Optimal Sequential Decision with Limited Attention

Yeon-Koo Che and Konrad Mierendorff

Columbia Cognition and Decision Lab Meeting
February 12, 2016

Introduction

- We revisit Wald's sequential decision problem: *DM decides sequentially on information acquisition before making a decision.*
- Classical feature: Information takes time to arrive, and earlier decision is better (discounting or flow cost)
- New feature: Different types of information are received, and the DM allocates limited attenion on them for processing.
- Applications.
  - Project selection
  - Recruiting
  - Selection of news media
  - Deliberation/research strategy: "Prove" or "disprove"?
  - Cognitive process (???)

Model

Baseline Model

- Two States: $s \in \{A, B\}$
- One DM — Two actions: a, b
- Payoffs conditional on state and action:

State:	A	B
a	u_a^A *	u_a^B
b	u_b^A	u_b^B *

  - Assume $u_a^A > u_b^A$, $u_b^B \geq u_a^B$.
- Prior probability of state A: $p_0 \in (0, 1)$.
- Continuous time $t \geq 0$, discount rate $r > 0$.
- At each point in time, the DM can take an action a or b, or **acquire information**.

Model

Information

- At each time t, DM allocates **unit of attention** between two Poisson-signals (or **news sources**):
  - A signal: the arrival rate of news is $(\lambda^A, \lambda^B) = (\lambda, 0)$.
  - B signal: the arrival rate of news is $(\lambda^A, \lambda^B) = (0, \lambda)$.
- $\alpha \in [0, 1]$, amount of attention directed to A signal:
- $\beta := 1 - \alpha$, amount of attention directed to B signal:
- Given "attention choice" (α, β), in state A, it is learned at the Poisson rate of $\lambda\alpha$ and in state B, it is learned at rate $\lambda\beta$,
- Posterior jumps to $p = 0$ or $p = 1$ after observing a signal.
- No signal — Bayesian updating:

$$\dot{p}_t = -\lambda(\alpha - \beta)p(1 - p).$$

Literature

- Wald's Sequential Decision Problem: Wald (1947), Arrow, Blackwell, and Girshick (1949), Moscarini and Smith (2001)
- Rational Inattention: Sims (2003), Shannon (1959), Matejka and McKay (2015), Steiner, Stewart and Matejka (2015)
- Poisson Bandit: Keller, Rady and Cripps (2005), Klein and Rady (2011), Francetich (2015)
- Mathematical Psychology ("DDM"): Many I have not read + Woodford (2014), Fudenberg, Strack and Strzalecki (2015)

Model—Benchmarks

- Immediate Action
  - Action a yields: $U_a(p) := pu_a^A + (1-p)u_a^B$
  - Action b yields: $U_b(p) := pu_b^A + (1-p)u_b^B$
  - Optimal action yields $U(p) := \max\{U_a(p), U_b(p)\}$
- First Best: $\overline{U}^*(p) = pu_a^A + (1-p)u_b^B$.
- Unlimited Attention: $\alpha = \beta = 1$. Upper bound for experimentation

$$\frac{\lambda \overline{U}^*(p)}{r+\lambda}$$

- Stationary attention strategy (with limited attention): $\alpha = \beta = \frac{1}{2}$ ($\Rightarrow \dot{p}_t = 0$) yields

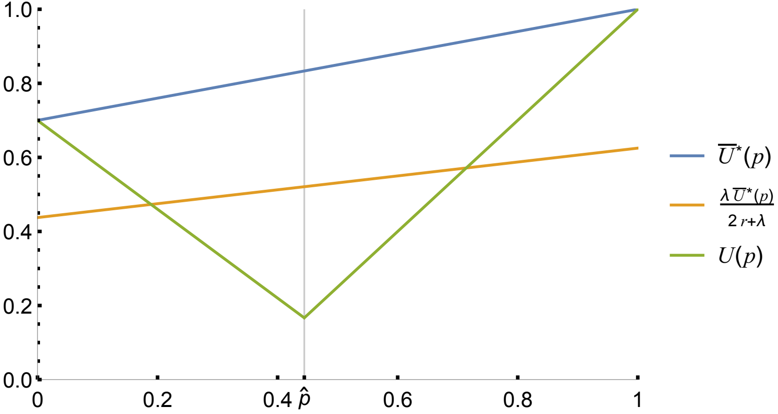
$$\frac{\lambda \overline{U}^*(p)}{2r+\lambda}.$$

