Timing decision in model-based persuasion

Jun Hyun Ji

University of Pittsburgh

December 3, 2024

Motivation for model-based persuasion

Persuasion often involves an expert providing an interpretation of observed data

—a model, a narrative, a relationship between state and data, etc.

- Debate on climate change/statistical facts/politics
- The defense and prosecution base their cases on the same evidence
- Engineers/mechanics provide different reasons for malfunctioning

DM seek/adopt expert's advice especially when data is surprising

- People initially interpret the data on their own
- Adopt a model that better fits the data

Does this mean Persuader should wait for unexpected data?

Motivation for timing decision

I focus on these aspects:

1. More data to come before DM chooses an action
2. Expert's opportunity to give advice is limited

Expert often decides when to persuade: before/after observing more data

3. Expert's interpretation needs to be consistent

   e.g., (X) "This time is different", "Last year was a special case", etc.

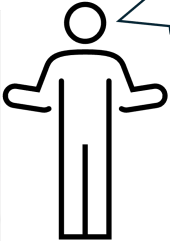
Expert needs to provide a single model to interpret all data

—a model = a data generating process

Example

Expert (an auto mechanic) – Client (Garrett's father)

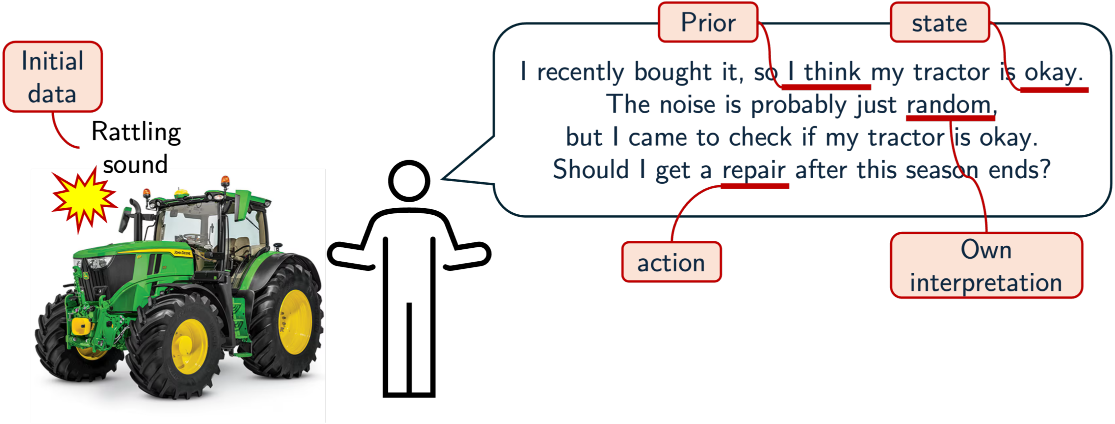
Rattling sound



I recently bought it, so I think my tractor is okay.
The noise is probably just random,
but I came to check if my tractor is okay.
Should I get a repair after this season ends?

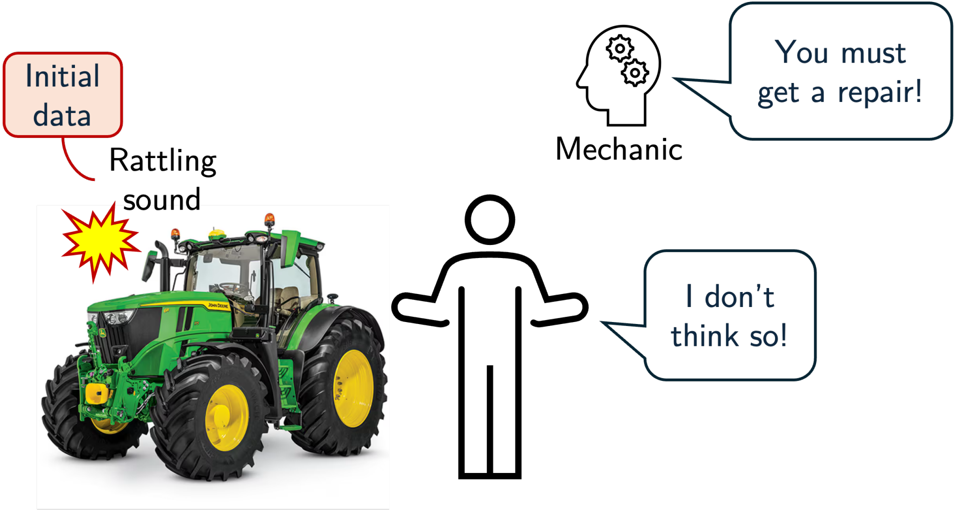
Example

Expert (an auto mechanic) – Client (Garrett's father)



