can construct a discrete time stochastic process with $k+1$. Define a new process $\langle\mu^i\rangle$ exactly as in the assumption for $i < k$. Now for any sample path in $\mathcal{F}_{k-1}$, applying the procedure in *step 2* to prior $\mu^{k-1}$ and distribution of $(\mu^k)$. Then I get $(\widetilde{q}^k_{i_k},\widetilde{\mu}^k_{i_k},\widetilde{\pi}^{k_{i_k}})$ satisfying:

$$\begin{cases}\sum \widetilde{q}^k_{i_k}\widetilde{\mu}^k_{i_k} = \mu^{k-1} \\ \int \widetilde{\pi}^k_{i_k}(\nu) = \widetilde{\mu}^k_{i_k} \\ \sum \widetilde{q}^k_{i_k}\widetilde{\pi}^k_{i_k}(\nu) \sim \mu^k|\mathcal{F}_{k-1} \\ \int \widetilde{\pi}^k_{i_k}(\nu)(H(\widetilde{\mu}^k_{i_k}) - H(\nu))\mathrm{d}\nu \equiv \frac{M}{2^k} = \sum_{I_K}\widetilde{q}^k_{i_k}\left(H(\mu^{k-1}) - H(\widetilde{\mu}^k_{i_k})\right),\end{cases} \tag{S.7}$$

In the new process, let $\mu^k|\mathcal{F}_{k-1} \sim (\widetilde{q}^k_{i_k},\widetilde{\mu}^k_{i_k})$ and $\mu^{k+1}|\mathcal{F}_{k-1},\widetilde{\mu}^k_{i_k} \sim \widetilde{\pi}^k_{i_k}$. Now let us verify Equation (S.6). The first condition is trivially satisfied for $i < k$. If $i = k$, by construction the support size of $\widetilde{\mu}^k_{i_i}$ is finite. The second condition is true for $i = k, k+1$ given Equation (S.7)'s first two properties. The third and forth condition are implied the last condition of Equation (S.7). $\mu^0$ is still $\mu$ and $\mu^{k+1} \sim \pi$ by third property of Equation (S.7).

Hence, for any positive $K$, a $\langle\mu^i\rangle_{i\leqslant K}$ is well defined. And by construction, for any $K_1 < K_2$, the two processes have exactly same path distribution for $i < K_1$. So except if I need to explicitly use the distribution of $\mu^K$, otherwise I refer to $\langle\mu^i\rangle$ as an infinite process.

*Step 4*. Extension to continuous time. Let $T_k = 1 - \frac{1}{2^k}$. The main idea is to define finite dimensional joint distribution at $T_k$'s according to $\langle\mu^k\rangle$. Then within each interval $[T_k, T_{k+1}]$, the process is defined using Lemma S.2. For any sequence of $\mu_0,\ldots,\mu_{T_k}$, define:

$$P(\mu_{T_k} = \mu^k|\mu_0,\ldots\mu_{T_{k-1}}) = \mathbb{P}(\mu^k\middle|\mu^l_{i_l} = \mu_{T_l}, \forall l < k)$$

Now for any $t_1,\ldots,t_k$ and $\mu_{t_1},\ldots,\mu_{t_k}$, I define the joint distribution of the sample path. First assume $t_k<1$. Then there exists a unique sequence of:

$$0\cdots\underbrace{T_{l_1},t_1\cdots t_{m_1},T_{l_1+1}}_{\text{Interval 1}}\cdots\underbrace{T_{l_2},t_{m_1+1}\cdots t_{m_2},T_{l_2+1}}_{\text{Interval 2}}\cdots\underbrace{T_{l_n},t_{m_{n-1}+1}\cdots t_k}_{\text{Interval n}}$$

Noticing that $t_{m_s+1}$ can be same as $T_{l_{s+1}}$ and $t_{m_s}$ can be same as $T_{l_s+1}$. Now for any sequence of $(\mu_{T_{l_s}})$, apply Lemma S.2 to prior $\mu_{T_{l_s}}$ and distribution $P(\mu_{T_{l_s+1}}|\mu_0,\ldots,\mu_{T_{l_s}})$. Lemma S.2 implies that there exists a space $(\Omega,\mathcal{F},\mathbb{P})$ and $\langle\widetilde{\mu}_t\rangle$ s.t. $\widetilde{\mu}_0=\mu_{T_{l_s}}$ and $\widetilde{\mu}_1\sim\mathbb{P}(\mu_{T_{l_s+1}})$ (the dependence of all terms on $(\mu_{T_{l_s}})$ is omitted for notational simplicity). Define:

$$\begin{aligned}&P\left(\mu_{t_{m_{s-1}+1}},\ldots,\mu_{t_{m_s}},\mu_{T_{l_s+1}}\middle|\mu_0,\ldots,\mu_{T_{l_s}}\right)\\=&\mathbb{P}\left(\widetilde{\mu}_{2^{l_s}\cdot(t_{m_{s-1}+1}-T_{l_s})}=\mu_{t_{m_{s-1}+1}},\ldots,\widetilde{\mu}_{2^{l_s}\cdot(t_{m_s}-T_{l_s})}=\mu_{t_{m_s}},\widetilde{\mu}_1=\mu_{T_{l_s+1}}\middle|\mu_0,\ldots,\mu_{T_{l_s}}\right)\end{aligned}$$

Now we can define the finite joint distribution of $\mu_{t_k}$:

$$\begin{aligned}&P(\mu_0,\ldots,\mu_{t_k})\\=&\int\prod_{s=1}^{n-1}\left[\left(\prod_{j=l_s+1}^{l_{s+1}}P(\mu_{T_j}|\mu_0,\ldots,\mu_{T_{j-1}})\right)\cdot P\left(\mu_{t_{m_s}},\ldots,\mu_{t_{m_{s+1}}},\mu_{T_{l_{s+1}+1}}\middle|\mu_0,\ldots\mu_{T_{l_{s+1}}}\right)\right]\mathrm{d}\mu_{T_1},\ldots\mu_{T_{l_n+1}}\end{aligned}$$

Noticing that by definition of $\langle\mu_k\rangle$, each $P(\mu_{T_j}|\mu_0,\ldots,\mu_{T_{j-1}})$ is a probability measure. By definition of $\langle\widetilde{\mu}_t\rangle$, each $P\left(\mu_{t_{m_s}},\ldots,\mu_{t_{m_{s+1}}},\mu_{T_{l_{s+1}+1}}\middle|\mu_0,\ldots\mu_{T_{l_{s+1}}}\right)$ is a joint probability measure. Therefore, $P(\mu_0,\ldots,\mu_{t_k})$ is a joint probability measure.

Now consider the case $t_k=1$. Since $t_{k-1}<1$, there must be some finite $T_l>t_{k-1}$. There exists a unique sequence of:

$$0\cdots\underbrace{T_{l_1},t_1\cdots t_{m_1},T_{l_1+1}}_{\text{Interval 1}}\cdots\underbrace{T_{l_2},t_{m_1+1}\cdots t_{m_2},T_{l_2+1}}_{\text{Interval 2}}\cdots,t_{k-1},T_{l_n+1}\cdots t_k$$

Pick $K=l_n+2$, for any given sequence of $\mu_0,\ldots,\mu_{T_{l_n+1}}$, $\forall S\subset\Delta(X)$ define:

$$P\left(S\middle|\mu_0,\ldots,\mu_{T_{l_n+1}}\right)=\mathbb{P}\left(\mu^K\in S\middle|\mu^l_{i_l}=\mu_{T_l},\forall l<K\right)$$

Now we can define the finite joint distribution of $\mu_{t_k}$:

$$\begin{aligned}&P(\mu_0,\ldots,\mu_{t_k})\\=&\int\prod_{s=1}^{n-1}\left[\left(\prod_{j=l_s+1}^{l_{s+1}}P(\mu_{T_j}|\mu_0,\ldots,\mu_{T_{j-1}})\right)\cdot P\left(\mu_{t_{m_s}},\ldots,\mu_{t_{m_{s+1}}},\mu_{T_{l_{s+1}+1}}\middle|\mu_0,\ldots\mu_{T_{l_{s+1}}}\right)\right]\\&\cdot P(S|\mu_0,\ldots,\mu_{T_{l_n+1}})\mathrm{d}\mu_{T_1},\ldots\mu_{T_{l_n+1}}\end{aligned}$$

Same as the previous case, each $P(\mu_{T_j}|\mu_0,\ldots,\mu_{T_{j-1}})$ and $P\left(\mu_{t_{m_s}},\ldots,\mu_{t_{m_{s+1}}},\mu_{T_{l_{s+1}+1}}\middle|\mu_0,\ldots\mu_{T_{l_{s+1}}}\right)$ are joint probability measures. Moreover, $P(S|\mu_0,\ldots,\mu_{T_{K-1}})$ is a probability measure. Importantly, by definition of $\langle\mu^k\rangle$, for $K_1<K$, $P$ is defined consistently:

$$P(S|\mu_0,\ldots,\mu_{K_1-1})=\int P(S|\mu_0,\ldots,\mu_{K-1})\mathrm{d}\mu_{K_1},\ldots,\mu_{K-1}$$

Therefore, $P(\mu_0,\ldots,\mu_{t_k})$ is a joint probability measure. To sum up, we defined a finite dimensional joint probability measure satisfying the conditions in Daniell-Kolmogorov theorem. Hence, there exists probability space $(\Omega,\mathcal{F},\mathbb{P})$ and $\langle\mu_t\rangle_{t\in[0,1]}$ satisfying all finite distributions.

*Step 5*. Verify that $\langle\mu_t\rangle$ satisfies Lemma S.3. $\mu_0=\mu$ is true by construction. $\forall S\subset\Delta(X)$:

$$\mathbb{P}(S)=\mathbb{P}(\mu^1\in S)=\pi(S)$$

So $\mu_1\sim\pi$. Now I verify property 1 and property 3. $\forall t_1,t_2\in[0,1]$ and $t_2>t_1$,

- *Case 1*. If $\exists k$ s.t. $T_k\leqslant t_1<t_2\leqslant T_{k+1}$, then by construction of $\langle\mu_t\rangle$ in $[T_j,T_{j+1}]$, $\mu_t$ is Markovian and $E[\mu_{t_2}|\mathcal{F}_{t_1}]=E[\mu_{t_2}|\mu_{t_1}]=\mu_{t_1}$.

$$\begin{aligned}E[H(\mu_{t_1})-H(\mu_{t_2})|\mu_{t_1}]=&2^k(t_2-t_1)E[H(\mu_{t_k})-H(\mu_{t_{k+1}})|\mathcal{F}_{t_k}]\\=&(t_2-t_1)2^k\cdot\frac{M}{2^k}\\=&(t_2-t_1)M\end{aligned}$$

- *Case 2*. There exists a unique sequence of:

$$0\cdots T_j,t_1,T_{j+1}\cdots T_k,t_2,T_{k+1}$$

By construction, for all path on $[0,t_1]$,

$$\begin{aligned}E[\mu_{t_2}|\mathcal{F}_{t_1}] &= E\Big[E[\mu_{t_2}|\mathcal{F}_{T_k}]\Big|\mathcal{F}_{t_1}\Big]\\ &= E\Big[\mu_{T_k}\Big|\mathcal{F}_{t_1}\Big]\\ &\vdots\\ &= E\Big[\mu_{T_{j+1}}|\mathcal{F}_{t_1}\Big]\\ &= \mu_{t_1}\end{aligned}$$

and

$$\begin{aligned}&H(\mu_{t_1}) - E[H(\mu_{t_2})|\mathcal{F}_{t_1}]\\ =&H(\mu_{t_1}) - E\Big[H(\mu_{T_{j+1}})|\mathcal{F}_{t_1}\Big] + E\Big[H(\mu_{T_{j+1}})|\mathcal{F}_{t_1}\Big] - E[H(\mu_{t_2})|\mathcal{F}_{t_1}]\\ =&2^j(T_{j+1}-t_1)\cdot\frac{M}{2^j} + E\Big[E\Big[H(\mu_{T_{j+1}}) - H(\mu_{t_2})|\mathcal{F}_{T_{j+1}}\Big]\Big|\mathcal{F}_{t_1}\Big]\\ =&(T_{j+1}-t_1)M + E\Big[H(\mu_{T_{j+1}}) - E\Big[H(\mu_{T_{j+2}})|\mathcal{F}_{T_{j+1}}\Big] + E\Big[H(\mu_{T_{j+2}}) - H(\mu_{t_2})|\mathcal{F}_{T_{j+1}}\Big]\Big|\mathcal{F}_{t_1}\Big]\\ =&(T_{j+1}-t_1)M + E\Big[\frac{M}{2^{j+2}} + E\Big[H(\mu_{T_{j+2}}) - H(\mu_{t_2})|\mathcal{F}_{T_{j+1}}\Big]\Big|\mathcal{F}_{t_1}\Big]\\ &\vdots\\ =&(1-\frac{1}{2^{j+1}}-t_1)M + \sum_{j+1}^{k-1}\frac{M}{2^{l+1}} + E\Big[E[H(\mu_{T_k}) - H(\mu_{t_2})|\mathcal{F}_{T_k}]\Big|\mathcal{F}_{t_1}\Big]\\ =&(1-\frac{1}{2^{j+1}}-t_1)M + \sum_{j+1}^{k-1}\frac{M}{2^{l+1}} + \left(t_2 - 1 + \frac{1}{2^k}\right)M\\ =&(t_2-t_1)M\end{aligned}$$

*Case 3.* $t_2=1$. Then there exists $k$ s.t $T_k \leqslant t_1 < T_{k+1}$. By construction, for all path on $[0,t_1]$:

$$\begin{aligned}E[\mu_1|\mathcal{F}_{t_1}] &= E[E[\mu_t|\mathcal{F}_{T_{k+1}}]|\mathcal{F}_{t_1}]\\ &= E[\mu_{T_{k+1}}|\mathcal{F}_{t_1}]\\ &= \mu_{t_1}\end{aligned}$$

and

$$\begin{aligned}&H(\mu_{t_1}) - E[H(\mu_1)|\mathcal{F}_{t_1}]\\ =&H(\mu_{t_1}) - E[H(\mu_{T_{k+1}})|\mathcal{F}_{t_1}] + E[E[H(\mu_{t_{K+1}}) - H(\mu_1)|\mathcal{F}_{T_{k+1}}]|\mathcal{F}_{t_1}]\\ =&\left(1-\frac{1}{2^{k+1}}-t_1\right)M + \frac{M}{2^{k+1}}\\ =&(t_2-t_1)M\end{aligned}$$

*Q.E.D.*

### *S1.2 Proof of Lemma 2*

**Proof.** I break the proof of Lemma 2 into three lemmas. Lemma S.4 shows that solving Equation (5′) is equivalent to solving Equation (S.8), which reduces the signal structure to be nested, and containing only action as direct signals and continuation signals. Then Lemma S.5 shows that solving Equation (S.8) is equivalent to solving Equation (S.9), which transforms signal process formulation to conditional distribution formulation. Then Lemma S.6 shows that solving functional equation Equation (S.10) is equivalent to solving sequential problem Equation (S.9) using the standard methodology. Finally, we apply Theorem S.1 to Equation (S.10) to further reduce the dimensionality of strategy space to Equation (6). *Q.E.D.*

**Lemma S.4** (Reduction of redundancy). $(\mathcal{S}^t,\mathcal{T},\mathcal{A}^\mathcal{T})$ *solves Equation (5′) if and only if there exists* $\left(\tilde{\mathcal{S}}^t,\mathcal{T},\mathcal{A}^\mathcal{T}\right)$ *solving :*

$$\sup_{\mathcal{S}^t,\mathcal{T},\mathcal{A}^\mathcal{T}}\sum_{t=0}^{\infty}\left(e^{-\rho dt}\mathbf{P}[\mathcal{T}=t]\left(E\left[u(\mathcal{A}^t,\mathcal{X})|\mathcal{T}=t\right]\right)-\mathbf{P}[\mathcal{T}>t]E\left[C_{dt}\left(I\left(\tilde{\mathcal{S}}^t;\mathcal{X}|\tilde{\mathcal{S}}^{\mathbf{t-1}}\right)\right)|\mathcal{T}>t\right]\right) \tag{S.8}$$

$$s.t.\ \tilde{\mathcal{S}}^t=\begin{cases}s_0 & when\ \mathcal{T}<t+1\\ \mathcal{A}^{t+1} & when\ \mathcal{T}=t+1\\ \mathcal{S}^t & when\ \mathcal{T}>t+1\end{cases}$$

*What's more, the optimal utility level is same in Equation (5′) and Equation (S.8).*

**Proof.** Suppose $(\mathcal{S}^t,\mathcal{T},\mathcal{A}^t)$ is a feasible strategy to Equation (5′). I first show that it is WLOO that the DM discards all information after taking an action: take given $\mathcal{T}$ and $\mathcal{A}^t$, let $s_0$ be a degenerate signal, define signal process $\hat{\mathcal{S}}^t$ as:

$$\hat{\mathcal{S}}^t=\begin{cases}\mathcal{S}^t & \text{when } \mathcal{T}\geqslant t+1\\ s_0 & \text{when } \mathcal{T}\leqslant t\end{cases}$$

By definition, $\hat{\mathcal{S}}^{\mathbf{t}}=\mathcal{S}^{\mathbf{t}}$ conditional on $\mathcal{T}\geqslant t+1$. Therefore:

$$I\left(\hat{\mathcal{S}}^t;\mathcal{X}|\hat{\mathcal{S}}^{\mathbf{t-1}},\mathbf{1}_{\mathcal{T}\leqslant t}\right)=\begin{cases}I(\mathcal{S}^t;\mathcal{X}|\mathcal{S}^{\mathbf{t-1}},\mathbf{1}_{\mathcal{T}\leqslant t}) & \text{when } \mathcal{T}\geqslant t+1\\ 0 & \text{when } \mathcal{T}\leqslant t\end{cases}$$

$$\mathcal{X}\to\hat{\mathcal{S}}^t\to\mathcal{A}^{t+1}\text{ conditional on }\mathcal{T}=t$$

1. By definition, when $\mathcal{T}\geqslant t+1$, $\hat{\mathcal{S}}^t=\mathcal{S}^t$. So conditional on $\mathcal{T}=t+1$, $\mathcal{X}\to\mathcal{S}^t\to\mathcal{A}^{t+1}$ implies $\mathcal{X}\to\hat{\mathcal{S}}^t\to\mathcal{A}^{t+1}$.

2. When $\hat{\mathcal{S}}^{t-1}=s_0$, $\mathbf{1}_{\mathcal{T}\leqslant t}=1$. When $\hat{\mathcal{S}}^{t-1}\neq s_0$:
$$\text{Prob}\left(\mathcal{T}>t|\mathcal{S}^{t-1}\right)=\text{Prob}\left(\mathcal{T}>t|\mathcal{S}^{t-1},\mathcal{X},\mathcal{T}\geqslant t\right)\text{Prob}\left(\mathcal{T}\geqslant t|\mathcal{S}^{t-1},\mathcal{X}\right)=\text{Prob}\left(\mathcal{T}>t|\mathcal{S}^{t-1},\mathcal{X},\mathcal{T}\geqslant t\right)$$
$$\Longrightarrow\text{Prob}\left(\mathcal{T}>t|\hat{\mathcal{S}}^{t-1},\mathcal{X}\right)=\text{Prob}\left(\mathcal{T}>t|\hat{\mathcal{S}}^{t-1}\right)$$
which is independent to realization of $\mathcal{X}$. So $\mathcal{X}\to\hat{\mathcal{S}}^{t-1}\to\mathbf{1}_{\mathcal{T}\leqslant t}$. The first equality is by the law of total probability (conditional on $\mathcal{T}\geqslant t$), $\mathcal{X}\to\mathcal{S}^{t-1}\to\mathbf{1}_{\mathcal{T}\leqslant t}$ and when $\hat{\mathcal{S}}^{t-1}\neq s_0$, $\text{Prob}(\mathcal{T}=t)=0$. The second equality is by when $\hat{\mathcal{S}}^{t-1}\neq s_0$, $\text{Prob}(\mathcal{T}\geqslant t)=1$.

3. Total information cost:
$$E\left[\sum_{t=0}^{\infty}e^{-\rho dt\cdot t}C_{dt}\left(I\left(\hat{\mathcal{S}}^t;\mathcal{X}|\hat{\mathcal{S}}^{\mathbf{t-1}},\mathbf{1}_{\mathcal{T}\leqslant t}\right)\right)\right]=E\left[\sum_{t=0}^{\mathcal{T}-1}e^{-\rho dt\cdot t}C_{dt}\left(I\left(\hat{\mathcal{S}}^t;\mathcal{X}|\hat{\mathcal{S}}^{\mathbf{t-1}},\mathbf{1}_{\mathcal{T}\leqslant t}\right)\right)\right]$$
$$=E\left[\sum_{t=0}^{\mathcal{T}-1}e^{-\rho dt\cdot t}C_{dt}(I(\mathcal{S}^t;\mathcal{X}|\mathcal{S}^{\mathbf{t-1}},\mathbf{1}_{\mathcal{T}\leqslant t}))\right]\leqslant E\left[\sum_{t=0}^{\infty}e^{-\rho dt\cdot t}C_{dt}(I(\mathcal{S}^t;\mathcal{X}|\mathcal{S}^{\mathbf{t-1}},\mathbf{1}_{\mathcal{T}\leqslant t}))\right]$$

The first equality is by $\hat{\mathcal{S}}^t$ being degenerate when $t\geqslant\mathcal{T}$. The second equality is from $\hat{\mathcal{S}}^t=\mathcal{S}^t$ when $\mathcal{T}>t$. Therefore, $\left(\hat{\mathcal{S}}^t,\mathcal{A}^t,\mathcal{T}\right)$ is a feasible strategy dominating $(\mathcal{S}^t,\mathcal{A}^t,\mathcal{T})$. Now we define $\tilde{\mathcal{S}}^t$:

$$\tilde{\mathcal{S}}^t=\begin{cases}s_0 & \text{when } \mathcal{T}<t+1\\ \mathcal{A}^{t+1} & \text{when } \mathcal{T}=t+1\\ \hat{\mathcal{S}}^t & \text{when } \mathcal{T}>t+1\end{cases}$$

Initial information $\tilde{\mathcal{S}}^{-1}$ is defined as a degenerate(uninformative) signal and induced belief is the prior. $\tilde{\mathcal{S}}^t$ replaces the signal defined in $\hat{\mathcal{S}}^t$ by a direct signal that suggests the corresponding action profile in next period when $\mathcal{T}=t+1$. Verify that the $\tilde{\mathcal{S}}^t$ satisfies the information processing constraints in Equation (5′) and improves utility:

1. When $\tilde{\mathcal{S}}^{t-1}\in\{s_0\}\bigcup A$, it's for sure that $\mathcal{T}\leqslant t$. Otherwise, $\mathcal{T}>t$. Therefore $\mathbf{1}_{\mathcal{T}\leqslant t}$ is a direct garbling of $\tilde{\mathcal{S}}^{t-1}$. So $\mathcal{X}\to\tilde{\mathcal{S}}^{t-1}\to\mathbf{1}_{\mathcal{T}\leqslant t}$.
2. When $\mathcal{T}=t$, $\mathcal{A}^t=\tilde{\mathcal{S}}^{t-1}$. Therefore $\mathcal{S}\to\hat{\mathcal{S}}^{t-1}\to\mathcal{A}^t$ implies $\mathcal{X}\to\tilde{\mathcal{S}}^{t-1}\to\mathcal{A}^t$ conditional on $\mathcal{T}=t$.

3. Information measure associated with $\left(\tilde{\mathcal{S}}^t, \mathcal{A}^t, \mathcal{T}\right)$ when $\mathcal{T} > t$:

$$\begin{aligned}
& I\left(\tilde{\mathcal{S}}^t; \mathcal{X} | \tilde{\mathcal{S}}^{\mathbf{t-1}}, \mathcal{T} > t\right) \\
=& \mathbf{1}_{\mathcal{T}=t+1} I\left(\mathcal{A}^{t+1}; \mathcal{X} | \tilde{\mathcal{S}}^{\mathbf{t-1}}, \mathcal{T} = t+1\right) + \mathbf{1}_{\mathcal{T}>t+1} I\left(\tilde{\mathcal{S}}^t; \mathcal{X} | \tilde{\mathcal{S}}^{\mathbf{t-1}}, \mathcal{T} > t+1\right) \\
=& \mathbf{1}_{\mathcal{T}=t+1} I\left(\mathcal{A}^{t+1}; \mathcal{X} | \hat{\mathcal{S}}^{\mathbf{t-1}}, \mathcal{T} = t+1\right) + \mathbf{1}_{\mathcal{T}>t+1} I\left(\hat{\mathcal{S}}^t; \mathcal{X} | \hat{\mathcal{S}}^{\mathbf{t-1}}, \mathcal{T} > t+1\right) \\
\leqslant& \mathbf{1}_{\mathcal{T}=t+1} I\left(\hat{\mathcal{S}}^t; \mathcal{X} | \hat{\mathcal{S}}^{\mathbf{t-1}}, \mathcal{T} = t+1\right) + \mathbf{1}_{\mathcal{T}>t+1} I\left(\hat{\mathcal{S}}^t; \mathcal{X} | \hat{\mathcal{S}}^{\mathbf{t-1}}, \mathcal{T} > t+1\right) \\
=& I\left(\hat{\mathcal{S}}^t; \mathcal{X} | \hat{\mathcal{S}}^{\mathbf{t-1}}, \mathcal{T} > 1\right)
\end{aligned}$$

First equality is simply rewriting two possible cases of $\mathcal{T}$. Second equality is from definition of $\tilde{\mathcal{S}}^t$ when $\mathcal{T} > t+1$. The inequality is from $\mathcal{X} \to \hat{\mathcal{S}}^t \to \mathcal{A}^{t+1}$ conditional on $\mathcal{T} = t+1$. Therefore, $\left(\tilde{\mathcal{S}}^t, \mathcal{T}, \mathcal{A}^t\right)$ dominates the original solution in Equation (5′) by achieving same action profile at lower costs. $\left(\tilde{\mathcal{S}}^t, \mathcal{T}, \mathcal{A}^t\right)$ is a feasible solution to Equation (S.8). Therefore solving Equation (S.8) yields a weakly higher utility than Equation (5′). What remains to be proved is that any $\left(\tilde{\mathcal{S}}^t, \mathcal{T}, \mathcal{A}^t\right)$ feasible in Equation (S.8) can be dominated by some strategy feasible in Equation (5′). It's not hard to see that the strategy is feasible in Equation (5′). Finally we show that the two formulation gives same utility:

$$\begin{aligned}
& E\left[e^{-\rho dt \cdot \mathcal{T}} E[u(\mathcal{A}^{\mathcal{T}}, \mathcal{X})] - \sum_{t=0}^{\infty} e^{-\rho dt \cdot t} C_{dt}\left(I\left(\tilde{\mathcal{S}}^t; \mathcal{X} | \tilde{\mathcal{S}}^{\mathbf{t-1}}, \mathbf{1}_{\mathcal{T} \leqslant t}\right)\right)\right] \\
=& \sum_{t=0}^{\infty}\left(e^{-\rho dt} \mathbf{P}[\mathcal{T} = t]\left(E\left[u(\mathcal{A}^t, \mathcal{X}) | \mathcal{T} = t\right]\right) - E\left[C_{dt}\left(I\left(\tilde{\mathcal{S}}^t; \mathcal{X} | \tilde{\mathcal{S}}^{\mathbf{t-1}}, \mathbf{1}_{\mathcal{T} \leqslant t}\right)\right)\right]\right) \\
=& \sum_{t=0}^{\infty}\left(e^{-\rho dt} \mathbf{P}[\mathcal{T} = t]\left(E\left[u(\mathcal{A}^t, \mathcal{X}) | \mathcal{T} = t\right]\right) - \mathbf{P}[\mathcal{T} > t] E\left[C_{dt}\left(I\left(\tilde{\mathcal{S}}^t; \mathcal{X} | \tilde{\mathcal{S}}^{\mathbf{t-1}}\right)\right) \middle| \mathcal{T} > t\right]\right)
\end{aligned}$$

First equality is from rewriting the utility part conditional on decision time $\mathcal{T} = t$. Second equality is from rewriting the information cost part conditional on decision time $\mathcal{T} \leqslant t$ and $\mathcal{T} > t$. Therefore, Equation (5′) is equivalent to Equation (S.8). *Q.E.D.*

**Lemma S.5** (Tranformation of space)**.** *With Assumption A satisfied,* $(\mathcal{S}^t, \mathcal{T}, \mathcal{A}^{\mathcal{T}})$ *solves Equation (5′) if and only if there exists* $p^t\left(\mu^{t+1} | \mu^t\right) : \Delta X \mapsto \Delta^2 X$ *and* $q_s^t(\mu^t) : \Delta X \mapsto [0,1]$ *solving:*

$$\begin{aligned}
\sup_{(p^t, q_s^t)} \sum_{t=0}^{\infty} e^{-\rho dt \cdot t} \int_{\Delta X} & \left[\left(\max_a \sum_j u(a, x_j) \cdot \mu_j\right) q_s^t(\mu^t) - C_{dt}\left(H(\mu^t) - \int_{\Delta X} H(\mu^{t+1}) p^t(\mu^{t+1} | \mu^t) \mathrm{d}\mu^{t+1}\right)(1 - q_s^t(\mu^t))\right] \\
& \left(\int_{\Delta X^{t-1}} \prod_{\tau=0}^{t-1} p^\tau(\mu^{\tau+1} | \mu^\tau)(1 - q_s^\tau(\mu^\tau)) \mathrm{d}\mu^1 \ldots \mu^{t-1}\right) \mathrm{d}\mu^t \qquad \text{(S.9)} \\
\text{s.t.} & \int_{\Delta X} \mu p^t(\mu | \mu^t) \mathrm{d}\mu = \mu^t
\end{aligned}$$

*What's more, the optimal utility level is same in Equation (5′) and Equation (S.9).*

**Proof.** Let $p^t(\cdot | \mu^t)$ be the distribution of posteriors generated by $\tilde{\mathcal{S}}^t|_{\mathcal{T} > t, \tilde{\mathcal{S}}^{t-1} = \tilde{s}^{t-1}}$, where $\mu^t$ is posterior belief associated with signal $\tilde{s}^{\mathbf{t-1}}$. Let $q_s^t(\mu^t) = \mathbf{P}\left[\mathcal{T} = t | \tilde{\mathcal{S}}^{\mathbf{t-1}} = \tilde{s}^{\mathbf{t-1}}, \mathcal{T} \geqslant t\right]$. Now we can explicitly represent the distribution of $\tilde{\mathcal{S}}, \mathcal{T}, \mathcal{A}$ with the conditional distributions. First, $\mathbf{P}[\mathcal{T} = t]$ and $\mathbf{P}[\mathcal{T} > t]$ can be calculated by integrating $q_s^t(\mu^t)$:

$$\begin{aligned}
\mathbf{P}[\mathcal{T} = t] =& E\left[\mathbf{P}[\mathcal{T} = t | \tilde{\mathcal{S}}^{-1}]\right] \\
=& E\left[\mathbf{P}\left[\mathcal{T} = t | \tilde{\mathcal{S}}^{-1}, \mathcal{T} > 0\right] \mathbf{P}\left[\mathcal{T} > 0 | \tilde{\mathcal{S}}^{-1}\right]\right] \\
=& (1 - q_s^0(\mu^0)) \mathbf{P}\left[\mathcal{T} = t | \mathcal{T} > 0\right] \\
=& (1 - q_s^0(\mu^0)) E\left[\mathbf{P}\left[\mathcal{T} = t | \mathcal{T} > 0, \tilde{\mathcal{S}}^0\right]\right]
\end{aligned}$$

$$
\begin{aligned}
&=(1-q_s^0(\mu^0))\int \mathbf{P}\Big[\mathcal{T}=t|\mathcal{T}\geqslant 1,\widetilde{\mathcal{S}}^0\Big]p^0(\mu^1|\mu^0)\mathrm{d}\mu^1\\
&=(1-q_s^0(\mu^0))\int \mathbf{P}\Big[\mathcal{T}=t|\mathcal{T}>1,\widetilde{\mathcal{S}}^0\Big]\mathbf{P}\Big[\mathcal{T}>1|\mathcal{T}\geqslant 1,\widetilde{\mathcal{S}}^0\Big]p^0(\mu^1|\mu^0)\mathrm{d}\mu^1\\
&=(1-q_s^0(\mu^0))\int E\Big[\mathbf{P}\Big[\mathcal{T}=t|\mathcal{T}>1,\widetilde{\mathcal{S}}^1\Big]|\mu^1\Big](1-q_s^1(\mu^1))p^0(\mu^1|\mu^0)\mathrm{d}\mu^1\\
&=\cdots\\
&=\int\prod_{\tau=0}^{t-1}p^\tau(\mu^{\tau+1}|\mu^\tau)(1-q_s^\tau(\mu^\tau))q_s^t(\mu^t)\mathrm{d}\mu^1\ldots\mu^t
\end{aligned}
$$

Similarly, we can get:

$$
\mathbf{P}[\mathcal{T}>t]=\int\prod_{\tau=0}^{t-1}p^\tau(\mu^{\tau+1}|\mu^\tau)(1-q_s^\tau(\mu^\tau))(1-q_s^t(\mu^t))\mathrm{d}\mu^1\ldots\mu^t
$$

Then we can calculate the joint distribution of $\mathcal{T}$ and $\mu^t$:

$$
\begin{cases}
\mathbf{P}[\mathcal{T}=t,\mu^t=\nu]=\displaystyle\int\prod_{\tau=0}^{t-1}p^\tau(\mu^{\tau+1}|\mu^\tau)(1-q_s^\tau(\mu^\tau))q_s^t(\mu^t)\mathrm{d}\mu^1\ldots\mu^{t-1}\\
\mathbf{P}[\mathcal{T}>t,\mu^t=\nu]=\displaystyle\int\prod_{\tau=0}^{t-1}p^\tau(\mu^{\tau+1}|\mu^\tau)(1-q_s^\tau(\mu^\tau))(1-q_s^t(\mu^t))\mathrm{d}\mu^1\ldots\mu^{t-1}
\end{cases}
$$

Therefore:

$$
\begin{cases}
\mathcal{A}^t|_{\mathcal{T}=t}\sim\dfrac{\int\prod_{\tau=0}^{t-1}p^\tau(\mu^{\tau+1}|\mu^\tau)(1-q_s^\tau(\mu^\tau))\mathrm{d}\mu^1\ldots\mu^{t-1}q_s^t(\mu^t)}{\int\prod_{\tau=0}^{t-1}p^\tau(\mu^{\tau+1}|\mu^\tau)(1-q_s^\tau(\mu^\tau))q_s^t(\mu^t)\mathrm{d}\mu^1\ldots\mu^t}\\
\widetilde{\mathcal{S}}^t|_{\mathcal{T}>t}\sim\dfrac{\int\prod_{\tau=0}^{t-1}p^\tau(\mu^{\tau+1}|\mu^\tau)(1-q_s^\tau(\mu^\tau))\mathrm{d}\mu^1\ldots\mu^{t-1}(1-q_s^t(\mu^t))}{\int\prod_{\tau=0}^{t-1}p^\tau(\mu^{\tau+1}|\mu^\tau)(1-q_s^\tau(\mu^\tau)(1-q_s^t(\mu^t))\mathrm{d}\mu^1\ldots\mu^t}
\end{cases}
$$

This implies:

$$
\begin{aligned}
&\mathbf{P}[\mathcal{T}=t]E\Big[u(\mathcal{A}^t,\mathcal{X})|\widetilde{\mathcal{S}}^{\mathbf{t-1}},\mathcal{T}=t\Big]\\
=&\int_{\Delta X}\max_a\sum_j u(a,x_j)\mu_j^t\int_{\Delta X^{t-1}}\prod_{\tau=0}^{t-1}p^\tau(\mu^{\tau+1}|\mu^\tau)(1-q_s^\tau(\mu^\tau))q_s^t(\mu^t)\mathrm{d}\mu^1\ldots\mu^{t-1}\mathrm{d}\mu^t\\
&\mathbf{P}[\mathcal{T}>t]E\Big[C_{dt}\Big(I\Big(\widetilde{\mathcal{S}}^t;\mathcal{X}|\widetilde{\mathcal{S}}^{\mathbf{t-1}}\Big)\Big)|\mathcal{T}>t\Big]\\
=&\int_{\Delta X}C_{dt}\left(H(\mu^t)-\int_{\Delta X}H(\mu^{t+1})p^t(\mu^{t+1}|\mu^t)\mathrm{d}\mu^{t+1}\right)\\
&\times\int\prod_{\tau=0}^{t-1}p^\tau(\mu^{\tau+1}|\mu^\tau)(1-q_s^\tau(\mu^\tau))(1-q_s^t(\mu^t))\mathrm{d}\mu^1\ldots\mu^{t-1}\mathrm{d}\mu^t
\end{aligned}
$$

To sum up, we showed that starting from $\widetilde{\mathcal{S}},\mathcal{T},\mathcal{A}$ solving Equation (S.8), we can construct $p^t,q_s^t$ such that the value of Equation (S.8) is achieved in Equation (S.9). Next, we start from $\left(p^t,q^t\right)$ solving Equation (S.9). We can easily define $\mathcal{T}$: $\mathcal{T}|_{\mathcal{T}\geqslant t,\mu^t}\sim B(q_s^t(\mu^t))$ conditionally independent across all $t,\mu^t$. $\widetilde{\mathcal{S}}^t|_{\mathcal{T}>t,\mu^t}\sim p^t(\cdot|\mu^t)$, $\mathcal{A}^t|_{\mathcal{T}=t,\mu^t}=\operatorname{argmax}\sum u(a,x_j)\mu_j^t$. Therefore, the previous calculation shows that the value of Equation (S.9) is also achieved in Equation (S.8). Combining with the previous result, we conclude that Equation (S.8) and Equation (S.9) are equivalent in the sense that $\left(\widetilde{\mathcal{S}},\mathcal{T},\mathcal{A}\right)$ solves Equation (S.8) if and only if the corresponding $(p^t,q_s^t)$ solves Equation (S.9). *Q.E.D.*

**Lemma S.6** (Recursive representation). *$V_{dt}(\mu)$ is the optimal utility level solving Equation (S.9) given initial belief $\mu$ if and only if $V_{dt}(\mu)$ satisfies the following functional equation:*

$$
\begin{aligned}
V_{dt}(\mu)=&\max\left\{\max_a E[u(a,x)|\mu],\sup_{p\in\Delta^2X}e^{-\rho dt}\int_{\Delta X}V_{dt}(\mu)p(\mu)\mathrm{d}\mu-C_{dt}\left(H(\mu)-\int_{\Delta X}H(\nu)p(\nu)\mathrm{d}\nu\right)\right\} &\text{(S.10)}\\
&s.t.\int_{\Delta X}\nu p(\nu)\mathrm{d}\nu=\mu
\end{aligned}
$$

**Proof.** We first derive the recursive representation of Equation (S.9). Consider the following functional equation:

$$V_{dt}(\mu)=\sup_{q_s(\mu),p(\cdot|\mu)} q_s(\mu)\left(\max_a \sum u(a,x_j)\mu_j\right)+(1-q_s(\mu))\left[\int_{\Delta X} V_{dt}(\nu)p(\nu|\mu)\mathrm{d}\nu - C_{dt}\left(H(\mu)-\int_{\Delta X} H(\nu)p(\nu|\mu)\mathrm{d}\nu\right)\right]$$

$$\text{s.t.} \int_{\Delta X} \nu p(\nu|\mu)\mathrm{d}\nu=\mu$$

Since RHS is linear in $q_s(\mu)$, it will be WLOG that we only consider boundary solution $q_s(\mu)\in\{0,1\}$. Therefore, it is be exactly the same as Equation (S.10).

Now consider the equivalence between the sequential problem and the recursive problem. By assumption $E[u(a,x)|\mu]$ is bounded above by $\max_{a,x} u(a,x)$. Therefore, $e^{-\rho dt\cdot t}E[u(a,x)|\mu]$ is uniformly (for all choice of $\mu,a$) converging to zero when $t\to\infty$. Then $V_{dt}(\mu)$ is the solution of Equation (S.9) by the standard theory of dynamic programming. *Q.E.D.*

### *S1.3 Convergence*

I first prove two useful lemmas. Lemma S.7 shows that optimal strategy has informativeness of signal in each period of same order of $dt$. Lemma S.8 shows that there exists a unique limit of $V_{dt}$ in $L_\infty$ norm.