- Preliminary observation:

$$V(p) \geq \max\left\{U(p), \frac{\lambda \overline{U}^*(p)}{2r+\lambda}\right\}$$

Model—Preliminary Analysis

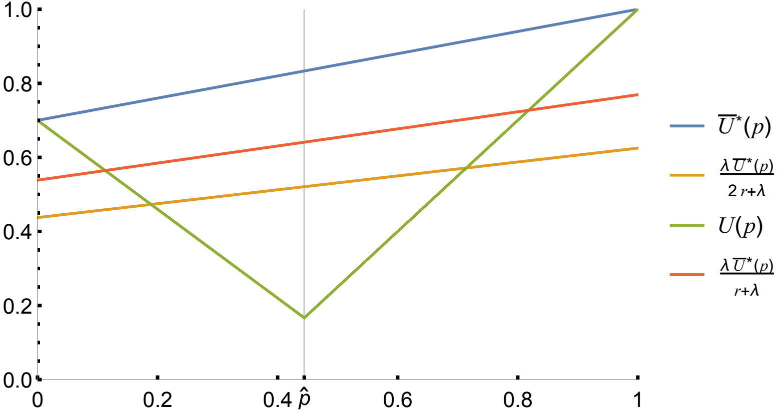
(Parameters: $\lambda = 1, r = \frac{3}{10},\ u_a^A = 1,\ u_b^B = .7,\ u_a^B = u_b^A = -\frac{1}{2}$)



- Define $\hat{p}$: $U_a(\hat{p}) = U_b(\hat{p})$

Model—Preliminary Analysis

(Parameters: $\lambda = 1, r = \frac{3}{10},\ u_a^A = 1,\ u_b^B = .7,\ u_a^B = u_b^A = -\frac{1}{2}$)



- Define $\hat{p}$: $U_a(\hat{p}) = U_b(\hat{p})$

Optimization Problem

$$V(p_0) := \max_{(\alpha_t, T)} \int_0^T e^{-rt} \left[p_t \lambda \alpha_t u_a^A + (1-p_t)\lambda(1-\alpha_t) u_b^B \right] Q_t(\alpha, p_0) dt$$
$$+ e^{-rT} Q_T(\alpha, p_0) U(p_T)$$

s.t.
$$\dot{p}_t = -\lambda(2\alpha_t - 1)p_t(1-p_t),$$
$$Q_t(\alpha, p) = \left(p_0 e^{-\lambda \mathcal{A}_t} + (1-p_0) e^{-\lambda \mathcal{B}_t} \right),$$

where

- T is the time at which DM takes an action if no signal arrives.
- $(\alpha_t)_{t \geq 0}$ is the attention (information) strategy
- $\mathcal{A}_t := \int_0^t \alpha_s ds$; $\mathcal{B}_t := t - \mathcal{A}_t$: "accumulated" attention.
- $Q_t(\alpha, p_0)$ is the probability that no signal arrives until t given α and p_0,

HJB Equation

$$(r+\lambda)V(p) = \max_{\alpha\in[0,1]} \Big\{ u_a^A \lambda \alpha p + u_b^B \lambda (1-\alpha)(1-p) \\ - \lambda\left[(2p-1)\alpha - p\right] V(p) - \lambda(2\alpha-1)p(1-p)V'(p) \Big\},$$

assuming the RHS is no less than $(r+\lambda)U(p)$, or else the RHS becomes $(r+\lambda)U(p)$.

- Objective is linear: **Bang-Bang** solution. Derivative:

$$u_a^A \lambda p - u_b^B \lambda(1-p) - \lambda(2p-1)V(p) - 2\lambda p(1-p)V'(p) \quad \text{(FOC)}$$

- If $\alpha = 0$

$$V_0'(p) = \frac{r+\lambda(1-p)}{\lambda p(1-p)} V_0(p) - \frac{u_b^B}{p}. \qquad \text{(ODE-0)}$$

- If $\alpha = 1$

$$V_1'(p) = -\frac{r+\lambda p}{\lambda p(1-p)} V_1(p) + \frac{u_a^A}{1-p}. \qquad \text{(ODE-1)}$$

Two Learning Strategies:

- Confirmatory strategy:
  - Trying to confirm what is likely
  - Choose $\alpha = 1$ for a high p and $\alpha = 0$ for a low p.
- Contradictory strategy:
  - Trying to rule out what is unlikely
  - Choose $\alpha = 0$ for a high p and $\alpha = 1$ for a low p.