#### *S1.3.1 Bounded flow cost*

**Lemma S.7** (Bounded flow cost)**.** *With Assumption 2 satisfied, there* $\exists\Delta\in\mathbb{R}^+$ *s.t.* $I^*_{dt}(\mu)\leqslant\Delta dt$. $\forall \mu, dt$. *Where* $I^*_{dt}(\mu)=\sum p_i(H(\mu)-H(\nu_i))$ *for optimal* $(p_i,\nu_i)$ *in Equation (6)*

**Proof.** $\forall\ (p_i,\nu_i)$ which solves Equation (6), assume the value is $V_{dt}(\mu)$ and $I^*_{dt}(\mu)=\sum p_i(H(\mu)-H(\nu_i))$. Now for $I<I^*_{dt}$, consider a different information acquisition strategy:

- At prior $\mu$, use the following information structure:

$$\begin{cases}\mu'_i=\nu_i & \text{with probability } \frac{I}{I^*_{dt}}p_i\\ \mu'_0=\mu & \text{with probability } 1-\frac{I}{I^*_{dt}}\end{cases}$$

  This information structure mixes uninformative signal into $(p_i,\nu_i)$ with probability $1-\frac{I}{I^*_{dt}}>0$. It is Bayes plausible by definition.
- At any posterior other than $\mu$, follow the original strategy.

Now let's calculate the expected utility of this strategy. The utility gain from experimentation is:

$$V'(\mu)=e^{-\rho dt}\frac{I}{I^*_{dt}}\sum p_i V(\nu_i)+e^{-\rho dt}\left(1-\frac{I}{I^*_{dt}}\right)V'(\mu)=\left(\sum p_i V(\nu_i)\right)\cdot\frac{e^{-\rho dt}\frac{I}{I^*_{dt}}}{1-e^{-\rho dt}\left(1-\frac{I}{I^*_{dt}}\right)}$$

$$\begin{aligned}\Longrightarrow \sum p_i V(\nu_i)-e^{\rho dt}V'(\mu)=&\left(\sum p_i V(\nu_i)\right)\cdot\frac{\left(1-e^{-\rho dt}\right)\left(1-\frac{I}{I^*_{dt}}\right)}{1-e^{-\rho dt}\left(1-\frac{I}{I^*_{dt}}\right)}\\ \leqslant&\max v\cdot\frac{\left(1-e^{-\rho dt}\right)\left(1-\frac{I}{I^*_{dt}}\right)}{1-e^{-\rho dt}\left(1-\frac{I}{I^*_{dt}}\right)}\\ \leqslant&\max v\cdot 2\rho dt\left(\frac{I^*_{dt}}{I}-1\right)\end{aligned}$$

The first inequality is from bounding $V$ with $\max v$. The second inequality is from $1-e^{-\rho dt}<2\rho dt$, $e^{-\rho dt}<1$ and $1-\frac{I}{I^*_{dt}}>0$. On the other hand, the discounted total cost of this strategy is:

$$Cost'(\mu)=C_{dt}(I)+e^{-\rho dt}\left(1-\frac{I}{I^*_{dt}}\right)Cost'(\mu)=\frac{C_{dt}(I)}{1-e^{-\rho dt}\left(1-\frac{I}{I^*_{dt}}\right)}$$

$$\begin{aligned}\Longrightarrow\ C_{dt}(I^*_{dt}(\mu))-Cost'(\mu)=&C_{dt}(I^*_{dt})-\frac{C_{dt}(I)}{1-e^{-\rho dt}\left(1-\frac{I}{I^*_{dt}}\right)}\\ \geqslant&C_{dt}(I^*_{dt})-\frac{C_{dt}(I)I^*_{dt}}{I}\end{aligned}$$

The inequality is from $e^{-\rho dt} < 1$ and $1-\frac{I}{I^*_{dt}} > 0$. Therefore, by optimality of $(p_i, \nu_i)$ at $\mu$, the new strategy I defined should not improve the expected utility:

$$\begin{aligned}
&e^{-\rho dt}\left(\sum p_i V(\nu_i)\right) - V'(\mu) - \left(C_{dt}(I^*_{dt}) - Cost'(\mu)\right) \geqslant 0\\
\Longrightarrow &\sum p_i V(\nu_i) - e^{\rho dt} V'(\mu) - (C_{dt}(I^*_{dt}) - Cost'(\mu)) \geqslant 0\\
\Longrightarrow &\max v \cdot 2\rho dt\left(\frac{I^*_{dt}}{I} - 1\right) \geqslant C_{dt}(I^*_{dt}) - \frac{C_{dt}(I) I^*_{dt}}{I}\\
\Longrightarrow &\max v \cdot 2\rho\left(1 - \frac{I}{I^*_{dt}}\right) \geqslant \frac{I}{I^*_{dt}} C\left(\frac{I^*_{dt}}{dt}\right) - C\left(\frac{I^*_{dt}}{dt} \cdot \frac{I}{I^*_{dt}}\right)
\end{aligned} \tag{S.11}$$

By Assumption 2, $\exists\, \Delta$ s.t. $\forall \frac{I^*_{dt}}{dt} \geqslant \Delta$, there exists $\alpha \in (0,1)$ s.t. $\alpha C\left(\frac{I^*_{dt}}{dt}\right) - C\left(\alpha \frac{I^*_{dt}}{dt}\right) > 2\rho \max v$. Let $I = \alpha I^*_{dt}$, then $I < I^*_{dt}$ and Equation (S.11) is violated. By contradiction, $I^*_{dt} \leqslant \Delta dt$. *Q.E.D.*

### S1.3.2 Convergence of $V_{dt}$

**Lemma S.8.** *With Assumption A and Assumption 2 satisfied. Let $\overline{V}(\mu) = \limsup_{dt\to 0} V_{dt}(\mu)$. Then $\lim_{dt\to 0} \left\| V_{dt}(\mu) - \overline{V}(\mu)\right\|_\infty = 0$.*

**Proof.** We break down the proof of Lemma Lemma S.8 into three steps:

- *Step 1*: Prove that if $V_{dt} = \limsup_{n\to\infty} V_{\frac{dt}{2^n}}$, then $\left\|V_{dt} - V_{\frac{dt}{2^n}}\right\| \to 0$. First $V_{\frac{dt}{2^n}}$ is an increasing sequence, because every experimentation strategy associated with $\frac{dt}{2^n}$ can be replicated in a problem with $\frac{dt}{2^{n+1}}$: the DM can always split the experiment into two stages with equal cost in two periods and get an identical distribution of posterior beliefs at the end of second period (Lemma S.3). Moreover, $V_{\frac{dt}{2^n}}$ is always bounded above by fully informed utility. Then existence of $V_{dt} = \lim V_{\frac{dt}{2^n}}$ is guaranteed by monotonic convergence theorem.

Now let's prove the convergence is uniform in sup norm, i.e. $V_{\frac{dt}{2^n}}$ is a Cauchy sequence under sup norm. $\forall m > n$, $\forall \mu_0$, consider the problem with $\frac{dt}{2^m}$, consider the optimal experimentation $(p_i(\mu), \nu_i(\mu))$ and associated action rule $A_T$, information measure $I_T$, the expected utility is:

$$\begin{aligned}
V_{\frac{dt}{2^m}}(\mu_0) &= \sum e^{-\rho T\frac{dt}{2^m}} E_{\mu_0}\left[u(A_T, X) - C_{\frac{dt}{2^m}}(I_T)\right].\\
&= \sum e^{-\rho T\frac{dt}{2^n}} \sum_{\tau=0}^{2^{m-n}-1} e^{-\rho\tau\frac{dt}{2^m}} E_{\mu_0}\left[u(A_{2^{m-n}T+\tau}, X) - C_{\frac{dt}{2^m}}(I_{2^{m-n}T+\tau})\right]
\end{aligned} \tag{S.12}$$

The second equality is get by rewriting $T = 2^{m-n}T' + \tau$. Then take summation first over $\tau$ then over $T'$ (and relabel $T'$ to be $T$).

Now we construct an experimentation strategy for problem with $\frac{dt}{2^n}$. We combine all experiments between $2^{m-n}T$ and $2^{m-n}(T+1)$, and get the joint distribution of posteriors. We use this as the signal structure in each period $T$. Given this construction, at the end of each $2^{m-n}T$, the posterior distribution will be exactly same as that using original experiment. Then we assign same action as before to each posterior. By construction this action profile satisfies Markov property of information (i.e. signal realization is a sufficient statistics for action). Therefore if we let $U(\mu_0)$ be the discounted expected utility associated with the aforementioned strategy at $\mu_0$:

$$\begin{aligned}
&V_{\frac{dt}{2^n}}(\mu_0)\\
\geqslant &U(\mu_0)\\
= &\sum e^{-\rho T\frac{dt}{2^n}}\left(\sum_{\tau=0}^{2^{m-n}-1} e^{-\rho\frac{dt}{2^n}} E_{\mu_0}[u(A_{2^{m-n}T+\tau}, X)] - E_{\mu_0}\left[C_{\frac{dt}{2^n}}\left(\sum_{\tau=0}^{2^{m-n}-1} E_{\mu_{2^{m-n}T}}[I_{2^{m-n}T+\tau}]\right)\right]\right)\\
\geqslant &\sum e^{-\rho T\frac{dt}{2^n}}\left(\sum_{\tau=0}^{2^{m-n}-1} e^{-\rho\frac{dt}{2^n}} E_{\mu_0}[u(A_{2^{m-n}T+\tau}, X)] - E_{\mu_0}\left[C_{\frac{dt}{2^n}}\left(\sum_{\tau=0}^{2^{m-n}-1} I_{2^{m-n}T+\tau}\right)\right]\right)\\
\geqslant &\sum e^{-\rho T\frac{dt}{2^n}}\left(\sum_{\tau=0}^{2^{m-n}-1} e^{-\rho\frac{dt}{2^n}} E_{\mu_0}[u(A_{2^{m-n}T+\tau}, X)] - E_{\mu_0}\left[\sum_{\tau=0}^{2^{m-n}-1} \frac{1}{2^{m-n}} C_{\frac{dt}{2^n}}\left(2^{m-n} \cdot I_{2^{m-n}T+\tau}\right)\right]\right)
\end{aligned}$$

$$=\sum e^{-\rho T\frac{dt}{2^n}}\left(\sum_{\tau=0}^{2^{m-n}-1} e^{-\rho\frac{dt}{2^n}}E_{\mu_0}[u(A_{2^{m-n}T+\tau},X)]-E_{\mu_0}\left[\sum_{\tau=0}^{2^{m-n}-1} C_{\frac{dt}{2^m}}(I_{2^{m-n}T+\tau})\right]\right) \tag{S.13}$$

$$=e^{-\rho\frac{dt}{2^n}}\sum e^{-\rho T\frac{dt}{2^n}}\sum_{\tau=0}^{2^{m-n}-1} E_{\mu_0}[u(A_{2^{m-n}T+\tau},X)]-\sum e^{-\rho T\frac{dt}{2^n}}\sum_{\tau=0}^{2^{m-n}-1} E_{\mu_0}C_{\frac{dt}{2^m}}(I_{2^{m-n}T+\tau})$$

$$>e^{-\rho\frac{dt}{2^n}}\sum e^{-\rho T\frac{dt}{2^n}}\sum_{\tau=0}^{2^{m-n}-1} e^{-\rho\tau\frac{dt}{2^m}}E_{\mu_0}[u(A_{2^{m-n}T+\tau},X)]-e^{\rho\frac{dt}{2^n}}\sum e^{-\rho T\frac{dt}{2^n}}\sum_{\tau=0}^{2^{m-n}-1} e^{-\rho\tau\frac{dt}{2^m}}E_{\mu_0}\left[C_{\frac{dt}{2^m}}(I_{2^{m-n}T+\tau})\right]$$

$$\geqslant V_{\frac{dt}{2^m}}(\mu_0)-\left(1-e^{-\rho\frac{dt}{2^n}}\right)\max v-\left(e^{\rho\frac{dt}{2^n}}-1\right)\max v$$

$$=V_{\frac{dt}{2^m}}(\mu_0)-\left(e^{\rho\frac{dt}{2^n}}-e^{-\rho\frac{dt}{2^n}}\right)\max v \tag{S.14}$$

Where $\max v$ is an upper bounded of total utility from action. The second and third inequalities are from concavity of $f$. Equation (S.13) is obtained by definition of $C_{dt}(\cdot)=dt\cdot C(\frac{\cdot}{dt})$. Noticing that Equation (S.13) is different from Equation (S.12) by only one term: the discounting term in inner summation ($e^{-\rho\frac{dt}{2^m}}$ instead of $e^{-\rho\frac{dt}{2^n}}$). This characterizes the experiment design in problem $\frac{dt}{2^n}$. In each period $T$, actions are all postponed to the end of period. Therefore they are discounted by at most $\frac{dt}{2^n}$, which is period length and costs are shifted to the beginning of each period. The next inequality is from $e^{-\rho\frac{dt}{2^m}}<1$ and $m>n$. By Lemma 2, both utility gain and information cost are uniformly bounded by $\max v$, then $\left\|V_{\frac{dt}{2^n}}-V_{\frac{dt}{2^m}}\right\|\leqslant\max v\left(e^{\rho\frac{dt}{2^n}}-e^{-\rho\frac{dt}{2^n}}\right)\to 0$ when $n\to 0$.

- *Step 2*: Prove that $\forall dt>0$, $V_{dt}$ are identical, WLOG we can call it $\overline{V}(\mu)$. $\forall dt,dt'>0,\forall n$, consider $V_{\frac{dt}{2^n}}$. Pick $m$ large enough that there exists $N$ s.t. $\frac{dt}{2^{n+1}}\leqslant N\frac{dt'}{2^m}\leqslant\frac{dt}{2^n}\leqslant(N+1)\frac{dt'}{2^m}$. Consider optimal experimentation and action associated with $\frac{dt}{2^n}$, we construct experimentation strategy for problem with $\frac{dt'}{2^m}$. For each time period $T$ in the original problem, split the experiment in period $T$ into $N+1$ periods and take any action at the end of $N+1$th period (apply Lemma S.3 recursively). In the new experiment strategy, the effective period length will increase from $\frac{dt}{2^n}$ to $(N+1)\frac{dt'}{2^m}$. First, suppose the information measure incurred in any period is $I$ in problem with $\frac{dt}{2^n}$. Then per-period information measure from the aforementioned strategy is $\frac{I}{N+1}\leqslant 2^{n-m}\frac{dt'}{dt}I$. This leads to per-period cost $\frac{dt'}{2^m}\cdot C\left(\frac{I\cdot 2^m}{(N+1)dt'}\right)\leqslant\frac{1}{N}\frac{dt}{2^n}\cdot C\left(\frac{2^n\cdot I}{dt}\right)$. Therefore, the total cost from experimentation will increase by no more than $\frac{N+1}{N}$ times and that will be bounded by $\frac{1}{N}\max v$. Second, since induced posterior distribution and action distribution are still the same, Markov property still holds. Finally:

$$\begin{aligned}
V_{\frac{dt'}{2^m}}(\mu_0)-V_{\frac{dt}{2^n}}(\mu_0)&\geqslant\sum e^{-\rho T(N+1)\frac{dt'}{2^m}}E_{\mu_0}[u(A_T,X)]-\sum e^{\rho T\frac{dt}{2^n}}E_{\mu_0}[u(A_T,X)]-\frac{1}{N}\max v\\
&=-\sum\left(e^{-\rho T\frac{dt}{2^n}}-e^{-\rho T(N+1)\frac{dt'}{2^m}}\right)E_{\mu_0}[u(A_T,X)]-\frac{1}{N}\max v\\
&\geqslant-\max v\left|\sum e^{-\rho T\frac{dt}{2^n}}-\sum e^{-\rho T(N+1)\frac{dt'}{2^m}}\right|-\frac{1}{N}\max v\\
&=-\max v\frac{e^{-\rho\frac{dt}{2^n}}-e^{-\rho(N+1)\frac{dt'}{2^m}}}{\left(1-e^{-\rho\frac{dt}{2^n}}\right)\left(1-e^{-\rho(N+1)\frac{dt'}{2^m}}\right)}-\frac{1}{N}\max v\\
&\geqslant-\max v\frac{e^{-\rho N\frac{dt'}{2^m}}-e^{-\rho(N+1)\frac{dt'}{2^m}}}{(1-e^{-\rho\frac{dt}{2^n}})^2}-\frac{1}{N}\max v\\
&=-\max v\frac{e^{-\rho N\frac{dt'}{2^m}}}{(1-e^{-\rho\frac{dt}{2^n}})^2}(e^{\rho\frac{dt'}{2^m}}-1)-\frac{1}{N}\max v\\
&\geqslant-\max v\left(\frac{e^{-\rho\frac{dt}{2^{n+1}}}}{(1-e^{-\rho\frac{dt}{2^n}})^2}(e^{\rho\frac{dt'}{2^m}}-1)+\frac{dt'}{dt\cdot 2^{m-n-1}}\right)
\end{aligned}$$

First inequality is from suboptimality of the constructed experiment and bound of cost difference. Second inequality is from $e^{-\rho T\frac{dt}{2^n}}\geqslant e^{-\rho T(N+1)\frac{dt'}{2^m}}$. Third inequality is from $\frac{dt}{2^n}\geqslant N\frac{dt'}{2^m}$. Last inequality is from $N\frac{dt'}{2^m}\geqslant\frac{dt}{2^{n+1}}$. Take $m\to\infty$ on both side, we have $V_{dt'}(\mu_0)\geqslant V_{\frac{dt}{2^n}}(\mu_0)$. Then take $n\to 0$ on both side $V_{dt'}(\mu_0)\geqslant V_{dt}(\mu_0)$. Since this holds for arbitrary

$dt$,$dt'$ and $\mu_0$, we conclude that $V_{dt} = V_{dt'}$.

- *Step 3*: $\left\|V_{dt} - \overline{V}\right\| \to 0$ when $dt \to 0$. Fix any $dt > 0$, then $\forall \varepsilon > 0$, there exists $N$ s.t. $\forall n \geqslant N$, $\left\|V_{\frac{dt}{2^n}} - \overline{V}\right\| < \frac{\varepsilon}{2}$. Then given the proof in last part, for any $dt' < \frac{dt}{2^n}$, suppose there exists $N$ s.t. $\frac{dt}{2^{n-1}} \leqslant Ndt' \leqslant \frac{dt}{2^n} \leqslant (N+1)dt'$, then the difference between $V_{\frac{dt}{2^n}}$ and $V_{dt'}$ will be bounded by:
$$\max v \left( \frac{e^{-\rho \frac{dt}{2^{n+1}}}}{(1-e^{-\rho \frac{dt}{2^n}})} (e^{\rho dt'} - 1) + \frac{2^{n+1}}{dt} dt' \right)$$
Actually such $N = \left[\frac{dt}{2^n dt'}\right]$ exists for any $dt' \leqslant \frac{dt}{2^n}$. Thus there exists $\delta$ s.t. $\forall dt' < \delta$, $\left\|V_{dt'} - V_{\frac{dt}{2^n}}\right\| < \frac{\varepsilon}{2}$, then $\left\|V_{dt'} - \overline{V}\right\| < \varepsilon$.

*Q.E.D.*

### S1.3.3 *Lemmas for Lemma 3*

**Lemma S.9.** *With Assumption A and Assumption 2 satisfied. Let $\overline{V}(\mu) = \lim_{dt \to 0} V_{dt}(\mu)$. Then $\overline{V} \in \mathcal{L}$ (pointwise Lipschitz function).*

**Proof.** We prove by induction on dimensionality of $\mu$. When $\mu = \delta_x$, supp$(\mu)$ is singleton. So Lemma S.9 is trivially satisifed. Now it is sufficient to prove that $\overline{V}$ is pointwise Lipschitz at any interior $\mu$.

First, since $\overline{V}$ is the uniform limit of continuous $V_{dt}$, $\overline{V}$ is continuous. $\forall \mu \in \Delta X^o$, suppose by contradiction $\overline{V}$ is not pointwise Lipschitz. Then $\exists\ \mu_n \to \mu$, $\frac{|\overline{V}(\mu_n) - \overline{V}(\mu)|}{\|\mu_n - \mu\|} \geqslant n$. There are two possibilities:

- $\frac{\overline{V}(\mu_n) - \overline{V}(\mu)}{\|\mu_n - \mu\|} \geqslant n$. Now let $\nu_n$ be a point in $\partial \Delta X$ s.t. $\mu_n$,$\mu$,$\nu_n$ are three ordered points on a straight line. Let $p_n$,$q_n$ be such that $p_n + q_n = 1$, $p_n \mu_n + q_n \nu_n = \mu$. Pick any $I$ s.t. $C(I) < \infty$ We have:
$$I \frac{\overline{V}(\nu_n) - \overline{V}(\mu) + \frac{\overline{V}(\mu_n) - \overline{V}(\mu)}{\|\mu_n - \mu\|} \|\nu_n - \mu\|}{H(\mu) - H(\nu_n) - \frac{H(\mu_n) - H(\mu)}{\|\mu_n - \mu\|} \|\nu_n - \mu\|} \geqslant I \frac{\overline{V}(\nu_n) - \overline{V}(\mu) + n \|\nu_n - \mu\|}{H(\mu) - H(\nu_n) - \frac{H(\mu_n) - H(\mu)}{\|\mu_n - \mu\|} \|\nu_n - \mu\|}$$
Noticing that the only difference between LHS and RHS is that $\frac{\overline{V}(\mu_n) - \overline{V}(\mu)}{\|\mu_n - \mu\|}$ is replaced with $n$ on RHS. Since the nominator is bounded, $\mu$ being interior suggesting $\|\nu_n - \mu\|$ is strictly positive in the limit. Take $n \to \infty$ on RHS, we observe that RHS goes to infinity. Therefore, there exists $N$ s.t. $\forall n \geqslant N$, RHS is larger than $3\rho \sup F + 2C(I)$.
$$I \frac{\overline{V}(\nu_n) - \overline{V}(\mu) + \frac{\overline{V}(\mu_n) - \overline{V}(\mu)}{\|\mu_n - \mu\|} \|\nu_n - \mu\|}{H(\mu) - H(\nu_n) - \frac{H(\mu_n) - H(\mu)}{\|\mu_n - \mu\|} \|\nu_n - \mu\|} \geqslant 3\rho \sup F + 2C(I)$$
$$\implies \frac{(\|\mu_n - \mu\|)\overline{V}(\nu_n) + \|\nu_n - \mu\|\overline{V}(\mu_n) - (\|\mu_n - \mu\| + \|\nu_n - \mu\|)\overline{V}(\mu)}{-\|\mu_n - \mu\| H(\mu_n) - \|\nu_n - \mu\| H(\mu_n) + (\|\mu_n - \mu\| + \|\nu_n - \mu\|) H(\mu)} \geqslant \frac{3\rho}{I} \sup F + \frac{2C(I)}{I}$$
$$\implies \frac{p_n \overline{V}(\mu_n) + q_n \overline{V}(\nu_n) - \overline{V}(\mu)}{-p_n H(\mu_n) - q_n H(\nu_n) + H(\mu)} \geqslant \frac{3\rho}{I} \sup F + \frac{2C(I)}{I}$$
$$\implies \frac{p_n \overline{V}(\mu_n) + q_n \overline{V}(\nu_n) - \overline{V}(\mu)}{I(\mu_n, \nu_n | \mu)} \geqslant \frac{3\rho}{I} \sup F + \frac{2C(I)}{I}$$
$$\implies p_n \overline{V}(\mu_n) + q_n \overline{V}(\nu_n) - V(\mu) \geqslant \frac{3\rho}{I} \sup F I(\mu_n, \nu_n | \mu) + 2C(I) \frac{I(\mu_n, \nu_n | \mu)}{I}$$
$$\implies p_n \overline{V}(\mu_n) + q_n \overline{V}(\nu_n) - 2C(I) \frac{I(\mu_n, \nu_n | \mu)}{I} \geqslant \overline{V}(\mu) \left(1 + 2\frac{\rho}{I} I(\mu_n, \nu_n | \mu)\right) + \sup F \frac{\rho}{I} I(\mu_n, \nu_n | \mu)$$
$$\implies p_n \overline{V}(\mu_n) + q_n \overline{V}(\nu_n) - 2C(I) \frac{I(\mu_n, \nu_n | \mu)}{I} \geqslant \overline{V}(\mu) e^{\frac{\rho}{I} I(\mu_n, \nu_n | \mu)} + \sup F \frac{\rho}{I} I(\mu_n, \nu_n | \mu)$$
Last inequality comes from $\forall x > 0, 1 + 2x > e^x$. Now we have:
$$e^{-\rho \frac{I(\mu_n, \nu_n | \mu)}{I}} (p_n \overline{V}(\mu_n) + q_n \overline{V}(\nu_n)) - 2e^{-\rho \frac{I(\mu_n, \nu_n | \mu)}{I}} C(I) \frac{I(\mu_n, \nu_n | \mu)}{I}$$
$$\geqslant \overline{V}(\mu) + e^{-\rho \frac{I(\mu_n, \nu_n | \mu)}{I}} \sup F \frac{\rho}{I} I(\mu_n, \nu_n | \mu)$$

Since $\mu_n$ are converging to $\mu$, $\lim_{n\to\infty} I(\mu_n,\nu_n|\mu)=0$. Then we can pick $N$ sufficiently large that $\forall n\geqslant N$:
$$e^{-\rho\frac{I(\mu_n,\nu_n|\mu)}{I}}(p_n\overline{V}\mu_n+q_n\overline{V}(\nu_n))-\frac{I(\mu_n,\nu_n|\mu)}{I}C(I)\geqslant \overline{V}(\mu)+\frac{\rho I(\mu_n,\nu_n|\mu)}{2I}\sup F$$
From now on, we keep $n$ fixed. Then we pick $dt=\frac{I(\mu_n,\nu_n|\mu)}{I}$ and $dt_m=\frac{dt}{2^m}$. $m$ is chosen sufficiently large that $\left|\overline{V}-V_{dt_m}\right|e^{\frac{\rho}{I}I(\mu_n,\nu_n|\mu)}<\frac{\rho I(\mu_n,\nu_n|\mu)}{8c}\sup F$, then:
$$e^{-\rho\frac{I(\mu_n,\nu_n|\mu)}{I}}(p_nV_{dt_m}(\mu_n)+q_nV_{dt_m}(\nu_n))-dtC\left(\frac{I(\mu_n,\nu_n|\mu)}{dt}\right)\geqslant V_{dt_m}(\mu)+\frac{\rho dt}{4}\sup F$$
We consider an experimentation strategy that divides information measure $I(\mu_n,\nu_n|\mu)$ into $2^m$ periods uniformly, and wait until the end of the $2^m$ periods to take action:
$$\begin{aligned}
&e^{-\rho dt}(p_nV_{dt_m}(\mu_n)+q_nV_{dt_m}(\nu_n))-\sum_{\tau=0}^{2^m-1}e^{-\rho\tau dt_m}dt_m\cdot C\left(\frac{I(\mu_n,\nu_n|\mu)/2^m}{dt_m}\right)\\
>&e^{-\rho dt}(p_nV_{dt_m}(\mu_n)+q_nV_{dt_m}(\nu_n))-\sum_{\tau=0}^{2^m-1}e^{-\rho dt}dt_m\cdot C\left(\frac{I(\mu_n,\nu_n|\mu)/2^m}{dt_m}\right)\\
=&e^{-\rho dt}\left(p_nV_{dt_m}(\mu_n)+q_nV_{dt_m}(\nu_n)-dt\cdot C\left(\frac{I(\mu_n,\nu_n|\mu)}{dt}\right)\right)\\
\geqslant&V_{dt_m}(\mu)+\frac{\rho dt}{4}\sup F
\end{aligned}$$
First line is expected utility from taking the aforementioned experiment at $\mu$. Second line is replacing all discounting in cost with a term larger than 1. Taking $m$ sufficiently large, last line will be strictly larger than $V_{dt_m}(\mu)$. Thus this experiment dominates optimal value of $dt_m$ problem at $\mu$. Contradiction.

- $\frac{\overline{V}(\mu_n)-\overline{V}(\mu)}{\|\mu_n-\mu\|}\leqslant -n$. Then pick $\nu_n\in\partial\Delta X$ s.t. $\mu,\mu_n,\nu_n$ are three ordered points on a straight line. Let $p_n,q_n$ be such that $p_n+q_n=1$, $p_n\mu+q_n\nu_n=\mu_n$. Pick any $I$ s.t. $C(I)<\infty$. We have:
$$I\frac{\overline{V}(\nu_n)-\overline{V}(\mu_n)+\frac{\overline{V}(\mu)-\overline{V}(\mu_n)}{\|\mu_n-\mu\|}\|\nu_n-\mu_n\|}{H(\mu_n)-H(\nu_n)-\frac{H(\mu)-H(\mu_n)}{\|\mu_n-\mu\|}\|\nu_n-\mu_n\|}\geqslant I\frac{\overline{V}(\nu_n)-\overline{V}(\mu_n)+n\|\nu_n-\mu_n\|}{H(\mu_n)-H(\nu_n)-\frac{H(\mu)-H(\mu_n)}{\|\mu_n-\mu\|}\|\nu_n-\mu_n\|}$$
Take $n\to\infty$ on RHS, we observe that RHS goes to infinity. Therefore, there exists $N$ s.t. $\forall n\geqslant N$, RHS is larger than $3\rho\sup F+2C(I)$.
$$\begin{aligned}
\implies p_n\overline{V}(\mu)+q_n\overline{V}(\nu_n)-2C(I)\frac{I(\mu,\nu_n|\mu_n)}{I}\geqslant&\overline{V}(\mu_n)+3\frac{\rho I(\mu,\nu_n|\mu_n)}{I}\sup F\\
\geqslant&e^{\rho\frac{I(\mu,\nu_n|\mu_n)}{I}}\overline{V}(\mu_n)+\frac{\rho I(\mu,\nu_n|\mu_n)}{I}\sup F
\end{aligned}$$
Similar to last part, $N$ can be chosen sufficiently large that:
$$e^{-\rho\frac{I(\mu,\nu_n|\mu_n)}{I}}(p_n\overline{V}(\mu)+q_n\overline{V}(\nu_n))-\frac{I(\mu,\nu_n|\mu_n)}{I}C(I)\geqslant\overline{V}(\mu_n)+\frac{\rho I(\mu,\nu_n|\mu_n)}{I}\sup F$$
Then pick $dt=\frac{I(\mu,\nu_n|\mu_n)}{I}$ and $dt_m=\frac{dt}{2^m}$. $m$ can be chosen sufficiently large that:
$$e^{-\rho dt}(p_nV_{dt_m}(\mu)+q_nV_{dt_m}(\nu_n))-dtC(I)\geqslant V_{dt_m}(\mu_n)+\frac{\rho dt}{2}\sup F$$
We consider a similar experimentation strategy as before that divides experiment uniformly:
$$e^{-\rho dt}(p_nV_{dt_m}(\mu)+q_nV_{dt_m}(\nu_n))-\sum_{\tau=0}^{2^m-1}e^{-\rho\tau dt_m}dt_m\cdot C\left(\frac{I(\mu,\nu_n|\mu_n)/2^m}{dt_m}\right)\geqslant V_{dt_m}(\mu_n)+\frac{\rho dt}{4}\sup F$$
RHS is strictly larger than $V_{dt_m}(\mu_n)$. This experiment dominates optimal experiment of $dt_m$ problem at $\mu_n$. Contradiction.

*Q.E.D.*

**Lemma S.10.** *$\forall f(x)$ differentiable on $(a,b)$. $\forall x,y\in(a,b)$,*
$$\frac{1}{2}\inf_{z\in(x,y)}D^2f(z,y)|y-x|^2\leqslant f(y)-f(x)-f'(x)(y-x)\leqslant\frac{1}{2}\sup_{z\in(x,y)}D^2f(z,y)|y-x|^2$$

**Proof.**

- First inequality: let $\underline{D}=\inf_{z\in(x,y)}D^2f(z,y)$. Suppose by contradiction the statement is not true, then there exists $\varepsilon$ s.t. $\frac{\underline{D}-\varepsilon}{2}|y-x|^2>f(y)-f(x)-f'(x)(y-x)$. Let $h(w)=f(w)-f(x)-f'(x)(w-x)-\frac{\underline{D}-\varepsilon}{2}(w-x)^2$. Then $h(x)=0$, $h'(x)=0$ and $h(y)<0$. Now consider $\max_z h(z)-\frac{h(y)}{y-x}(z-x)$. By continuity of $h$, maximizer $z^*$ exists in $[x,y]$. FOC implies $h'(z^*)=\frac{h(y)}{y-x}$ so $z^*\neq x$. The objective function is 0 at both $x,y$ so $z^*\neq y$. Then optimality of $z^*$ implies $\forall dz$ sufficiently small:
$$h(z^*+dz)-\frac{h(y)}{y-x}(z^*+dz-x)\leqslant h(z^*)-\frac{h(y)}{y-x}(z^*-x)$$
$$\Longrightarrow f(z^*+dz)-f(z^*)-f'(x)dz-\frac{\underline{D}-\varepsilon}{2}(2z^*-2x+dz)dz\leqslant dz(f'(z^*)-f'(x)-(\underline{D}-\varepsilon)(z^*-x))$$
$$\Longrightarrow\frac{f(z^*+dz)-f(z^*)-f'(z^*)dz}{dz^2}\leqslant\frac{\underline{D}-\varepsilon}{2}$$
$$\Longrightarrow D^2f(z^*,y)<\underline{D}$$
Contradiction.
- Second inequality: let $\overline{D}=\sup_{z\in(x,y)}D^2(z,y)$. Suppose by contradiction the statement is not true, then there exists $\varepsilon$ s.t. $\frac{\overline{D}+\varepsilon}{2}|y-x|^2<f(y)-f(x)-f'(x)(y-z)$. Let $h(w)=f(w)-f(x)-f'(x)(w-x)-\frac{\overline{D}+\varepsilon}{2}(w-x)^2$. Then $h(x)=0,h'(x)=0$ and $h(y)>0$. Now consider $\min_z h(z)-\frac{h(y)}{y-x}(z-x)$. By continuity of $h$, minimizer $z^*$ exists in $[x,y]$. FOC implies $h'(z^*)=\frac{h(y)}{y-z}$ so $z^*\neq x$. Then optimality of $z^*$ implies $\forall dz$ sufficiently small:
$$h(z^*+dz)-\frac{h(y)}{y-x}(z^*+dz-x)\geqslant h(z^*)-\frac{h(y)}{y-x}(z^*-x)$$
$$\Longrightarrow f(z^*+dz)-f(z^*)-f'(x)dz-\frac{\overline{D}+\varepsilon}{2}(2z^*-2x+dz)dz\geqslant dz(f'(z^*)-f'(x)-(\overline{D}+\varepsilon)(z^*-x))$$
$$\Longrightarrow\frac{f(z^*+dz)-f(z^*)-f'(z^*)dz}{dz^2}\geqslant\frac{\overline{D}+\varepsilon}{2}$$
$$\Longrightarrow D^2f(z^*,y)>\overline{D}$$
Contradiction.

*Q.E.D.*

## S2 Proofs in Section 6

### *S2.1 Proof and lemmas for Theorem 2*

*Proof of smoothness in Theorem 2*

I first show that there exists a set of $\mu_0$ such that on each interval when $V(\mu)>F(\mu)$, $V(\mu)$ is defined a $V_{\mu_0}$. Then I utilize this result to show that $V$ is $C^{(1)}$ smooth on $[0,1]$.

**Proof.** This is true when $\mu\leqslant\mu^{**}$ by definition of $V_{\mu^*}$. So I prove this for $\mu>\mu^{**}$. First prove some useful lemmas:

**Lemma S.11.** *$\forall k$, there exists $\mu_0\in\Omega$ s.t. $V_{\mu_0}(\underline{\mu}_k)>F(\underline{\mu}_k)$.*

**Proof.** Suppose $F=F_{k-1}$ at $\mu^{**}$. Equation (32) implies $\underline{\mu}_k>\mu^{**}>\underline{\mu}_{k-1}$. Consider:
$$U_k(\mu)=\max_{\nu\geqslant\mu}\frac{c}{\rho}\frac{F(\nu)-F(\mu)-F'^{-}(\mu)(\nu-\mu)}{J(\mu,\nu)}$$
$U_k$ is continuous by maximum theorem on $[\mu^{**},\underline{\mu}_k)$. Since $U_k(\mu^{**})=F(\mu^{**})$, $\lim_{\mu\to\underline{\mu}_k}U_k(\mu)=+\infty$, there exists $\mu_0$ s.t. $U_k(\mu_0)=F(\mu_0)$ and $U_k(\mu)>F(\mu)\ \forall\mu\in(\mu_0,\underline{\mu}_k)$. Now consider $V_{\mu_0}(\mu)$. I claim that $V_{\mu_0}(\mu)>F(\mu)\ \forall\mu\in(\mu_0,\underline{\mu}_k)$. Suppose not, then by intermediate value theorem, there exists $\mu'$ s.t. $V_{\mu_0}(\mu')\leqslant F(\mu)$ and $V'_{\mu_0}(\mu')\leqslant F(\mu)$. However, this implies
$$V_{\mu_0}(\mu')=\max_{\nu\geqslant\mu}\frac{c}{\rho}\frac{F(\nu)-V_{\mu_0}(\mu')-V'_{\mu_0}(\nu-\mu')}{J(\mu',\nu)}\geqslant U_k(\mu')>F(\mu')$$
Contradiction. Now assume $V_{\mu_0}$ hits $F$ at $\mu'_0$. Then $U_{k+1}(\mu'_0)\leqslant 0$ and $lim_{\mu\to\underline{\mu}_{k+1}}(\mu)=+\infty$, so we can find $V_{\mu_1}(\underline{\mu}_{k+1})>F(\underline{\mu}_{k+1})$. By induction on $k$, Lemma S.11 is true. *Q.E.D.*

**Lemma S.12.** *$\forall\mu_0\leqslant\mu_1\in\Omega$, let $I_i=\{\mu|V_{\nu_i}(\mu)>F(\mu)\}$. Then either $I_0\bigcap I_1=\varnothing$, or $I_1\subset I_0$ and $V_{\mu_0}\geqslant V_{\mu_1}$.*

**Proof.** The only possible contradiction of Lemma S.12 is that $\exists \mu' \in I_0 \bigcap I_1$ s.t. $V_{\mu_1}(\mu') > V_{\mu_0}(\mu')$. Since at $\mu_1$, $V_{\mu_0}(\mu_1) > V_{\mu_1}(\mu_1) = F(\mu_1)$, by intermediate value theorem, there exists $\xi \in (\mu_1, \mu')$ s.t. $V_{\mu_1}(\xi) > V_{\mu_0}(\xi)$ and $V'(\mu_1)(\xi) > V'(\mu_0)(\xi)$. Since $\xi \in I_1$, there exists $\nu, m$ solving Equation (32) for $V_{\mu_1}(\xi)$:

$$V_{\mu_0}(\xi) \geqslant \frac{c}{\rho}\frac{F_m(\nu) - V_{\mu_0}(\xi) - V'_{\mu_0}(\xi)(\nu - \xi)}{J(\xi, \nu)} > \frac{c}{\rho}\frac{F_m(\nu) - V_{\mu_1}(\xi) - V'_{\mu_1}(\xi)(\nu - \xi)}{J(\xi, \nu)} = V_{\mu_1}(\xi) > V_{\mu_0}(\xi)$$

Contradiction. So Lemma S.12 is true. *Q.E.D.*

**Lemma S.13.** $\mathbb{V} = \left\{\max_{i=1}^{n}\{V_{\nu_i}\}\right\}_{\nu_i \in \Omega, n \in \mathbb{N}}$ *is totally bounded and equi-continuous on* $[\mu^{**}, 1]$.

**Proof.** $V_{\mu^\diamond}$ is bounded above by $\sup F(\mu)$ and below by $\inf F(\mu)$. Consider $V'_{\mu^\diamond}$. When $V_{\mu^\diamond}(\mu) = F(\mu)$, obviously derivative is bounded by $\max|F'|$. When $V_{\mu^\diamond}(\mu) > F(\mu)$. Suppose $V'_{\mu^\diamond}(\mu) > \max|F'|$, then $F(\nu) - V_{\mu^\diamond}(\mu) - V'_{\mu^\diamond}(\mu)(\nu - \mu) < F(\nu) - F(\mu) - F'(\nu)(\nu - \mu) \leqslant 0$, contradiction. By Lemma B.2, $V'_{\mu^\diamond} \geqslant 0$. So $V'_{\mu^\diamond}$ are uniformly bounded in $[0, \max|F'|]$.

Now consider $\forall n$, $\forall \nu_i \in \Omega$, $V_{\nu_i} \in [\inf F, \sup F] \implies \max_i\{V_{\nu_i}\} \in [\inf F, \sup F]$. By Lemma S.12, $\max\{V_{\nu_i}\}$ is piecewisely defined as $V_{\nu_i}$ on finite disjoint intervals. So its derivative is piecewisely defined as $V'_{\nu_i}$, therefore bounded in $[0, \max|F'|]$. Therefore $\mathbb{V}$ is totally bounded and equi-continuous on $[\mu^{**}, 1]$. *Q.E.D.*

**Lemma S.14.** *There exists* $\Delta$ *s.t.* $\forall \nu_i \in \Omega$, *on* $\{\mu | V_{\nu_i}(\mu) > F(\mu)\}$, $V'(\mu)$ *has Lipschitz parameter* $\Delta$.

**Proof.** $\forall \mu \in (\hat{\mu}_{k+1}, \hat{\mu}_k)$, $\nu$ is smooth in $\mu$ and $V'_{\nu_i} > 0$, by envelope theorem:

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

$$\implies V''_{\nu_i}(\mu) + \frac{\rho}{c}V_{\nu_i}(\mu)H''(\mu) < 0$$

$V_{\nu_i}(\mu)$ is bounded by $\sup F$. It is easy to see that $\sup\Omega < \underline{\mu}_n$ (where $n$ is the largest index). By Lemma S.11, there is $\mu_0 \in \Omega$ s.t. $V_{\mu_0}(\underline{\mu}_n) > F(\underline{\mu}_n)$. By Lemma S.12, $\sup\Omega = \sup\{\mu | V_{\mu_0}(\mu) > F(\mu)\} < \nu(\mu_0) < 1$. Therefore, $\mu$ is bounded away from 1. Then by Assumption 3, $-H''(\mu)$ is bounded above. Therefore, $\Delta$ exists for all such $\mu$.

Then consider $\mu = \hat{\mu}_k$, since $V''_{\nu_i}$ is bounded on both side by $\Delta$, $V''_{\nu_i}(\mu) \leqslant \Delta$. Therefore at $\mu$ $V'_{\nu_i}$ has Lipschitz parameter $\Delta$ by Kirszbraun theorem. *Q.E.D.*

- *Step 1*: prove $V \in C[\mu^{**}, 1]$.
  Sort all rational numbers in $[\mu^{**}, 1]$ as $\{r_n\}$. $\forall N$, there exists $\mu_{n,M} \in \Omega$ s.t. $V(r_n) - V_{\mu_{n,M}}(r_n) \leqslant \frac{1}{N}$. Let $V_N = \max_n\{V_{\mu_{n,N}}\}$, then $\{V_N\} \subset \mathbb{V}$ and $V_N$ converges to $V$ pointwisely on $\{r_n\}$. Let $\hat{V} = \lim V_n$, by Lemma S.13, $\hat{V} \in C[\mu^{**}, 1]$. By definition $\hat{V} \leqslant V$. Suppose $\hat{V}(\mu) < V(\mu)$, then there exists $V_{\mu_0}(\mu) > \hat{V}(\mu)$. Since both $V_{\mu_0}$ and $\hat{V}$ are continuous, $V_{\mu_0} > \hat{V}$ on an open interval, containing some $r_n$. Contradiction. So $\hat{V} = V \in C[\mu^{**}, 1]$. Let $\{\mu \geqslant \mu^{**} | V(\mu) > F(\mu)\} = \bigcup I_m$ where $I_m$ are disjoint open intervals.

- *Step 2*: prove $\forall I_m$, exists $\mu_n \in \Omega$ s.t. $V(\mu) = \lim V_{\mu_n}(\mu)$ and $V'(\mu) = \lim V'_{\mu_n}(\mu)$ on $I_m$.
  Pick any $\mu \in I_m$. Let $\Theta(\mu) = \left\{\mu^\diamond \in \Omega | V_{\mu^\diamond}(\mu) > F(\mu)\right\}$. Then by definition of $V(\mu)$, $\Theta(\mu)$ is non-empty. Let $\tilde{V} = \sup_{\mu^\diamond \in \Theta(\mu)} V_{\mu^\diamond}$. $\forall N$, there exists $\mu_{n,M} \in \Theta(\mu)$ s.t. $\tilde{V}(r_n) - V_{\mu_{n,N}}(r_n) \leqslant \frac{1}{N}$. Since $V_{\mu_{n,N}}(\mu) > F(\mu)$, by Lemma S.12, there exists $V_{\mu_N} = \max\{V_{\mu_{n,N}}\}$. Therefore, $\lim V_{\mu_N} = \tilde{V}$ on $\{r_n\}$. By Lemma S.13 $\tilde{V} = \lim V_{\mu_N} \in C[\mu^{**}, 1]$. Now suppose $V(\mu) > \tilde{V}(\mu)$, then there exists $V_{\mu^\diamond}(\mu) > V_{\mu_n}(\mu) > F(\mu)$. Then $\mu^\diamond \in \Theta(\mu)$ by Lemma S.12, contradiction. Therefore, $\lim V_{\mu_n} = V$ on $I_m$.

  Let $I_m = (a_m, b_m)$. Now consider $\{V'_{\mu_n}\}$. $V'_{\mu_n}(a_m) = F'(a_m)$. Lemma S.14 implies that $V'_{\mu_n}$ are totally bounded and equi-continuous on $I_m$. Therefore, there exists subsequence $V'_{\mu_n}$ being Cauchy w.r.t. sup norm on $[a_m, b_m]$. So $V$ as limit of $V_{\mu_n}$ is differentiable on $[a_m, b_m]$ and $V' = \lim V'_{\mu_n}$.[1]

- *Step 3*: prove $\forall I_m$, exists $\mu^m \in \Omega$ s.t. $V(\mu) = V_{\mu^m}$ on $I_m$.
  Let $\mu^m = \inf I_m$. By step 2, it is easy to verify that $\mu_n \to \mu^m$. Then since Equation (33) is continuous in $\mu$, it is satisfied at $\mu^m$ and $\mu^m \in \Omega$. Since both $V_{\mu_n}$ and $V'_{\mu_n}$ converges on $I_m$, Equation (32) is satisfied for $V$ on $I_m$. Let $F(\mu^m) = F_k(\mu^m)$.

[1] This result is ex. 14.2.7 from Tao, 2016.

As an intermediate step, I first prove that Equation (32) is solved for $k' > k$ in a non-degenerate neighbour of $\mu^m$. Take any $\mu' > \mu^m$ s.t. $V(\mu') > F(\mu')$, since $V(\mu^m) = F_k(\mu^m)$, there exists $\mu^\star \in (\mu^m, \mu')$ and $\varepsilon > 0$ s.t $\forall \mu \in (\mu^m, \mu^\star)$ $V(\mu) - F_k(\mu) < V(\mu') - F_k(\mu') - \varepsilon$. I claim that Equation (32) is solved at all $\mu \in (\mu^m, \mu^\star)$ with $k' > k$. Suppose not, then for $n$ sufficiently large:

$$V_{\mu_n}(\mu) = \frac{c}{\rho} \frac{F_k(\nu) - V_{\mu_n}(\mu) - V'_{\mu_n}(\mu)(\nu - \mu)}{J(\mu,\nu)} \leqslant \frac{c}{\rho} \frac{F_k(\nu) - F_k(\mu) - V'_{\mu_n}(\mu)(\nu - \mu)}{J(\mu,\nu)} = (F'_k - V'_{\mu_n}(\mu)) \frac{\nu - \mu}{J(\mu,\nu)}$$

Therefore $F'_k \geqslant V'_{\mu_n}(\mu)$. By construction of $V_{\mu_n}$ at any $\mu'' \geqslant \mu$ Equation (32) is solved with $k$, therefore $F'_k \geqslant V_{\mu_n}(\mu'')$ holds for all $\mu'' \geqslant \mu$. This implies $\forall \mu'' \geqslant \mu$, $V_{\mu_n}(\mu'') - F_k(\mu'') \leqslant V_{\mu_n}(\mu) - F_k(\mu) < V(\mu') - F_k(\mu') - \varepsilon$. Take $n \to \infty$ and $\mu'' = \mu'$, contradiction. Therefore, Equation (32) is solved at all $\mu \in (\mu^m, \mu^\star)$ for $V(\mu)$ with $k' > k$.

Now consider $V_{\mu^m}(\mu)$. By my construction, suppose $V_{\mu^m}$ is updated up to action $k+1$. I claim that $V_{\mu^m} = V$ when $\mu \in [\mu^m, \mu^\star)$. Suppose not true, then there exists $\mu$ at which $V_{\mu^m}(\mu) < V(\mu)$, $V'_{\mu^m}(\mu) < V'(\mu)$. It is easy to verify that Equation (32) is violated at $V_{\mu_m}(\mu)$. Therefore, if $V_{\mu^m} \neq V$, it must happen in $(\mu^\star, b_m)$. Again we can find $\mu \in (\mu^\star, b_m)$ s.t. $V_{\mu^m}(\mu) < V(\mu)$, $V'_{\mu^m}(\mu) < V'(\mu)$, which is not possible. So $V(\mu) = V_{\mu^m}(\mu)$ on $I_m$.

To sum up, $V$ can be represented as:

$$V(\mu) = \begin{cases} V_{\mu^*}(\mu) & \text{if } \mu \in [\mu^*, \mu^{**}] \\ V_{\mu^m}(\mu) & \text{if } \mu \in I^m \\ F(\mu) & \text{otherwise} \end{cases}$$

Now I prove smoothness of $V(\mu)$ on $[\mu^*, 1]$. By Lemma S.14, $\forall \mu \in I_m$:

$$F'(a_m) - \Delta|\mu - a_m| \leqslant V'(\mu) \leqslant F'(a_m) + \Delta|\mu - a_m|$$
$$F'(b_m) + \Delta|\mu - b_m| \geqslant V'(\mu) \geqslant F'(b_m) - \Delta|\mu - b_m|$$

Therefore $|V'(\mu) - F'(\mu)|$ is bounded by $\Delta|I_n|$. Define:

$$V_n(\mu) = \begin{cases} V_{\mu^m}(\mu) & \text{when } \mu \in I_m, m \leqslant n \\ F(\mu) & \text{otherwise} \end{cases}$$

Then $V_n(\mu) \to V(\mu)$. By Lemma S.11, we can without loss assume first $n$ $V_{\mu^m}$ have $I_m$ covering $\underline{\mu}_m$. Fix $n$, $\forall \mu$, $\forall m \geqslant n$, if $\mu \in I_m$ and $m \leqslant n$ or $\mu \notin \bigcup I_m$, then $V'_n(\mu) = V'_m(\mu)$, else if $\mu \in I_m$, $m > n$, then $|V'_n(\mu) - F'(\mu)|$ and $|V'_m(\mu) - F'(\mu)|$ are all bounded by $\Delta|I_m|$. Therefore, $V'_n(\mu)$ is a Cauchy sequence. Then $V'_n(\mu) \to V'(\mu)$ pointwise. Since each $V'_n$ is continuous, $V$ is a smooth function on $[0,1]$ and $V' = F'$ when $V = F$. *Q.E.D.*

*Other lemmas for Theorem 2*

**Lemma S.15.** *$\forall \delta, \eta > 0$, $\forall \mu, \nu$ s.t. $\mu, \nu \in (\delta, 1-\delta)$, $|F_m(\mu)| > \eta$,*

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

*$L(\mu,\nu)$ is uniformly Lipschtiz continuous in $\nu$ and continuous in $\mu$.*

**Proof.** There exists $\sigma, \Delta > 0$ s.t. $\forall \mu \in (\delta, 1-\delta)$

$$\begin{cases} \Delta \geqslant |F_m(\mu)| \geqslant \eta \\ \Delta \geqslant |H''(\mu)| \geqslant \varepsilon \\ \Delta \geqslant |H'(\mu)|, |H(\mu)|, |F'_m| \end{cases}$$

Since $[\delta, 1-\delta]$ is compact, $H''$ is Lipschitz continuous on $[\delta, 1-\delta]$ with Lipschitz parameter $\Delta$. Then:

$$\begin{aligned} &\left|L(\mu,\nu) - L(\mu,\nu')\right| \\ \leqslant &\left|\frac{J(\nu,\mu)}{(\nu-\mu)F_m(\mu)}\right| \left| \frac{F'_m(1 + \frac{\rho}{c} J(\mu,\nu)) + \frac{\rho}{c} F_m(\nu)(H'(\nu) - H'(\mu))}{H''(\nu)} - \frac{F'_m(1 + \frac{\rho}{c} J(\mu,\nu')) + \frac{\rho}{c} F_m(\nu')(H'(\nu') - H'(\mu))}{H''(\nu')} \right| \\ &+ \left| \frac{J(\nu,\mu)}{\nu - \mu} - \frac{J(\nu',\mu)}{\nu' - \mu} \right| \left| \frac{F'_m(1 + \frac{\rho}{c} J(\mu,\nu')) + \frac{\rho}{c} F_m(\nu')(H'(\nu') - H'(\mu))}{H''(\nu')} \right| \\ \leqslant &\frac{2\Delta}{\eta\varepsilon} \left| \begin{matrix} F'_m(1 + \frac{\rho}{c} J(\mu,\nu)) + \frac{\rho}{c} F_m(\nu)(H'(\nu) - H'(\mu)) \\ -F'_m(1 + \frac{\rho}{c} J(\mu,\nu')) + \frac{\rho}{c} F_m(\nu')(H'(\nu') - H'(\mu)) \end{matrix} \right| + \frac{2\Delta}{\eta} \left| \frac{H''(\nu') - H''(\mu)}{H''(\nu')H''(\mu)} \right| \left| F'_m\left(1 + \frac{\rho}{c} J(\mu,\nu')\right) + \frac{\rho}{c} F_m(\nu')(H'(\nu') - H'(\mu)) \right| \end{aligned}$$

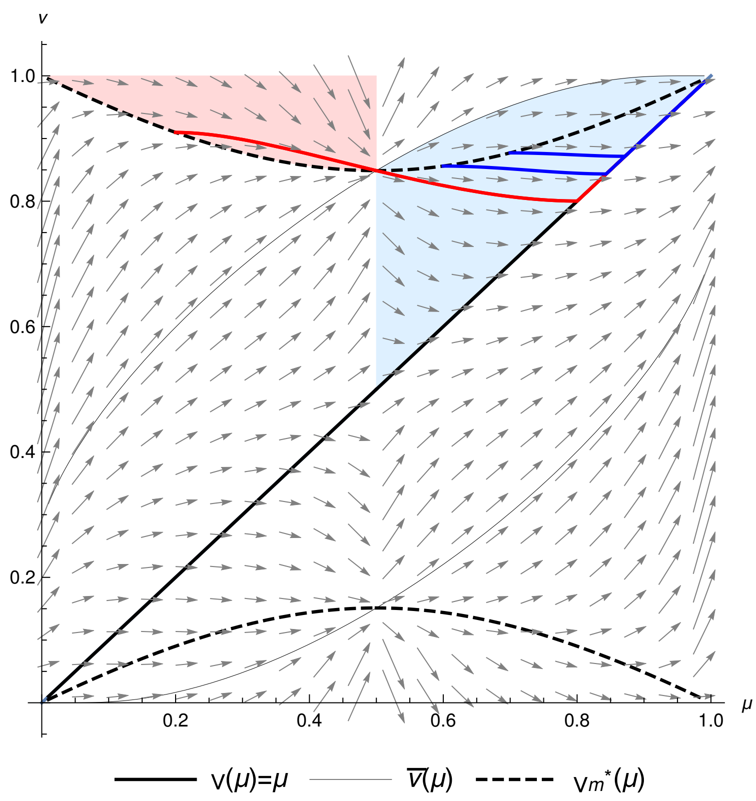
$$
\begin{aligned}
&+\left|\frac{J(\nu,\mu)}{\nu-\mu}-\frac{J(\nu',\mu)}{\nu'-\mu}\right|\left|\frac{F_m'\left(1+\frac{\rho}{c}J(\mu,\nu')\right)+\frac{\rho}{c}F_m(\nu')(H'(\nu')-H'(\mu))}{H''(\nu')}\right| \\
&\leqslant \frac{2\Delta}{\eta\varepsilon}\left|\begin{matrix} F_m'\left(1+\frac{\rho}{c}J(\mu,\nu)\right)+\frac{\rho}{c}F_m(\nu)(H'(\nu)-H'(\mu)) \\ -F_m'\left(1+\frac{\rho}{c}J(\mu,\nu')\right)+\frac{\rho}{c}F_m(\nu')(H'(\nu')-H'(\mu))\end{matrix}\right| + \frac{2\Delta}{\eta}\left|\frac{H''(\nu')-H''(\mu)}{H''(\nu')H''(\mu)}\right|\left(\Delta+\frac{\rho}{c}5\Delta^2\right)+\left|\frac{J(\nu,\mu)}{\nu-\mu}-\frac{J(\nu',\mu)}{\nu'-\mu}\right|\frac{\Delta+\frac{\rho}{c}5\Delta^2}{\varepsilon} \\
&\leqslant \frac{2\Delta}{\eta\varepsilon}\left|F_m'\left(\frac{\rho}{c}H'(\nu)\right)+\frac{\rho}{c}F_m(\mu)H''(\tilde{\nu})\right||\nu-\nu'|+\frac{2\Delta^2+10\frac{\rho}{c}\Delta^3}{\eta\varepsilon^2}\Delta|\nu'-\nu| \\
&+\frac{\Delta+\frac{\rho}{c}5\Delta^2}{\varepsilon}\left|-H''(\tilde{\nu}-\frac{J(\tilde{\nu},\mu)}{(\tilde{\nu}-\mu)^2})\right||\nu'-\mu| \\
&\leqslant \frac{2\Delta}{\eta\varepsilon}\left|2\frac{\rho}{c}\Delta^2\right||\nu-\nu'|+\frac{2\Delta^2+10\frac{\rho}{c}\Delta^3}{\eta\varepsilon^2}\Delta|\nu'-\nu|+\frac{\Delta+\frac{\rho}{c}5\Delta^2}{\varepsilon}\left|-H''(\tilde{\nu})+\frac{1}{2}H''(\tilde{\tilde{\nu}})\right||\nu'-\mu| \\
&\leqslant \left(\frac{4\frac{\rho}{c}\Delta^3}{\eta\varepsilon}+\frac{2\Delta^3+10\frac{\rho}{c}\Delta^4}{\eta\varepsilon^2}+\frac{2\Delta^2+10\frac{\rho}{c}\Delta^3}{\varepsilon}\right)|\nu'-\nu|
\end{aligned}
$$

Therefore, $L(\mu,\nu)$ is uniformly Lipschitz continuous in $\nu$. It is easy to see that $L(\mu,\nu)$ is continuous in $\mu$ when $\mu$ is bounded away from $\nu$. Now we only need to consider $\mu \to \nu$:

$$
\left|\frac{L(\mu,\nu)}{\nu-\mu}\right|=\left|\frac{(\mu-\nu)H''(\tilde{\nu})\frac{\rho}{c}F_m(\tilde{\nu}')(\nu-\mu)}{\frac{\rho}{c}(\nu-\mu)^2F_m(\mu)H''(\nu)}\right|\leqslant\frac{\Delta^2}{\eta}
$$

Therefore, $L(\nu,\mu)$ is uniformly Lipschitz continuous in $\nu$ and continuous in $\mu$.

*Q.E.D.*




$\bar{\nu}(\mu)$ is defined by: $\frac{\rho}{c}J(\bar{\nu}(\mu),\mu)=1$.
$\nu_m^*(\mu)$ is defined by: $F_m'\left(1+\frac{\rho}{c}J(\mu,\nu_m^*(\mu))\right)+\frac{\rho}{c}F_m(\nu_m^*(\mu))(H'(\nu^*(\mu))-H'(\mu))=0$.
The red line and blue lines are solution path of ODE $\dot{\mu}=L(\mu,\nu)$ with initial value satisfying Lemma S.16.

Figure S.1: Phase diagram of $(\dot{\mu},\dot{\nu})$.

**Lemma S.16.** *Assume $\mu_0\in[\mu^*,1)$, $F_m(\mu_0)\neq 0$, $F_m'\geqslant 0$, $\nu_0\in[\mu_0,1)$ satisfies:*

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

*Then there is a continuous function $\nu$ on $[\mu_0,1]$ satisfying initial condition $\nu(\mu_0)=\nu_0$. On $\{\mu|\nu(\mu)>\mu\}$, $\nu$ is differentiable, strictly decreasing and satisfis ODE:*

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

**Proof.** Before we proceed to solving the ODE, we characterize the dynamics of $(\mu,\nu)$ on $[0,1]^2$. Figure S.1 shows the phase diagram of $\dot{\mu},\dot{\nu}$ on $[0,1]^2$ and some important functions that determines the dynamics of $(\mu,\nu)$. The horizongtal axis is $\mu$ and vertical axis is $\nu$. The black line is $\nu=\mu$. The two thin black lines characterizes $\bar{\nu}(\mu)$ as the solutions to:

$$1-\frac{\rho}{c}J(\bar{\nu}(\mu),\mu)=0$$

The two dashed black lines characterizes $\nu^*(\mu)$ as the two solutions to:

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

Since we are discussing the case $\nu\geqslant\mu$, we only focus on the upper left half of the graph:

- $F(\mu_0)<0$. This corresponds to the left half of the graph.

$$F_m'\left(1+\frac{\rho}{c}J(\mu_0,\nu_0)\right)+\frac{\rho}{c}F_m(\nu_0)(H'(\nu_0)-H'(\mu_0))\leqslant 0$$
$$\Longrightarrow \nu_0\geqslant\nu^*(\mu_0)$$

  Therefore our initial condition means $(\mu_0,\nu_0)$ lies in the red region. $\dot{\nu}=0$ when $\nu(\mu)=\nu^*$. otherwise $\dot{\nu}<0$. When $F(\mu)$ is close to 0, $\dot{\nu}$ goes to negative infinity if $\nu>\nu^*(\mu)$. So the dynamics of $\nu$ in this region must have $\nu$ strictly decreasing and reaches $\nu^*$ when $F(\mu)=0$. Intuitively, $\nu$ will never reach the region $\nu>\nu_0$. Then uniform Lipschtiz continuity of $L(\mu,\nu)$ on $\nu\in[\mu,\nu_0]$, for $\mu\in[\mu_0,F^{-1}(-\eta)]$ will be enough to guarantee existence of solution.

- $F(\mu_0)>0$. This corresponds to the right half of the graph.

$$F_m'\left(1+\frac{\rho}{c}J(\mu_0,\nu_0)\right)+\frac{\rho}{c}F_m(\nu_0)(H'(\nu_0)-H'(\mu_0))\geqslant 0$$
$$\Longrightarrow \nu_0\leqslant\nu^*(\mu_0)$$

  Our intial condition will lie below the dashed line in blue region. $L(\mu,\nu)<0$ in this region and $L(\mu,\nu^*)=0$. So the dynamics of $\nu$ in this region must have $\nu$ strictly decreasing until it reaches $\nu=\mu$. Then uniform Lipschtiz continuity of $L(\mu,\nu)$ on $\nu\in[\mu,\nu_0]$ for $\mu\in[\mu_0,1]$ will be sufficient ot guanrantee existence of solution.

Then we characterize formally the solution of ODEs:

- $F_m(\mu_0)>0$. Our conjecture is that solution $\nu$ will be no larger than $\nu_0$ within the region: $\mu\in[\mu_0,\nu_0]$, $\nu\in[\mu_0,\nu_0]$. Therefore, we modify $L(\mu,\nu)$ to define $\tilde{L}(\mu,\nu)$ on the whole space:

$$\tilde{L}(\mu,\nu)=L(\max\{\min\{\mu,\nu_0\},\mu_0\},\max\{\min\{\nu,\nu_0\},\mu_0\})$$

  It's not hard to see that $\tilde{L}$ is uniformly Lipschtiz continuous w.r.t $\nu\in\mathbb{R}$ for $\mu\in[0,1]$ and continuous in $\mu\in[0,1]$. We can apply Picard-Lindelof to solve for ODE $\dot{\tilde{\nu}}=\tilde{L}(\mu,\tilde{\nu})$ on the space with initial condition $\tilde{\nu}(\mu_0)=\nu_0$.

  - Consider $\tilde{\nu}$ on $[\mu_0,1]$, it starts at $\nu_0>\mu_0$. It first reaches $\nu=\mu$ at $\bar{\mu}\in(\mu_0,1]$ (we define it to be 1 when it doesn't exist). Then for $\mu\in(\mu_0,\bar{\mu})$, we must have $L(\mu,\tilde{\nu})<0$. Suppose not, then there exists $\tilde{\nu}(\mu)\geqslant\nu_m^*(\mu)>\nu_0$. We pick a smallest $\mu$ such that this is true. Then this $\mu$ must be strictly larger than $\mu_0$ because $L(\mu_{,0},\nu_0)=0<\dot{\nu}_m^*(\mu_0)$. Then at $\mu$, $\dot{\tilde{\nu}}(\mu)=0$ but $\dot{\nu}_m^*(\mu)>0$. It's impossible that $\tilde{\nu}$ crosses $\nu_m^*$ from below. Contradiction. Then $\dot{\tilde{\nu}}<0$ until it hits $\nu=\mu$.
  - $\bar{\mu}<\nu_0$. Suppose $\bar{\mu}\geqslant\mu_0$, since $\dot{\tilde{\nu}}<0$ on $(\mu_0,\bar{\mu})$, $\tilde{\nu}(\bar{\mu})<\nu_0$. Contradiction. Therefore, $\tilde{\nu}$ on $[\mu_0,\bar{\mu}]$ will be with region $[\mu_0,\nu_0]$.

  In the region $[\mu_0,\bar{\mu}]\times[\mu_0,\nu_0]$, $\tilde{L}$ coincides $L$. Therefore, $\tilde{\nu}$ is a solution to original ODE Equation (44). We define $\nu$:

$$\nu(\mu)=\begin{cases}\tilde{\nu}(\mu) & \text{if } \mu\in[\mu_0,\bar{\mu}]\\ \mu & \text{if } \mu\in[\bar{\mu},1]\end{cases}$$

  It's easy to verify that $\nu$ satisfies Lemma S.16. The blue line on Figure S.1 illustrates a solution in this case.

- $F_m(\mu_0) < 0$. Define $\mu^0 = F^{-1}(0)$, our conjecture is that solution $\nu$ will be decreasing on $[\mu_0, \mu^0)$. $\forall \eta > 0$, define $\mu^\eta = F_{-1}(-\eta)$, we modify $L(\mu,\nu)$ to define $\tilde{L}(\mu,\nu)$ on the whole space:
$$\tilde{L}(\mu,\nu) = L(\max(\min(\mu,\mu^\eta),\mu_0), \max\{\min\{\nu,\nu_0\},\nu_m^*(\mu)\})$$
It's not hard to see that $\tilde{L}$ is uniformly Lipschtiz continuous w.r.t. $\nu \in \mathbb{R}$ for $\mu \in [0,1]$ and continuous in $\mu \in [0,1]$. We can apply Picard-Lindelof to solve for ODE $\dot{\tilde{\nu}} = \tilde{L}(\mu,\tilde{\nu})$ on the space with initial condition $\tilde{\nu}(\mu_0) = \nu_0$. $\tilde{\nu}$ will be strictly decreasing on $(\mu_0, \mu^\eta]$. Because when $\tilde{\nu}$ first touches $\nu_m^*$ is must crosses from below and this is not possible. Then, when $\mu \in [\mu_0, \mu^\eta]$, we have $L(\mu,\tilde{\nu}) = \tilde{L}(\mu,\tilde{\nu})$. Therefore $\tilde{\nu}$ is a solution to original ODE Equation (44).

Then we extend $\tilde{\nu}$ to $[\mu_0, \mu^0)$ by taking $\eta \to 0$ and define:
$$\nu(\mu) = \begin{cases} \tilde{\nu}(\mu) & \text{if } \mu \in [\mu_0, \mu^0) \\ \overline{\lim}_{\mu \to F^{-1}(0)} \tilde{\nu}(\mu) & \text{if } \mu = F^{-1}(0) \end{cases}$$
First since $\tilde{\nu}$ is decreasing, the sup limit will actually be the limit and $\nu \in C[\mu_0, \mu^0]$. Then we show that this extension is left differentiable at $\mu^0$. Consider:
$$V(\mu) = \frac{F_m(\mu)}{1 - \frac{\rho}{c} J(\nu(\mu),\mu)}$$
By Equation (45), we know that on $[\mu_0, \mu^0)$ sign of $V'$ is determined by sign of $1 - \frac{\rho}{c} J(\nu(\mu),\mu)$. At initial value, $V_0 \geqslant 0 \implies 1 - \frac{\rho}{c} J(\nu_0, \mu_0) > 0$. On the other hand, $V(\mu)$ will be bounded above by $\overline{V}$. So $1 - \frac{\rho}{c} J(\nu(\mu),\mu)$ as a continuous function of $\mu$ has to stay above 0. Therefore $V'(\mu) > 0$ on $[\mu_0, \mu^0)$. By monotonic convergence, there exists $\lim_{\mu \to \mu^{0-}} V(\mu)$. Define it as $V(\mu^0)$. We define:
$$\dot{\nu}(\mu^0) = \frac{\frac{F_m'}{V(\mu_0)} + \frac{\rho}{c}\left(H'(\nu(\mu^0)) - H'(\mu^0)\right)}{\frac{\rho}{c} H''(\nu(\mu^0))(\nu(\mu^0) - \mu^0)}$$
Now we show that $\dot{\nu}(\mu^0) = \lim_{\mu \to \mu^0} \frac{\nu(\mu) - \nu(\mu^0)}{\mu - \mu^0}$. Suppose not, there exists $\varepsilon > 0$, $\mu_n \to \mu^0$ s.t. $\left|\dot{\nu}(\mu^0) - \frac{\nu(\mu_n) - \nu(\mu^0)}{\mu_n - \mu^0}\right| > \varepsilon$. Suppose $\nu(\mu_n) > \nu(\mu^0) + \left(\dot{\nu}(\mu^0) - \varepsilon\right)(\mu_n - \mu^0)$:
$$\begin{aligned}
V(\mu_n) &< \frac{F_m(\mu)}{1 - \frac{\rho}{c} J(\nu^0 + (\dot{\nu}(\mu^0) - \varepsilon)(\mu_n - \mu^0), \mu_n)} \\
\implies \lim_{n \to \infty} V(\mu_n) &\leqslant \frac{F_m'}{\frac{\rho}{c}(-H'(\nu(\mu^0)) + H'(\mu^0) + H''(\nu(\mu^0))(\nu(\mu^0) - \mu^0)(\dot{\nu}(\mu^0) - \varepsilon))} \\
&< \frac{F_m'}{\frac{\rho}{c}(-H'(\nu(\mu^0)) + H'(\mu^0) + H''(\nu(\mu^0))(\nu(\mu^0) - \mu^0)\dot{\nu}(\mu^0))} \\
&= V(\mu^0)
\end{aligned}$$
First strict inequality is from $1 - \frac{\rho}{c} J(\nu,\mu)$ strictly increasing in $\nu$. When $F_m(\mu) < 0$, $\frac{F_m(\mu)}{1 - \frac{\rho}{c} J(\nu,\mu)}$ will be decreasing in $\nu$. Second inequality is by taking limit of lower bounded of $V(\mu_n)$ with L'Hospital rule. Third strict inequality is from $\varepsilon > 0$, $H'' < 0$. Last equality is from definition of $\dot{\nu}(\mu^0)$. We get contradiction. Similarly, we cna rule out $\nu(\mu_n) < \nu(\mu^0) + (\dot{\nu}(\mu^0) + \varepsilon)(\mu_n - \mu^0)$. Therefore, we extended $\nu$ to $[\mu_0, \mu^0]$ such that it's differentiable on $[\mu_0, \mu^0]$ and smooth on $(\mu_0, \mu^0)$.

Let $\mu_0 = \mu^0$, $\nu_0 = \nu(\mu^0)$, $\nu_0' = \dot{\nu}(\mu^0)$, then $\nu_0 > \mu_0$ and
$$\begin{cases} 1 - \frac{\rho}{c} J(\nu_0, \mu_0) = 0 \\ 0 < \frac{F_m'}{\frac{\rho}{c}(H'(\mu_0) - H'(\nu_0) + H''(\nu_0)(\nu_0 - \mu_0)\nu_0')} = V(\mu_0) \leqslant \overline{V}(\mu_0) \end{cases}$$
Then by Lemma S.17, we can solve for $\nu(\mu)$ on $[\mu^0, 1]$ satisfying the conditions in Lemma S.17. Moreover, $\dot{\nu}(\mu^0) = \nu_0$, then $\nu$ is differentiable at $\mu^0$. For any other points in $\{\mu | \nu(\mu) > \mu\}$, $\nu$ is $C^{(1)}$ smooth. Since $\nu_0' < 0$, then the solved $\nu$ will be strictly decreasing.

*Q.E.D.*

**Lemma S.17.** *Assume $F_m(\mu_0) = 0$, $F_m' > 0$, $\nu_0 \in [\mu_0, 1)$, $\nu_0'$ satisfies*
$$\begin{cases} 1 - \frac{\rho}{c} J(\nu_0, \mu_0) = 0 \\ 0 < \frac{F_m'}{\frac{\rho}{c}\left(H'(\mu_0) - H'(\nu_0) + H''(\nu_0)(\nu_0 - \mu_0)\nu_0'\right)} \leqslant \overline{V}(\mu_0) \end{cases}$$

*Then there is a continuous function $\nu$ on $[\mu_0,1]$ satisfying initial condition $\nu(\mu_0)=\nu_0$, $\dot{\nu}(\mu_0)=\nu_0'$. On $\{\mu|\nu(\mu)>\mu\}$, $\nu$ is differentiable, strictly decreasing and satisfies ODE:*

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

**Proof.** $\forall\ \mu_1\in(\mu_0,1)$, $\forall \nu_1\in[\mu_1,\nu_m^*(\mu_1))$, we consider the solution of ODE with initial condition $(\mu_0,\nu_0)$. $\forall\eta>0$, define $\mu^\eta=F^{-1}(\eta)$. Then like the proof of Lemma S.16, we can solve for a smooth $\nu$ on $[\mu^\eta,\bar{\mu}]$. $\nu$ will be strictly decreasing below $\nu_m^*$ and strictly increasing over $\nu_m^*$. Consider the slope of $\bar{\nu}$:

$$\dot{\bar{\nu}}=\frac{H'(\bar{\nu})-H'(\mu)}{H''(\bar{\nu})(\bar{\nu}-\mu)}=L(\mu,\bar{\nu})$$

$\bar{\nu}$ itself satisfies ODE Equation (44), then uniqness of solution to ODE implies $\nu<\bar{\nu}\ \forall\mu\in[\mu^\eta,\bar{\mu}]$. So solution must lies in the blue region in Figure S.1. Let

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

When $\nu_1\to\bar{\nu}(\mu_1)$, $1-\frac{\rho}{c}J(\nu(\mu),\mu)\to 0$. Thus $V(\mu)\to\infty$. On the other hand, when $\mu_1\to\mu_0$, $\nu_1=\mu_1$, $V(\mu)\to F_m(\mu_0)=0$. Define

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

I want to show that there exists there exists $\mu_1,\nu_1$ s.t. $V(\mu)\to V_0$ when $\mu\to\mu_0$.