Structure of Value Function and Optimal Policy

Theorem

The optimal solution has one of the following forms:

1. *No information acquisition:* $V(p) = U(p)$ *for all* p.
2. *Only "contradictory evidence":*
   - *There are cutoffs* $0 < \underline{p}^* < \check{p} < \bar{p}^* < 1$ *such that the optimal structure is*

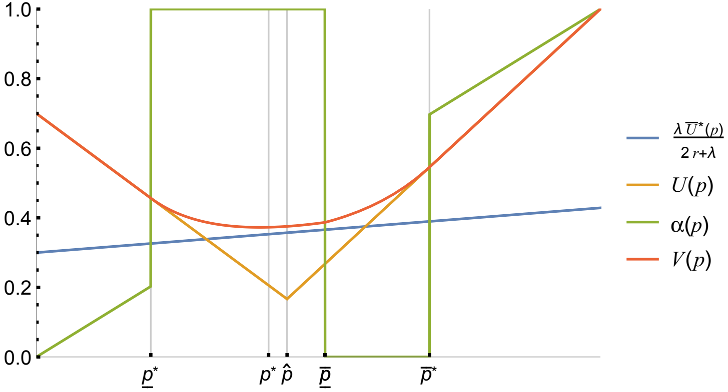
$$\Big|\underbrace{\text{— — —}}_{b}\ \underline{p}^*\ \underbrace{\longleftarrow\longleftarrow\longleftarrow}_{\alpha=1}\ \check{p}\ \underbrace{\longrightarrow\longrightarrow\longrightarrow}_{\alpha=0}\ \bar{p}^*\ \underbrace{\text{— — —}}_{a}\Big|$$

3. *"Contradictory" and "Confirmatory" evidence:*
   - *There are cutoffs* $0 < \underline{p}^* < \underline{p} < p^* < \bar{p} < \bar{p}^* < 1$ *such that the optimal structure is*

$$\Big|\text{—}_{b}\text{—}\ \underline{p}^*\ \overbrace{\underset{\alpha=1}{\longleftarrow\longleftarrow}}^{\textit{contradictory}}\ \underline{p}\ \overbrace{\underset{\alpha=0}{\longrightarrow\longrightarrow}\ p^*\ \underset{\alpha=1}{\longleftarrow\longleftarrow}}^{\textit{confirmatory}}\ \bar{p}\ \overbrace{\underset{\alpha=0}{\longrightarrow\longrightarrow}}^{\textit{contradictory}}\ \bar{p}^*\ \text{—}_{a}\text{—}\Big|$$

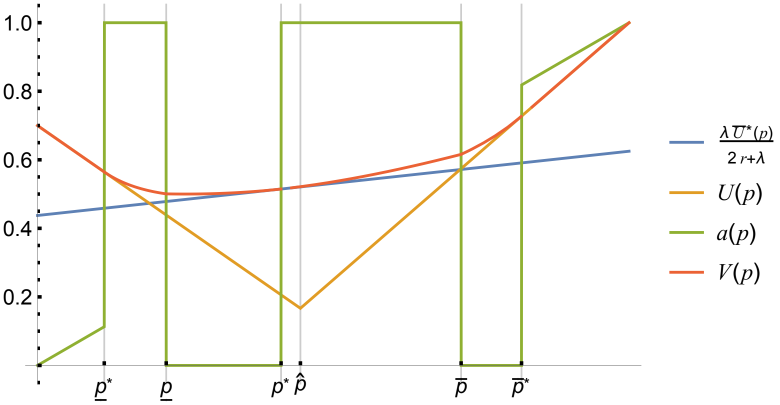
Example (Case 2: Only Contradictory)