Index $V(\mu^0)$ by initial value $(\mu_1,\nu_1)$: $V_0(\mu_1,\nu_1)$. I claim that $V_0(\mu_1,\nu_1)$ is continuous in $(\mu_1,\nu_1)$. Suppose not, then there exists $\lim_{\mu_1^n,\nu_1^n\to\mu_1,\nu_1}V_0(\mu_1^n,\nu_1^n)\neq V_0(\mu_1,\nu_1)$. On the other hand, index $V(\mu^\eta)$ by initial value $(\mu_1,\nu_1)$: $V_\eta(\mu_1,\nu_1)$, then continuous dependence of ODE guanrantees that $\lim_{\mu_1^n,\nu_1^n\to\mu_1,\nu_1}V_\eta(\mu_1^n,\nu_1^n)=V_\eta(\mu_1,\nu_1)$. Therefore, $\forall N$, there exists $\eta$ s.t.

$$\frac{\left|\lim_{\mu_1^n,\nu_1^n\to\mu_1,\nu_1}V_0(\mu_1^n,\nu_1^n)-V_0(\mu_1,\nu_1)\right|}{|\mu^0-\mu^\eta|}>3N$$

Then by continuity, we can have $\eta$ sufficiently small that:

$$\frac{\left|\lim_{\mu_1^n,\nu_1^n\to\mu_1,\nu_1}V_0(\mu_1^n,\nu_1^n)-V_\eta(\mu_1,\nu_1)\right|}{|\mu^0-\mu^\eta|}>2N$$

Then we can have $n$ sufficiently large that:

$$\frac{\left|V_0(\mu_1^n,\nu_1^n)-V_\eta(\mu_1^n,\nu_1^n)\right|}{|\mu_0-\mu_\eta|}>N$$

Then there must exists $\tilde{\mu}_N$ s.t. $|V'(\tilde{\mu}_N)|>N$. On the other hand, $|V'|$ must be bounded because:

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

When $V'$ going to positive infinity, $V(\mu)$ will go to $F_m(\mu)$. When $V'$ going to negative infinity, $V(\mu)$ will go to positive infinity. Both cases are impossible. Therefore, $V_0(\mu_1,\nu_1)$ will be a continuous function on initial value. There exists $\mu_1,\nu_1$ such that $\lim_{\eta\to 0}V(\mu^\eta)=V_0$. Apply L'Hospital rule to $V(\mu)=\frac{F_m(\mu)}{1-\frac{\rho}{c}J(\nu(\mu),\mu)}$, we ge that:

$$V_0=\frac{F_m'}{\frac{\rho}{c}\left(H'(\mu_0)-H'(\nu_0)+H''(\nu_0)\lim_{\mu\to\mu_0}\nu'(\mu)\right)}\implies\lim_{\mu\to\mu_0}\nu'(\mu)=\nu_0'$$

Smoothly extend $\nu(\mu)$ to $\mu_0$. Therefore, $\nu(\mu)$ associated with initial value $(\mu_1,\nu_1)$ satisfies $\dot{\nu}(\mu^0)=\nu_0'$. Since $\bar{\nu}$ satisfies $\frac{F_m'}{\frac{\rho}{c}(H'(\mu_0)-H'(\bar{\nu}(\mu_0)))}=\overline{V}(\mu_0)$, the assumption in Lemma S.17 implies $\nu_0'\leqslant 0$. *Q.E.D.*

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

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

*Then there exists a $C^{(1)}$ smooth and strictly decreasing $V(\mu)$ defined on $[0,\mu_0]$ satisfying*

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

*and initial condition $V(\mu_0)=V_0,V'(\mu_0)=V_0'$.*

**Lemma S.16′.** *Assume $\mu_0 \in (0,\mu^*]$, $F_m(\mu_0) \neq 0$, $F'_m \leqslant 0$, $\nu_0 \in (0,\mu_0]$ satisfies:*

$$F_m(\mu_0)\Big(-F'_m\Big(1+\frac{\rho}{c}J(\mu_0,\nu_0)\Big)+\frac{\rho}{c}F_m(\nu_0)(H'(\mu_0)-H'(\nu_0))\Big) \geqslant 0$$

*Then $\exists \nu \in C[0,\mu_0]$ satisfying initial condition $\nu(\mu_0) = \nu_0$. On $\{\mu|\nu(\mu) > \nu\}$, $\nu$ is differentiable, strictly decreasing and satifies ODE:*

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

**Lemma S.17′.** *Assume $F_m(\mu_0) = 0$, $F'_m < 0$, $\nu_0 \in (0,\mu_0]$, $\nu'_0$ satisfies*

$$\begin{cases} 1-\frac{\rho}{c}J(\nu_0,\mu_0) = 0 \\ 0 > \frac{\rho}{c}(H'(\mu_0)-H'(\nu_0)+J''(\nu_0)(\nu_0-\mu_0)\nu_0) \geqslant \frac{\overline{V}(\mu_0)}{F'_m} \end{cases}$$

*Then $\exists\, \nu \in C[0,\mu_0]$ satifying initial condition $\nu(\mu_0) = \nu_0$, $\dot{\nu}(\mu_0) = \nu'_0$. On $\{\mu|\nu(\mu) > \mu\}$, $\nu$ is differentiable, strictly decreasing and satisfies ODE:*

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

**Lemma S.18.** *Suppose at $\mu_0$, $V_0, V'_0$, $k \geqslant 1$ satisfies:*

$$\begin{cases} V_0 = \max\limits_{\nu \geqslant \mu_0}\frac{c}{\rho}\frac{F_{m-k}(\nu)-V_0-V'_0(\nu-\mu_0)}{J(\mu_0,\nu)} \geqslant \max\limits_{\nu \geqslant \mu_0}\frac{c}{\rho}\frac{F_m(\nu)-V_0-V'_0(\nu-\mu_0)}{J(\mu_0,\nu)} \\ \overline{V}(\mu_0) \geqslant V_0 \geqslant F_{m-k}(\mu_0) \end{cases}$$

*$V_{m-k}$ is the solution as defined in Lemma B.2 with initial condition $V_0, V'_0$, then $\forall \mu \in [\mu_0, \nu(\mu_0)]$:*

$$V_{m-k}(\mu) \geqslant \max_{\nu \geqslant \mu, m' \in [m-k,m]}\frac{c}{\rho}\frac{F_{m'}(\nu)-V_{m-k}(\mu)-V'_{m-k}(\mu)(\nu-\mu)}{J(\mu,\nu)}$$

**Proof.** I first claim that:

$$V_0 \geqslant \max_{\nu \in \left[\mu_0,\underline{\mu}_m\right]}\frac{c}{\rho}\frac{V_{m-k}(\nu)-V_0-V'_0(\nu-\mu_0)}{J(\mu_0,\nu)}$$

Suppose not, then there exists $\mu'$ s.t.

$$V_0 < \frac{c}{\rho}\frac{V_{m-k}(\mu')-V_0-V'_0(\mu'-\mu_0)}{J(\mu_0,\mu')} \tag{S.15}$$

By definition of $V_0$, we must have $V_{m-k}(\mu') > F_{m-k}(\mu')$. The inequality is trivial because if $F_{m-k}(\mu') = V_{m-k}(\mu')$, then choosing $\mu'$ will be suboptimal. Therefore $\nu(\mu') > \mu'$. Optimality implies Equation (41) and Equation (40) at $\mu = \mu_0$:

$$\begin{cases} F'_{m-k}+\frac{\rho}{c}V_0H'(\nu(\mu)) = V'_0+\frac{\rho}{c}V_0H'(\mu) \\ \Big(F_{m-k}(\nu(\mu))+\frac{\rho}{c}V_0H(\nu(\mu))\Big)-\Big(V_0+\frac{\rho}{c}V_0H(\mu)\Big) = \Big(V'_0+\frac{\rho}{c}V_0H'(\mu)\Big)(\nu(\mu)-\mu) \end{cases}$$

We define $L(V,\lambda,\mu)(\mu')$ as a linear function of $\mu'$:

$$L(V,\lambda,\mu)(\mu') = (V(\mu)+\lambda H(\mu))+(V'(\mu)+\lambda H'(\mu))(\mu'-\mu) \tag{S.16}$$

Define $G(V,\lambda)(\mu)$ as a function of $\mu$:

$$G(V,\lambda)(\mu) = V(\mu)+\lambda H(\mu) \tag{S.17}$$

Then $G(F_{m-k},\frac{\rho}{c}V_{m-k}(\mu_0))(\mu')$ is a concave function of $\mu'$. Consider:

$$L\Big(V_{m-k},\frac{\rho}{c}V_{m-k}(\mu_0),\mu_0\Big)(\mu')-G\Big(F_{m-k},\frac{\rho}{c}V_{m-k}(\mu_0)\Big)(\mu')$$

This is a convex function and have unique minimum. Therefore, the minimum will be determined by FOC. Simple calculation shows that it is minimized at $\nu(\mu_0)$ and the minimal value is 0.

$$\text{FOC}: V'_{m-k}(\mu_0)+\frac{\rho}{c}V_{m-k}(\mu_0)H'(\mu_0) = F'_{m-k}+\frac{\rho}{c}V_{m-k}(\mu_0)H'(\mu')$$

It's easy to see that this equation is identical to the FOC for $\nu(\mu_0)$. Now consider:

$$\begin{aligned} &L\Big(V_{m-k},\frac{\rho}{c}V_{m-k}(\mu_0),\mu_0\Big)(\mu')-G\Big(V_{m-k},\frac{\rho}{c}V_{m-k}(\mu_0)\Big)(\mu') \\ =&\Big(V_{m-k}(\mu_0)+\frac{\rho}{c}V_{m-k}(\mu_0)H(\mu_0)\Big)+\Big(V'_{m-k}(\mu_0)+\frac{\rho}{c}V_{m-k}(\mu_0)H'(\mu_0)\Big)(\mu'-\mu_0) \\ &-\Big(V_{m-k}(\mu')+\frac{\rho}{c}V_{m-k}(\mu_0)H(\mu')\Big) \end{aligned}$$

$$=-\left(V_{m-k}(\mu')-V_{m-k}(\mu_0)-V'_{m-k}(\mu_0)(\mu'-\mu_0)-\frac{\rho}{c}V_{m-k}(\mu_0)J(\mu_0,\mu')\right)<0$$

The last inequality is from rewriting Equation (S.15). Therefrore, $L(V_{m-k},\frac{\rho}{c}V_{m-k}(\mu_0),\mu_0)(\mu')-G(V_{m-k},\frac{\rho}{c}V_{m-k}(\mu_0))(\mu')$ will have minimum strictly negative. Suppose it's minimized at $\mu'$ (Since $L(\mu_0)-G(\mu_0)=0$, $\mu'$ must be bounded away from $\mu_0$). Then FOC implies:

$$V'_{m-k}(\mu_0)+\frac{\rho}{c}V_{m-k}(\mu_0)H'(\mu_0)=V'_{m-k}(\mu')+\frac{\rho}{c}V_{m-k}(\mu_0)H(\mu')$$

Consider:

$$\begin{aligned}
&L\left(V_{m-k},\frac{\rho}{c}V_{m-k}(\mu_0),\mu'\right)(\nu(\mu'))-G\left(F_{m-k},\frac{\rho}{c}V_{m-k}(\mu_0)\right)(\nu(\mu'))\\
=&L\left(V_{m-k},\frac{\rho}{c}V_{m-k}(\mu_0),\mu_0\right)(\nu(\mu'))-G\left(F_{m-k},\frac{\rho}{c}V_{m-k}(\mu_0)\right)(\nu(\mu'))\\
&+V_{m-k}(\mu')-V_{m-k}(\mu_0)+\frac{\rho}{c}V_{m-k}(\mu_0)(H(\mu')-H(\mu_0))-(V'_{m-k}(\mu_0)+\frac{\rho}{c}H'(\mu_0))(\mu'-\mu_0)\\
\geqslant&V_{m-k}(\mu')-V_{m-k}(\mu_0)+\frac{\rho}{c}V_{m-k}(\mu_0)(H(\mu')-H(\mu_0))-(V'_{m-k}(\mu_0)+\frac{\rho}{c}H'(\mu_0))(\mu'-\mu_0)\\
=&G\left(V_{m-k},\frac{\rho}{c}V_{m-k}(\mu_0)\right)(\mu')-L\left(V_{m-k},\frac{\rho}{c}V_{m-k}(\mu_0),\mu_0\right)(\mu')>0
\end{aligned}$$

In the first equality we used FOC. In the first inequality we used suboptimality of $\nu(\mu')$ at $\mu_0$. However:

$$\begin{aligned}
0=&L\left(V_{m-k},\frac{\rho}{c}V_{m-k}(\mu'),\mu'\right)(\nu(\mu'))-G\left(F_{m-k},\frac{\rho}{c}V_{m-k}(\mu')\right)(\nu(\mu'))\\
=&L\left(V_{m-k},\frac{\rho}{c}V_{m-k}(\mu_0),\mu'\right)(\nu(\mu'))-G\left(F_{m-k},\frac{\rho}{c}V_{m-k}(\mu_0)\right)(\nu(\mu'))\\
&+\frac{\rho}{c}(V_{m-k}(\mu')-V_{m-k}(\mu))(H(\mu')-H(\nu(\mu'))+H'(\mu')(\nu(\mu')-\mu'))\\
>&\frac{\rho}{c}(V_{m-k}(\mu')-V_{m-k}(\mu))J(\mu',\nu(\mu'))>0
\end{aligned}$$

Contradiction.

Now we show Lemma S.18. Suppose that it is not true, then there exists $\mu'\in(\mu_0,\nu(\mu_0))$ and $\mu''\geqslant\underline{\mu}_{m'}$ s.t.:

$$V_{m-k}(\mu')<\frac{c}{\rho}\frac{F_{m'}(\mu'')-V_{m-k}(\mu')-V'_{m-k}(\mu')(\mu''-\mu')}{J(\mu',\mu'')}$$

Then by definition:

$$\begin{aligned}
0\leqslant&L\left(V_{m-k},\frac{\rho}{c}V_{m-k}(\mu_0),\mu_0\right)(\mu'')-G\left(F_{m'},\frac{\rho}{c}V_{m-k}(\mu_0)\right)(\mu'')\\
=&L\left(F_{m-k},\frac{\rho}{c}V_{m-k}(\mu_0),\nu(\mu_0)\right)(\mu'')-G\left(F_{m'},\frac{\rho}{c}V_{m-k}(\mu_0)\right)(\mu'')\\
0\leqslant&L\left(V_{m-k},\frac{\rho}{c}V_{m-k}(\mu_0),\mu_0\right)(\mu')-G\left(V_{m-k},\frac{\rho}{c}V_{m-k}(\mu_0)\right)(\mu')\\
=&L\left(F_{m-k},\frac{\rho}{c}V_{m-k}(\mu_0),\nu(\mu_0)\right)(\mu')-G\left(V_{m-k},\frac{\rho}{c}V_{m-k}(\mu_0)\right)(\mu')\\
\implies&L\left(F_{m-k},\frac{\rho}{c}V_{m-k}(\mu'),\nu(\mu_0)\right)(\mu'')-G\left(F_{m'},\frac{\rho}{c}V_{m-k}(\mu')\right)(\mu'')\\
=&L\left(F_{m-k},\frac{\rho}{c}V_{m-k}(\mu_0),\nu(\mu_0)\right)(\mu'')-G\left(F_{m'},\frac{\rho}{c}V_{m-k}(\mu_0)\right)(\mu'')\\
&+\frac{\rho}{c}(V_{m-k}(\mu')-V_{m-k}(\mu_0))J(\mu_0,\mu'')\\
>&0\\
&L\left(F_{m-k},\frac{\rho}{c}V_{m-k}(\mu'),\nu(\mu_0)\right)(\mu')-G\left(V_{m-k},\frac{\rho}{c}V_{m-k}(\mu')\right)(\mu')\\
=&L\left(F_{m-k},\frac{\rho}{c}V_{m-k}(\mu_0),\nu(\mu_0)\right)(\mu')-G\left(V_{m-k},\frac{\rho}{c}V_{m-k}(\mu_0)\right)(\mu')\\
&+\frac{\rho}{c}(V_{m-k}(\mu')-V_{m-k}(\mu_0))J(\mu_0,\mu')\\
>&0
\end{aligned}$$

Now we consider $L\left(V_{m-k},\frac{\rho}{c}V_{m-k}(\mu'),\mu'\right)(\cdot)$:

$$\begin{cases} L\left(V_{m-k},\frac{\rho}{c}V_{m-k}(\mu'),\mu'\right)(\mu')=G\left(V_{m-k},\frac{\rho}{c}V_{m-k}(\mu')\right)(\mu') \\ L\left(V_{m-k},\frac{\rho}{c}V_{m-k}(\mu'),\mu'\right)(\nu(\mu_0))\geqslant G\left(V_{m-k},\frac{\rho}{c}V_{m-k}(\mu')\right)(\nu(\mu_0)) \\ L\left(F_{m-k},\frac{\rho}{c}V_{m-k}(\mu'),\nu(\mu_0)\right)(\mu')>G\left(V_{m-k},\frac{\rho}{c}V_{m-k}(\mu')\right)(\mu') \\ L\left(F_{m-k},\frac{\rho}{c}V_{m-k}(\mu'),\nu(\mu_0)\right)(\nu(\mu_0))=G\left(V_{m-k},\frac{\rho}{c}V_{m-k}(\mu')\right)(\nu(\mu_0)) \end{cases}$$

$$\Longrightarrow\begin{cases} L\left(V_{m-k},\frac{\rho}{c}V_{m-k}(\mu'),\mu'\right)(\nu(\mu_0))\geqslant L\left(F_{m-k},\frac{\rho}{c}V_{m-k}(\mu'),\nu(\mu_0)\right)(\nu(\mu_0)) \\ L\left(V_{m-k},\frac{\rho}{c}V_{m-k}(\mu'),\mu'\right)(\mu')<L\left(F_{m-k},\frac{\rho}{c}V_{m-k}(\mu'),\nu(\mu_0)\right)(\mu') \end{cases}$$

The two equalities are directly from definition of $L$ and $G$. First inequality is from suboptimality, second inequality is from previous calculation. Therefore $L\left(V_{m-k},\frac{\rho}{c}V_{m-k}(\mu'),\mu'\right)(\cdot)$ is lower at $\mu'$ and $L\left(F_{m-k},\frac{\rho}{c}V_{m-k}(\mu'),\nu(\mu_0)\right)(\cdot)$ is lower at $\nu(\mu_0)$. Since both of them are linear fuinctions, then $L\left(V_{m-k},\frac{\rho}{c}V_{m-k}(\mu'),\mu'\right)(\cdot)$ must be higher at any $\mu''>\nu(\mu_0)$. Therefore, this implies:

$$L\left(V_{m-k},\frac{\rho}{c}V_{m-k}(\mu'),\mu'\right)(\mu'')>G\left(F_{m'},\frac{\rho}{c}V_{m-k}(\mu')\right)(\mu'')$$

Contradicting that $\mu''$ is superior than $\nu(\mu')$. *Q.E.D.*

**Lemma S.18′.** *Suppose at $\mu_0$, $V_0,V_0'$, $k\geqslant 1$ satisfies:*

$$\begin{cases} V_0=\max_{\nu\leqslant\mu_0}\frac{c}{\rho}\frac{F_{m-k}(\nu)-V_0-V_0'(\nu-\mu_0)}{J(\mu_0,\nu)}\leqslant\max_{\nu\geqslant\mu_0}\frac{c}{\rho}\frac{F_m(\nu)-V_0-V_0'(\nu-\mu_0)}{J(\mu_0,\nu)} \\ \overline{V}(\mu_0)\geqslant V_0\geqslant F_{m+k}(\mu_0) \end{cases}$$

*$V_{m+k}$ is the solution as defined in Lemma B.2 with initial condition $V_0,V_0'$, then $\forall\mu\in[\nu(\mu_0),\mu_0]$:*

$$V_{m+k}(\mu)\geqslant\max_{\nu\leqslant\mu,m'\in[m,m+k]}\frac{c}{\rho}\frac{F_{m'}(\nu)-V_{m-k}(\mu)-V_{m-k}'(\mu)(\nu-\mu)}{J(\mu,\nu)}$$

### S2.2 *Proof of Theorem 3*

**Proof.** In this part, we introduce the algorithm to construct $V(\mu)$ and $\nu(\mu)$. We only discuss the case $\mu\geqslant\mu^*$ and the case $\mu\leqslant\mu^*$ will follow by a symmetric method.

**Algorithm:**

- *Step 1*: Define:

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

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

  $\overline{V}^+$ is increasing and $\overline{V}^-$ is decreasing. There exists $\mu^*\in[0,1]$ s.t. $\overline{V}^+(\mu)\geqslant\overline{V}^-(\mu)$ when $\mu\geqslant\mu^*$ and $\overline{V}^-(\mu)\leqslant\overline{V}^-(\mu)$ when $\mu\leqslant\mu^*$ (See Lemma S.20). Define $\overline{V}(\mu)=\max\left\{\overline{V}^+(\mu),\overline{V}^-(\mu)\right\}$.

- *Step 2*: I construct the first piece of $V(\mu)$ to the right of $\mu^*$. By Lemma S.20, there are three possible cases of $\mu^*$ to discuss ($\mu^*=1$ is omitted by symmetry).

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

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

  With initial condition $(\mu_0=\mu^*,V_0=\overline{V}(\mu^*),V_0'=0)$, we solve for $V_m(\mu)$ on $[\mu^*,1]$ as defined by Lemma S.22. Define

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

  Be Lemma S.22, 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. Let $\hat{\mu}_m$ be the smallest $\mu \geqslant \mu^*$ that:

$$V_m(\mu) = \max_{\nu \geqslant \mu, I} \frac{I}{\rho} \frac{F_{m-1}(\nu) - V_m(\mu) - V_m'(\mu)(\nu-\mu)}{J(\mu,\nu)} - \frac{C(I)}{\rho}$$

If $V_m(\hat{\mu}_m) > F_{m-1}(\hat{\mu}_m)$ we solve for $V_{m-1}$ with initial condition $\mu_0 = \hat{\mu}_m, V_0 = V_m(\hat{\mu}_m), V_0' = V_m'(\hat{\mu}_m)$. Then redefine $V_{\mu^*}(\mu)|_{\mu \geqslant \hat{\mu}_m} = V_{m-1}(\mu)$. Otherwise skip to looking for $\hat{\mu}_{m-1}$. If $m-1 > \overline{m}$, we continue this procedure by looking for $\hat{\mu}_{m-1}$ until $m = \overline{m}$. Now suppose $V_{\overline{m}}$ first hits $F(\mu)$ at $\mu^{**} > \mu^*$. $V_{\mu^*}$ is a smooth function on $[\mu^*, \mu^{**}]$ such that:

$$V_{\mu^*}(\mu) = \begin{cases} F(\mu) \text{ if } \mu \leqslant \mu^* \text{ or } \mu \geqslant \mu^{**} \\ V_k(\mu) \text{ if } \mu \in [\hat{\mu}_k, \hat{\mu}_{k-1}]^2 \end{cases}$$

By construction, optimal posterior $\nu_{\mu^*}(\mu)$ is smoothly decreasing on each $(\hat{\mu}_{k+1}, \hat{\mu}_k)$ and jumps down at each $\hat{\mu}_k$. By Lemma S.23 and our construction, $\forall \mu \in [\mu^*, \mu^{**}]$:

$$V_{\mu^*}(\mu) = \max_{\nu \geqslant \mu, k, I} \frac{I}{\rho} \frac{F_k(\nu) - V_{\mu^*}(\mu) - V_{\mu^*}'(\mu)(\nu-\mu)}{J(\mu,\nu)} - \frac{C(I)}{\rho} \tag{S.18}$$

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

*Case 3*: Suppose $\mu^* = 0$, consider:

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

Define $\mu^{**} = \inf\left\{\mu | \widetilde{V}(\mu) > F_1(\mu)\right\} > 0$.

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

$$F(\mu^\diamond) = \max_{\nu \geqslant \mu, k} \frac{I}{\rho} \frac{F_k(\nu) - F(\mu^\diamond) - F'^-(\mu^\diamond)(\nu-\mu^\diamond)}{J(\mu^\diamond,\nu)} - \frac{C(I)}{\rho} \tag{S.19}$$

Let $m$ be the index of optimal action. Solve for $V_m$ with initial condition $\mu_0 = \mu^\diamond, V_0 = F(\mu^\diamond), V_0' = F'^-(\mu^\diamond)$. Then take same steps in *Step 3* and solve for $\hat{\mu}_k$ and $V_{k-1}$ sequentially until $V_{m_0}$ first hits $F$. This step refers to Figure 22-4,5. Now suppose $V_{m_0}$ first hits $F(\mu)$ at some point $\mu^{\diamond\diamond}$ (can potentially be $\mu$), define:

$$V_{\mu^\diamond}(\mu) = \begin{cases} F(\mu) & \text{if } \mu < \mu^\diamond \text{ or } \mu > \mu^{\diamond\diamond} \\ V_k(\mu) & \text{if } \mu \in [\hat{\mu}_{k+1}, \hat{\mu}_k]^3 \end{cases}$$

By Lemma S.21, $V_\mu$ is piecewise smooth are pasted smoothly. So $V_\mu$ is a smooth function on $[\mu, \mu'']$. Optimal posterior $\nu_{\mu^\diamond}(\mu)$ is smoothly decreasing on each $(\hat{\mu}_{k+1}, \hat{\mu}_k)$ and jumps down at each $\hat{\mu}_k$. By Lemma S.23 and our construction, $\forall \mu \in [\mu^\diamond, \mu^{\diamond\diamond}]$:

$$V_\mu(\mu) = \max_{\nu \geqslant \mu^\diamond, k} \frac{I}{\rho} \frac{F_k(\nu) - V_{\mu^\diamond}(\mu) - V_{\mu^\diamond}'(\mu)(\nu-\mu)}{J(\mu,\nu)} - \frac{C(I)}{\rho} \tag{S.18}$$

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

- *Step 4:* Define:

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

**Smoothness:**

I want to show that $V(\mu)$ is piecewisely defined as $V_{\mu_0}$'s. This is true when $\mu \leqslant \mu^{**}$ by definition of $V_{\mu^*}$. So I prove this for $\mu > \mu^{**}$. First it is easy to verify that Lemmas S.11, S.12 and S.13 still hold. The original proof directly applies by replacing Equation (32) with Equation (S.18) and Lemma B.2 with Lemma S.21.

**Lemma S.19.** *There exists $\Delta$ s.t. $\forall \mu_i \in \Omega$, on $\{\mu | V_{\mu_i}(\mu) > F(\mu)\}$, $V'(\mu)$ has Lipschitz parameter $\Delta$.*

**Proof.** $\forall \mu \in (\hat{\mu}_{k+1}, \hat{\mu}_k)$, $\nu$ is smooth in $\mu$ and $V_{\mu_i}' > 0$, by envelope theorem:

$$V_{\mu_i}'(\mu) = -\frac{I}{\rho} \frac{\nu - \mu}{J(\mu,\nu)} \left( V_{\mu_i}''(\mu) + C'(I) H''(\mu) \right) > 0$$
$$\implies V_{\mu_i}''(\mu) + C'(I) H''(\mu) < 0$$

[1]Define $\hat{\mu}_{m+1} = \mu^*$ and $\hat{\mu}_{\overline{m}} = \mu^{**}$ for consistency.
[3]Define $\hat{\mu}_{m+1} = \mu^\diamond$ and $\hat{\mu}_{m_0} = \mu^{\diamond\diamond}$ for consistency.

$C'(I)$ is bounded since $C(I)$ is bounded by $\sup F$. It is easy to see that $\sup\Omega < \underline{\mu}_n$ (where $n$ is the largest index). By Lemma S.11, there is $\mu_0 \in \Omega$ s.t. $V_{\mu_0}(\underline{\mu}_n) > F(\underline{\mu}_n)$. By Lemma S.12, $\sup\Omega = \sup\{\mu | V_{\mu_0}(\mu) > F(\mu)\} < \nu(\mu_0) < 1$. Therefore, $\mu$ is bounded away from 1. Then by Assumption 3, $-H''(\mu)$ is bounded above. Therefore, $\Delta$ exists for all such $\mu$.

Then consider $\mu = \widehat{\mu}_k$, since $V''_{\mu_i}$ is bounded on both side by $\Delta$, $V''_{\mu_i}(\mu) \leqslant \Delta$. Therefore at $\mu$ $V'_{\mu_i}$ has Lipschitz parameter $\Delta$ by Kirszbraun theorem. *Q.E.D.*

- *Step 1*: $V \in C[\mu^{**},1]$. Sort all rational numbers in $[\mu^{**},1]$ as $\{r_n\}$. $\forall N$, there exists $\mu_{n,M} \in \Omega$ s.t. $V(r_n) - V_{\mu_{n,M}}(r_n) \leqslant \frac{1}{N}$. Let $V_N = \max_n\{V_{\mu_{n,N}}\}$, then $\{V_N\} \subset \mathbb{V}$ and $V_N$ converges to $V$ pointwisely on $\{r_n\}$. Let $\widehat{V} = \lim V_n$, by Lemma S.13, $\widehat{V} \in C[\mu^{**},1]$. By definition $\widehat{V} \leqslant V$. Suppose $\widehat{V}(\mu) < V(\mu)$, then there exists $V_{\mu_0}(\mu) > \widehat{V}(\mu)$. Since both $V_{\mu_0}$ and $\widehat{V}$ are continuous, $V_{\mu_0} > \widehat{V}$ on an open interval, containing some $r_n$. Contradiction. So $\widehat{V} = V \in C[\mu^{**},1]$. Let $\{\mu \geqslant \mu^{**} | V(\mu) > F(\mu)\} = \bigcup I_m$ where $I_m$ are disjoint open intervals.

- *Step 2*: $\forall I_m$, exists $\mu_n \in \Omega$ s.t. $V(\mu) = \lim V_{\mu_n}(\mu)$ and $V'(\mu) = \lim V'_{\mu_n}(\mu)$ on $I_m$. Pick any $\mu \in I_m$. Let $\Theta(\mu) = \{\mu^\diamond \in \Omega | V_{\mu^\diamond}(\mu) > F(\mu)\}$. Then by definition of $V(\mu)$, $\Theta(\mu)$ is non-empty. Let $\widetilde{V} = \sup_{\mu^\diamond \in \Theta(\mu)} V_{\mu^\diamond}$. $\forall N$, there exists $\mu_{n,M} \in \Theta(\mu)$ s.t. $\widetilde{V}(r_n) - V_{\mu_{n,N}}(r_n) \leqslant \frac{1}{N}$. Since $V_{\mu_{n,N}}(\mu) > F(\mu)$, by Lemma S.12, there exists $V_{\mu_N} = \max\{V_{\mu_{n,N}}\}$. Therefore, $\lim V_{\mu_N} = \widetilde{V}$ on $\{r_n\}$. By Lemma S.13 $\widetilde{V} = \lim V_{\mu_N} \in C[\mu^{**},1]$. Now suppose $V(\mu) > \widetilde{V}(\mu)$, then there exists $V_{\mu^\diamond}(\mu) > V_{\mu_n}(\mu) > F(\mu)$. Then $\mu^\diamond \in \Theta(\mu)$ by Lemma S.12, contradiction. Therefore, $\lim V_{\mu_n} = V$ on $I_m$.

  Let $I_m = (a_m, b_m)$. Now consider $\{V'_{\mu_n}\}$. $V'_{\mu_n}(a_m) = F'(a_m)$. Lemma S.19 implies that $V'_{\mu_n}$ are totally bounded and equi-continuous on $I_m$. Therefore, there exists subsequence $V'_{\mu_n}$ being Cauchy w.r.t. sup norm on $[a_m, b_m]$. So $V$ as limit of $V_{\mu_n}$ is differentiable on $[a_m, b_m]$ and $V' = \lim V'_{\mu_n}$.

- *Step 3* $\forall I_m$, exists $\mu^m \in \Omega$ s.t. $V(\mu) = V_{\mu^m}$ on $I_m$. Let $\mu^m = \inf I_m$. By step 2, it is easy to verify that $\mu_n \to \mu^m$. Then since Equation (S.19) is continuous in $\mu$, it is satisfied at $\mu^m$ and $\mu^m \in \Omega$. Since both $V_{\mu_n}$ and $V'_{\mu_n}$ converges on $I_m$, Equation (S.18) is satisfied for $V$ on $I_m$. Let $F(\mu^m) = F_k(\mu^m)$.

  As an intermediate step, I first prove that Equation (S.18) is solved for $k' > k$ in a non-degenerate neighbour of $\mu^m$. Take any $\mu' > \mu^m$ s.t. $V(\mu') > F(\mu')$, since $V(\mu^m) = F_k(\mu^m)$, there exists $\mu^\star \in (\mu^m, \mu')$ and $\varepsilon > 0$ s.t $\forall \mu \in (\mu^m, \mu^\star)$ $V(\mu) - F_k(\mu) < V(\mu') - F_k(\mu') - \varepsilon$. I claim that Equation (S.18) is solved at all $\mu \in (\mu^m, \mu^\star)$ with $k' > k$ and $I$. Suppose not, then for $n$ sufficiently large:
  $$V_{\mu_n}(\mu) = \frac{I}{\rho}\frac{F_k(\nu) - V_{\mu_n}(\mu) - V'_{\mu_n}(\mu)(\nu - \mu)}{J(\mu,\nu)} - \frac{C(I)}{\rho} \leqslant \frac{I}{\rho}\frac{F_k(\nu) - F_k(\mu) - V'_{\mu_n}(\mu)(\nu - \mu)}{J(\mu,\nu)} = (F'_k - V'_{\mu_n}(\mu))\frac{\nu - \mu}{J(\mu,\nu)}$$
  Therefore $F'_k \geqslant V'_{\mu_n}(\mu)$. By construction of $V_{\mu_n}$ at any $\mu'' \geqslant \mu$ Equation (S.18) is solved with $k$, therefore $F'_k \geqslant V_{\mu_n}(\mu'')$ holds for all $\mu'' \geqslant \mu$. This implies $\forall \mu'' \geqslant \mu$, $V_{\mu_n}(\mu'') - F_k(\mu'') \leqslant V_{\mu_n}(\mu) - F_k(\mu) < V(\mu') - F_k(\mu') - \varepsilon$. Take $n \to \infty$ and $\mu'' = \mu'$, contradiction. Therefore, Equation (S.18) is solved at all $\mu \in (\mu^m, \mu^\star)$ for $V(\mu)$ with $k' > k$.

  Now consider $V_{\mu^m}(\mu)$. By my construction, suppose $V_{\mu^m}$ is updated up to action $k+1$. I claim that $V_{\mu^m} = V$ when $\mu \in [\mu^m, \mu^\star)$. Suppose not true, then there exists $\mu$ at which $V_{\mu^m}(\mu) < V(\mu)$, $V'_{\mu^m}(\mu) < V'(\mu)$. It is easy to verify that Equation (S.18) is violated at $V_{\mu_m}(\mu)$. Therefore, if $V_{\mu^m} \neq V$, it must happen in $(\mu^\star, b_m)$. Again we can find $\mu \in (\mu^\star, b_m)$ s.t. $V_{\mu^m}(\mu) < V(\mu)$, $V'_{\mu^m}(\mu) < V'(\mu)$, which is not possible. So $V(\mu) = V_{\mu^m}(\mu)$ on $I_m$.

To sum up, $V$ can be represented as:
$$V(\mu) = \begin{cases} V_{\mu^*}(\mu) & \text{if } \mu \in [\mu^*, \mu^{**}] \\ V_{\mu^m}(\mu) & \text{if } \mu \in I^m \\ F(\mu) & \text{otherwise} \end{cases}$$
Now I prove smoothness of $V(\mu)$ on $[\mu^*,1]$. By Lemma S.19 $|V'(\mu) - F'(\mu)|$ is bounded by $\Delta|I_n|$. Define:
$$V_n(\mu) = \begin{cases} V_{\mu^m}(\mu) & \text{when } \mu \in I_m, m \leqslant n \\ F(\mu) & \text{otherwise} \end{cases}$$
Then $V_n(\mu) \to V(\mu)$. By Lemma S.11, we can without loss assume first $n$ $V_{\mu^m}$ have $I_m$ covering $\underline{\mu}_m$. Fix $n$, $\forall \mu$, $\forall m \geqslant n$, if $\mu \in I_m$ and $m \leqslant n$ or $\mu \notin \bigcup I_m$, then $V'_n(\mu) = V'_m(\mu)$, else if $\mu \in I_m$, $m > n$, then $|V'_n(\mu) - F'(\mu)|$ and $|V'_m(\mu) - F'(\mu)|$ are all bounded by $\Delta|I_m|$. Therefore, $V'_n(\mu)$ is a Cauchy sequence. Then $V'_n(\mu) \to V'(\mu)$ pointwise. Since each $V'_n$ is continuous, $V$ is a smooth function on $[0,1]$ and $V' = F'$ when $V = F$.

**Unimprovability**

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

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

$$V(\mu)=\max\left\{F(\mu),\max_{\nu,m,I}\frac{I}{\rho}\frac{F_m(\nu)-V(\mu)-V'(\mu)(\nu-\mu)}{J(\mu,\nu)}-\frac{C(I)}{\rho}\right\} \tag{P-C1}$$

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

We still focus on the case $\mu\geqslant\mu^*$. For the case $\mu\leqslant\mu^*$, a totally symmetric argument applies by referring to Lemma S.23′. Equation (P-C1) is implied by Equation (S.18) for $\mu\in E$. So it is sufficient to prove Equation (P-C1) for $\mu\in E^C$. Suppose there exists $\mu\geqslant\mu^*$ s.t. Equation (P-C) is violated. Let $F(\mu)=F_k(\mu)$. Then without loss we can assume that:

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

By Lemma S.11, there exists $I_k$ s.t. $\underline{\mu}_k\in I_k$. At $b_k=\sup I_k$, $U(b_k)\leqslant F_k(b_k)$. Therefore, since $U(\mu)$ is continuous there exists largest $\mu'<\mu$ s.t .$U(\mu')=F_k(\mu')$. Then Equation (S.19) is satisfied at $\mu'$ so consider $V_{\mu'}$. Sicne $V_{\mu'}(\mu)\leqslant V(\mu)=F_k(\mu)$, there exists $\mu''\in(\mu',\mu)$ s.t. $V_{\mu'}(\mu'')\leqslant F_k(\mu)$ and $V'_{\mu'}(\mu'')\leqslant F_k(\mu)$. Therefore $U(\mu'')>F_k(\mu'')$ implies $V_{\mu'}(\mu'')>F_k(\mu'')$, contradiction. Apply a symmetric argument to $\mu\leqslant\mu^*$, I proved Equation (P-C).

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

$$V(\mu)=\max\left\{F(\mu),\max_{\nu,I}\frac{I}{\rho}\frac{V(\nu)-V(\mu)-V'(\mu)(\nu-\mu)}{J(\mu,\nu)}-\frac{C(I)}{\rho}\right\} \tag{P-D1}$$

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

Suppose not, then there exists:

$$\begin{aligned}\widetilde{V}=&\max_{\nu\geqslant\mu,I}\frac{I}{\rho}\frac{F_m(\nu)-V(\mu)-V'(\mu)(\nu-\mu)}{J(\mu,\nu)}-\frac{C(I)}{\rho}\\ \leqslant V(\mu)<&\frac{I''}{\rho}\frac{V(\mu'')-V(\mu)-V'(\mu)(\mu''-\mu)}{J(\mu,\mu'')}-\frac{C(I'')}{\rho}\end{aligned}$$

Suppose the optimizer is $\nu,m,I$. Optimality implies Equation (S.22):

$$\frac{V(\nu)-V(\mu)-V'(\mu)(\nu-\mu)}{J(\mu,\nu)}=C'(I')$$

Together with Equation (S.20), we have $I'C'(I')=\rho\widetilde{V}+C(I')$. Then combine with Equation (S.21), we get:

$$\begin{cases}F'_m+C'(I)H'(\nu)=V'(\mu)+C'(I')H'(\mu)\\ (F_m(\nu)+C'(I')H(\nu))-(V(\mu)+C'(I')H(\mu))=(V'(\mu)+C'(I')H'(\mu))(\nu-\mu)\end{cases}$$

We define $L$ and $G$ as in Theorem 2. Then $L$ will be linear and $G(F_m,C'(I'))(\nu)$ will be a concave function of $\nu$. Consider:

$$L(V,C'(I'),\mu)(\nu)-G(F_m,C'(I'))$$

FOC implies that it will be convex and attains minimum 0 at $\nu$. For any $m'$ other than $m$,

$$L(V,C'(I'))(\nu)-G(F_{m'},C'(I'))(\nu)$$

will be convex and weakly larger than zero. However:

$$\begin{aligned}&L(V,C'(I'),\mu)(\mu'')-G(V,C'(I'))(\mu'')\\ =&-(V(\mu'')-V(\mu)-V'(\mu)(\mu''-\mu)-C'(I')J(\mu,\mu''))\\ <&0\end{aligned}$$

The inequality is from definition of $I'$:

$$\begin{aligned}&I'C'(I')-C(I')<I''C'(I'')-C(I'')\\ \Longrightarrow&C'(I')<C'(I'')\\ \Longrightarrow&C'(I')<\frac{V(\mu'')-V(\mu)-V'(\mu)(\mu''-\mu)}{J(\mu,\mu'')}\end{aligned}$$

Therefore, $L(V,C'(I'),\mu)(\cdot)-G(V,C'(I'))(\cdot)$ will have a strictly negative minimum. Suppose it's minimized at $\widetilde{\mu}$, Then FOC implies:

$$V'(\mu)+C'(I')H'(\mu)=V'(\widetilde{\mu})+C'(I')H'(\widetilde{\mu})$$

Consider:

$$\begin{aligned}
&L(V,C'(I'),\widetilde{\mu})(\nu(\widetilde{\mu}))-G(F_m,C'(I'))(\widetilde{\nu})\\
=&L(V,C'(I'),\mu)(\nu(\widetilde{\mu}))-G(F_m,C'(I'))(\nu(\widetilde{\mu}))\\
&+V(\widetilde{\mu})-V(\mu)+C'(I')(H(\widetilde{\mu})-H(\mu))-(V'(\mu)+C'(I')H(\mu))(\widetilde{\mu}-\mu)\\
\geqslant&V(\widetilde{\mu})-V(\mu)+C'(I')(H(\widetilde{\mu})-H(\mu))-(V'(\mu)+C'(I')H'(\mu))(\widetilde{\mu}-\mu)\\
=&G(V,C'(I'))(\widetilde{\mu})-L(V,C'(I'),\mu)(\widetilde{\mu})\\
>&0
\end{aligned}$$

Let $m',\nu(\widetilde{\mu}),\widetilde{I}$ be maximizer at $\widetilde{\mu}$, $\widetilde{I}C'(\widetilde{I})=\rho V(\widetilde{\mu})+C(\widetilde{I})$:

$$\begin{aligned}
0=&L(V,C'(\widetilde{I}),\widetilde{\mu})(\nu(\widetilde{\mu}))-G(F_{m'},C'(\widetilde{I}))(\nu(\widetilde{\mu}))\\
=&L(V,C'(I'),\widetilde{\mu})(\nu(\widetilde{\mu}))-G(F_{m'},C'(I'))(\nu(\widetilde{\mu}))\\
&+(C'(\widetilde{I})-C'(I'))J(\widetilde{\mu},\nu(\widetilde{\mu}))\\
>&(C'(\widetilde{I})-C'(I'))J(\widetilde{\mu},\nu(\widetilde{\mu}))
\end{aligned}$$

Since $\widetilde{\mu}>\mu$, we have $C'(\widetilde{I})-C'(I')>0$. Contradiction. Therefore we proved Equation (P-D1).

- *Step 3*: We show that $V$ satisfies Equation (4). First, since $V$ is smooth, envelope theorem implies:

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

Therefore, allocating to diffusion experiment will always be suboptimal. What's more, consider:

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

$V^-(\mu^*)=V(\mu^*)$ and whenever $V(\mu)=V^-(\mu)$, we will have $V^{-\prime}(\mu)<0$. Therefore, $V^-(\mu)$ can never cross from below, that is to say:

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

*Q.E.D.*

**Lemma S.20.** *Define $\overline{V}^+$ and $\overline{V}^-$ :*

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

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

**Proof.** We define function $U_m^+$ and $U_m^-$ as following:

$$\begin{aligned}
\overline{U}_m^+(\mu)=&\max_{\nu\geqslant\mu,I}\frac{IF_m(\nu)-C(I)J(\mu,\nu)}{I+\rho J(\mu,\nu)}\\
\overline{U}_m^-(\mu)=&\max_{\nu\leqslant\mu,I}\frac{IF_m(\nu)-C(I)J(\mu,\nu)}{I+\rho J(\mu,\nu)}
\end{aligned}$$

Since $C(I)$,$F_m(\mu)$ and $J(\mu,\nu)$ are all smooth functions, the objective function will be smooth. First consider FOCs and SOCs:

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

$$\text{FOC-}I: \rho F_m(\nu)+C(I)-C'(I)(I+\rho J(\mu,\nu))=0$$

$$\text{SOC}: \quad H=\begin{bmatrix} I(\rho F_m(\nu)+C(I))(I+\rho J(\mu,\nu))H''(\nu) & 0 \\ 0 & -J(\mu,\nu)(I+\rho J(\mu,\nu))^2C''(I)\end{bmatrix}$$

Noticing that SOC is evaluated at the pairs $(\nu,I)$ at which FOC holds.

*Remark.* Details of calculation of second derivatives:

- $H_{\nu,\nu}$:

$$\begin{aligned}
&\frac{\partial^2}{\partial\nu^2}\frac{IF_m(\nu)-C(I)J(\mu,\nu)}{I+\rho J(\mu,\nu)}\\
&=\frac{1}{(I+\rho J(\mu,\nu))^3}\Big[2\rho^2(IF_m(\nu)-C(I)J(\mu,\nu))(H'(\mu)-H'(\nu))^2\\
&-2\rho(I+\rho J(\mu,\nu))(H'(\mu)-H'(\nu))(IF_m'-C(I)(H'(\mu)-H'(\nu)))\\
&+\rho(I+\rho J(\mu,\nu))(IF_m(\nu)-C(I)J(\mu,\nu))H''(\nu)\\
&+(I+\rho J(\mu,\nu))^2C(I)H''(\nu)\Big]
\end{aligned}$$

$$\text{FOC-}\nu \implies F_m'=\frac{(C(I)+\rho F_m(\nu))(H'(\mu)-H'(\nu))}{I+\rho J(\mu,\nu)}$$

$$\begin{aligned}
&\implies \frac{\partial^2}{\partial\nu^2}\frac{IF_m(\nu)-C(I)J(\mu,\nu)}{I+\rho J(\mu,\nu)}\\
&=\frac{1}{(I+\rho J(\mu,\nu))^3}\Big[2\rho^2(IF_m(\nu)-C(I)J(\mu,\nu))(H'(\mu)-H'(\nu))^2\\
&+2\rho(I+\rho J(\mu,\nu))C(I)(H'(\mu)-H'(\nu))^2\\
&-2\rho(C(I)+\rho F_m(\nu))(H'(\mu)-H'(\nu))^2\\
&+\rho(I+\rho J(\mu,\nu))(IF_m(\nu)-C(I)J(\mu,\nu))H''(\nu)\\
&+(I+\rho J(\mu,\nu))^2C(I)H''(\nu)\Big]\\
&=\frac{1}{(I+\rho J(\mu,\nu))^3}\Big[\rho(I+\rho J(\mu,\nu))(IF_m(\nu)-C(I)J(\mu,\nu))H''(\nu)\\
&+(I+\rho J(\mu,\nu))^2C(I)H''(\nu)\Big]\\
&=(I+\rho J(\mu,\nu))H''(\nu)(\rho IF_m(\nu)-\rho C(I)J(\mu,\nu)+IC(I)+\rho C(I)J(\mu,\nu))\\
&=\frac{I(\rho F_m(\nu)+C(I))(I+\rho J(\mu,\nu))H''(\nu)}{(I+\rho J(\mu,\nu))^3}
\end{aligned}$$

- $H_{I,I}$:

$$\begin{aligned}
&\frac{\partial^2}{\partial I^2}\frac{IF_m(\nu)-C(I)J(\mu,\nu)}{I+\rho J(\mu,\nu)}\\
&=\frac{1}{(I+\rho J(\mu,\nu))^3}\Big[2(IF_m(\nu)-C(I)J(\mu,\nu))\\
&-2(I+\rho J(\mu,\nu))(F_m(\nu)-C'(I)J(\mu,\nu))\\
&-J(\mu,\nu)(I+\rho J(\mu,\nu))^2C''(I)\Big]
\end{aligned}$$

$$\text{FOC-}I \implies IF_m(\nu)-C(I)J(\mu,\nu)=(I+\rho J(\mu,\nu))(F_m(\nu)-C'(I)J(\mu,\nu))$$

$$\begin{aligned}
&\implies \frac{\partial^2}{\partial I^2}\frac{IF_m(\nu)-C(I)J(\mu,\nu)}{I+\rho J(\mu,\nu)}\\
&=\frac{1}{(I+\rho J(\mu,\nu))^3}\Big[2(I+\rho J(\mu,\nu))(F_m(\nu)-C'(I)J(\mu,\nu))\\
&-2(I+\rho J(\mu,\nu))(F_m(\nu)-C'(I)J(\mu,\nu))\\
&-J(\mu,\nu)(I+\rho J(\mu,\nu))^2C''(I)\Big]\\
&=\frac{-J(\mu,\nu)(I+\rho J(\mu,\nu))^2C''(I)}{(I+\rho J(\mu,\nu))^3}
\end{aligned}$$

- $H_{\nu,I}$:

$$\begin{aligned}
&\frac{\partial^2}{\partial I\partial \nu}\frac{IF_m(\nu)-C(I)J(\mu,\nu)}{I+\rho J(\mu,\nu)}\\
=&\frac{1}{(I+\rho J(\mu,\nu))^3}\Big[2\rho(IF_m(\nu)-C(I)J(\mu,\nu))(H'(\mu)-H'(\nu))\\
&-\rho(I+\rho J(\mu,\nu))(F_m(\nu)-C'(I)J(\mu,\nu))(H'(\mu)-H'(\nu))\\
&-(I+\rho J(\mu,\nu))(IF_m'-C(I)(H'(\mu-H'(\nu))))\\
&+(I+\rho J(\mu,\nu))^2(F_m'-C'(I)(H'(\mu)-H'(\nu)))\Big]\\
=&\frac{1}{(I+\rho J(\mu,\nu))^3}\Big[2\rho(IF_m(\nu)-C(I)J(\mu,\nu))(H'(\mu)-H'(\nu))\\
&-\rho(IF_m(\nu)-C(I)J(\mu,\nu))(H'(\mu)-H'(\nu))\\
&-(I+\rho J(\mu,\nu))(I\frac{(C(I)+\rho F_m(\nu))(H'(\mu)-H'(\nu))}{I+\rho J(\mu,\nu)}C(I)(H'(\mu-H'(\nu))))\\
&+(I+\rho J(\mu,\nu))^2\left(\frac{(C(I)+\rho F_m(\nu))(H'(\mu)-H'(\nu))}{I+\rho J(\mu,\nu)}-C'(I)(H'(\mu)-H'(\nu))\right)\Big]\\
=&\frac{H'(\mu)-H'(\nu)}{(I+\rho J(\mu,\nu))^3}(\rho IF_m(\nu)-\rho C(I)J(\mu,\nu)-I(C(I)+\rho F_m(\nu))\\
&+(I+\rho J(\mu,\nu))C(I)+(I+\rho J(\mu,\nu)(C(I)+\rho F_m(\nu)))-(I+\rho J(\mu,\nu))^2C'(I))\\
=&0
\end{aligned}$$

The only term we don't know its sign is

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

Therefore, $H$ will be ND if $\nu>\mu$ and $F_m'>0$, or $\nu<\mu$ and $F_m'<0$. In these cases, FOC uniquely characterizes the maximum. Suppose $\nu>\mu$ and $F_m'<0$ or $\nu<\mu$ and $F_m'>0$, the $H$ will never be ND, and choice of $\nu$ will be on boundary. What's more, simple calculation shows that choosing $\nu=\mu$ will dominate choosing $\nu=0,1$. Therefore:

$$\overline{U}_m^+(\mu)=F_m(\mu) \text{ when } F_m'<0$$
$$\overline{U}_m^-(\mu)=F_m(\mu) \text{ when } F_m'>0$$

When $F_m'>0$, envelope condition implies:

$$\frac{d}{d\mu}\overline{U}_m^+(\mu)=\frac{-H''(\mu)(\nu-\mu)\left(C(I)+\frac{\rho}{I}F_m(\nu)\right)}{\left(1+\frac{\rho}{I}J(\mu,\nu)\right)^2}>0$$

Similarly, when $F_m'<0$, envelope condition implies:

$$\frac{d}{d\mu}\overline{U}_m^-(\mu)=\frac{-H''(\mu)(\nu-\mu)\left(C(I)+\frac{\rho}{I}F_m(\nu)\right)}{\left(1+\frac{\rho}{I}J(\mu,\nu)\right)^2}<0$$

Therefore, $\overline{U}_m^+$ and $\overline{U}_m^-$ have exactly the same properties as in Lemma B.1, the rest of proofs simply follow Lemma B.1. What's more, we define $\nu_m^*$ and $I_m^*$ as the maximizer in this problem.

Now I prove that $\overline{V}(\mu^*)>0$. We know that $\overline{V}(\mu^*)$ solves:

$$\overline{V}(\mu^*)=\max_{\nu\geqslant\mu^*,I}\frac{F(\nu)-\frac{C(I)}{I}J(\mu^*,\nu)}{1+\frac{\rho}{I}J(\mu^*,\nu)}=\max_{\nu\leqslant\mu^*,I}\frac{F(\nu)-\frac{C(I)}{I}J(\mu^*,\nu)}{1+\frac{\rho}{I}J(\mu^*,\nu)}$$

Consider the following term:

$$\underline{V}=\max_{\mu_i,p_i,I}\frac{p_1F(\mu_1)+p_2F(\mu_2)-\frac{C(I)}{I}I(\mu_i|\mu^*)}{1+\frac{\rho}{I}I(\mu_i|\mu^*)}$$

Suppose $\mu_i,p_i,I$ solves $\underline{V}$. Then:

$$\frac{\rho}{I}\underline{V}+\frac{C(I)}{I}=\frac{p_1F(\mu_1)+p_2F(\mu_2)-\underline{V}}{-p_1H(\mu_1)-p_2H(\mu_2)+H(\mu^*)}$$

I want to claim that $\underline{V}\leqslant\overline{V}(\mu^*)$. Suppose not, then:

$$\frac{p_1F(\mu_1)+p_2F(\mu_2)-\overline{V}(\mu^*)}{-p_1H(\mu_1)-p_2H(\mu_2)+H(\mu^*)}>\frac{p_1F(\mu_1)+p_2F(\mu_2)-\underline{V}}{-p_1H(\mu_1)-p_2H(\mu_2)+H(\mu^*)}$$
$$\geqslant\frac{\rho}{I}\overline{V}(\mu^*)+\frac{C(I)}{I}$$

Then at least one of the following:

$$\frac{F(\mu_1)-\overline{V}(\mu^*)}{-H(\mu_1)+H(\mu^*)+H'(\mu^*)(\mu_1-\mu^*)};\frac{F(\mu_2)-\overline{V}(\mu^*)}{-H(\mu_2)+H(\mu^*)+H'(\mu^*)(\mu_2-\mu^*)}$$

is larger than $\frac{\rho}{I}\overline{V}(\mu^*)+\frac{C(I)}{I}$. Suppose the fisrt term does, then:

$$\rho\overline{V}(\mu^*)<I\frac{F(\mu_1)-\overline{V}(\mu^*)}{J(\mu^*,\mu_1)}-C(I)$$

Contradicting optimality of $\overline{V}(\mu^*)$. Same argument applies to the second term. So $\overline{V}(\mu^*)\geqslant\underline{V}$. However:

$$\lim_{c\to0}p_1F(\mu_1)+p_2F(\mu_2)-\frac{C(I)}{I}I(\mu_i|\mu^*)=p_1F(\mu_1)+p_2F(\mu_2)-C'(0)I(\mu_i|\mu)>0$$

Therefore, $\overline{V}(\mu^*)\geqslant\underline{V}>0$. *Q.E.D.*

**Lemma S.21.** *Assume $\mu_0\geqslant\mu^*$, $F_m'\geqslant0$, $V_0,V_0'$ satisfies:*

$$\begin{cases}\overline{V}(\mu_0)\geqslant V_0>F_m(\mu_0)\\ V_0=\max_{\nu\geqslant\mu_0,I}\frac{I}{\rho}\frac{F_m(\nu)-V_0-V_0'(\nu-\mu)}{J(\mu,\nu)}-\frac{C(I)}{\rho}\end{cases}$$

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

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

*and initial condition $V(\mu_0)=V_0$, $V'(\mu_0)=V_0'$.*

**Proof.** We start from deriving FOC and SOC for Equation (39-c):

$$\text{FOC-}\nu\text{: }\frac{I}{\rho}\left(\frac{F_m'-V'(\mu)}{J(\mu,\nu)}+\frac{F_m(\nu)-V(\mu)-V'(\mu)(\nu-\mu)}{J(\mu,\nu)^2}(H'(\nu)-H'(\mu))\right)=0$$
$$\text{FOC-}I\text{: }\frac{1}{\rho}\left(\frac{F_m(\nu)-V(\mu)-V'(\mu)(\nu-\mu)}{J(\mu,\nu)}-C'(I)\right)=0$$
$$\text{SOC: }\quad H=\begin{bmatrix}\frac{-2(H'(\mu)-H'(\nu))(\text{FOC-}\nu)}{J(\mu,\nu)}+\frac{I}{\rho}\frac{(F_m(\nu)-V(\mu)-V'(\mu)(\nu-\mu))H''(\nu)}{J(\mu,\nu)^2} & \frac{1}{I}\text{FOC-}\nu\\ \frac{1}{I}\text{FOC-}\nu & -\frac{C''(I)}{\rho}\end{bmatrix}$$

Noticing that $H_{I,I}<0$, therefore $I$ satisfying FOC will be unique given $\mu,\nu$. On the other hand, FOC-$\nu$ is independent of $I$. $H_{\nu,\nu}<0$ when FOC-$\nu\geqslant0$. Therefore, solution of F)C-$\nu$ will be unique. When FOCs are satisfied, $H$ is strictly ND, then the solution of FOCs are going to be maximizer. Therefore, FOC-$\nu$ and FOC-$I$ uniquely characterize optimal choice of $\nu,I$. Now we impose feasibility:

$$V(\mu)=\frac{I}{\rho}\frac{F_m(\nu)-V(\mu)-V'(\mu)(\nu-\mu)}{J(\mu,\nu)}-\frac{C(I)}{\rho}\tag{S.20}$$

FOCs reduces to:

$$\text{FOC-}\nu\text{:}(F_m'-V'(\mu))+\frac{\rho V(\mu)+C(I)}{I}(H'(\nu)-H'(\mu))=0\tag{S.21}$$

$$\text{FOC-}I\text{:}\, IC'(I)=\rho V(\mu)+C(I) \tag{S.22}$$

Differentiate FOC-$I$, we get:

$$\begin{cases} V(\mu)=\dfrac{IC'(I)-C(I)}{\rho} \\ V'(\mu)=\dfrac{IC''(I)}{\rho}\dot{I} \end{cases} \tag{S.23}$$

Plug Equation (S.23) into Equation (S.21) and Equation (S.20):

$$\begin{cases} \dot{I}=\dfrac{\rho}{IC''(I)}(F_m'+C'(I)(H'(\nu)-H'(\mu))) \\ J(\nu,\mu)=\dfrac{1}{\rho}\left(I-\dfrac{C(I)+\rho F_m(\mu)}{C'(I)}\right) \end{cases} \tag{S.24}$$

We obtained an equation system with one ODE of $(c,\dot{I})$ and one regular equation for $\nu$. Since $J(\nu,\mu)$ is strictly monotonic for $\nu\geqslant\mu$, we can also define an implicit inverse function $M$ to eliminate $\nu$ in the equation.

$$J(M(y,\mu),\mu)=y$$

Therefore we get an ODE:

$$\dot{I}=\frac{\rho}{IC''(I)}\left(F_m'+C'(I)\left(H'\left(M\left(\frac{1}{\rho}\left(I-\frac{C(I)+\rho F_m(\mu)}{C'(I)}\right)\right),\mu\right)-H'(\mu)\right)\right) \tag{S.25}$$

We define $\underline{I}_m(\mu_0)C'(\underline{I}_m(\mu_0))-C(\underline{I}_m(\mu_0))=\rho F_m(\mu)$ when this equation has solution and $\underline{I}_m(\mu)=0$ when $\rho F_m(\mu)$ is so small that this equation has no solution. Since $F_m(\mu)$ is increasing in $\mu$, $\underline{I}_m(\mu)$ is increasing and strictly increasing when $\underline{I}_m(\mu)>0$. We consider the initial conditions:

$$F_m(\mu_0)<V_0=\frac{I_0C'(I_0)-C(I_0)}{\rho}\leqslant\overline{V}(\mu_0)$$

$$\Longrightarrow \underline{I}_m(\mu_0)<I_0\leqslant I_m^*(\mu_0)$$

Then Lemma S.22 guaranteed the existence of an increasing function $I(\mu)$ on $[\mu_0,1]$. *Q.E.D.*

**Lemma S.22.** *Define $M$ as $J(M(y,\mu),\mu)=y$. Assume $\mu_0\in[\mu^*,1)$, $I_0$ satisfies:*

$$\underline{I}_m(\mu_0)<I_0\leqslant I_m^*(\mu_0)$$

*Then there exists a $C^{(1)}$ and strictly increasing $I$ on $[\mu_0,1]$ satisfying initial condition $I(\mu_0)=I_0$. On $\{\mu|I(\mu)>\underline{I}_m(\mu)\}$, $I$ solves:*

$$\dot{I}=\frac{\rho}{IC''(I)}\left(F_m'+C'(I)\left(H'\left(M\left(\frac{1}{\rho}\left(I-\frac{C(I)+\rho F_m(\mu)}{C'(I)}\right)\right),\mu\right)-H'(\mu)\right)\right) \tag{S.25}$$

**Proof.** We first characterize some useful properties of the ODE. We denote the ODE by $\dot{I}=R(\mu,I)$.

- *Domain*: By definition of $\underline{I}_m(\mu)$, $\forall\mu\in(0,1)$

  $$\underline{I}_m(\mu)-\frac{C(\underline{I}_m(\mu))+\rho F_m(\mu)}{C'(\underline{I}_m(\mu))}=0$$

  Since $\underline{I}_m\geqslant 0$, then $C(\underline{I}_m(\mu))+\rho F_m(\mu)\geqslant 0$. Therefore at $I=\underline{I}_m(\mu)$:

  $$\frac{\partial}{\partial I}\left(I-\frac{C(I)+\rho F_m(\mu)}{C'(I)}\right)=\frac{C(I)+\rho F_m(\mu)}{C'(I)^2}C''(I)>0$$

  Therefore, $\forall I\geqslant\underline{I}_m(\mu)$, $I-\frac{C(I)+\rho F_m(\mu)}{C'(I)}\geqslant 0$. Strictly inequality holds when $I>\underline{I}_m(\mu)$. On the other hand, when $I<\underline{I}_m(\mu)$, if $F_m(\mu)\geqslant 0$, then $I-\frac{C(I)+\rho F_m(\mu)}{C'(I)}<0$. Else if $F_m(\mu)\leqslant 0$, then $\underline{I}_m(\mu)=0$. Since $M$ only applies to non-negative reals, we know that the ODE is only well defined in the region: $\{I|I\geqslant\underline{I}_m(\mu)\}$.

- *Continuity*: It is not hard to verify that the ODE is well behaved (satisfying Picard-Lindelof) when $\mu$ is strictly bounded away from $\{0,1\}$, $I$ is uniformly bounded away from $\underline{I}_m(\mu)$. One just need to calculate $M_y(y,\mu)$ by implied function theorem:

  $$\frac{\partial}{\partial y}M_y(y,\mu)=-\frac{1}{H''(M(y,\mu))(M(y,\mu)-\mu)}$$

  $M(y,\mu)=\mu$ implies $J(\nu,\mu)=0$, implies $\frac{1}{\rho}\left(I-\frac{C(I)+\rho F_m(\mu)}{C'(I)}\right)=0$. Since $I$ is uniformly bounded away from $\underline{I}_m(\mu)$, then $M(y,\mu)-\mu$ is uniformly bounded away from 0.

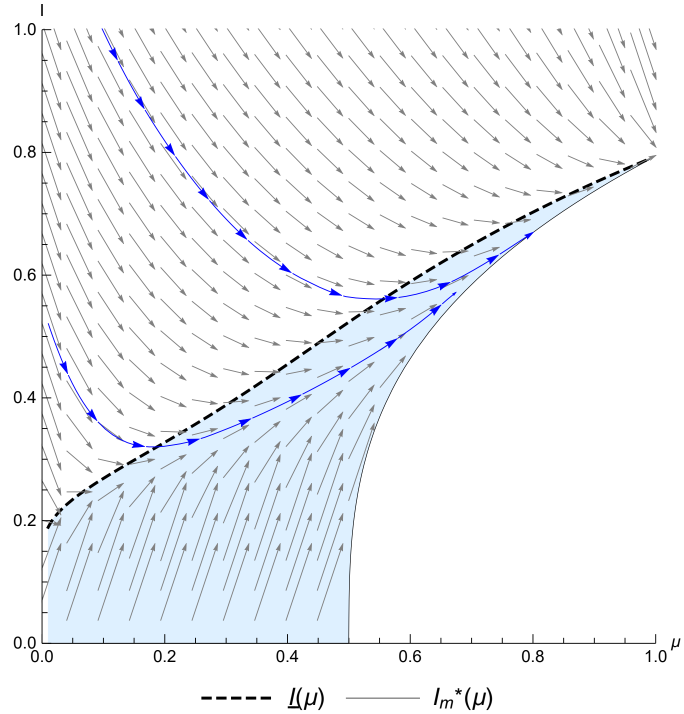



Figure S.2: Phase diagram of $\left(\dot{\mu},\dot{I}\right)$.

- *Monotonicity*: When $I = I_m^*(\mu)$, $\dot{I} = 0$. This can be shown by considering FOC on $I_m^*$:

$$\begin{cases} F_m' - C'(I)(H'(\mu) - H'(\nu)) = 0 \\ (I + \rho J(\mu,\nu))C'(I) = C(I) + \rho F_m(\nu) \end{cases}$$
$$\Longrightarrow (I - \rho J(\nu,\mu))C'(I) = C(I) + \rho F_m(\mu) + \rho F_m'(\nu-\mu) + C'(I)(H'(\nu) - H'(\mu))(\nu - \mu)$$
$$\Longrightarrow (I - \rho J(\nu,\mu))C'(I) = C(I) + \rho F_m(\mu)$$
$$\Longrightarrow F_m' + C'(I)\left(H'\left(M\left(\frac{1}{\rho}\left(I - \frac{C(I)+\rho F_m(\mu)}{C'(I)}\right),\mu\right)\right) - H'(\mu)\right) = 0$$
$$\Longrightarrow \dot{I} = R(\mu,I) = 0$$

Then we consider the monotonicity of $R(\mu,I)$:

$$\frac{\partial}{\partial I}R(\mu,I) = C''(I)(H'(M) - H'(\mu)) + C'(I)\frac{H''(M)}{H''(M)(\mu - M)}\frac{1}{\rho}\frac{C(I)+\rho F_m(\mu)}{C'(I)^2}C''(I) < 0$$

Therefore, $R(\mu,I)$ will be positive in $\{\underline{I}_m(\mu) < I \leqslant I_m^*(\mu)\}$. This refers to the blue region in Figure S.2.

$\forall \delta > 0$, we consider solving the ODE $\dot{I} = R(\mu,I)$ in region: $\mu \in [\delta, 1-\delta], I \in [\underline{I}_m(\mu) + \delta, I_m^*(\mu)]$. The initial condition $(\mu_0, I_0)$ is in the blue region of Figure S.2. (When $H'$ is finite, we can take $\mu \in [0,1]$.) Picard-Lindelof guarantees a unique solution satisfying the ODE in the region. What's more, it's straight forward that the solution $I(\mu)$ will be increasing. A solution is a blue line with arrows in Figure S.2. A solution $I(\mu)$ will lie between $\underline{I}_m(\mu)$ and $I_m^*(\mu)$ until it hits the boundary of region.

Now we can take $\delta \to 0$ and extend $I(\mu)$ towards the boundary. Since the end point of $I(\mu)$ has both $\mu, I$ monotonically increasing, there is a limit $\bar{I}, \bar{\mu}$ with $\underline{I}_m(\bar{\mu}) = \bar{I}$. Then since $R(\mu,I)$ has a limit $\frac{\rho F_m'}{\bar{I}h''(\bar{I})}$, we actually have $\lim_{\mu \to \bar{\mu}} V'(\mu) = F_m'$ by Equation (S.23). So the resulting $V(\mu)$ calculated from

$$V(\mu) = \frac{I(\mu)C'(I(\mu)) - C(I(\mu))}{\rho}$$

will be smooth on $[\mu_0, 1]$. *Q.E.D.*

**Lemma S.21′.** *Assume $\mu_0 \leqslant \mu^*$, $F_m' \geqslant 0$, $V_0, V_0'$ satisfies:*

$$\begin{cases} \overline{V}(\mu_0) \geqslant V_0 > F_m(\mu_0) \\ V_0 = \max_{\nu \leqslant \mu_0, I} \frac{I}{\rho} \frac{F_m(\nu) - V_0 - V_0'(\nu - \mu)}{J(\mu,\nu)} - \frac{C(I)}{\rho} \end{cases}$$

*Then there exists a $C^{(1)}$ smooth and strictly decreasing $V(\mu)$ defined on $[0,\mu_0]$ satisfying:*

$$V(\mu) = \max_{\nu \leqslant \mu, I} \frac{I}{\rho} \frac{F_m(\nu) - V(\mu) - V'(\mu)(\nu - \mu)}{J(\mu,\nu)} - \frac{C(I)}{\rho} \tag{39-I′}$$

*and initial condition $V(\mu_0) = V_0$, $V'(\mu_0) = V_0'$.*

**Lemma S.22′.** *Define $M$ as $J(M(y,\mu),\mu) = y$. Assume $\mu_0 \in (0,\mu^*]$, $I_0$ satisfies:*

$$\underline{I}_m(\mu_0) < I_0 \leqslant I_m^*(\mu_0)$$

*Then there exists a $C^{(1)}$ and strictly decreasing $I$ on $[0,\mu_0]$ satisfying initial condition $c(\mu_0) = I_0$. On $\{\mu | I(\mu) > \underline{I}_m(\mu)\}$, $I$ solves:*

$$\dot{I} = \frac{\rho}{IC''(I)} \left( F_m' + C'(I) \left( H' \left( M \left( \frac{1}{\rho} \left( I - \frac{C(I) + \rho F_m(\mu)}{C'(I)} \right) \right), \mu \right) - H'(\mu) \right) \right) \tag{S.25′}$$

**Lemma S.23.** *Suppose at $\mu_0$, $V_0, V_0', k \geqslant 1$ satisfies:*

$$\begin{cases} V_0 = \max_{\nu \geqslant \mu_0, I} \frac{I}{\rho} \frac{F_{m-k}(\nu) - V_0 - V_0'(\nu - \mu_0)}{J(\mu_0,\nu)} - \frac{C(I)}{\rho} \geqslant \max_{\nu \geqslant \mu_0, I} \frac{I}{\rho} \frac{F_m(\nu) - V_0 - V_0'(\nu - \mu_0)}{J(\mu_0,\nu)} - \frac{C(I)}{\rho} \\ \overline{V}(\mu_0) \geqslant V_0 \geqslant F_{m-k}(\mu_0) \end{cases}$$

*$V_{m-k}$ is the solution as defined in Lemma S.22 with initial condition $\mu_0, V_0, V_0'$, then $\forall \mu \in [\mu_0, \nu(\mu_0)]$:*

$$V_{m-k}(\mu) \geqslant \max_{\nu \geqslant \mu, m' \in [m-k,m], I} \frac{I}{\rho} \frac{F_{m'}(\nu) - V_{m-k}(\mu) - V_{m-k}'(\mu)(\nu - \mu)}{J(\mu,\nu)} - \frac{C(I)}{\rho}$$

**Proof.** We first show that:

$$V_0 \geqslant \max_{\nu \in [\mu_0, \overline{\mu}_m], I} \frac{I}{\rho} \frac{V_{m-k}(\nu) - V_0 - V_0'(\nu - \mu_0)}{J(\mu_0,\nu)} - \frac{C(I)}{\rho}$$

Suppose not, then there exists $\nu, I'$ s.t.

$$\begin{cases} V_0 < \frac{I'}{\rho} \frac{V_{m-k}(\nu) - V_0 - V_0'(\nu - \mu_0)}{J(\mu_0,\nu)} - \frac{C(I')}{\rho} \\ \frac{V_{m-k}(\nu) - V_0 - V_0'(\nu - \mu_0)}{J(\mu_0,\nu)} = C'(I') \end{cases} \tag{S.26}$$

Let $I_0 C'(I_0) = \rho V_0 + C(I_0)$, then optimality implies Equation (S.20) and Equation (S.21) at $\mu = \mu_0$:

$$\begin{cases} F_{m-k}' + C'(I_0) H'(\nu(\mu)) = V_0' + C'(I_0) H'(\mu) \\ (F_{m-k}(\nu(\mu)) + C'(I_0) H(\nu(\mu))) - (V_0 + C'(I_0) H(\mu)) = (V_0' + C'(I_0) H'(\mu))(\nu(\mu) - \mu) \end{cases}$$

We define $L(V,\lambda,\mu)(\nu)$ and $G(V,\lambda)(\mu)$ as Equation (S.16), Equation (S.17). Consider:

$$L(V_{m-k}, C'(I_0), \mu_0)(\nu) - G(F_{m-k}, C'(I_0))(\nu)$$

$L$ is a linear function and $G$ is a concave function. Therefore this is a convex function and have unique minimum determined by FOC. Simple calculation shows that it is minimized at $\nu(\mu_0)$ and the minimal value is 0. Now consider

$$\begin{aligned} & L(V_{m-k}, C'(I_0), \mu_0)(\nu) - G(V_{m-k}, C'(I_0))(\nu) \\ = & -(V_{m-k}(\nu) - V_{m-k}(\mu_0) - V_{m-k}'(\mu_0)(\nu - \mu_0) - C'(I_0) J(\mu_0,\nu)) \\ < & 0 \end{aligned}$$

The inequality is from Equation (S.26) and definition of $I_0$:

$$\begin{aligned} & I_0 C'(I_0) - C(I_0) < I' C'(I') - C(I') \\ \Longrightarrow & C'(I_0) < C'(I') \\ \Longrightarrow & C'(I_0) < \frac{V_{m-k}(\nu) - V_0 - V_0'(\nu - \mu_0)}{J(\mu_0,\nu)} \end{aligned}$$

Therefore $L(V_{m-k}, C'(I_0), \mu_0)(\nu) - G(V_{m-k}, C'(I_0))(\nu)$ will be strictly negative at $\nu$ and will have minimum strictly negative. Suppose it's minimized at $\mu''$ ($\mu'' > \mu_0$), then FOC implies:

$$V_{m-k}'(\mu_0) + C'(I_0) H'(\mu_0) = V_{m-k}'(\mu'') + C'(I_0) H(\mu'')$$

Let $I'' C'(I'') = \rho V_{m-k}(\mu'') + C(I'')$, then we have $I'' > I_0$ and $C'(I'') > C'(I_0)$. Consider:

$$L(V_{m-k}, C'(I_0), \mu'')(\nu(\mu'')) - G(F_{m-k}, C'(I_0))(\nu(\mu''))$$

$$
\begin{aligned}
&= L(V_{m-k}, C'(I_0), \mu_0)(\nu(\mu'')) - G(F_{m-k}, C'(I_0))(\nu(\mu'')) \\
&\quad + V_{m-k}(\mu'') - V_{m-k}(\mu_0) + C'(I_0)(H(\mu'') - H(\mu_0)) - (V'(\mu_0) + C'(I_0))(\mu'' - \mu_0) \\
&\geqslant V_{m-k}(\mu'') - V_{m-k}(\mu_0) + C'(I_0)(H(\mu'') - H(\mu_0)) - (V'(\mu_0) + C'(I_0))(\mu'' - \mu_0) \\
&= G(V_{m-k}, C'(I_0))(\mu'') - L(V_{m-k}, C'(I_0), \mu_0)(\mu'') > 0
\end{aligned}
$$

However:

$$
\begin{aligned}
0 &= L(V_{m-k}, C'(I''), \mu'')(\nu(\mu'')) - G(F_{m-k}, C'(I''))(\nu(\mu'')) \\
&= L(V_{m-k}, C'(I_0), \mu'')(\nu(\mu'')) - G(F_{m-k}, C'(I_0))(\nu(\mu'')) \\
&\quad + (C'(\mu'') - C'(I_0))(H(\mu'') - H(\nu(\mu'')) + H'(\mu'')(\nu(\mu'') - \mu'')) \\
&> (C'(I'') - C'(I_0)) J(\mu'', \nu(\mu'')) > 0
\end{aligned}
$$

Contradiction.

Now we show Lemma S.23. Suppose it's not true, then there exists $\nu \in (\mu_0, \nu(\mu_0))$, $\mu'' \geqslant \underline{\mu}_{m'}$, and $I''$ s.t.

$$
\begin{cases}
V_{m-k}(\nu) < \dfrac{I''}{\rho} \dfrac{F_{m'}(\mu'') - V'_{m-k}(\nu) - V'_{m-k}(\nu)(\mu'' - \nu)}{J(\nu, \mu'')} - \dfrac{C(I'')}{\rho} \\
\dfrac{F_{m'}(\mu'') - V'_{m-k}(\nu) - V'_{m-k}(\nu)(\mu'' - \nu)}{J(\nu, \mu'')} = C'(I'')
\end{cases}
$$

If we let $I'C'(I') = \rho V(\nu) + C(I')$, then $I' > I_0$ and $C'(I') > C'(I_0)$. By definition:

$$
\begin{aligned}
0 &\leqslant L(V_{m-k}, C'(I_0), \mu_0)(\mu'') - G(F_{m'}, C'(I_0))(\mu'') \\
&= L(F_{m-k}, C'(I_0), \nu(\mu_0))(\mu'') - G(F_{m'}, C'(I_0))(\mu'') \\
0 &\leqslant L(V_{m-k}, C'(I_0), \mu_0)(\nu) - G(F_{m'}, C'(I_0))(\nu) \\
&= L(F_{m-k}, C'(I_0), \nu(\mu_0))(\nu) - G(F_{m'}, C'(I_0))(\nu) \\
\Longrightarrow\ & L(F_{m-k}, C'(I'), \nu(\mu_0))(\mu'') - G(F_{m'}, C'(I'))(\mu'') \\
&= L(F_{m-k}, C'(I_0), \nu(\mu_0))(\mu'') - G(F_{m'}, C'(I_0))(\mu'') \\
&\quad + (C'(I') - C'(I_0)) J(\mu_0, \mu'') \\
&> 0 \\
& L(F_{m-k}, C'(I'), \nu(\mu_0))(\mu'') - G(F_{m'}, C'(I'))(\mu'') \\
&= L(F_{m-k}, C'(I_0), \nu(\mu_0))(\nu) - G(F_{m'}, C'(I_0))(\nu) \\
&\quad + (C'(I') - C'(I_0)) J(\mu_0, \nu) > 0
\end{aligned}
$$

No we consider $L(V_{m-k}, C'(I'), \nu)(\cdot)$ and $L(F_{m-k}, C'(I'), \nu(\mu_0))(\cdot)$:

$$
\begin{cases}
L(V_{m-k}, C'(I'), \nu)(\nu) = G(V_{m-k}, C'(I'))(\nu) \\
L(v_{m-k}, C'(I'), \nu)(\nu(\mu_0)) \geqslant G(V_{m-k}, C'(I'))(\nu(\mu_0)) \\
L(F_{m-k}, C'(I'), \nu(\mu_0))(\nu) > G(V_{m-k}, C'(I'))(\nu) \\
L(F_{m-k}, C'(I'), \nu(\mu_0))(\nu(\mu_0)) = G(V_{m-k}, C'(I'))(\nu(\mu_0))
\end{cases}
$$

$$
\Longrightarrow
\begin{cases}
L(V_{m-k}, C'(I'), \nu)(\nu(\mu_0)) \geqslant L(F_{m-k}, C'(I'), \nu(\mu_0))(\nu(\mu_0)) \\
L(V_{m-k}, C'(I'), \nu)(\nu) < L(F_{m-k}, C'(I'), \nu(\mu_0))(\nu)
\end{cases}
$$

Since both functions are linear: $\frac{\mathrm{d}}{\mathrm{d}\mu} L(V_{m-k}, C'(I'), \nu)(\mu) > \frac{\mathrm{d}}{\mathrm{d}\mu} L(F_{m-k}, C'(I')\nu(\mu_0))(\mu)$, then $L(V_{m-k}, C'(I'), \nu)(\cdot)$ must be larger than $L(F_{m-k}, C'(I'), \nu(\mu_0))(\cdot)$ at any $\mu'' \geqslant \nu(\mu_0)$. This implies:

$$
L(V_{m-k}, C'(I'), \nu)(\mu'') > G(F_{m'}, C'(I'))(\mu'')
$$

Contradicting the assumption. *Q.E.D.*

**Lemma S.23′.** *Suppose at $\mu_0$, $V_0, V'_0, k \geqslant 1$ satisfies:*

$$
\begin{cases}
V_0 = \max\limits_{\nu \leqslant \mu_0, I} \dfrac{I}{\rho} \dfrac{F_{m+k}(\nu) - V_0 - V'_0(\nu - \mu)}{J(\mu_0, \nu)} - \dfrac{C(I)}{\rho} \geqslant \max\limits_{\nu \leqslant \mu_0, I} \dfrac{I}{\rho} \dfrac{F_m(\nu) - V_0 - V'_0(\nu - \mu_0)}{J(\mu_0, \nu)} - \dfrac{C(I)}{\rho} \\
\overline{V}(\mu_0) \geqslant V_0 \geqslant F_{m+k}(\mu_0)
\end{cases}
$$

*$V_{m+k}$ is the solution as defined in Lemma S.22 with initial condition $\mu_0, V_0, V'_0$, then $\forall \mu \in [\nu(\mu_0), \mu_0]$:*

$$
V_{m+k}(\mu) \geqslant \max_{\nu \leqslant \mu, \nu \in [m, m+k], I} \frac{I}{\rho} \frac{F_{m'}(\nu) - V_{m-k}(\mu) - V'_{m-k}(\mu)(\nu - \mu)}{J(\mu, \nu)} - \frac{C(I)}{\rho}
$$

## S3 Proofs in Section 7

### S3.1 Linear delay cost

#### S3.1.1 Proof of Theorem 4

**Proof.** $\forall \langle\mu\rangle \in \mathbb{M}, \tau$,

$$E\left[\int_0^\tau C(I_t)\mathrm{d}t\right] \geqslant C\left(\frac{E\left[\int_0^\tau I_t \mathrm{d}t\right]}{E\left[\int_0^\tau \mathrm{d}t\right]}\right) E\left[\int_0^\tau \mathrm{d}t\right] = C\left(\frac{E\left[\int_0^\tau - E\left[\frac{\mathrm{d}H(\mu_t)}{\mathrm{d}t}\middle|\mathcal{F}_t\right]\mathrm{d}t\right]}{E[\tau]}\right) E[\tau]$$
$$= C\left(\frac{E\left[-\int_0^\tau \frac{\mathrm{d}H(\mu_t)}{\mathrm{d}t}\mathrm{d}t\right]}{E[\tau]}\right) E[\tau] = C\left(\frac{E[H(\mu)-H(\mu_\tau)]}{E[\tau]}\right) E[\tau]$$

First inequality is by Jensen's inequality. First equality is by definition of $I_t$. Second inequality is by iterated law of expectation. Last equality is straight forward. Since $\langle\mu_t\rangle \in \mathbb{M}$, $E[\mu_\tau] = \mu$. Let $\mu_\tau \sim P$ and $\lambda = E[H(\mu)-H(\mu_\tau)]/E[\tau]$, then:

$$E\left[F(\mu_\tau) - m\tau - \int_0^\tau C(I_t)\mathrm{d}t\right] \leqslant E_P[F(\nu)] - \frac{E_P[H(\mu)-H(\nu)]}{\lambda}(m + C(\lambda))$$
$$\Longrightarrow V(\mu) \leqslant \sup_{P\in\Delta^2(X), \lambda>0} E_P[F(\nu)] - \frac{m+C(\lambda)}{\lambda} E_P[H(\mu)-H(\nu)]$$