(Parameters: $\lambda = 1, \mathbf{r} = \mathbf{\frac{2}{3}}, u_a^A = 1, u_b^B = .7, u_a^B = u_b^A = -\frac{1}{2}$)



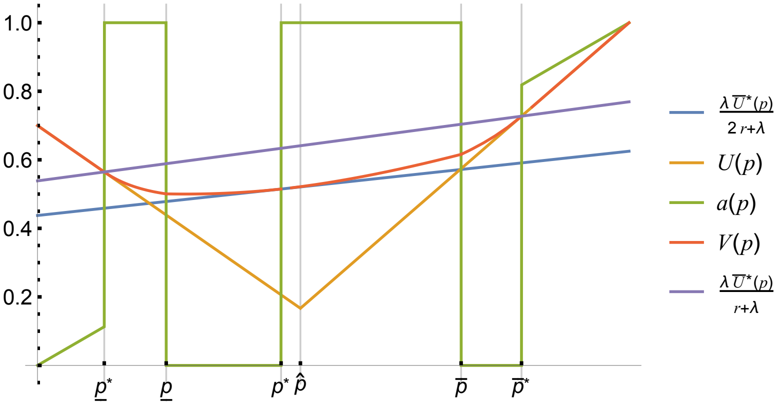
Example (Case 3: Confirmatory and Contradictory)

(Parameters: $\lambda = 1, r = \frac{3}{10},\ u_a^A = 1,\ u_b^B = .7,\ u_a^B = u_b^A = -\frac{1}{2}$)



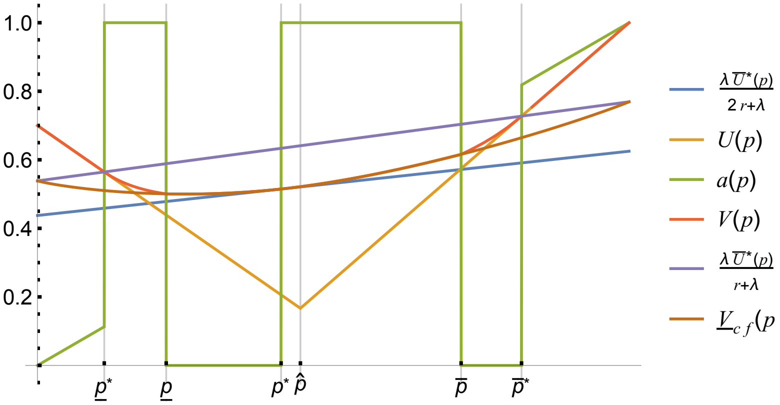
Example (Case 3: Confirmatory and Contradictory)

(Parameters: $\lambda = 1, r = \frac{3}{10},\ u_a^A = 1,\ u_b^B = .7,\ u_a^B = u_b^A = -\frac{1}{2}$)



Example (Case 3: Confirmatory and Contradictory)

(Parameters: $\lambda = 1, r = \frac{3}{10}$, $u_a^A = 1$, $u_b^B = .7$, $u_a^B = u_b^A = -\frac{1}{2}$)



Intuition

- For an "extremely certain" DM: *Immediate action is clearly optimal.*
- For a "moderately certain or highly convinced" DM: *"Little value of acquiring info that will not change your action; Better look for a surprise, which will either change your action or rule out the unlikely and convince you of the likely more than before."*
- For an "uncertain" DM: *"Surprise value of learning the unlikely is small when one is uncertain, and ruling out the less likely takes a long time; may as well learn the likely."*

Characterizing the cut-offs $\underline{p}^*, \overline{p}^*$

Lemma

- *If we are in Case 2 or 3 we have*

$$U_b(\underline{p}^*) = \frac{\lambda \overline{U}^*(\underline{p}^*)}{r+\lambda} \qquad \textit{and} \qquad U_a(\overline{p}^*) = \frac{\lambda \overline{U}^*(\overline{p}^*)}{r+\lambda}$$

Proof.

At $\underline{p}^*$ the DM is indifferent between b and $\alpha = 1$:

$$(r+\lambda)U_b(\underline{p}^*) = u_a^A \lambda \underline{p}^* + \lambda(1-\underline{p}^*)U_b(\underline{p}^*) - \lambda \underline{p}^*(1-\underline{p}^*)U_b'(\underline{p}^*)$$
$$\iff (r+\lambda)U_b(\underline{p}^*) = \lambda \overline{U}^*(\underline{p}^*)$$

□

- From ODEs, we obtain two branches $\underline{V}_{ct}(p)$ and $\overline{V}_{ct}(p)$ that define the Value function for contradictory evidence.

Characterizing the cut-off p^*

Lemma

$$p^* = \frac{u_b^B}{u_a^A + u_b^B}, \; V_{cf}(p^*) = \frac{\lambda \overline{U}^*(p^*)}{2r+\lambda}, \; \textit{and } V'_{cf}(p^*) = \frac{\lambda \overline{U}^{*\prime}(p^*)}{r+\lambda}$$

Proof.

At p^* the DM is indifferent between $a = 1$ and $a = 0$:

- Optimality of $a = \frac{1}{2}$ implies $V_{cf}(p^*) = \frac{\lambda \overline{U}^*(p^*)}{2r+\lambda}$.
- From ODEs, we obtain two branches $\underline{V}_{cf}(p)$ and $\overline{V}_{cf}(p)$.
- $\underline{V}'_{cf}(p^*) = \overline{V}'_{cf}(p^*) = \frac{\lambda \overline{U}^{*\prime}(p^*)}{2r+\lambda}$ pins down p^*.

□

When is (some) Information Acquisition Optimal?

Proposition

Fix the parameters $u_a^A, u_a^B, u_b^A, u_b^B$.

- *If* $u_a^A u_b^B < u_a^B u_b^A$ *("mistakes are very costly"), then some experimentation (contradictory) is optimal.*
- *Otherwise, the experimentation is valuable if*

$$\theta := \frac{r}{\lambda} < \left(u_a^A - u_b^A\right) \max\left\{ \frac{\left(u_b^B - u_a^B\right)}{u_a^A u_b^B - u_a^B u_b^A}, \frac{1}{u_b^A - u_b^B} \right\}$$

When is (some) Confirmatory Evidence Optimal?

Proposition

- *If u_b^A, u_a^B are negative and sufficiently large in absolute values ("mistakes are sufficiently costly"), then a confirmation regions exists.*
- *Otherwise, there exists $\underline{\theta} > 0$ such that a confirmatory region exists if*

$$\theta = \frac{r}{\lambda} < \underline{\theta}.$$

Proof.

The condition is obtained from

$$V_{cf}(p^*) > \underline{V}_{ct}(p^*)$$

□

General Model: Rich signals

Continuum of signals

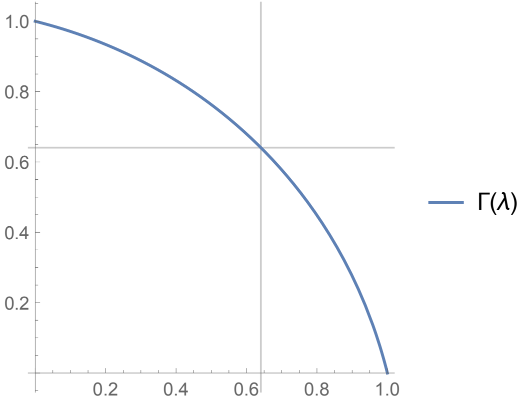
Signals indexed by $\lambda \in [0, 1]$:

- Given $\lambda \in [0, 1]$, two states are learned (conclusively) at Poisson arrival rates of $(\lambda^A, \lambda^B) = (\lambda, \Gamma(\lambda))$
- $\Gamma(\lambda)$ is decreasing and concave, symmetric ($\Gamma(\lambda) = 1 - \Gamma(1 - \lambda)$), and $\Gamma(1) = 0, \Gamma(0) = 1, \Gamma(\gamma) = \gamma$, for some $\gamma > 1/2$.
- The DM picks $\lambda \in (0, 1)$ at each moment. Absent any news, the belief is updated via

$$\dot{p} = -(\lambda - \Gamma(\lambda))p(1 - p).$$

Example for Richer Set of News Sources

$$\alpha\left(\lambda + \Gamma(\lambda)\right) + (1-\alpha)\sqrt{\lambda^2 + \Gamma(\lambda)^2} = 1$$