On the other hand, $\forall P \in \Delta^2(X), \lambda > 0$, let $\langle\mu_t\rangle$ be a compound Poisson process which realizes according to $P$ with Poisson rate $\frac{\lambda}{E_p[H(\mu)-H(\nu)]}$, $\tau$ is jump time of $\langle\mu_t\rangle$. Then it is easy to verify that:

$$E\left[F(\mu_\tau) - m\tau - \int_0^\tau C(I_t)\mathrm{d}t\right] = E_P[F(\nu)] - \frac{m+C(\lambda)}{\lambda} E_p[H(\mu)-H(\nu)]$$

*Q.E.D.*

### S3.2 General information measure

#### S3.2.1 Proof of Theorem 5

**Proof.** Consider Equation (13), it's sasy to see that both the inner maximization problem and the constraint are linear in $p_i$ and $\sigma^2$. Therefore, Equation (13) can be written equivalently as choosing either one posterior or a diffusion experiment:

$$\rho V(\mu) = \max\left\{\rho F(\mu), \sup_\nu \frac{c(V(\nu)-V(\mu)-V'(\mu)(\nu-\mu))}{J(\mu,\nu)}, \frac{cV''(\mu)}{J''_{\nu\nu}(\mu,\mu)}\right\}$$

Now suppose $\mu \in D$ and $\rho V(\mu) = c\frac{V''(\mu)}{J''_{\nu\nu}(\mu,\mu)}$. This is saying, the maximization problem:

$$\sup_\nu \frac{c(V(\nu)-V(\mu)-V'(\mu)(\nu-\mu))}{J(\mu,\nu)}$$

will be solved for $\nu \to \mu$. Therefore, consider the FOC:

$$\text{FOC: } \frac{V'(\nu)-V'(\mu)}{J(\mu,\nu)} - \frac{J'_\nu(\mu,\nu)}{J(\mu,\nu)^2}(V(\nu)-V(\mu)-V'(\mu)(\nu-\mu))$$

It must be $\leqslant 0$ when $\nu \to \mu^+$ and $\geqslant 0$ when $\nu \to \mu^-$. Otherwise, the diffusion experiment will be locally dominated by some Poisson experiment. When $\nu \to \mu$, $J(\mu,\nu) \to 0$, $V'(\nu) \to V'(\mu)$, $V(\nu) - V(\mu) - V'(\mu)(\nu-\mu) \to 0$. Therefore, we can apply L'Hospital's rule:

$$\lim_{\nu\to\mu}\text{FOC} = \frac{\lim_{\nu\to\mu}\left(V''(\nu) - J''_{\nu\nu}(\mu,\nu)\frac{V(\nu)-V(\mu)-V'(\mu)(\nu-\mu)}{J(\mu,\nu)} - J'_\nu(\mu,\nu)\cdot FOC\right)}{\lim_{\nu\to\mu} J'_\nu(\mu,\nu)}$$
$$= \frac{1}{2}\frac{\lim_{\nu\to\mu}\left(V''(\nu) - J''_{\nu\nu}(\mu,\nu)\frac{V(\nu)-V(\mu)-V'(\mu)(\nu-\mu)}{J(\mu,\nu)}\right)}{\lim_{\nu\to\mu} J'_\nu(\mu,\nu)}$$
$$= \frac{1}{2}\frac{\lim_{\nu\to\mu}\left(V^{(3)}(\nu) - J^{(3)}_{\nu\nu\nu}(\mu,\nu)\frac{V(\nu)-V(\mu)(\nu-\mu)}{J(\mu,\nu)} - J''_{\nu\nu}(\mu,\nu)\cdot\text{FOC}\right)}{\lim_{\nu\to\mu} J''_{\nu\nu}(\mu,\nu)}$$
$$= \frac{1}{3}\frac{V^{(3)}(\mu) - J^{(3)}_{\nu\nu\nu}(\mu,\mu)\frac{V''(\mu)}{J''_{\nu\nu}(\mu,\mu)}}{J''_{\nu\nu}(\mu,\mu)} \tag{S.27}$$

Now consider $V(\mu)-\frac{c}{\rho}\frac{V''(\mu)}{J''_{\nu\nu}(\mu,\mu)}$. By assumption, it's non-negative and achieves 0 at $\mu$. Therefore it is locally minimized at $\mu$:

$$\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}\mu}\left(V(\mu)-\frac{c}{\rho}\frac{V''(\mu)}{J''_{\nu\nu}(\mu,\mu)}\right)=0\\
\Longrightarrow\ &\frac{\rho}{c}V'(\mu)-\frac{V^{(3)}(\mu)}{J''_{\nu\nu}(\mu,\mu)}+\frac{V''(\mu)}{J''_{\nu\nu}(\mu,\mu)^2}\left(J^{(3)}_{\nu\nu\nu}(\mu,\mu)+J^{(3)}_{\nu\nu\mu}(\mu,\mu)\right)=0\\
\Longrightarrow\ &\frac{V^{(3)}(\mu)-J^{(3)}_{\nu\nu\nu}(\mu,\mu)\frac{V''(\mu)}{J''_{\nu\nu}(\mu,\mu)}}{J''_{\nu\nu}(\mu,\mu)}=\frac{\rho}{c}V'(\mu)+V''(\mu)\frac{J^{(3)}_{\nu\nu\mu}(\mu,\mu)}{J''_{\nu\nu}(\mu,\mu)^2}\\
\Longrightarrow\ &\frac{V^{(3)}(\mu)-J^{(3)}_{\nu\nu\nu}(\mu,\mu)\frac{V''(\mu)}{J''_{\nu\nu}(\mu,\mu)}}{J''_{\nu\nu}(\mu,\mu)}=\frac{\rho}{c}V'(\mu)+\frac{\rho}{c}V(\mu)\frac{J^{(3)}_{\nu\nu\mu}(\mu,\mu)}{J''_{\nu\nu}(\mu,\mu)}
\end{aligned}\tag{S.28}$$

By smoothness of $V$ and $J$, for FOC to be non-positive when $\nu\to\mu^+$ and non-negative when $\nu\to\mu^-$. So Equations (S.27) and (S.28) implies:

$$V'(\mu)J''_{\nu\nu}(\mu,\mu)+V(\mu)J^{(3)}_{\nu\nu\mu}(\mu,\mu)=0$$

Now suppose there exists $\mu_n\to\mu$ s.t. $\rho V(\mu_n)=c\frac{V''(\mu_n)}{J''_{\nu\nu}(\mu_n,\mu_n)}$, we have:

$$V'(\mu_n)J''_{\nu\nu}(\mu_n,\mu_n)+V(\mu_n)J^{(3)}_{\nu\nu\mu}(\mu_n,\mu_n)=0$$

By differentiability of the whole term, we have:

$$\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}\mu}\left(V'(\mu)J''_{\nu\nu}(\mu,\mu)+V(\mu)J^{(3)}_{\nu\nu\mu}(\mu,\mu)\right)=0\\
\Longrightarrow\ &V''(\mu)J''_{\nu\nu}(\mu,\mu)+V'(\mu)\left(2J^{(3)}_{\nu\nu\mu}(\mu,\mu)+J^{(3)}_{\nu\nu\nu}(\mu,\mu)\right)+V(\mu)\left(J^{(4)}_{\nu\nu\nu\mu}(\mu,\mu)+J^{(4)}_{\nu\nu\mu\mu}(\mu,\mu)\right)=0\\
\Longrightarrow\ &\frac{\rho}{c}V(\mu)J''_{\nu\nu}(\mu,\mu)^2-\frac{V(\mu)}{J''_{\nu\nu}(\mu,\mu)}\left(2J^{(3)}_{\nu\nu\mu}(\mu,\mu)^2+J^{(3)}_{\nu\nu\nu}(\mu,\mu)J^{(3)}_{\nu\nu\mu}(\mu,\mu)\right)\\
&+V(\mu)\left(J^{(4)}_{\nu\nu\nu\mu}(\mu,\mu)+J^{(4)}_{\nu\nu\mu\mu}(\mu,\mu)\right)=0\\
\Longrightarrow\ &\frac{\rho}{c}J''_{\nu\nu}(\mu,\mu)^2-\frac{2J^{(3)}_{\nu\nu\mu}(\mu,\mu)^2+J^{(3)}_{\nu\nu\nu}(\mu,\mu)J^{(3)}_{\nu\nu\mu}(\mu,\mu)}{J''_{\nu\nu}(\mu,\mu)}+J^{(4)}_{\nu\nu\nu\mu}(\mu,\mu)+J^{(4)}_{\nu\nu\mu\mu}(\mu,\mu)=0
\end{aligned}$$

By assumption, $\mu\in D$, therefore the differential equation must not be satisfied. This implies that there doesn't exist such $\mu_n\to\mu$. So the set:

$$\left\{\mu\in D\middle|\rho V(\mu)=c\frac{V''(\mu)}{J''_{\nu\nu}(\mu,\mu)}\right\}$$

is a closed set (closed w.r.t. D) containing no limiting point. That is to say, within any compact subset of $D$, this set is finite. This set is a Borel set, thus Lebesgue measurable. By definition of Lebesgue measure, the measure of a set can be approximated by compact subsets from below. Therefore, this set has zero-measure. *Q.E.D.*

### *S3.2.2 Construction of a special cost function*

Take any general cost structure $J(\mu,\nu)$ and $\kappa(\mu,\sigma)$ that satisfies Assumption 4. In the section, I introduce the method to construct a cost structure such that (i) the cost of Gaussian learning is $\kappa(\mu,\sigma)$ and (ii) the DM is exactly indifferent between using Guassian learning and Poisson learning.

**Step 1**. Let $g(\mu)=J''_{\nu\nu}(\mu,\mu)$ (then $\kappa(\mu,\sigma)=\frac{1}{2}g(\mu)\sigma^2$). Restrict the DM to using only Gaussian learning, then Equation (13) becomes:

$$\rho V(\mu)=\max\left\{\rho F(\mu),\frac{c}{g(\mu)}V''(\mu)\right\}\tag{S.29}$$

Equation (S.29) can be solved by solving the following ODE and applying smooth pasting:

$$V(\mu)=\frac{c}{\rho}\frac{V''(\mu)}{g(\mu)}\tag{S.30}$$

Change parameter and let $v(\mu)=\frac{\mathrm{d}}{\mathrm{d}\mu}\log(V(\mu))$, then $v(\mu)$ satisfies the following ODE:

$$v'(\mu)+v(\mu)^2=\frac{\rho}{c}g(\mu)\tag{S.31}$$

By Assumption 4, $g(\mu)$ is a smooth function on $(0,1)$. Therefore it is easy to verify that on any closed sub-interval of $(0,1)$, Picard-Lindelöf is satisfied that there exist unique solution to Equation (S.31) given initial condition. Let $v(\mu,C_1)$

be the solution indexed by free parameter $C_1$, then $V(\mu)=C_2e^{\int_0^\mu v(C_1,\nu)\mathrm{d}\nu}$. The two free parameters $(C_1,C_2)$ can be pinned down by two smooth pasting conditions:

$$\begin{cases} C_2e^{\int_0^{\mu_1} v(C_1,\nu)\mathrm{d}\nu}=F(\mu_1) \\ C_2e^{\int_0^{\mu_1} v(C_1,\nu)\mathrm{d}\nu}v(C_1,\mu_1)=F'(\mu_1) \\ C_2e^{\int_0^{\mu_2} v(C_1,\nu)\mathrm{d}\nu}=F(\mu_2) \\ C_2e^{\int_0^{\mu_1} v(C_1,\nu)\mathrm{d}\nu}v(C_1,\mu_2)=F'(\mu_2) \end{cases}$$

Notice that smooth pasting need to be checked for at most $C_{|A|}^2$ pairs of actions, index all solutions by $V_i$. Then $V(\mu)=\max\{V_i(\mu)\}$ solves Equation (S.29). Let $E=\{\mu|V(\mu)>F(\mu)\}$.

**Step 2**. $\forall\mu\in E$, define $J_0(\mu,\nu)$ as:

$$J_0(\mu,\nu)=\frac{c}{\rho}\frac{V(\nu)-V(\mu)-V'(\mu)(\nu-\mu)}{V(\mu)}$$

It is easy to verify that $J_0>0$. Now let us verify the solution of Equation (13):

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

$$\text{s.t. } pJ_0(\mu,\nu)+\frac{1}{2}J_{0\nu\nu}(\mu,\mu)\sigma^2\leqslant c$$

First of all, by definition, $V(\mu)$, $p\equiv 0$ and $\sigma^2=\frac{c}{J_{0\nu\nu}(\mu,\mu)}$ is feasible and satisfies the equality condition. Now $\forall\nu\in[0,1]$:

$$c\frac{V(\nu)-V(\mu)-V'(\mu)(\nu-\mu)}{J_0(\mu,\nu)}=\rho V(\mu)$$

Therefore, any Poisson learning strategy is as good as the Gaussian learning strategy. By definition of $J_0$, $\frac{1}{2}J_{0\nu\nu}(\mu,\mu)\sigma^2=\frac{1}{2}g(\mu)\sigma^2=\kappa(\mu,\sigma)$.

**Step 3**. Smooth extension of $J_0$. So far, $J_0$ is only defined on $E$. $J_0$ can be extended smoothly onto $[0,1]$ satisfying $\forall\mu\in E^C$:

$$\begin{cases} J_{0\nu\nu}(\mu,\mu)=g(\mu) \\ J_0(\mu,\nu)\geqslant\dfrac{c}{\rho}\dfrac{V(\nu)-V(\mu)-V'(\mu)(\nu-\mu)}{V(\mu)},\forall\nu \end{cases}$$

The extension is intuitively simple but quite technical, which is ommited here.

Such a $J_0(\mu,\nu)$ is uniquely defined for $\mu\in E$. For $\mu\in E^C$, there can be many degrees of freedom because Gaussian learning is anyway strictly dominated by stopping. So it is sufficienly to make Poisson learning also dominated. Now, suppose $J(\mu,\nu)$ and $\kappa(\mu,\sigma)$ are such that Gaussian learning is weakly optimal. Then $J(\mu,\nu)$ must be pointwise weakly higher than $J_0(\mu,\nu)$, $\forall\mu\in E$. On the other hand, since $J_{\nu\nu}(\mu,\mu)=J_{0\nu\nu}(\mu,\mu)$, this implies $J_{\nu^3}(\mu,\mu)=J_{0\nu^3}(\mu,\mu)$. That is to say, assuming Gaussian learning being weakly optimal is imposing an additional third derivative constraint on $J(\mu,\nu)$ on the constriants in Assumption 4, making the set of cost functions non-generic.

### S3.3 *Linear cost function*

#### S3.3.1 *Proof of Theorem 6*

**Proof.** I first prove a result in discrete time. Take any information acquisition strategy $(\mathcal{S}^t,\mathcal{A}^t,\mathcal{T})$ that satisfies the constraints in Equation (5′). The achieved expected utility will be:

$$E\left[e^{-\rho dt\cdot\mathcal{T}}u(\mathcal{A}^\mathcal{T},\mathcal{X})-\sum_{t=0}^{\infty}e^{-\rho dt\cdot t}\lambda I(\mathcal{S}^t;\mathcal{X}|\mathcal{S}^{\mathbf{t-1}},\mathbf{1}_{\mathcal{T}\leqslant t})\right]$$

We can separate the utility gain part and information cost part. Utility gain part is:

$$E\left[e^{-\rho dt\cdot\mathcal{T}}u\left(\mathcal{A}^\mathcal{T},\mathcal{X}\right)\right]$$

It's easy to see that this is determined only by action time $\mathcal{T}$ and action process $\mathcal{A}^\mathcal{T}$. Let $\widetilde{\mathcal{S}}^{t-1}=(\mathbf{1}_{\mathcal{T}=t},\mathcal{A}^t|_{\mathcal{T}=t})$. Then by information processing constraint in Equation (5′), we have:

$$\begin{aligned}\text{Prob}\left(\widetilde{\mathcal{S}}^{\mathbf{t}}|\mathcal{S}^{\mathbf{t}},\mathcal{X}\right)=&\text{Prob}\left(\mathbf{1}_{\mathcal{T}=\mathbf{t+1}},\mathcal{A}^{\mathbf{t+1}}|\mathcal{S}^{\mathbf{t}},\mathcal{X}\right)=\text{Prob}\left(\mathbf{1}_{\mathcal{T}\leqslant\mathbf{t+1}},\mathcal{A}^{\mathbf{t+1}}|\mathcal{S}^{\mathbf{t}},\mathcal{X}\right)\\=&\text{Prob}\left(\mathbf{1}_{\mathcal{T}\leqslant\mathbf{t+1}},\mathcal{A}^{\mathbf{t+1}}|\mathcal{S}^{\mathbf{t}}\right)=\text{Prob}\left(\widetilde{\mathcal{S}}^{\mathbf{t}}|\mathcal{S}^{\mathbf{t}}\right)\\\Longrightarrow&\mathcal{X}\to\mathcal{S}^{\mathbf{t}}\to\widetilde{\mathcal{S}}^{\mathbf{t}}\end{aligned}$$

Therefore:

$$
\begin{aligned}
&\sum_{t=0}^{\infty} e^{-\rho dt\cdot t}\lambda E\left[I\left(\mathcal{S}^t;\mathcal{X}|\mathcal{S}^{\mathbf{t-1}},\mathbf{1}_{\mathcal{T}\leqslant t}\right)\right]\\
=&\lambda\sum_{t=0}^{\infty} e^{-\rho dt\cdot t}E\left[I\left(\mathcal{S}^{\mathbf{t}},\mathbf{1}_{\mathcal{T}\leqslant t};\mathcal{X}\right)-I\left(\mathcal{S}^{\mathbf{t-1}},\mathbf{1}_{\mathcal{T}\leqslant t};\mathcal{X}\right)\right]\\
=&\lambda\sum_{t=0}^{\infty} e^{-\rho dt\cdot t}E\left[I\left(\mathcal{S}^{\mathbf{t}};\mathcal{X}\right)+I\left(\mathbf{1}_{\mathcal{T}\leqslant t};\mathcal{X}|\mathcal{S}^{\mathbf{t}}\right)-I\left(\mathcal{S}^{\mathbf{t-1}};\mathcal{X}\right)-I\left(\mathbf{1}_{\mathcal{T}\leqslant t};\mathcal{X}|\mathcal{S}^{\mathbf{t-1}}\right)\right]\\
=&\lambda\sum_{t=0}^{\infty} e^{-\rho dt\cdot t}E\left[I\left(\mathcal{S}^{\mathbf{t}};\mathcal{X}\right)-I\left(\mathcal{S}^{\mathbf{t-1}};\mathcal{X}\right)\right]\\
=&\lambda\sum_{t=0}^{\infty} e^{-\rho dt\cdot t}E\left[I\left(\widetilde{\mathcal{S}}^{\mathbf{t}};\mathcal{X}\right)+I\left(\mathcal{S}^{\mathbf{t}};\mathcal{X}|\widetilde{\mathcal{S}}^{\mathbf{t}}\right)-I\left(\widetilde{\mathcal{S}}^{\mathbf{t}};\mathcal{X}|\mathcal{S}^{\mathbf{t}}\right)\right]\\
&-\lambda\sum_{t=0}^{\infty} e^{-\rho dt\cdot t}E\left[I\left(\widetilde{\mathcal{S}}^{\mathbf{t-1}};\mathcal{X}\right)+I\left(\mathcal{S}^{\mathbf{t-1}};\mathcal{X}|\widetilde{\mathcal{S}}^{\mathbf{t-1}}\right)-I\left(\widetilde{\mathcal{S}}^{\mathbf{t-1}};\mathcal{X}|\mathcal{S}^{\mathbf{t-1}}\right)\right]\\
=&\lambda\sum_{t=0}^{\infty} e^{-\rho dt\cdot t}E\left[I\left(\widetilde{\mathcal{S}}^{\mathbf{t}};\mathcal{X}\right)+I\left(\mathcal{S}^{\mathbf{t}};\mathcal{X}|\widetilde{\mathcal{S}}^{\mathbf{t}}\right)\right]\\
&-\lambda\sum_{t=0}^{\infty} e^{-\rho dt\cdot t}E\left[I\left(\widetilde{\mathcal{S}}^{\mathbf{t-1}};\mathcal{X}\right)+I\left(\mathcal{S}^{\mathbf{t-1}};\mathcal{X}|\widetilde{\mathcal{S}}^{\mathbf{t-1}}\right)\right]\\
=&\sum_{t=0}^{\infty} e^{-\rho dt\cdot t}\lambda E\left[I\left(\widetilde{\mathcal{S}}^t;\mathcal{X}|\widetilde{\mathcal{S}}^{\mathbf{t-1}},\mathbf{1}_{\mathcal{T}\leqslant t}\right)\right]+\lambda\sum e^{-\rho dt\cdot t}\left(1-e^{-\rho dt}\right)E\left[I\left(\mathcal{S}^{\mathbf{t}};\mathcal{X}|\widetilde{\mathcal{S}}^{\mathbf{t}}\right)\right]\\
\geqslant&\sum_{t=0}^{\infty} e^{-\rho dt\cdot t}\lambda E\left[I\left(\widetilde{\mathcal{S}}^t;\mathcal{X}|\widetilde{\mathcal{S}}^{\mathbf{t-1}},\mathbf{1}_{\mathcal{T}\leqslant t}\right)\right]
\end{aligned}
$$

Therefore, by replacing signal process $\mathcal{S}^t$ with $\widetilde{\mathcal{S}}^t$, the DM can achieve the same utility gain and pay a weakly lower information cost. Now consider:

$$
\begin{aligned}
&E\left[e^{-\rho dt\cdot\mathcal{T}}u\left(\mathcal{A}^{\mathcal{T}},\mathcal{X}\right)\right]-\lambda\sum_{t=0}^{\infty} e^{-\rho dt\cdot t}E\left[I\left(\widetilde{\mathcal{S}}^t;\mathcal{X}|\widetilde{\mathcal{S}}^{\mathbf{t-1}},\mathbf{1}_{\mathcal{T}\leqslant t}\right)\right]\\
=&\text{Prob}(\mathcal{T}=0)E\left[u\left(\mathcal{A}^0,\mathcal{X}\right)\right]\\
&+\text{Prob}(\mathcal{T}\geqslant 1)E\left[e^{-\rho dt\cdot\mathcal{T}}u\left(\mathcal{A}^{\mathcal{T}},\mathcal{X}\right)|\mathcal{T}\geqslant 1\right]\\
&-\lambda I\left(\widetilde{\mathcal{S}}^0;\mathcal{X}|\mu\right)-\lambda\sum_{t=1}^{\infty} e^{-\rho dt\cdot t}E\left[I\left(\widetilde{\mathcal{S}}^t;\mathcal{X}|\widetilde{\mathcal{S}}^{\mathbf{t-1}},\mathbf{1}_{\mathcal{T}\leqslant t}\right)\right]\\
=&\text{Prob}(\mathcal{T}=0)E\left[u\left(\mathcal{A}^0,\mathcal{X}\right)\right]\\
&+\text{Prob}(\mathcal{T}=1)E\left[e^{-\rho dt}u\left(\mathcal{A}^1,\mathcal{X}\right)|\mathcal{T}=1\right]-\lambda I\left(\widetilde{\mathcal{S}}^0;\mathcal{X}|\mu,\mathcal{T}>0\right)\\
&+\text{Prob}(\mathcal{T}\geqslant 2)E\left[e^{-\rho dt\cdot\mathcal{T}}u\left(\mathcal{A}^{\mathcal{T}},\mathcal{X}\right)|\mathcal{T}\geqslant 2\right]-\lambda\sum_{t=1}^{\infty} e^{-\rho dt\cdot t}E\left[I\left(\widetilde{\mathcal{S}}^t;\mathcal{X}|\widetilde{\mathcal{S}}^{\mathbf{t-1}},\mathbf{1}_{\mathcal{T}\leqslant t}\right)\right]
\end{aligned}
$$

Suppose the term:

$$
\text{Prob}(\mathcal{T}\geqslant 2)E\left[e^{-\rho dt\cdot\mathcal{T}}u\left(\mathcal{A}^{\mathcal{T}},\mathcal{X}\right)|\mathcal{T}\geqslant 2\right]-\lambda\sum_{t=1}^{\infty} e^{-\rho dt\cdot t}E\left[I\left(\widetilde{\mathcal{S}}^t;\mathcal{X}|\widetilde{\mathcal{S}}^{\mathbf{t-1}},\mathbf{1}_{\mathcal{T}\leqslant t}\right)\right] \tag{S.32}
$$

is negative, then discard all actions and information after first period will give the DM higher expected utility. This information and action process satisfies this theorem. Therefore, WLOG we assume Equation (S.32), as well as all continuation payoffs are non-negative. Then:

$$
E\left[e^{-\rho dt\cdot\mathcal{T}}u\left(\mathcal{A}^{\mathcal{T}},\mathcal{X}\right)\right]-\lambda\sum_{t=0}^{\infty} e^{-\rho dt\cdot t}E\left[I\left(\widetilde{\mathcal{S}}^t;\mathcal{X}|\widetilde{\mathcal{S}}^{\mathbf{t-1}},\mathbf{1}_{\mathcal{T}\leqslant t}\right)\right]
$$

$$
\begin{aligned}
&=\text{Prob}(\mathcal{T}=0)E[u(\mathcal{A}^0,\mathcal{X})]\\
&\quad+\text{Prob}(\mathcal{T}=1)\left(E\left[e^{-\rho dt}u\left(\mathcal{A}^1,\mathcal{X}\right)|\mathcal{T}=1\right]-\lambda I\left(\tilde{\mathcal{S}}^0;\mathcal{X}|\mu,\mathcal{T}>0\right)\right)\\
&\quad+\text{Prob}(\mathcal{T}\geqslant 2)E\left[e^{-\rho dt\cdot\mathcal{T}}u\left(\mathcal{A}^{\mathcal{T}},\mathcal{X}\right)|\mathcal{T}\geqslant 2\right]-\lambda\sum_{t=1}^{\infty}e^{-\rho dt\cdot t}E\left[I\left(\tilde{\mathcal{S}}^t;\mathcal{X}|\tilde{\mathcal{S}}^{\mathbf{t-1}},\mathbf{1}_{\mathcal{T}\leqslant t}\right)\right]\\
&\leqslant\text{Prob}(\mathcal{T}=0)E[u(\mathcal{A}^0,\mathcal{X})]\\
&\quad+\text{Prob}(\mathcal{T}=1)\left(E\left[e^{-\rho dt}u\left(\mathcal{A}^1,\mathcal{X}\right)|\mathcal{T}=1\right]-\lambda I\left(\tilde{\mathcal{S}}^0;\mathcal{X}|\mu,\mathcal{T}>0\right)\right)\\
&\quad+\text{Prob}(\mathcal{T}\geqslant 2)E\left[e^{-\rho dt\cdot(\mathcal{T}-1)}u\left(\mathcal{A}^{\mathcal{T}},\mathcal{X}\right)|\mathcal{T}\geqslant 2\right]-\lambda\sum_{t=1}^{\infty}e^{-\rho dt\cdot(t-1)}E\left[I\left(\tilde{\mathcal{S}}^t;\mathcal{X}|\tilde{\mathcal{S}}^{\mathbf{t-1}},\mathbf{1}_{\mathcal{T}\leqslant t}\right)\right]\\
&=\text{Prob}(\mathcal{T}=0)E[u(\mathcal{A}^0,\mathcal{X})]\\
&\quad+\text{Prob}(\mathcal{T}=1)\left(E\left[e^{-\rho dt}u\left(\mathcal{A}^1,\mathcal{X}\right)|\mathcal{T}=1\right]-\lambda I\left(\tilde{\mathcal{S}}^0;\mathcal{X}|\mu,\mathcal{T}>0\right)\right)\\
&\quad+\text{Prob}(\mathcal{T}=2)\left(E\left[e^{-\rho dt}u(\mathcal{A}^2,\mathcal{X})|\mathcal{T}=2\right]-\lambda I\left(\tilde{\mathcal{S}}^1;\mathcal{X}|\tilde{\mathcal{S}}^0,\mathcal{T}>1\right)\right)\\
&\quad+\text{Prob}(\mathcal{T}\geqslant 3)E\left[e^{-\rho dt\cdot(\mathcal{T}-1)}u\left(\mathcal{A}^{\mathcal{T}},\mathcal{X}\right)|\mathcal{T}\geqslant 3\right]-\lambda\sum_{t=2}^{\infty}e^{-\rho dt\cdot(t-1)}E\left[I\left(\tilde{\mathcal{S}}^t;\mathcal{X}|\tilde{\mathcal{S}}^{\mathbf{t-1}},\mathbf{1}_{\mathcal{T}\leqslant t}\right)\right]\\
&\leqslant\text{Prob}(\mathcal{T}=0)E[u(\mathcal{A}^0,\mathcal{X})]\\
&\quad+\text{Prob}(\mathcal{T}=1)\left(E\left[e^{-\rho dt}u\left(\mathcal{A}^1,\mathcal{X}\right)|\mathcal{T}=1\right]-\lambda I\left(\tilde{\mathcal{S}}^0;\mathcal{X}|\mu,\mathcal{T}>0\right)\right)\\
&\quad+\text{Prob}(\mathcal{T}=2)\left(E\left[e^{-\rho dt}u(\mathcal{A}^2,\mathcal{X})|\mathcal{T}=2\right]-\lambda I\left(\tilde{\mathcal{S}}^1;\mathcal{X}|\tilde{\mathcal{S}}^0,\mathcal{T}>1\right)\right)\\
&\quad+\text{Prob}(\mathcal{T}\geqslant 3)E\left[e^{-\rho dt\cdot(\mathcal{T}-2)}u\left(\mathcal{A}^{\mathcal{T}},\mathcal{X}\right)|\mathcal{T}\geqslant 3\right]-\lambda\sum_{t=2}^{\infty}e^{-\rho dt\cdot(t-2)}E\left[I\left(\tilde{\mathcal{S}}^t;\mathcal{X}|\tilde{\mathcal{S}}^{\mathbf{t-1}},\mathbf{1}_{\mathcal{T}\leqslant t}\right)\right]\\
&\quad\vdots\\
&\leqslant\text{Prob}(\mathcal{T}=0)E[u(\mathcal{A}^0,\mathcal{X})]\\
&\quad+\text{Prob}(\mathcal{T}\geqslant 1)\left(E\left[e^{-\rho dt}u\left(\mathcal{A}^{\mathcal{T}};\mathcal{X}\right)|\mathcal{T}\geqslant 1\right]-\lambda\sum_{t=0}^{\infty}E\left[I\left(\tilde{\mathcal{S}}^t;\mathcal{X}|\tilde{\mathcal{S}}^{\mathbf{t-1}},\mathbf{1}_{\mathcal{T}\leqslant t},\mathcal{T}\geqslant 1\right)\right]\right)\\
&=\text{Prob}(\mathcal{T}=0)E[u(\mathcal{A}^0,\mathcal{X})]\\
&\quad+\text{Prob}(\mathcal{T}\geqslant 1)\left(E\left[e^{-\rho dt}u\left(\mathcal{A}^{\mathcal{T}};\mathcal{X}\right)|\mathcal{T}\geqslant 1\right]-\lambda\lim_{t\to\infty}I\left(\tilde{\mathcal{S}}^{\mathbf{t}};\mathcal{X}|\mathcal{T}\geqslant 1\right)\right)\\
&=\text{Prob}(\mathcal{T}=0)E[u(\mathcal{A}^0,\mathcal{X})]\\
&\quad+\text{Prob}(\mathcal{T}\geqslant 1)\left(E\left[e^{-\rho dt}u\left(\mathcal{A}^{\mathcal{T}};\mathcal{X}\right)|\mathcal{T}\geqslant 1\right]-\lambda I\left(\mathcal{A}^{\mathcal{T}};\mathcal{X}|\mathcal{T}\geqslant 1\right)\right)
\end{aligned}
$$

Therefore:

$$
\begin{aligned}
&E\left[e^{-\rho dt\cdot\mathcal{T}}u\left(\mathcal{A}^{\mathcal{T}},\mathcal{X}\right)\right]-\lambda\sum_{t=0}^{\infty}e^{-\rho dt\cdot t}E\left[I\left(\tilde{\mathcal{S}}^t;\mathcal{X}|\tilde{\mathcal{S}}^{\mathbf{t-1}},\mathbf{1}_{\mathcal{T}\leqslant t}\right)\right]\\
\leqslant&\,\text{P}[\mathcal{T}=0]F(\mu)+(1-\text{P}[\mathcal{T}=0])\left(E\left[e^{-\rho dt}u\left(\mathcal{A}^{\mathcal{T}},\mathcal{X}\right)\right]-\lambda I\left(\mathcal{A}^{\mathcal{T}};\mathcal{X}|\mu\right)\right)\\
\leqslant&\max\left\{F(\mu),\sup_{\mathcal{A}}E\left[e^{-\rho dt}u(\mathcal{A},\mathcal{X})\right]-\lambda I(\mathcal{A};\mathcal{X})\right\}
\end{aligned}
$$

Therefore, we showed that any dynamic information acquisition strategy solving Equation (5′) will have weakly lower expected utility level than a static information acquisition strategy solving Equation (14).

Now let us consider the continuous time problem. It is clear by Lemma S.5 that any discretization of Equation (1) can be implemented by Equation (5′). Hence,

$$
V(\mu)\leqslant\lim V_{dt}(\mu)\leqslant\overline{\lim}\max\left\{F(\mu),\sup_{\mathcal{A}}E\left[e^{-\rho dt}u(\mathcal{A},\mathcal{X})\right]-\lambda I(\mathcal{A};\mathcal{X})\right\}
$$

$$=\sup_{\mathcal{A}} E[u(\mathcal{A},\mathcal{X})]-\lambda I(\mathcal{A};\mathcal{X})$$
$$=\sup_{P\in\Delta^2(X)} E_P[F(\nu)-\lambda(H(\mu)-H(\nu))]$$

On the other hand, take any $P$ and $dt>0$, by Lemma S.3, there exists $\langle\mu_t\rangle\in\mathbb{M}$ such that:

$$E\left[e^{-\rho dt}F(\mu_{dt})-\int_0^{dt} e^{-\rho t}\lambda\frac{E_P[H(\mu)-H(\nu)]}{dt}\mathrm{d}t\right]=E_P[e^{-\rho dt}F(\nu)]-\frac{1-e^{-\rho dt}}{\rho dt}\lambda E_P[H(\mu)-H(\nu)]$$
$$\Longrightarrow V(\mu)\geqslant \sup_{P\in\Delta^2(X)}\lim_{dt\to 0}E_P[e^{-\rho dt}F(\nu)]-\frac{1-e^{-\rho dt}}{\rho dt}\lambda E_P[H(\mu)-H(\nu)]$$
$$\Longrightarrow V(\mu)\geqslant \sup_{P\in\Delta^2(X)} E_P[F(\nu)-\lambda(H(\mu)-H(\nu))]$$

Combining the two inequalities, Equation (14) is proved. *Q.E.D.*

## S4 Proofs in Section 8

### S4.1 *Choice accuracy and response time: proof of Proposition 1*

**Proof.** Since both $H_0(\mu)$ and $F(\mu)$ are symmetric functions around $\mu_0=0.5$, by symmetry and quasi-concavity of value function (Theorem 2), $\forall c_k$, $\mu^*=\mu_0$. Let the expected utility of the action favoring beliefs $>0.5$ be $F_r(\mu)$, and the utility of the other action $F_l(\mu)$. $\forall c_k$, by the proof of Lemma B.1, there exists unique $\nu_k^l$ and $\nu_k^r=1-\mu_k^r$ s.t.

$$\begin{cases}\nu_k^r\in\operatorname{argmax}_{\nu\geqslant\mu}\dfrac{F_r(\nu)}{1+\frac{\rho}{c_k}J_0(\mu_0,\nu)}\\ \nu_k^l\in\operatorname{argmax}_{\nu\leqslant\mu}\dfrac{F_l(\nu)}{1+\frac{\rho}{c_k}J_0(\mu_0,\nu)}\end{cases}$$

Where $J_0(\mu,\nu)=H_0(\mu)-H_0(\nu)+H_0'(\mu)(\nu-\mu)$. Now I determine the location of $\{\nu_k^l,\nu_k^r\}$ by studying the following cross derivative:

$$\left.\frac{\mathrm{d}^2}{\mathrm{d}c,\mathrm{d}\nu}\frac{F_r(\nu)}{1+\frac{\rho}{c}J_0(\mu,\nu)}\right|_{\nu=\nu^{r*},\mu=\mu_0}$$
$$=\frac{\rho^2}{c^3}\left.\frac{F_r(\nu)(H_0'(\mu)-H_0'(\nu))(F_r(\mu)(H_0'(\mu)-H_0'(\nu))+F_r'J_0(\nu,\mu))}{F_r'\left(1+\frac{\rho}{c}J_0(\mu,\nu)\right)}\right|_{\nu=\nu^{r*},\mu=\mu_0}>0$$

The equality is by plug the FOC determining $\nu^*$ into the cross derivative. The inequality follows by $H_0(\mu)$ being strictly concave, $F_r'>0$ and $F_r(\mu_0)=F(\mu_0)>0$. Since the cross derivative w.r.t. $\nu$ and $c$ is strictly positive at $\nu^{r*}$, the standard comparative statics analysis suggests that the optimal belief $\nu^{r*}$ is strictly increasing in parameter $c$. A completely symmetric argument applies to $\nu^{l*}$ and $\nu^{l*}$ is strictly decreasing in parameter $c$. Therefore, all the $\{\nu_k^r,\nu_k^l\}$ are ordered on $[0,1]$:

$$0,\nu_k^l,\cdots,\nu_1^l,\mu_0,\nu_1^r,\cdots,\nu_k^r,1$$

Moreover, $\forall c\in(c_i,c_{i+1})$, $\nu^{r*}(c)\in(\nu_i^r,\nu_{i+1}^r)$ and $\nu^{l*}(c)\in(\nu_{i+1}^l,\nu_i^l)$. Now assume that the goal is to make the sign of $\mu^*-\mu_0$ strictly positive (negative) when $c\in(c_{2i},c_{2i+1})$ ($c\in(c_{2i-1},c_{2i})$). To achieve this goal, define $H(\mu)$ based on $H_0(\mu)$. Let $M_{a,b}(\mu)$ be a function on $\mathbb{R}$ with the following properties:

- Parameter $a,b\in\mathbb{R}$ and $a<b$.
- $\forall a,b,\mu$, $M_{a,b}(\mu)<0$ if $\mu\in(a,b)$ and $M_{a,b}=0$ otherwise.
- $\forall a,b$, $M_{a,b}(\mu)$ is $C^{(2)}$ smooth on $\mathbb{R}$ and $|M_{a,b}''(\mu)|$ is bounded by 1.