(Plot Parameter $\alpha = \frac{1}{4}$)

$\alpha > 0$ avoids $\Gamma'(0) = 0$ and $\Gamma'(1) = -\infty$

Sketch of Analysis

- HJB can be written as:

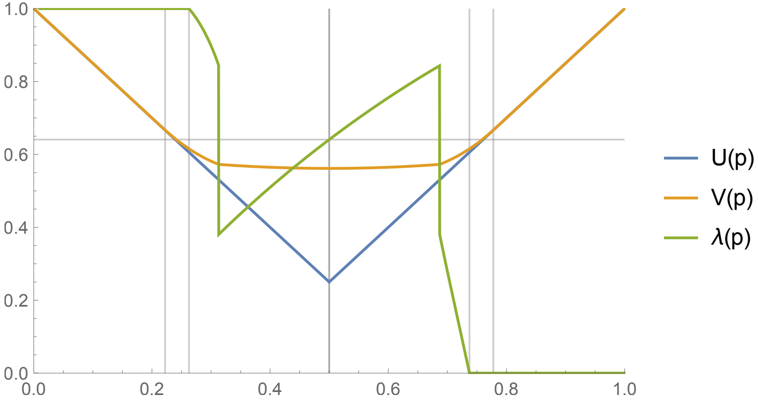
$$
\begin{aligned}
&(r+\lambda)V(p) \\
&= \max_{\lambda\in[0,1]} \Big\{ u_a^A\,\lambda p + u_b^B\,\Gamma(\lambda)(1-p) - (\lambda p + \Gamma(\lambda)(1-p))V(p) \\
&\qquad\qquad - (\lambda - \Gamma(\lambda))p(1-p)V'(p) \Big\},
\end{aligned}
$$

- The overall structure of the optimal policy is preserved:

$$
|\!-\!\!-\underset{b}{\ }-\!\!-\ \underline{p}^* \overbrace{\underset{\lambda=1}{\longleftarrow}\ \underline{p}^{**}\ \underset{\lambda>\gamma}{\longleftarrow}}^{\text{contradictory}} \underline{p} \overbrace{\underset{\lambda<\gamma,\text{rising}}{\longrightarrow\longrightarrow}\ p^*\ \underset{\lambda>\gamma,\text{rising}}{\longleftarrow\longleftarrow}}^{\text{confirmatory}} \bar{p} \overbrace{\underset{\lambda<\gamma}{\longrightarrow}\ \bar{p}^{**}\ \underset{\lambda=0}{\longrightarrow}}^{\text{contradictory}} \bar{p}^*\ -\!\!-\underset{a}{\ }-\!\!-|
$$

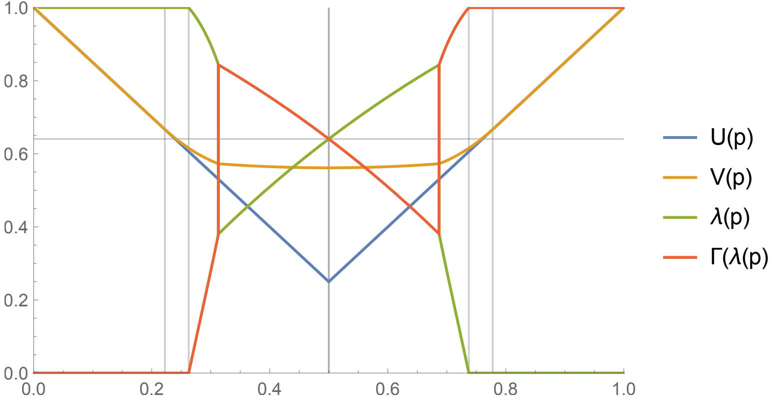
Example Rich News: Confirmatory and Contradictory

(Parameters: $r = \frac{1}{2}$, $u_a^A = u_b^B = 1$, $u_a^B = u_b^A = -\frac{1}{2}$,$\alpha = \frac{1}{4}$)