The choice of function $M$ can be quite arbitrary. For example, it is not hard to verify that:

$$M_{a,b}(\mu)=-\mathbf{1}_{a<\mu<b}\frac{(b-a)^4}{256e^{-\frac{16}{(b-a)^2}}}e^{-\left(\frac{1}{\mu-a}+\frac{1}{b-\mu}\right)^2}$$

satisfies these properties. Define $\nu_{k+1}^r=\frac{1+\nu_k^r}{2}$ and $\nu_{k+1}^l=\frac{\nu_k^l}{2}$. Since $H_0(\mu)$ satisfies Assumption 2-a, there exists $\varepsilon$ s.t. $\forall\mu\in[\nu_{k+1}^l,\nu_{k+1}^r]$, $H''(\mu)\leqslant-2\varepsilon$. Now define $H(\mu)$:

$$H(\mu)=H_0(\mu)+\sum M_{\nu_{2i}^r,\nu_{2i+1}^r}(\mu)+\sum M_{\nu_{2i}^l,\nu_{2i-1}^l}(\mu)$$

I verify that $H(\mu)$ satisfies the conditions in Proposition 1. It is easy to verify that $J(\mu_0,\nu)=J(\mu_0,\nu)$ when $\nu\notin\bigcup(\nu^r_{2i},\nu^r_{2i+1})$ or $\bigcup(\nu^l_{2i}),\nu^l_{2i-1}$. $J(\mu_0,\nu)>J_0(\mu_0,\nu)$ otherwise. First, when $c\in\{c_k\}$:

$$\sup_{\nu\geqslant\mu}\frac{F_r(\nu)}{1+\frac{\rho}{c}J(\mu_0,\nu)}\geqslant\sup_{\nu\geqslant\mu}\frac{F_r(\nu)}{1+\frac{\rho}{c}J_0(\mu_0,\nu)}\geqslant\frac{F_r(\nu^r_i)}{1+\frac{\rho}{c_i}J(\mu_0,\nu^r_i)}=\frac{F_r(\nu^r_i)}{1+\frac{\rho}{c_i}J_0(\mu_0,\nu^r_i)}\geqslant\sup_{\nu\geqslant\mu}\frac{F_r(\nu)}{1+\frac{\rho}{c}J(\mu_0,\nu)}$$

$$\sup_{\nu\leqslant\mu}\frac{F_l(\nu)}{1+\frac{\rho}{c}J(\mu_0,\nu)}\geqslant\sup_{\nu\leqslant\mu}\frac{F_l(\nu)}{1+\frac{\rho}{c}J_0(\mu_0,\nu)}\geqslant\frac{F_l(\nu^l_i)}{1+\frac{\rho}{c_i}J(\mu_0,\nu^l_i)}=\frac{F_l(\nu^l_i)}{1+\frac{\rho}{c_i}J_0(\mu_0,\nu^l_i)}\geqslant\sup_{\nu\leqslant\mu}\frac{F_l(\nu)}{1+\frac{\rho}{c}J(\mu_0,\nu)}$$

$$\Longrightarrow\frac{F_r(\nu^r_i)}{1+\frac{\rho}{c_i}J(\mu_0,\nu^r_i)}=\frac{F_l(\nu^l_i)}{1+\frac{\rho}{c_i}J(\mu_0,\nu^l_i)}=\sup_{\nu\geqslant\mu}\frac{F_r(\nu)}{1+\frac{\rho}{c}J_0(\mu_0,\nu)}=\sup_{\nu\leqslant\mu}\frac{F_l(\nu)}{1+\frac{\rho}{c}J_0(\mu_0,\nu)}$$

Therefore, $\mu^*=\mu_0$ and the optimal strategy at $\mu_0$ is the same as that with $H_0(\mu)$.

Now I prove that the sign of $\mu^*-\mu_0$ strictly positive (negative) when $c\in(c_{2i},c_{2i+1})$ ($c\in(c_{2i-1},c_{2i})$). For the first case $c\in(c_{2i},c_{2i+1})$, since $J(\mu_0,\nu^{r*}(c))>J_0(\mu,\nu^{r*}(c))$ and $J(\mu_0,\nu^{l*}(c))=J_0(\mu_0,\nu^{l*}(c))$,

$$\overline{V}^+(\mu_0)=\max_{\nu\geqslant\mu}\frac{F_r(\nu)}{1+\frac{\rho}{c}J(\mu_0,\nu)}<\max_{\nu\geqslant\mu}\frac{F_r(\nu)}{1+\frac{\rho}{c}J_0(\mu_0,\nu)}$$

$$\overline{V}^-(\mu_0)=\max_{\nu\leqslant\mu}\frac{F_l(\nu)}{1+\frac{\rho}{c}J(\mu_0,\nu)}=\max_{\nu\leqslant\mu}\frac{F_r(\nu)}{1+\frac{\rho}{c}J_0(\mu_0,\nu)}$$

The first strict inequality is from uniqueness of optimal $\nu^{r*}(c)$, $J(\mu_0,\nu^{r*}(c))>J_0(\mu,\nu^{r*}(c))$ and continuity of the objective function in $\nu$. Therefore, $\overline{V}^+(\mu_0)<\overline{V}^-(\mu_0)$. Since $\overline{V}^+(\mu)$ is increasing in $\mu$ and $\overline{V}^-(\mu)$ is decreasing in $\mu$, their crossing point $\mu^*>\mu_0$. For the other case $c\in(c_{2i-1},c_{2i})$, $J(\mu_0,\nu^{r*}(c))=J_0(\mu,\nu^{r*}(c))$ and $J(\mu_0,\nu^{l*}(c))<J_0(\mu_0,\nu^{l*}(c))$. Therefore,

$$\overline{V}^+(\mu_0)=\max_{\nu\geqslant\mu}\frac{F_r(\nu)}{1+\frac{\rho}{c}J(\mu_0,\nu)}=\max_{\nu\geqslant\mu}\frac{F_r(\nu)}{1+\frac{\rho}{c}J_0(\mu_0,\nu)}$$

$$\overline{V}^-(\mu_0)=\max_{\nu\leqslant\mu}\frac{F_l(\nu)}{1+\frac{\rho}{c}J(\mu_0,\nu)}<\max_{\nu\leqslant\mu}\frac{F_r(\nu)}{1+\frac{\rho}{c}J_0(\mu_0,\nu)}$$

Therefore, $\overline{V}^+(\mu_0)>\overline{V}^-(\mu_0)$. Since $\overline{V}^+(\mu)$ is increasing in $\mu$ and $\overline{V}^-(\mu)$ is decreasing in $\mu$, their crossing point $\mu^*<\mu_0$. *Q.E.D.*

### *S4.2 Radical innovation: proof of Propositions 2 and 3*

**Proof.** Consider the solution to the problem of firm $L$, where the payoff to riskless arm is $u_L(P_s)$. By Theorem 2, the policy function $\nu_L(\nu)$ is a strictly decreasing function on experimentation region $E_L$. $\nu_L(\mu)$ is piecewise smooth. Each discrete point of $\nu_L(\mu)$ corresponds to a critical point where the DM is indifferent between confirming two beliefs associated with two different actions. Now I call the set of those critical beliefs $\{\mu_j\}_{j=1}^J$, where $\mu_J$ is the smallest and $\mu_1$ is the largest. I first prove the following useful lemma.

**Lemma S.24.** *At each $\mu_j$, let $\underline{\nu}^j_L<\bar{\nu}^j_L$ be the smallest and largest optimal posterior beliefs for firm L. Then either $\nu_S(\mu_j)<\underline{\nu}^j_L$ or $\nu_S(\mu_j)>\bar{\nu}^j_L$.*

**Proof.** Define $L(V,\lambda,\mu)(\nu)$ and $G(V,\lambda,\mu)(*)\nu$ and $G(V,\lambda)(\mu)$ as Equation (36). Consider:

$$L\Big(V_L,\frac{\rho}{c}V_L(\mu_j),\mu_j\Big)(\nu)-G\Big(F,\frac{\rho}{c}V_L(\mu_j)\Big)(\nu) \tag{S.33}$$

Optimality condition Equations (40) and (41) implies that it attains minimum 0 at both $\underline{\nu}^j_L$ and $\bar{\nu}^j_L$ (and at no other beliefs outside of the range $(\underline{\nu}^j_L),\bar{\nu}^j_L)$. Now consider:

$$L\Big(V_S,\frac{\rho}{c}V_L(\mu_j),\mu_j\Big)(\nu)-G\Big(F,\frac{\rho}{c}V_L(\mu_j)\Big)(\nu)$$

Since in $L$'s experimentation region $V_L>V_S$, the term is strictly positive for $\nu>\mu_j$. Now the optimality condition implies that

$$L\Big(V_S,\frac{\rho}{c}V_S(\mu_j),\mu_j\Big)(\nu)-G\Big(F,\frac{\rho}{c}V_S(\mu_j)\Big)(\nu) \tag{S.34}$$

attains minimum 0 at $\nu_S(\mu_j)$. Notice that Equation (S.34) is equivalent to:

$$\begin{aligned}&L\Big(V_S,\frac{\rho}{c}V_L(\mu_j),\mu_j\Big)(\nu)-G\Big(F,\frac{\rho}{c}V_L(\mu_j)\Big)(\nu)\\=&L\Big(V_L,\frac{\rho}{c}V_L(\mu_j),\mu_j\Big)(\nu)-G\Big(F,\frac{\rho}{c}V_L(\mu_j)\Big)(\nu)\end{aligned}$$

$$+\left(V_S(\mu_j)-V_L(\mu_j)+(V_S'(\mu_j)-V_L'(\mu_j))(\nu-\mu_j)\right)$$
$$+\frac{\rho}{c}(V_S(\mu_j)-V_L(\mu_j))(H(\mu)-H(\nu)+H'(\mu)(\nu-\mu))$$

Notice that the second term is linear in $\nu$ and the third term is concave in $\mu$. Since Equation (S.33) is minimized at $\underline{\nu}_L^j$ and $\bar{\nu}_L^j$, the two points share the same supporting hyperplane. Now Equation (S.34) equals Equation (S.33) plus a strictly concave term. As a result, Equation (S.34) has positive first derivative at $\underline{\nu}_L^j$ and negative first derivative at $\overline{V}_L^j$. Therefore, the posterior belief that minimizes Equation (S.34) is either strictly less than $\underline{\nu}_L^j$ or strictly larger than $\overline{\nu_L^j}$. *Q.E.D.*

*Step 1.* I prove that in the region $\mu \geqslant \mu_1$, there exists critical belief $\mu_c$ that the satisfies the property of Proposition 3. By Lemma S.24, there are two possible cases.
The first case is that $\nu_S(\mu_1) < \underline{\nu}_L^1$. Now if at belief $\mu_1$, firm $S$'s optimal posterior is already associated with a less risky action, then since by definition of $\mu_1$ firm $L$ doesn't use any action less risky than that associated with $\underline{\mu}_L^1$ at all. So $\mu_c = \mu_1$. If otherwise firm $S$'s and firm $L$'s optimal beliefs are associated with the same action, then by the previous analysis, $\nu_S(\mu_j) < \underline{\nu}_L^j$. Now I prove that $\nu_S(\mu) < \nu_L(\mu)$ for all $\mu \geqslant \mu_j$. This can be easily seen from the phase diagram Figure S.1. Since the two firms are using the same action, their optimal belief is characterized by the same set of ODEs (except for different in initial value). Since we know that $V_L > V_S$, then the policy function $\nu_L$ must touch the diagonal line later than $\nu_S$. By Picard-Lindelof solution to ODE doesn't cross, $\nu_L(\mu) > \nu_S(\mu)$ for $\mu \geqslant \mu_0$ and $\nu_L(\mu) < \nu_S(\mu)$ for $\mu \leqslant \mu_0$ ($\mu_0$ is the critical belief the action giving zero expected payoff). Giving this single crossing property, since $\nu_L(\mu_1^+) > \nu_S(\mu_1^+)$, $\nu_L(\mu) > \nu_S(\mu)$ for all $\mu \geqslant \mu_1$.
The second case is that $\nu_S(\mu_1) > \bar{\nu}_L^1$. Now for some $\mu > \mu_1$, $\nu_S(\mu)$ stays above $\nu_L(\mu)$ whenever it corresponds to a more risky action. However, the analysis in the first case shows that when firm $S$ switches action, it either jumps to a strictly less risky action, or stays at the same action as firm $L$ for some beliefs (but $\nu_L(\mu)$ and $\nu_S(\mu)$ crosses once). In either cases, the single crossing property holds. So there exists such critical belief $\mu_c$.
Notice that the analysis in this region is already sufficient to prove Proposition 2.

*Step 2.* I prove Proposition 3 by induction. I prove the following statement that *if* for $\mu_j$, there are two possibles cases: $\nu_S(\mu_j) < \underline{\nu}_L^j$ and $\nu_S(\mu) < \nu_L(\mu)\ \forall \mu \geqslant \mu_j$; or $\nu_S(\mu_j) > \bar{\nu}_L^j$ and there exists $\mu_c > \mu_j$, *then* the same statement is true for $\mu_{j+1}$.
If $\nu_S(\mu_{j+1}) < \underline{\nu}_L^{j+1}$, then the argument is simple. Case 1 is that firm $S$ has already switched to a less risky action, then before the firm $H$ switches, $\nu_L(\mu) > \nu_S(\mu)$ for sure, up to $\mu_j$. Then $\nu_L(\mu) > \nu_S(\mu)$ for $\mu \geqslant \mu_j$ as well by assumption in induction. Case 2 is that firm $S$ is using the same action as firm $L$. The by the argument in step 1, before either firm switches to a less risky action, $\nu_L(\mu) > \nu_S(\mu)$. Suppose by contradiction that firm $L$ first switches to a less risky action at $\mu_j$, then by Lemma S.24, $\nu_S(\mu_j^-) > \nu_L(\mu_j^-)$, contradiction. Therefore, to sum up $\nu_S(\mu) < \mu_L(\mu)\ \forall \mu > \underline{\nu}_L^{j+1}$.
If $\nu_S(\mu_{j+1}) > \bar{\nu}_L^{j+1}$, the we only need to discuss that firm $S$ ever uses the same action as firm (because otherwise either single crossing happens and we are in the case $\nu_S(\mu_j) < \underline{\nu}_L^j$, then the induction assumptions shows $\nu_S(\mu) < \nu_L(\mu)$ for all $\mu \geqslant \mu_j$; or crossing doesn't happen, then the induction assumptions shows that single crossing happens for $\mu_c > \mu_j$). In this case, the analysis in step 1 shows that $\nu_L(\mu)$ and $\nu_S(\mu)$ crosses at most once, and afterwards, the induction assumptions shows $\nu_S(\mu) < \nu_L(\mu)$ for all $\mu \geqslant \mu_j$. To sum up, I prove that $\nu_L$ and $\nu_S$ crosses at most once in firm $L$'s experimentation region. Notice that $V_L > V_S$, therefore, there exists $\mu$ in $L$'s experimentation region where $V_S(\mu) = F(\mu)$ already. Obviously for such belief $\nu_L(\mu) > \nu_S(\mu)$. On the other hand, on the left end of $L$'s experimentation region:

$$\frac{F(\nu)-V_L(\mu)}{H(\mu)-H(\nu)+H'(\mu)(\nu-\mu)} < \frac{F(\nu)-V_S(\mu)}{H(\mu)-H(\nu)+H'(\mu)(\nu-\mu)} \leqslant V_S(\mu)$$

So since $V_L(\mu) > V_S(\mu)$ it must be that $V_S'(\mu) > 0$. Therefore, there exist $\mu$ in $S$'s experimentation region where $V_L(\mu) = F(\mu)$. This proves that $E_0 \bigcap (0,\mu_c) \neq \varnothing$ and $E_0 \bigcap (\mu_c, 1) \neq \varnothing$. *Q.E.D.*

## S5 Proofs in Appendix A

### S5.1 *Convergence of policy*

#### S5.1.1 *Proof of Theorem 7*

**Proof.** The original statement in Theorem 7 is equivalent to: $\forall \varepsilon > 0$, there exists $\delta$ s.t. $\forall dt \leqslant \delta$, $\forall \mu \in [0,1]$, there exists $|\mu' - \mu| \leqslant \varepsilon$ and any optimal posterior induced in discrete time problem with period length $dt$ will be within either $[\mu' - \varepsilon, \mu' + \varepsilon]$ or $[\nu(\mu') - \varepsilon, \nu(\mu') + \varepsilon]$.Now pick any $\varepsilon > 0$, let's discuss two cases separately:

*Case 1:* $\mu\in[0,1]\backslash E$. I first prove the case with Assumption A and Assumptions 2-a and 3. I will show that for any $dt$, any informative experiment is suboptimal. Suppose not, and there exists $\nu_i\neq\mu$ s.t.:

$$e^{-\rho dt}\sum p_i V_{dt}(\nu_i)\geqslant V_{dt}(\mu)$$

and

$$\begin{cases}\sum p_i\nu_i=\mu\\ H(\mu)-\sum p_i H(\nu_i)\leqslant cdt\end{cases}$$

Now consider a problem with $\frac{dt}{2}$. Consider the following strategy: mix experiment $p_i,\nu_i$ and prior with probability $\frac{1}{2}$. Then obviously Bayes plausibility and capacity constraint are satisfied. The expected utility from this strategy is:

$$\begin{aligned}V_{\frac{dt}{2}}(\mu)\geqslant&\sum_{t=1}e^{-\rho\frac{dt}{2}\cdot t}\sum p_i V_{\frac{dt}{2}}(\nu_i)\cdot\frac{1}{2^t}=\frac{1}{2-e^{-\frac{\rho dt}{2}}}\sum p_i V_{\frac{dt}{2}}(\nu_i)\\ >&e^{-\rho dt}\sum p_i V_{\frac{dt}{2}}(\nu_i)\\ \geqslant&e^{-\rho dt}\sum p_i V_{dt}(\nu_i)\\ \geqslant&F(\mu)\end{aligned}$$

First inequality is from optimality of $V_{\frac{dt}{2}}$. Second inequality is from $\frac{1}{2-x}>x^2$ for $x\in(0,1)$. Third inequality is from $V_{\frac{dt}{2}}\geqslant V_{dt}$. Last inequality is from assumption. Therefore $F(\mu)=V(\mu)\geqslant V_{\frac{dt}{2}}>F(\mu)$. Contradiction. So for $\mu\in[0,1]$, $N_{dt}(\mu)=\{\mu\}$ for any $dt>0$. Noticing that this satisfies Theorem 7 independent of choice of $dt$ and $\varepsilon$.

Then consider the case with Assumption A and Assumptions 2-b and 3. Suppose not true, and there exists $\nu_i\neq\mu$ s.t.:

$$e^{-\rho dt}\sum p_i V_{dt}(\nu_i)-dt\cdot C\left(\frac{I}{dt}\right)\geqslant V_{dt}(\mu)$$

and

$$\begin{cases}\sum p_i\nu_i=\mu\\ H(\mu)-\sum p_i H(\nu_i)=I\end{cases}$$

Now consider a problem with $\frac{dt}{2}$. Consider the following strategy: mix experiment $p_i,\nu_i$ and prior with probability $\frac{1}{2}$. Then obviously Bayes plausibility and capacity constraint are satisfied. The expected utility from this strategy is:

$$\begin{aligned}&V_{\frac{dt}{2}}(\mu)\\ \geqslant&\sum_{t=1}e^{-\rho\frac{dt}{2}\cdot t}\sum p_i V_{\frac{dt}{2}}(\nu_i)\cdot\frac{1}{2^t}-\sum_{t=0}e^{-\rho\frac{dt}{2}t}\frac{1}{2^t}\frac{dt}{2}C\left(\frac{I}{dt}\right)\\ =&\frac{e^{\rho dt}}{2-e^{-\frac{\rho dt}{2}}}\left(e^{-\rho dt}\sum p_i V_{\frac{dt}{2}}(\nu_i)-e^{-\frac{3\rho dt}{2}}dtC\left(\frac{I}{dt}\right)\right)\\ >&e^{-\rho dt}\sum p_i V_{\frac{dt}{2}}(\nu_i)-dtC\left(\frac{I}{dt}\right)\\ \geqslant&e^{-\rho dt}\sum p_i V_{dt}(\nu_i)-dtC\left(\frac{I}{dt}\right)\\ \geqslant&F(\mu)\end{aligned}$$

First inequality is from optimality of $V_{\frac{dt}{2}}$. Second inequality is from $\frac{1}{2-x}>x^2$ for $x\in(0,1)$. Third inequality is from $V_{\frac{dt}{2}}\geqslant V_{dt}$. Last inequality is from assumption. Therefore $F(\mu)=V(\mu)\geqslant V_{\frac{dt}{2}}>F(\mu)$. Contradiction. So for $\mu\in[0,1]$, $N_{dt}(\mu)=\{\mu\}$ for any $dt>0$. Noticing that this satisfies Theorem 7 independent of choice of $dt$ and $\varepsilon$.

*Case 2*: $\mu\in E$. Suppose Theorem 7 is not true. Then there exists $\varepsilon$ s.t. $\forall dt$, there exists $\mu_{dt}\in E$ s.t. $\exists\nu_{dt}\in N_{dt}(\mu_{dt})$ and $\forall\mu\in B_\varepsilon(\mu_{dt})$, $d_H(\nu_{dt},N(\mu))>\varepsilon$. Now pick $dt_n=2^{-n}\to0$. Since $(\mu_{dt_n},\nu_{dt_n})$ is an infinite sequence in compact space $[0,1]^2$, we can WLOG assume $(\mu_{dt_n},\nu_{dt_n})\to(\mu,\nu)$. $\forall\mu'\in B_{\frac{\varepsilon}{2}}(\mu)$, there exists $N$ sufficiently large that $\forall n\geqslant N$, $\mu'\in B_\varepsilon(\mu_{dt_n})$ and $d_H(\nu_{dt_n},N(\mu'))>\varepsilon$. $N$ can be picked sufficiently large that $|\nu_{dt_n}-\nu|<\frac{\varepsilon}{2}$. Therefore $d_H(\nu,N(\mu'))>\frac{\varepsilon}{2}$. To sum up, we find a converging sequence $(\mu_{dt_n},\nu_{dt_n})$ to $(\mu,\nu)$, which is bounded away by $\frac{\varepsilon}{2}$ from the graph of $N(\cdot)$.

Let $\widetilde{\nu}$ be the non-empty set of optimal posteriors (including $\mu$ itself) at $\mu$ solving Equation (4). Let

$$\lambda=\begin{cases}\frac{\rho}{c}V(\mu) & \text{with Assumption 2-a}\\ C'(I(\mu)) & \text{with Assumption 2-b}\end{cases}$$

Consider:

$$G(\cdot)=V(\cdot)+\lambda H(\cdot)$$

Then optimality condition implies that:

$$\begin{cases} G(\nu)=G(\mu)+G'(\mu)(\nu-\mu) & \forall \nu\in\tilde{\nu} \\ G(\nu)<G(\mu)+G'(\mu)(\nu-\mu) & \text{otherwise} \end{cases} \tag{S.35}$$

By Theorem S.1, $\forall dt$, there exists $\lambda_{dt_n}$ s.t. Equation (6) is solved by concavifying $G_n=V_{dt_n}+\lambda_{dt_n}H$ at $\mu_{dt_n}$.

Obviously $\lambda_{dt_n}$ is non-negative. Suppose it diverges to $+\infty$. Then consider function $G_{dt_n}(\cdot)=V_{dt_n}(\cdot)+\lambda_{dt_n}H(\cdot)$. Let $\nu_{dt_n,1}$ and $\nu_{dt_n,2}$ be two optimal posterior. Let $\nu'_{dt_n}=\frac{1}{2}(\nu_{dt_n,1}+\nu_{dt_n,2})$. By $|\nu_{dt_n}-\mu_{dt_n}|>\frac{\varepsilon}{2}$, we know that $|\nu_{dt_n,1}-\nu_{dt_n,2}|>\frac{\varepsilon}{2}$ by Lemma S.25.

$$\begin{aligned} & G_{dt_n}(\nu'_{dt_n})-\frac{1}{2}(G_{dt_n}(\nu_{dt_n,1})+G_{dt_n}(\nu_{dt_n,2})) \\ =& V_{dt_n}(\nu'_{dt_n})-\frac{1}{2}(V_{dt_n}(\mu_{dt_n,1})+V_{dt_n}(\nu_{dt_n,1})) \\ & +\lambda_{dt_n}\left(H(\nu'_{dt_n})-\frac{1}{2}(H(\nu_{dt_n,1})+H(\nu_{dt_n,2}))\right) \\ \leqslant & \sup F+\lambda_{dt_n}\frac{1}{8}\sup H''(\nu_{dt_n,1}-\nu_{dt_n,2})^2\to-\infty \end{aligned}$$

So for $n$ sufficiently large, $\mu'_{dt_n}$ will be higher than the connected straight line of $\mu_{dt_n}$ and $\nu_{dt_n}$ on $G_{dt_n}$. Contradicting optimality of $\nu_{dt_n}$. So $\lambda_{dt_n}$ is a bounded sequence.

Suppose there exists convergent subsequence $\lim\lambda_{dt_n}<\lambda$. If $\mu\neq\mu^*$ then pick $\mu'$ such that $\mu'$ is in the same interval as $\mu$ in $E$ and $\lim\lambda_{dt_n}<\lambda(\mu')<\lambda$, let $\lambda'=\lambda(\mu')$. If $\mu=\mu^*$ then let $\lambda(\mu^*)>\lambda'>\lim\lambda_{dt_n}$. Now consider concavifying $V+\lambda'H$. By monotonicity in Theorem 2 and definition of $\mu^*$, we know that optimal posteriors are bounded away from $\mu$. Moreover, $V(\mu)+\lambda'H(\mu)<\text{cov}(V+\lambda'H)(\mu)$. Pick $\varepsilon>0$ sufficiently small such that optimal posteriors of $V+\lambda'H$ are bounded away from $\mu$ by $\varepsilon$ and $V(\mu)+\lambda'H(\mu)+\varepsilon<\text{cov}(V+\lambda'H)(\mu)$. Let $\nu_1<\mu<\nu_2$ be two optimal posterios for $V+\lambda'H$ closest to $\mu$. By continuity, there exists $\delta$ s.t $\forall|\mu''-\mu|<\delta$, $V(\mu'')+\lambda'H(\mu'')+\frac{\varepsilon}{2}<\text{cov}(V+\lambda'H)(\mu'')$. Pick $dt_n$ s.t. $\|V_{dt_n}-V\|<\frac{\varepsilon}{8}$, then

$$\begin{aligned} & V_{dt_n}(\mu'')+\lambda'H(\mu'') \\ <& V(\mu'')+\lambda'H(\mu'')+\frac{\varepsilon}{8} \\ <& \text{cov}(V+\lambda'H)(\mu'')-\frac{3\varepsilon}{8} \\ \leqslant & \text{cov}(V_{dt_n}+\lambda'H)(\mu'')-\frac{\varepsilon}{4} \end{aligned}$$

The last inequality comes from the fact that any convex combination of points on $V+\lambda'H$ is less than $\frac{\varepsilon}{8}$ higher than convex combination of those points on $V_{dt_n}+\lambda'H$, therefore less than $\text{cov}(V_{dt_n}+\lambda'H)+\frac{\varepsilon}{8}$. Therefore, we showed that any point $\mu''$ within $\delta$ ball of $\mu$ can't be on supporting hyperplane of $V_{dt_n}+\lambda'H$. So any optimal posterior of $V_{dt_n}+\lambda'H$ is bounded away from $\mu$ by $\delta$. Pick $N$ sufficently large than $\forall n\geqslant N$, $|\mu_{dt_n}-\mu|<\frac{\delta}{2}$. Then, optimal posterior of $V_{dt_n}+\lambda'H$ is bounded away from $\mu_{dt_n}$ by $\frac{\delta}{2}$. By definition of $\lambda'$, $N$ can be picked sufficently large that $\forall n\geqslant N$ $\lambda_{dt_n}\leqslant\lambda'$. Therefore, by Lemma S.25, optimal posteriors are even further from $\mu_{dt_n}$. To sum up, we found $N$ s.t. $\forall n\geqslant N$, the optimal posteriors from concavifying $V_{dt_n}+\lambda_{dt_n}H$ are bounded away from $\mu_{dt_n}$ by $\frac{\delta}{2}$. The experimentation cost of any such information structure is:

$$\begin{aligned} & \sum p_i(H(\mu_{dt_n})-H(\nu_{dt_n,i})) \\ =& \sum p_i\left(H'(\mu_{dt_n})(\mu_{dt_n}-\nu_{dt_n,i})-\frac{1}{2}H''(\xi_i)(\nu_{dt_n,i}-\mu_{dt_n})^2\right) \\ \geqslant & -\sup H''\frac{\delta^2}{4}>0 \end{aligned}$$

Therefore, for sufficently large $n$, experimentation cost will eventually exceed $cdt_n$. Contradiction.

Suppose there exists subsequence $\lim\lambda_{dt_n}=\lambda'\geqslant\lambda$. By definition, there exists linear function $L_n(\mu)$ s.t.

$$\begin{cases} G_n(\nu_{dt_n})=L_n(\nu_{dt_n}) \\ G_n(\mu_{dt_n})=L_n(\mu_{dt_n}) \\ G_n(\nu)\leqslant L_n(\nu) \end{cases}$$

Since $\lambda_{dt_n} \to \lambda'$, $G_n$ is bounded at $\mu_{dt_n}$ and $\nu_{dt_n}$. Therefore, $L_n$ has bounded slope and constant term. It's easy to see that $L_n$ will converge uniformly to linear function $L_\infty$ on belief space $\Delta X$. Moreover, $\forall \nu \in \Delta X$:

$$\begin{aligned} G_n(\nu) &= V_{dt_n}(\nu) + \lambda_{dt_n} H(\nu) \to V(\nu) + \lambda' H(\nu) = \widetilde{G}(\nu) \leqslant L_\infty(\nu) \\ G_n(\mu_{dt_n}) &= V_{dt_n}(\mu_{dt_n}) + \lambda_{dt_n} H(\mu_{dt_n}) \to \widetilde{G}(\mu) = L_\infty(\mu) \\ G_n(\nu_{dt_n}) &= V_{dt_n}(\nu_{dt_n}) + \lambda_{dt_n} H(\nu_{dt_n}) \to \widetilde{G}(\nu) = L_\infty(\nu) \end{aligned} \tag{S.36}$$

Second and third convergence comes from $V_{dt_n}$ uniformly convergent and $V$ continuous. $\widetilde{G}(\mu) = G(\mu) + (\lambda' - \lambda) H(\mu)$. Equation (S.36) implies that $L_\infty$ is a supporting hyperplane of graph of $\widetilde{G}$, tangents $\widetilde{G}$ at $\mu$ and $\nu$. Since $\widetilde{G}$ is a smooth function, we know that $\widetilde{G}'(\mu) = \frac{\widetilde{G}(\nu) - \widetilde{G}(\mu)}{\nu - \mu}$. On the other hand, Equation (S.35) implies that:

$$\begin{aligned} & G(\nu) < G(\mu) + G'(\mu)(\nu - \mu) \\ \Longrightarrow & \left(\widetilde{G}(\nu) - \widetilde{G}(\mu)\right) - (\lambda' - \lambda)(H(\nu) - H(\mu)) < \left(\widetilde{G}'(\mu) - (\lambda' - \lambda) H'(\mu)\right)(\nu - \mu) \\ \Longrightarrow & \widetilde{G}(\nu) - \widetilde{G}(\mu) < \widetilde{G}'(\mu)(\nu - \mu) \end{aligned}$$

Contradiction. Last inequality is from concavity of $H: H(\nu) - H(\mu) < H'(\mu)(\nu - \mu)$. Therefore Theorem 7 is true.

*Q.E.D.*

**Lemma S.25.** *Let $X$ be closed interval in $\mathbb{R}$. Let $V$ be a continuous function on $X$, $H$ be a concave function on $X$. Let $E_\lambda = \{x \in X | cov(V + \lambda H)(x) > V(x) + \lambda H(x)\}$. Then $\{E_\lambda\}$ are ordered monotonically as $\lambda$ by set inclusion: if $\lambda \geqslant \lambda'$, then $\forall$ interval $I$ in $E_\lambda$, there exists interval $I'$ in $E_{\lambda'}$ s.t. $I \subset I'$.*

**Proof.** $\forall \lambda$, take any $I \in E_\lambda$. Let $I = [x, y]$. Define:

$$L(z) = V(x) + \lambda H(x) + \frac{V(y) - V(x) + \lambda H(y) - \lambda H(x)}{y - x}(z - x)$$

Then $\forall z \in X$:

$$\begin{cases} L(x) = V(x) + \lambda H(x) \\ L(y) = V(y) + \lambda H(y) \\ L(z) > V(z) + \lambda H(z) & \text{if } z \in (x, y) \\ L(z) \geqslant V(z) + \lambda H(z) & \text{if } z < x \text{ or } z > y \end{cases}$$

Now take any $\lambda' < \lambda$ and consider $V + \lambda' H$. Let:

$$\begin{aligned} \widetilde{L}(z) &= V(x) + \lambda' H(x) + \frac{V(y) - V(x) + \lambda' H(y) - \lambda' H(x)}{y - x}(z - x) \\ &= L(z) + (\lambda' - \lambda)\left(H(x) + \frac{H(y) - H(x)}{y - x}(z - x)\right) \\ & \begin{cases} \geqslant L(z) + (\lambda' - \lambda) H(z) & \text{if } z \in [x, y] \\ \leqslant L(z) + (\lambda' - \lambda) H(z) & \text{if } z \notin [x, y] \end{cases} \\ \Longrightarrow & \begin{cases} \widetilde{L}(x) = V(x) + \lambda' H(x) \\ \widetilde{L}(y) = V(y) + \lambda' H(y) \\ \widetilde{L}(z) > V(z) + \lambda' H(z) \text{ if } z \in (x, y) \end{cases} \end{aligned}$$

Therefore, $\forall z \in (x, y)$, $\text{cov}(V + \lambda' H)(z) > V(z) + \lambda' H(z)$. So there exists interval $I' \in E_{\lambda'}$ s.t. $I \subset I'$. *Q.E.D.*

### S5.2 Continuum of actions

#### S5.2.1 Proof of Lemma A.1

**Proof.** We prove with two steps:

*Step 1*: We first show that if we let $\mathcal{V}_{dt}(F)$ be the solution to Equation (6), then $\mathcal{V}_{dt}$ is Lipschitz continuous in $F$ under $L_\infty$ norm. $\forall F_1, F_2$ convex and with bounded subdifferentials, consider $\overline{F} = \max\{F_1, F_2\}$, $\underline{F} = \min\{F_1, F_2\}$. Then by properties of convex functions, $\overline{F}, \underline{F}$ are convex. $\partial \underline{F}(\mu), \partial \overline{F}(\mu) \subset \partial F_1(\mu) \bigcup \partial F_2(\mu)$. Therefore $\overline{F}$ and $\underline{F}$ are both within the domain of convex and bounded subdifferential functions with the following quantitative property:

$$\begin{cases} \overline{F} \geqslant F_1, F_2 \geqslant \underline{F} \\ |\overline{F} - \underline{F}| = |F_1 - F_2| \end{cases}$$

It's not hard to see that $\mathcal{V}$ is monotonicaly increasing in $F$. Therefore, we have:

$$\mathcal{V}_{dt}(\underline{F}) \leqslant \mathcal{V}_{dt}(F_1), \mathcal{V}_{dt}(F_2) \leqslant \mathcal{V}_{dt}(\overline{F})$$

Now let $(p_i,\mu_i)$ be the policy solving $\mathcal{V}_{dt}(\overline{F})$. Let $\overline{V}_{dt}=\mathcal{V}_{dt}(\overline{F}), \underline{V}_{dt}=\mathcal{V}_{dt}(\underline{F})$. Let $C$ be total expected cost associate with this strategy. Then consider:

$$\begin{aligned}
\underline{V}_{dt}(\mu) \geqslant & \mathbf{1}_{\overline{V}_{dt}(\mu)\leqslant\overline{F}(\mu)}\underline{F}(\mu)+\mathbf{1}_{\overline{V}_{dt}(\mu)>\overline{F}(\mu)}e^{-\rho dt}\sum p_i^1(\mu)\underline{V}_{dt}(\mu_i^1)-C \\
\geqslant & \mathbf{1}_{\overline{V}_{dt}(\mu)\leqslant\overline{F}(\mu)}\underline{F}(\mu)+\mathbf{1}_{\overline{V}_{dt}(\mu)>\overline{F}(\mu)}e^{-\rho dt}\sum p_i^1(\mu)\mathbf{1}_{\overline{V}_{dt}(\mu_i^1)\leqslant\overline{F}(\mu_i^1)}\underline{F}(\mu_i^1)-C \\
& +\mathbf{1}_{\overline{V}_{dt}(\mu)>\overline{F}(\mu)}e^{-2\rho dt}\sum p_i^1(\mu)\mathbf{1}_{\overline{V}_{dt}(\mu_i^1)>\overline{F}(\mu_i^1)}\sum p_i^2(\mu_i^1)\underline{V}_{dt}(\mu_i^2)-C \\
\geqslant & \cdots \\
= & \sum_t e^{-\rho t\cdot dt}\sum_{i^1,\dots,i^{t-1}}\prod p_i^\tau(\mu_i^{\tau-1})\mathbf{1}_{\overline{V}_{dt}(\mu_i^\tau)>\overline{F}(\mu_i^\tau)}\sum p_i^t(\mu_i^{t-1})\mathbf{1}_{\overline{V}_{dt}(\mu_i^t)\leqslant\overline{F}(\mu_i^t)}\underline{F}(\mu_i^t)-C \\
\geqslant & \sum_t e^{-\rho t\cdot dt}\sum_{i^1,\dots,i^{t-1}}\prod p_i^\tau(\mu_i^{\tau-1})\mathbf{1}_{\overline{V}_{dt}(\mu_i^\tau)>\overline{F}(\mu_i^\tau)}\sum p_i^t(\mu_i^{t-1})\mathbf{1}_{\overline{V}_{dt}(\mu_i^t)\leqslant\overline{F}(\mu_i^t)}\overline{F}(\mu_i^t)-C \\
& -\sum_t\sum_{i^1,\dots,i^{t-1}}\prod p_i^\tau(\mu_i^{\tau-1})\mathbf{1}_{\overline{V}_{dt}(\mu_i^\tau)>\overline{F}(\mu_i^\tau)}\sum p_i^t(\mu_i^{t-1})\mathbf{1}_{\overline{V}_{dt}(\mu_i^t)\leqslant\overline{F}(\mu_i^t)}\left|\overline{F}-\underline{F}\right| \\
= & \overline{V}_{dt}(\mu)-\left|\overline{F}-\underline{F}\right|
\end{aligned}$$

Therefore, $\left|\overline{V}_{dt}-\underline{V}_{dt}\right|\leqslant\left|\overline{F}-\underline{F}\right| \Longrightarrow |\mathcal{V}_{dt}(F_1)-\mathcal{V}_{dt}(F_2)|\leqslant|F_1-F_2|$. $\mathcal{V}_{dt}(F)$ has Lipschitz parameter 1.

*Step 2*: $\forall F_1,F_2$, $\forall\varepsilon>0$, by Lemma 3, there exists $dt$ s.t. $|\mathcal{V}(F_i)-\mathcal{V}_{dt}(F_i)|\leqslant\varepsilon|F_1-F_2|$. Therefore:

$$\begin{aligned}
|\mathcal{V}(F_1)-\mathcal{V}(F_2)| \leqslant & |\mathcal{V}(F_1)-\mathcal{V}_{dt}(F_1)|+|\mathcal{V}(F_2)-\mathcal{V}_{dt}(F_2)|+|\mathcal{V}_{dt}(F_1)-\mathcal{V}_{dt}(F_2)| \\
\leqslant & (1+2\varepsilon)|F_1-F_2|
\end{aligned}$$

Take $\varepsilon\to0$, since LHS is not a function of $\varepsilon$, we conclude that $\mathcal{V}(F)$ is Lipschitz continuous in $F$ with Lipschitz parameter 1. *Q.E.D.*

### S5.2.2 *Proof of Theorem 8*

**Proof.** We prove the three main results in following steps:

- *Lipschitz continuity*. By Lemma A.1, we directly get Lipschitz continuity of operator $\mathcal{V}$ on $\{F_n,F\}$ and the Lipschitz parameter being 1.
- *Convergence of derivatives*. Let $V_n=\mathcal{V}(F_n)$, $V=\mathcal{V}(F)$, we show that $\forall\mu$ s.t. $V(\mu)>F(\mu)$, $V'(\mu)=\lim V_n'(\mu)$. Since $V(\mu)>F(\mu)$, by continuity strict inequality holds in an closed interval $[\mu_1,\mu_2]$ around $\mu$. Then by Lemma S.27, $\lim_{n\to\infty}V_n'(\mu')$ exists $\forall\mu'\in[\mu_1,\mu_2]$. Now consider function $V_n'(\mu)$. Since $V_n''(\mu)$ is uniformly bounded for all $n$, $V_n'(\mu)$ are uniformly Lipschitz continuous, thus equicontinuous and totally bounded. Therefore by lemma Arzela-Ascolli, $V_n'$ converges uniformly to $\lim V_n'$. By convergence theorem of derivatives, $V'=\lim V_n'$ on $[\mu_1,\mu_2]$. Therefore, $V'(\mu)=\lim_{n\to\infty}V_n'(\mu)$.
- *Feasibility*. For $\mu$ s.t. $V(\mu)=F(\mu)$, feasibility is trivial. Now we discuss the case $V(\mu)>F(\mu)$. We only prove for $\mu>\mu^*$ and $\mu=\mu^*$, the case $\mu<\mu^*$ follows by symmetry. If $\mu>\mu^*$, there exists $N$ s.t. $\forall n\geqslant N$, $\mu>\mu_n^*$. $N$ can be picked large enough that in a closed interval around $\mu$, $V_n(\mu)>F_n(\mu)$. Therefore, there exists $\nu_n$ being maximizer for $V_n(\mu)$ bounded away from $\mu$ and satisfying:

  $$V_n(\mu)=\frac{c}{\rho}\frac{F_n(\nu_n)-V_n(\mu)-V_n'(\mu)(\nu_n-\mu)}{J(\mu,\nu_n)}$$

  Pick a converging subsequence $\nu_n\to\nu$:

  $$\begin{aligned}
  & \frac{c}{\rho}\frac{F(\nu)-V(\mu)-V'(\mu)(\nu-\mu)}{J(\mu,\nu_n)} \\
  = & \lim_{n\to\infty}\frac{c}{\rho}\frac{F_n(\nu_n)-V_n(\nu)-V_n'(\nu)(\nu_n-\mu)}{J(\mu,\nu_n)} \\
  = & \lim_{n\to\infty}V_n(\mu) \\
  = & V(\mu)
  \end{aligned}$$

  Therefore $V(\mu)$ is feasible in Equation (18).