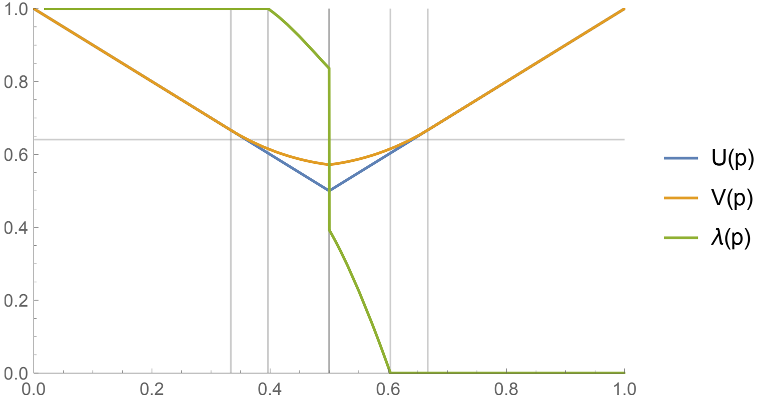
Example Rich News: Confirmatory and Contradictory

(Parameters: $r = \frac{1}{2},\ u_a^A = u_b^B = 1,\ u_a^B = u_b^A = -\frac{1}{2}, \alpha = \frac{1}{4}$)



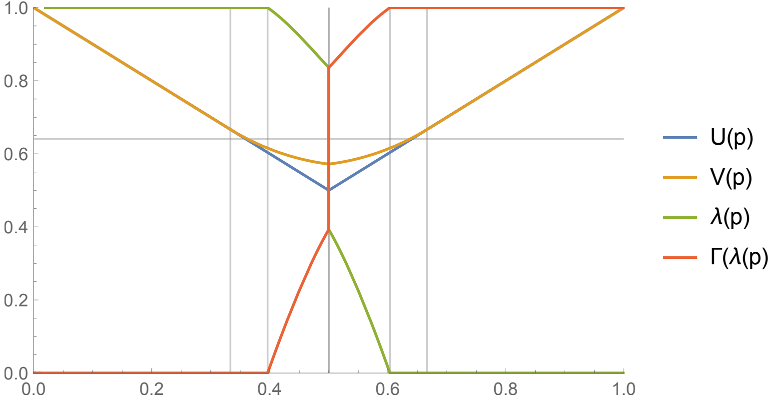
Example Rich News: Only Contradictory

(Parameters: $r = \frac{1}{2}$, $u_a^A = u_b^B = 1$, $\mathbf{u_a^B = u_b^A = 0}$, $\alpha = \frac{1}{4}$)



Example Rich News: Only Contradictory

(Parameters: $r = \frac{1}{2}$, $u_a^A = u_b^B = 1$, $\mathbf{u_a^B = u_b^A = 0}$, $\alpha = \frac{1}{4}$)



General Model: Non-conclusive signals with flow cost

Posterior choice model

A signal, or news source, is indexed by a posterior $q \in [0, 1]$.

- Any news q arrives at the Poisson arrival rate of $\lambda > 0$.
- Once news q arrives, it becomes the DM's posterior belief.
- The DM picks q at cost $c(q, p)$ given current belief p.
- $c(p, p) = c_q(p, p) = 0, c_{qq}(\cdot, \cdot) \geq 0, c_{qp}(\cdot, \cdot) \leq 0$.
  $c_{qqp}(q, p) = 0$.
- Special case: Entropy. If the cost is mutual information on the experiment, the flow cost becomes
  $c(q, p) = \mu D(q||p) = \mu(q \ln(q/p) + (1 - q) \ln(1 - q)/(1 - p))$.
- Absent news, the belief updates according to: $\dot{p} = -\lambda(q - p)$.

General Model: Non-conclusive signals with flow cost

Benchmark: $r = 0$

- Closed form solution, characterized by two posteriors: $q_+ > \hat{p} > q_-$, such that, an immediate action is chosen for $p \leq q_-$ and for $q \geq p_+$, and some mix of q_- and q_+ is chosen for $p \in [q_-, q_+]$.
- The outcome coincides with the RI type discrete choice prediction (e.g., Matejka-McKay (2015)). [An implication of the "chain rule" property of the entropy; but note it applies to "a little" beyond the entropy model.]
- The optimal dynamic attention path not uniquely pinned down: contradictory, confirmatory, or stationary all consistent with optimal decision rule.
- Away from $r = 0$, we have a unique optimal decision. $\Rightarrow$ Sequential decision foundation for the RI choice.

General Model: Non-conclusive signals with flow cost

General Case: $r > 0$

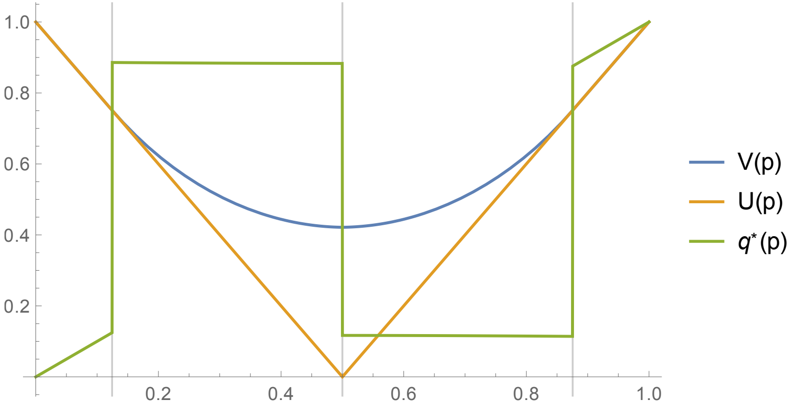
- The HJB equation:

$$(r + \lambda)V(p) = \max_q \left[\lambda U(q) - c(q,p) - \lambda(q-p)V'(p)\right],$$

- The RHS has two local optima: $q_+(p) > \hat{p} > q_-(p)$.
- The structure of the optimal policy is same as before: $V(p) = \max\{U(p), V_{ct}(p), V_{cf}(p)\}$, where
  - $V_{ct}(p)$ = value of contradictory strategy: Choose $q_+(p) > \hat{p}$ for low p and $q_-(p)$ for high p.
  - $V_{cf}(p)$ = value of confirmatory strategy: Choose $q_+(p) > \hat{p}$ for high p and $q_-(p)$ for low p.
- For the entropy model, the confirmatory region never arises in all numerical exercises.

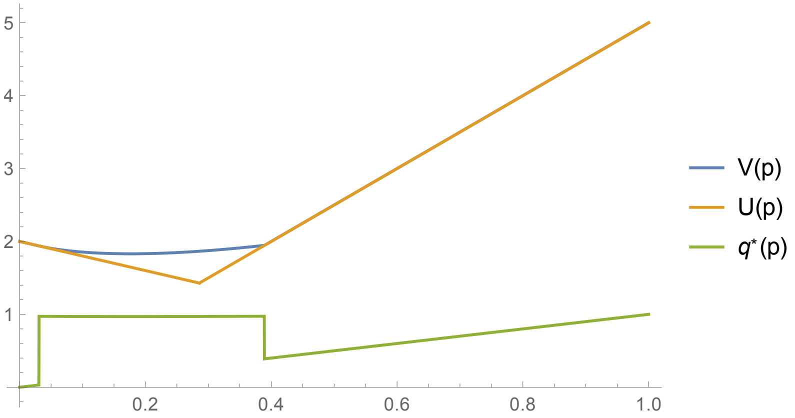
Example: Symmetric Payoffs

(Parameters: $\lambda = \mu = 1,\ r = 0.05,\ u_a^A = u_b^B = 1,\ u_a^B = u_b^A = -1$)



Example: Asymmetrically Risky Actions

(Parameters: $\lambda = \mu = 1$, $r = 0.8$, $u_a^A = 5$, $u_b^B = 2$, $u_a^B = u_b^A = 0$)



Summary

- In a broad set of Poisson bandit signal environments, the ptimal learning strategy combines
  - immediate action
  - contradictory learning
  - confirmatory learning
- Uncertain DM tends to seek confirmatory evidence.
- Moderately certain DM becomes "skeptical" and seeks contradictory evidence.
- Extremely certain DM takes immediate action.

Extensions

- Response time?!?: Yet to be explored.
- More states/actions
  - What is the multi-dimensional equivalent of "contradictory evidence"?
  - Uncertain state may involve different values of λ.
- Applications...
- Experiment... (we have worked out a finite-period version of our model.)