Suppose $\mu=\mu^*$. Then there exists a subsequence of $\mu_n^*$ converging from one side of $\mu^*$. Suppose they are converging from left. Then $\mu\geqslant\mu_n^*$. Previous proof still works. Essentially, what we showed is that the limit of strategy in discrete action problem achieves $V(\mu)$ in the continuous action limit.

- *Unimprovability*. First, when $\mu\in\{0,1\}$, information provides no value but discounting is costly, therefore $V(\mu)$ is unimprovable. We now show unimprovability on $(0,1)$ by adding more feasible information acquisition strategies in several steps.

  - *Step 1*. Poisson experiments at $V(\mu)>F(\mu)$. In this step, we show that $\forall\mu\geqslant\mu^*$ and $V(\mu)>F(\mu)$:
  $$\rho V(\mu)=\max_{\nu\geqslant\mu} c\frac{F(\nu)-V(\mu)-V'(\mu)(\nu-\mu)}{J(\mu,\nu)}$$
  Suppose not true, then there exists $\nu$ s.t.:
  $$\begin{aligned}\lim_{n\to\infty}\rho V_n(\mu)=&\rho V(\mu)\\ &<c\frac{F(\nu)-V(\mu)-V'(\nu)(\nu-\mu)}{J(\mu,\nu)}\\ &=\lim_{n\to\infty}c\frac{F_n(\nu)-V_n(\mu)-V_n'(\mu)(\nu-\mu)}{J(\mu,\nu)}\\ &\leqslant\lim_{n\to\infty}\rho V_n(\mu)\end{aligned}$$
  Second line is by the contradictory assumption. Third line is by convergence of $F_n$ by assumption, convergence of $V_n$ by Lemma A.1 and convergence of $V_n'$ by Lemma S.27. Last inequality is by suboptimality of $\nu$.
  Similarly, for the case $\mu\leqslant\mu^*$, we can apply a symmetric argument to prove.

  - *Step 2*. Poisson experiments at $V(\mu)=F(\mu)$. In this step, we show that $\forall\mu\geqslant\mu^*$ and $V(\mu)=F(\mu)$ (The symmetric case $\mu\leqslant\mu^*$ is ommited).
  First of all, we show that $V$ is differentiable at $\mu$ and $V'(\mu)=F'(\mu)$. Suppose not, then since $V(\mu)=F(\mu)$ and $V\geqslant F$, we know that $V-F$ is locally minimized at $\mu$. Therefore $DV_+(\mu)>DV_-(\mu)$. By Definition 2, there exists $\varepsilon>0$, $\mu_1^n\nearrow\mu$ and $\mu_2^n\searrow\mu$ s.t. $\frac{V(\mu_2^n)-V(\mu)}{\mu_2^n-\mu}\geqslant\varepsilon+\frac{V(\mu)-V(\mu_1^n)}{\mu-\mu_1^n}$. Let $\delta_1^n=\mu-\mu_1^n,\delta_2^n=\mu_2^n-\mu$, this implies:
  $$\begin{aligned}&\frac{\mu-\mu_1^n}{\mu_2^n-\mu_1^n}(V(\mu_2^n)-V(\mu))+\frac{\mu_2^n-\mu}{\mu_2^n-\mu_1^n}(V(\mu_1^n)-V(\mu))\geqslant\varepsilon\frac{(\mu_2^n-\mu)(\mu-\mu_1^n)}{\mu_2^n-\mu_1^n}\\ \Longrightarrow&\frac{\mu-\mu_1^n}{\mu_2^n-\mu_1^n}V(\mu_2^n)+\frac{\mu_2^n-\mu}{\mu_2^n-\mu_1^n}V(\mu_1^n)\geqslant V(\mu)+\varepsilon\cdot\min\{\delta_1^n,\delta_2^n\}\end{aligned}$$
  On the other hand:
  $$\begin{aligned}&\frac{\mu-\mu_1^n}{\mu_2^n-\mu_1^n}(H(\mu)-H(\mu_2^n))+\frac{\mu_2^n-\mu}{\mu_2^n-\mu_1^n}(H(\mu)-H(\mu_1^n))\\ =&\frac{\mu-\mu_1^n}{\mu_2^n-\mu_1^n}\left(H'(\mu)(\mu-\mu_2^n)+\frac{1}{2}H''(\xi_2^n)(\mu-\mu_2^n)\right)\\ &+\frac{\mu_2^n-\mu}{\mu_2^n-\mu_1^n}\left(H'(\mu)(\mu-\mu_1^n)+\frac{1}{2}H''(\xi_1^n)(\mu-\mu_1^n)^2\right)\\ =&\frac{1}{2}\frac{(\mu_2^n-\mu)(\mu-\mu_1^n)}{\mu_2^n-\mu_1^n}(H''(\xi_2^n)(\mu_2^n-\mu)+H''(\xi_1^n)(\mu-\mu_1^n))\end{aligned}$$
  $\xi_1^n$ and $\xi_2^n$ are determined by applying intermediate value theorem on $H'$. Now we can choose $N$ s.t. $\forall n\geqslant N$, $\max_{\mu'\in[\mu_1^n,\mu_2^n]}\{H''(\mu')\}\leqslant 2H''(\mu)$. Therefore:
  $$\begin{aligned}&\frac{\mu-\mu_1^n}{\mu_2^n-\mu_1^n}(H(\mu)-H(\mu_2^n))+\frac{\mu_2^n-\mu}{\mu_2^n-\mu_1^n}(H(\mu)-H(\mu_1^n))\\ \leqslant&H''(\mu)(\mu_2^n-\mu)(\mu-\mu_1^n)\\ =&H''(\mu)\delta_1^n\delta_2^n\end{aligned}$$
  Now we consider a stationary experiment at $\mu$ that takes any experiment with posteriors $(\mu_1^n,\mu_2^n)$ with flow probability $\frac{c}{H''(\mu)\delta_1^n\delta_2^n}$. Then by definition the flow cost of this information acquisition strategy is less than $c$, thus is

feasible. The expected utility is:

$$\begin{aligned}\widetilde{V}(\mu)=&\frac{c}{\rho}\frac{\frac{\mu-\mu_1^n}{\mu_2^n-\mu_1^n}V(\mu_2^n)+\frac{\mu_2^n-\mu}{\mu_2^n-\mu_1^n}V(\mu_1^n)-\widetilde{V}(\mu)}{\frac{\mu-\mu_1^n}{\mu_2^n-\mu_1^n}\left(H(\mu)-H(\mu_2^n)\right)+\frac{\mu_2^n-\mu}{\mu_2^n-\mu_1^n}\left(H(\mu)-H(\mu_1^n)\right)}\\ \geqslant&\frac{V(\mu)-\widetilde{V}(\mu)+\varepsilon\min\{\delta_1^n,\delta_2^n\}}{H''(\mu)\delta_1^n\delta_2^n}\\ \implies\widetilde{V}(\mu)\geqslant&\frac{V(\mu)+\varepsilon\min\{\delta_1^n,\delta_2^n\}}{1+\frac{\rho}{c}H''(\mu)\delta_1^n\delta_2^n}\\ =&V(\mu)+\frac{\varepsilon\min\{\delta_1^n,\delta_2^n\}-\frac{\rho}{c}H''(\mu)\delta_1^n\delta_2^n}{1+\frac{\rho}{c}H''(\mu)\delta_1^n\delta_2^n}\\ =&V(\mu)+\min\{\delta_1^n,\delta_2^n\}\frac{\varepsilon-H''(\mu)\max\{\delta_1^n,\delta_2^n\}}{1+\frac{\rho}{c}H''(\mu)\delta_1^n\delta_2^n}\end{aligned}$$

$n$ can be pick large enough that $\varepsilon-H''(\mu)\max\{\delta_1^n,\delta_2^n\}$ is positive. Therefore $\widetilde{V}(\mu)>V(\mu)$. Now fix $n$ and define:

$$\begin{aligned}\widetilde{V}_m(\mu)=&\frac{c}{\rho}\frac{\frac{\mu-\mu_1^n}{\mu_2^n-\mu_1^n}V_m(\mu_2^n)+\frac{\mu_2^n-\mu}{\mu_2^n-\mu_1^n}V_m(\mu_1^n)-\widetilde{V}_m(\mu)}{\frac{\mu-\mu_1^n}{\mu_2^n-\mu_1^n}\left(H(\mu)-H(\mu_2^n)\right)+\frac{\mu_2^n-\mu}{\mu_2^n-\mu_1^n}\left(H(\mu)-H(\mu_1^n)\right)}\\ \implies\lim_{m\to\infty}\widetilde{V}_m(\mu)=&\widetilde{V}(\mu)>\lim_{m\to\infty}V_m(\mu)\end{aligned}$$

There exists $m$ large enough that $\widetilde{V}_m(\mu)>V_m(\mu)$, violating optimality of $V_m$. Contradiction. Therefore, we showed that $V'(\mu)=F'(\mu)$.

Next we show unimperovability. Suppose not, then $\exists\nu$ s.t.:

$$F(\mu)<\frac{c}{\rho}\frac{F(\nu)-F(\mu)-F'(\mu)(\nu-\mu)}{J(\mu,\nu)}$$

By continuity of $V$, $\exists\varepsilon$ and a neighbourhood $\mu\in O$, $\forall\mu'\in O$:

$$V(\mu')+\varepsilon\leqslant\frac{c}{\rho}\frac{F(\nu)-V(\mu')-F'(\mu)(\nu-\mu')}{J(\mu',\nu)}$$

By uniform convergence of $F_n$ and $V_n$, there exists $\varepsilon>0$ and $N$ s.t. $\forall n\geqslant N$:

$$\begin{aligned}V_n(\mu')+\frac{\varepsilon}{2}\leqslant&\frac{c}{\rho}\frac{F_n(\nu)-V_n(\mu')-F'(\mu)(\nu-\mu')}{J(\mu',\nu)}\\ \implies\frac{c}{\rho}\frac{F_n(\nu)-V_n(\mu')-V_n'(\mu')(\nu-\mu')}{J(\mu',\nu)}+\frac{\varepsilon}{2}\leqslant&\frac{c}{\rho}\frac{F_n(\nu)-V_n(\mu')-F'(\mu)(\nu-\mu')}{J(\mu',\nu)}\\ \implies V_n'(\mu')\geqslant&F'(\mu)+\frac{\rho\varepsilon}{2c}\frac{J(\mu',\nu)}{\nu-\mu'}\end{aligned}$$

In an interval around $\mu$, $V_n'(\mu')-F'(\mu)\geqslant\frac{\rho\varepsilon}{2c}\frac{J(\mu',\nu)}{\nu-\mu'}$, which is a positive number independent of $n$ and uniformly bounded away from 0 for all $\mu'$. Then it's impossible that $V'(\mu)=F'(\mu)$. Contradiction.

What's more, since $V'$ is Lipschitz continuous at any $V(\mu)>F(\mu)$, it can be extended smoothly to the boundary. Since $V'=F'$ at $V(\mu)=F(\mu)$, then the limit of this smooth extension has $\lim V'(\mu)=F'(\mu)$. Therefore $V$ is $C^{(1)}$ smooth on $[0,1]$.

- *Step 3*. Repeated experiments and contradictory experiments. With the convergence result we have on hand, we can apply similar proof by contradiction method in step 1 and 2 to rule out these two cases. We omitted the proofs here. Therefore:

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

- *Step 4*. Diffusion experiments. Suppose at $\mu$, diffusion experiment is strictly optimal:

$$V(\mu)<-\frac{c}{\rho}\frac{D^2V(\mu)}{H''(\mu)}$$

Then by Definition 2, there exists $\varepsilon$, $\delta_1$ s.t.:

$$V(\mu)+\varepsilon\leqslant\frac{c}{\rho}\frac{V(\mu+\delta_1)-V(\mu)-V'(\mu)\delta_1}{H(\mu)-H(\mu+\delta_1)+H'(\mu)\delta_1}$$

Then by definition of derivative, there exists $\delta_2$ s.t.:

$$V(\mu)+\frac{\varepsilon}{2}\leqslant\frac{c}{\rho}\frac{\frac{\delta_2}{\delta_1+\delta_2}(V(\mu+\delta_1)-V(\mu))+\frac{\delta_2}{\delta_1+\delta_2}(V(\mu-\delta_2)-V(\mu))}{\frac{\delta_2}{\delta_1+\delta_2}(H(\mu)-H(\mu+\delta_1))+\frac{\delta_2}{\delta_1+\delta_2}(H(\mu)-H(\mu-\delta_2))}$$

By convergence of $V_n$, there exists $n$ s.t.:

$$\begin{aligned}&V_n(\mu)+\frac{\varepsilon}{4}\leqslant\frac{c}{\rho}\frac{\frac{\delta_2}{\delta_1+\delta_2}(V_n(\mu+\delta_1)-V_n(\mu))+\frac{\delta_2}{\delta_1+\delta_2}(V_n(\mu-\delta_2)-V_n(\mu))}{\frac{\delta_2}{\delta_1+\delta_2}(H(\mu)-H(\mu+\delta_1))+\frac{\delta_2}{\delta_1+\delta_2}(H(\mu)-H(\mu-\delta_2))}\\
\Longrightarrow&\frac{\delta_2}{\delta_1+\delta_2}V_n(\mu+\delta_1)+\frac{\delta_1}{\delta_1+\delta_2}V_n(\mu-\delta_2)\\
&\geqslant V_n(\mu)\left(1+\frac{\rho}{c}\left(H(\mu)-\frac{\delta_2}{\delta_1+\delta_2}H(\mu+\delta_1)-\frac{\delta_1}{\delta_1+\delta_2}H(\mu-\delta_2)\right)\right)\\
&\quad+\frac{\rho}{c}\left(H(\mu)-\frac{\delta_2}{\delta_1+\delta_2}H(\mu+\delta_1)-\frac{\delta_1}{\delta_1+\delta_2}H(\mu-\delta_2)\right)\frac{\varepsilon}{4}\end{aligned}$$

If we consider the experiment with posterior beliefs $\mu+\delta_1$, $\mu-\delta_2$ at $\mu$. Taking this experiment at $\mu$ with flow probability:

$$\frac{c}{H(\mu)-\frac{\delta_2}{\delta_1+\delta_2}H(\mu+\delta_1)-\frac{\delta_1}{\delta_1+\delta_2}H(\mu-\delta_2)}$$

Then the flow cost constraint will be satisfied and the utility gain is:

$$\begin{aligned}\widetilde{V}_n(\mu)=&\frac{\frac{\delta_2}{\delta_1+\delta_2}V_n(\mu+\delta_1)+\frac{\delta_1}{\delta_1+\delta_2}V_n(\mu-\delta_2)}{1+\frac{\rho}{c}\left(H(\mu)-\frac{\delta_2}{\delta_1+\delta_2}H(\mu+\delta_1)-\frac{\delta_1}{\delta_1+\delta_2}H(\mu-\delta_2)\right)}\\
\geqslant&V_n(\mu)+\frac{\frac{\rho}{c}\left(H(\mu)-\frac{\delta_2}{\delta_1+\delta_2}H(\mu+\delta_1)-\frac{\delta_1}{\delta_1+\delta_2}H(\mu-\delta_2)\right)}{1+\frac{\rho}{c}\left(H(\mu)-\frac{\delta_2}{\delta_1+\delta_2}H(\mu+\delta_1)-\frac{\delta_1}{\delta_1+\delta_2}H(\mu-\delta_2)\right)}\frac{\varepsilon}{4}\\
>&V_n(\mu)\end{aligned}$$

Contradiction.

To sum up, we proved that $V(\mu)$ solves Equation (18).

*Q.E.D.*

**Lemma S.26** (Convergence of $\mu^*$)**.** *Suppose Assumption A and Assumptions 3, 2-a and 5 are satisfied. Let $F_n$ be piecewise linear function on [0,1] satisfying:*

1. $|F_n-F|\to0$;
2. $\forall\mu\in[0,1], \lim F_n'(\mu)=F'(\mu)$.

*Let $\mu_n^*$ be as defined in Lemma B.1 associated with $F_n$. Suppose $\mu^*=\lim\mu_n^*$. Then,*

1. $\forall\mu>\mu^*, \exists N$ *s.t.* $\forall n\geqslant N, \nu_n(\mu)\geqslant\mu$.
2. $\forall\mu<\mu^*, \exists N$ *s.t.* $\forall n\geqslant N, \nu_n(\mu)\leqslant\mu$.

**Proof.** $\forall\mu>\mu^*$, by definition $\lim\mu_n^*=\mu^*$, there exists $N$ s.t. $\forall n\geqslant N$: $|\mu_n^*-\mu^*|<|\mu-\mu^*|$. Therefore $\mu>\mu_n^*$ and thus $\nu_n(\mu)\geqslant\mu$. Same argument applies to $\mu<\mu^*$. *Q.E.D.*

**Lemma S.27.** *Suppose Assumption A and Assumptions 3, 2-a and 5 are satisfied. Let $F_n$ be piecewise linear function on [0,1] satisfying:*

1. $|F_n-F|\to0$;
2. $\forall\mu\in[0,1], \lim F_n'(\mu)=F'(\mu)$.

*Define $V_n=\mathcal{V}(F_n)$ and $V=\mathcal{V}(F)$. Then: $\forall\mu\in[0,1]$ s.t. $V(\mu)>F(\mu)$, $\exists\lim V_n'(\mu)$.*

**Proof.** With Lemma S.26, we can define $\mu^*\in[0,1]$ (we pick an arbitrary limiting point when there are multiple ones). First by assumption $\lim F_n'(\mu)=F'(\mu)$, and $V_n'=F_m'$ on the boundary by construction in Theorem 2, the statement is automatically true for $\mu\in\{0,1\}$. We discuss three possible cases for different $\mu\in(0,1)$ separately.

- *Case 1*: $\mu > \mu^*$. If $V(\mu) > F(\mu)$, then by convergence in $L_\infty$ norm, there exists $N$ and neighbourhood $\mu \in O$ s.t. $\forall n \geqslant N$, $\mu' \in O$, $V_n(\mu') > F_n(\mu')$. We know that by no-repeated-experimentation property of solution $\nu_n(\mu)$ to problem with $F_n$, $\nu_n(\mu) > \sup O$. Now consider $V_n'(\mu)$. Suppose $V_n'(\mu)$ have unlimited limiting point. Then exists subsequence $\lim V_n'(\mu) = \infty$ or $-\infty$. If $\lim V_n'(\mu) = \infty$, consider $\nu = 0$, else if $\lim V_n'(\mu) = -\infty$, consider $\nu = 1$:
$$\begin{aligned} V(\mu) &= \lim_{n\to\infty} V_n(\mu) \\ &\geqslant \lim_{n\to\infty} \frac{c}{\rho} \frac{F_n(\nu) - V_n(\mu) - V_n'(\mu)(\nu-\mu)}{J(\mu,\nu)} \\ &= \frac{c}{\rho}\frac{F(\nu)-V(\mu)}{J(\mu,\nu)} - \frac{c}{\rho}\lim_{n\to\infty} V_n'(\mu)\frac{\nu-\mu}{J(\mu,\nu)} \\ &= +\infty \end{aligned}$$
Contradiction. Therefore we know that $V_n'(\mu)$ must have finite limiting points. Now suppose $V_n'(\mu)$ doesn't converge, then there exists two subsequences $\lim V_n'(\mu) = V_1'$ and $\lim V_m'(\mu) = V_2'$, $V_1' \neq V_2' \in \mathbb{R}$. Suppose $V_1' > V_2'$. Now take a converging subsequence of optimal policy at $\mu$ $\nu_{n_k} \to \nu^1$. By previous result $\nu^1 \geqslant \sup O$. Therefore $\nu^1$ will be bounded away from $\mu$. Consider:
$$\begin{aligned} V(\mu) &= \lim_{k\to\infty} V_{n_k}(\mu) \\ &\geqslant \lim_{k\to\infty} \frac{c}{\rho} \frac{F_{m_k}(\nu^1) - V_{m_k}(\mu) - V_{m_k}'(\mu)(\nu^1-\mu)}{J(\mu,\nu^1)} \\ &= \frac{c}{\rho}\frac{F(\nu^1)-V(\mu)-V_2'(\nu^1-\mu)}{J(\mu,\nu^1)} \\ &= \lim_{k\to\infty} \frac{F_{n_k}(\nu_{n_k}) - V_{n_k}(\mu) - V_{n_K}'(\mu)(\nu_{n_k}-\mu)}{J(\mu,\nu_{n_k})} + \frac{(V_1'-V_2')(\nu^1-\mu)}{J(\mu,\nu^1)} \\ &> V(\mu) \end{aligned}$$
Contradiction. Therefore, limit point of $V_n'(\mu)$ must be unique. Such limit point exists since $V_n'$ are uniformly bounded. To sum up, there exists $\lim V_n'(\mu)$.

- *Case 2*: $\mu = \mu^*$. Since $V(\mu^*) > F(\mu^*)$. This implies that $\exists N$ s.t. $\forall n \geqslant N$, $V_n(\mu^*) > F_n(\mu^*)$. In this case, by Lemma B.1, $\mu_n^*$ are unique. Since $\mu_n^*$ is the unique intersection of $U^{n+}$ and $U^{n-}$ (Definition of $U^{n+}$,$U^{n-1}$ are as in Lemma B.1, $n$ is index), we can first establish convergence of $\mu^*$ through convergence of $U^{n+}$ and $U^{n-1}$. By definition:
$$U^+(\mu) = \max_{\mu' \geqslant \mu, m \geqslant \underline{\mu}} \frac{F_m(\mu')}{1+\frac{\rho}{c}J(\mu,\mu')}$$
Therefore, suppose the maximizer for index $n$ is $\nu_n$, $m_n$, then for index $n'$:
$$\begin{aligned} U^{n'+}(\mu) &\geqslant \frac{F_{n'}(\nu_n)}{1+\frac{\rho}{c}J(\mu,\nu_n)} \\ &\geqslant U^{n+}(\mu) + \frac{F_n(\nu_n)-F_{n'}(\nu_n)}{1+\frac{\rho}{c}J(\mu,\nu_n)} \\ &\geqslant U^{n+}(\mu) - |F_n - F_{n'}| \end{aligned}$$
Since $n$ and $n'$ are totally symmetric, we actually showed that the functional map from $F_n$ to $U^{n+}$ is Lipschitz continuous in $F_n$ with Lipschitz parameter 1. Symmetric argument shows that same property for $U^{n-}$. Since by assumption $F_n$ is uniformly converging, we can conclude that $U^{n+}$ and $U^{n-}$ are Cauchy sequence with $L_\infty$ norm. Therefore converging. Then $U^{n+} - U^{n-}$ uniformly converges and their roots will be UHC when $n \to \infty$. To show convergence of $\mu_n^*$, it's sufficient to show that such limit is unique. This is not hard to see by applying envelope theory to $U^{n+}$ and $U^{n-}$: $\frac{\mathrm{d}}{\mathrm{d}\mu}U^{n+}(\mu) = -\frac{\rho}{c}\frac{F(\nu_n)H''(\mu)(\nu_n-\mu)}{J(\mu,\nu_n)^2}$. Therefore $U^{n+} - U^{n-1}$ will have slope bounded below from zero, therefore the limit will also be strictly increasing. So $\mu^*$ is unique.

  Since $\mu_n^* \to \mu$, and $V_n''(\mu)$ are all bounded from above:
$$\begin{aligned} V_n'(\mu^*) &= V_n'(\mu_n^*) + V_n''(\xi_n)(\mu^* - \mu_n^*) \\ &= V_n''(\xi_n)(\mu^*-\mu_n^*) \to 0 \end{aligned}$$

- *Case 3*: $\mu < \mu^*$. We can apply exactly the symmetric proof of case 1.

*Q.E.D.*

### S5.3 General State Space

#### S5.3.1 Proof of Theorem 9

**Proof.** $\forall \mu \in E$, consider $X = \mathrm{supp}(\mu)$ (This is without loss since we can always focus on only the support of $\mu$). Let $(p,\nu,\Sigma)$ be optimal policy at $\mu$.

*Step 1*. Derive optimality condition. Suppose $p \neq 0$:

$$\rho V(\mu) = -c\frac{V(\nu)-V(\mu)-\nabla V(\mu)(\nu-\mu)}{H(\nu)-H(\mu)-\nabla H(\mu)(\nu-\mu)} \tag{S.37}$$

Now let $p = -\frac{c}{H(\nu)-H(\mu)-\nabla H(\mu)(\nu-\mu)}$. As an analog to Equation (8), first order condition implies:

$$\mathrm{FOC}-\nu: \nabla V(\nu)-\nabla V(\mu)+\lambda(\nabla H(\nu)-\nabla H(\mu))=0$$

$$\mathrm{FOC}-p: V(\nu)-V(\mu)-\nabla V(\mu)(\nu-\mu)+\lambda((H(\nu)-H(\mu)-\nabla H(\mu)(\nu-\mu))=0$$

$$\xRightarrow{G=V+\lambda H}\begin{cases}\nabla G(\nu)=\nabla G(\mu)\\ G(\nu)-G(\mu)-\nabla G(\mu)(\nu-\mu)=0\end{cases} \tag{S.38}$$

Feasibility condition Equation (S.37) implies $\lambda=\frac{\rho}{c}V(\mu)$.Moreover, optimality implies $\forall \nu' \in \Delta(X)$:

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

$$\Longrightarrow G(\nu')-G(\mu)-\nabla G(\mu)(\nu'-\mu)\leqslant 0 \tag{S.39}$$

Suppose $p=0$, then $\Sigma \neq 0$. Pick any non-zero row $\sigma$, then feasibility condition of Equation (15) implies:

$$\rho V(\mu) = -c\frac{\sigma^T \mathrm{H}V(\mu)\sigma}{\sigma^T \mathrm{H}H(\mu)\sigma}$$

Optimality condition also implies Equation (S.39).

*Step 2*. Prove $V(\nu) > V(\mu)$. Suppose by contradiction that $V(\nu) \leqslant V(\mu)$. Consider $V(\mu_\alpha)$ where $\mu_\alpha = \alpha\nu + (1-\alpha)\mu$, $\alpha \in (0,1)$. Since $\Delta(X)$ is convex, $\mu_\alpha \in \Delta(X)$. Now by Equation (S.39), $G(\mu_\alpha) \leqslant G(\mu) + \nabla G(\mu)(\mu_\alpha - \mu)$. For $\alpha$ sufficiently small, $\mu_\alpha \in E$. $\forall \lambda' < \lambda$, let $G' = V + \lambda' H$. Then since $H$ is strictly concave, $G'$ is more convex that $G$, therefore

$$\begin{cases}G'(\mu_\alpha)-G'(\mu)-\nabla G'(\mu)(\mu_\alpha-\mu)<0\\ G'(\mu_\alpha)-G'(\nu)-\nabla G'(\nu)(\mu_\alpha-\nu)<0\end{cases}$$

$$\Longrightarrow G'(\mu_\alpha)+\nabla G'(\mu_\alpha)(\mu-\mu_\alpha)<G'(\mu)$$

$$\text{or } G'(\mu_\alpha)+\nabla G'(\mu_\alpha)(\nu-\mu_\alpha)<G'(\nu)$$

So optimality condition is not satisfied at $\mu_\alpha$. Suppose $V(\mu_\alpha)$ is achieved with non-zero $p_i$, Then $\lambda$ characterizing FOC at $\mu_\alpha$ must be strictly larger than $\lambda$. Therefore $V(\mu_\alpha) > V(\mu)$. Suppose $V(\mu_\alpha)$ is achieved with zero $p_i$. Then $V(\mu_\alpha) \leqslant V(\mu)$ again implies Equation (S.39) violated. So $V(\mu_\alpha) > V(\mu)$. This implies

$$\frac{\mathrm{d}}{d_\alpha}V(\mu_\alpha)\geqslant 0$$

$$\Longleftrightarrow \nabla V(\mu)(\nu-\mu)\geqslant 0$$

$$\Longrightarrow V(\nu)-V(\mu)-\nabla V(\mu)(\nu-\mu)\leqslant 0$$

Contradicting Equation (S.37).

*Step 3*. Prove $V(\nu) = F(\nu)$. Suppose by contradiction that $V(\nu) > F(\nu)$. By the analysis in step 2, let $\lambda = \frac{\rho V(\mu)}{c}$ and $G = V + \lambda H$. Let $\lambda' = \frac{\rho V(\nu)}{c}$ and $G' = V + \lambda' H$. Then $\forall \nu' \in \Delta(X)$, $\nu' \neq \nu$:

$$\begin{aligned} G(\nu') &\leqslant G(\nu)+\nabla G(\nu)(\nu'-\nu)\\ \Longrightarrow G'(\nu') &= G(\nu')+(\lambda'-\lambda)H(\nu')\\ &\leqslant G(\nu)+\nabla G(\nu)(\nu'-\nu)+(\lambda'-\lambda)H(\nu')\\ &< G(\nu)+\nabla G(\nu)(\nu'-\nu)+(\lambda'-\lambda)H(\nu)+\nabla H(\nu)(\nu'-\nu)\\ &= G'(\nu)+\nabla G'(\nu)(\nu'-\nu)\end{aligned}$$

On the other hand, $\forall \nu'$, $G(\nu') \leqslant G(\nu) + \nabla G(\nu)(\nu'-\nu)$ implies $\mathrm{H}G(\nu)$ being negative semi-definite. Then $\forall \sigma$, $\sigma^T \mathrm{H}G(\nu)\sigma \leqslant 0$. Therefore, $\forall \sigma$, $\sigma^T \mathrm{H}G(\nu)\sigma + (\lambda'-\lambda)\sigma^T \mathrm{H}H(\nu)\sigma < 0 \Longrightarrow \frac{\rho}{c}V(\nu) < -\frac{\sigma^T \mathrm{H}V(\nu)\sigma}{\sigma^T \mathrm{H}H(\nu)\sigma}$. Contradicting $V(\nu)$ being sovled in Equation (15).

*Step 4.* Prove that the set of $\mu$ at which $\frac{\rho}{c}V(\mu)=-\frac{\sigma^T\text{H}V(\mu)\sigma}{\sigma^T\text{H}H(\mu)\sigma}$ is no where dense. Suppose by contradiction that there exists an open ball $O\subset E$ on which $\forall\mu$, $\frac{\rho}{c}V(\mu)=\max_\sigma-\frac{\sigma^T\text{H}V(\mu)\sigma}{\sigma^T\text{H}H(\mu)\sigma}$. Let $\underline{O}$ be a non-degenerate closed ball contained in $O$. Since $V$ is continuous on $V$, there exists $\mu^*\in\text{argmin}_{\mu\in\underline{O}}V(\mu)$. $\forall\mu\in\underline{O}$, by definition $\text{H}V(\mu)+\frac{\rho V(\mu)}{c}\text{H}H(\mu)$ is negative semi-definite. Therefore, $\text{H}V(\mu)+\frac{\rho V(\mu^*)}{c}\text{H}H(\mu)$ is negative semi-definite. Now consider $G(\mu)=V(\mu)+\frac{\rho}{c}V(\mu^*)H(\mu)$ on $\underline{O}$. $G(\mu)$ has pointwise negative semi-definite Hessian. So $G(\mu)$ is a convex function. On the other hand, optimality of Gaussian signal at $\mu^*$ implies $G(\mu)$ to be concave. Therefore $G(\mu)$ is linear on $\underline{O}$. So $V(\mu)=L(\mu)-\frac{\rho}{c}V(\mu^*)H(\mu)$ on $\underline{O}$, where $L(\mu)$ is a linear function.
Now I show that $V(\mu)$ is a constant on $\underline{O}$. Suppose not, $V(\mu)>V(\mu^*)$. Then $V(\cdot)+\frac{\rho V(\mu)}{c}H(\cdot)=L(\cdot)+\frac{\rho}{c}(V(\mu)-V(\mu^*))H(\cdot)$ has negative-definit Hessian at $\mu$. So there exists no $\sigma$ s.t. $\sigma^T\text{H}V(\mu)\sigma+\frac{\rho V(\mu)}{c}\sigma^T\text{H}H(\mu)\sigma=0$. Contradiction. However, $V(\mu)$ being a constant on $\underline{O}$ implies $\text{H}V(\mu)\equiv0$ on $\underline{O}$, contradiction.

*Step 5.* Prove that $\forall\mu\in E$, exists $\nu\in E^C$ satisfying Equation (S.37). Suppose $p>0$, then as discussed in step 1, proof is done. Now suppose $p=0$. Then by step 4, there is a converging sequence of $\mu_n\to\mu$ and $\nu_n$ satisfying Equation (S.37) for each $\mu_n$. By step 3, $\nu_n\in E^C$ so $\nu_n$ are bounded away from $\mu_n$ by positive distance. Since $\nu_n\in E^C$ and $E^C$ is closed subset of $\Delta(X)$, there exists converging subsequence $\nu_n\to\nu\in E^C$. Therefore, by smoothness of $V$ and $H$,

$$V(\mu)=\lim_{n\to\infty}V(\mu_n)=\lim_{n\to\infty}-c\frac{V(\nu_n)-V(\mu_n)-\nabla V(\mu_n)(\nu_n-\mu_n)}{H(\nu_n)-H(\mu_n)-\nabla H(\mu_n)(\nu_n-\mu_n)}=-c\frac{V(\nu)-V(\mu)-\nabla V(\mu)(\nu-\mu)}{H(\nu)-H(\mu)-\nabla H(\mu)(\nu-\mu)}$$

*Step 6.* Prove the strict inequality. Define $K=\left\{\mu\middle|\rho V(\mu)=\sup_\sigma-c\frac{\sigma^T\text{H}V(\mu)\sigma}{\sigma^T\text{H}H(\mu)\sigma}\right\}$. Then by *step 4*, $K$ is a nowhere dense set and the inequality in property 4 is satisfied by construction. Now I prove property 1 on $E\backslash K$:

$$\begin{aligned}D_{\nu(\mu)-\mu}V(\mu)=&(\nu(\mu)-\mu)^T\cdot\nabla V(\mu)=(\nu(\mu)-\mu)^T\cdot\frac{\partial}{\partial\mu}\left(-\frac{c}{\rho}\frac{F(\nu(\mu))-V(\mu)-\nabla V(\mu)(\nu(\mu)-\mu)}{H(\nu(\mu))-H(\mu)-\nabla H(\mu)(\nu(\mu)-\mu)}\right)\\=&(\nu(\mu)-\mu)^T\left(-\frac{c}{\rho}\frac{-\text{H}V(\mu)(\nu(\mu)-\mu)+\frac{\rho}{c}V(\mu)(-\text{H}H(\mu)(\nu(\mu)-\mu))}{H(\nu(\mu))-H(\mu)-\nabla H(\mu)(\nu(\mu)-\mu)}\right)>0\end{aligned}$$

Now I prove property 3 on $E\backslash K$: Define $J(\mu,\nu)=H(\mu)-H(\nu)+\nabla H(\mu)(\nu-\mu)$. Then Equations (S.37) and (S.38) implies

$$\begin{cases}V(\mu)=\frac{F(\nu)-(\nu-\mu)^T\cdot\nabla V(\mu)}{1+\frac{\rho}{c}J(\mu,\nu)}\\\nabla V(\mu)=\left((\nabla H(\nu)-\nabla H(\mu))(\nu-\mu)^T+(1+\frac{\rho}{c}J(\mu,\nu))I\right)^{-1}(F(\nu)(\nabla H(\nu)-\nabla H(\mu))+(1+\frac{\rho}{c}J(\mu,\nu))\nabla F)\end{cases}$$

$$\Longrightarrow\begin{cases}V(\mu)=\frac{F(\mu)}{1-\frac{\rho}{c}J(\nu,\mu)}\\\nabla V(\mu)=\nabla F+\frac{\frac{\rho}{c}F(\mu)(\nabla H(\nu)-\nabla H(\mu))}{1-\frac{\rho}{c}J(\nu,\mu)}\end{cases}$$

Then $\nu=\nu(\mu)$ satisfies the following PDE $\forall\alpha$:

$$\begin{aligned}&\alpha^T\cdot\frac{\partial}{\partial\mu}\frac{F(\mu)}{1-\frac{\rho}{c}J(\nu,\mu)}+D_\alpha\nu\frac{\partial}{\partial\nu}\frac{F(\mu)}{1-\frac{\rho}{c}J(\nu,\mu)}=\alpha^T\cdot\left(\nabla F+\frac{\frac{\rho}{c}F(\mu)(\nabla H(\nu)-\nabla H(\mu))}{1-\frac{\rho}{c}J(\nu,\mu)}\right)\\\Longrightarrow&F(\mu)D_\alpha\nu\cdot\text{H}H(\nu)(\nu-\mu)=J(\nu,\mu)\left(\alpha^T\nabla F\left(1-\frac{\rho}{c}J(\nu,\mu)\right)+\frac{\rho}{c}F(\mu)\alpha^T(\nabla H(\nu)-\nabla H(\mu))\right)\\\Longrightarrow&D_\alpha\nu\cdot\text{H}H(\nu)(\nu-\mu)=\frac{J(\nu,\mu)}{F(\mu)(1-\frac{\rho}{c}J(\nu,\mu))}D_\alpha V(\mu)\\\Longrightarrow&D_{\mu-\nu}\nu\cdot\text{H}H(\nu)(\nu-\mu)=\frac{J(\nu,\mu)}{F(\mu)(1-\frac{\rho}{c}J(\nu,\mu))}\left(-D_{\nu-\mu}V(\mu)\right)<0\end{aligned}$$

The inequality comes from $V(\mu)>0$ and $D_{\nu-\mu}V(\mu)>0$. *Q.E.D.*

### S5.4 *Axiom for posterior separability*

#### S5.4.1 *Proof of Theorem 10*

**Proof.** Let $\mathcal{S}_0$ be a fully revealing information structure i.e. with any prior belief $\mu$, each signal induces posterior belief $\delta_{x_i}$ with probability $\mu(x_i)$. $\forall\mu\in\Delta X$, define:

$$H(\mu)=I(\mathcal{S}_0;\mathcal{X}|\mu)$$

$\forall \mathcal{S}$ which induces $\nu$ with probability $h(\nu)$ with prior $\mu$:

$$\begin{aligned} I(\mathcal{S}_0;\mathcal{X}|\mu) &= I(\mathcal{S};\mathcal{X}|\mu) + E[I(\mathcal{S}_0;\mathcal{X}|\mathcal{S},\mu)] \\ &= I(\mathcal{S};\mathcal{X}|\mu) + \int I(\mathcal{S}_0;\mathcal{X}|\nu)h(\nu)\mathrm{d}\nu \\ &= H(\mathcal{S};\mathcal{X}|\mu) + E_h[I(\mathcal{S}_0;\mathcal{X}|\nu)] \\ \implies I(\mathcal{S};\mathcal{X}|\mu) &= H(\mu) - E_h[H(\nu)] \end{aligned}$$

Moreover, $H(E_h[\nu]) - E_h[H(\nu)] \geqslant 0$ for all distribution $h$ implies that $H$ is a concave function on $\Delta X$. *Q.E.D.*