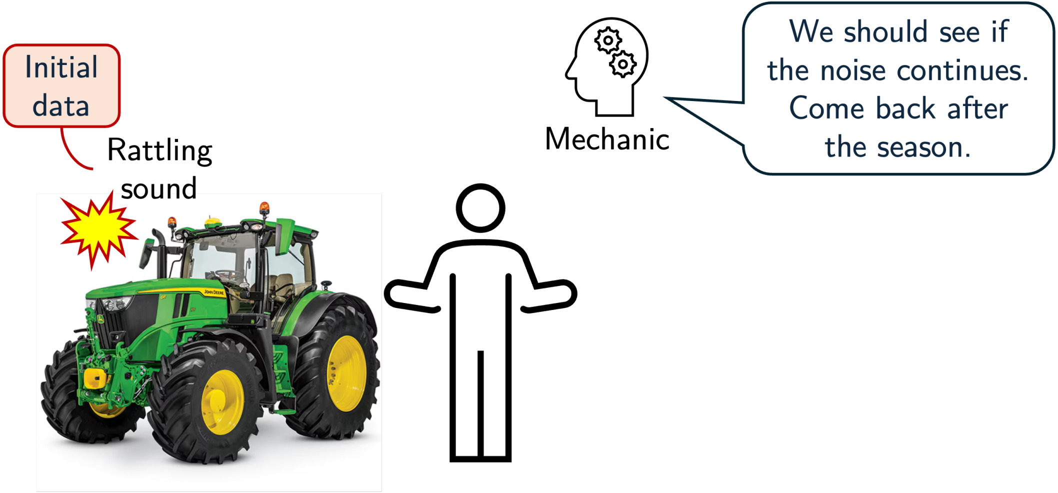
Example

Expert (an auto mechanic) – Client (Garrett's father)



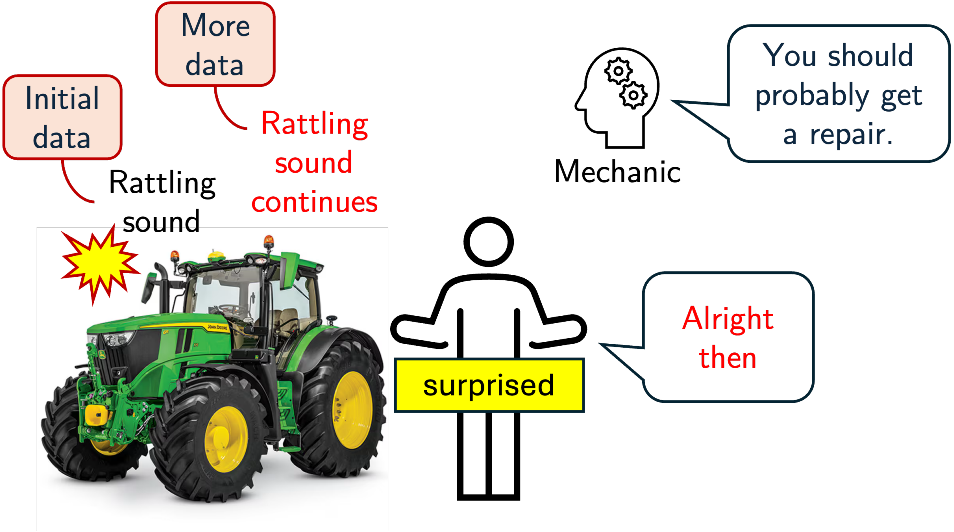
Example

Expert (an auto mechanic) – Client (Garrett's father)



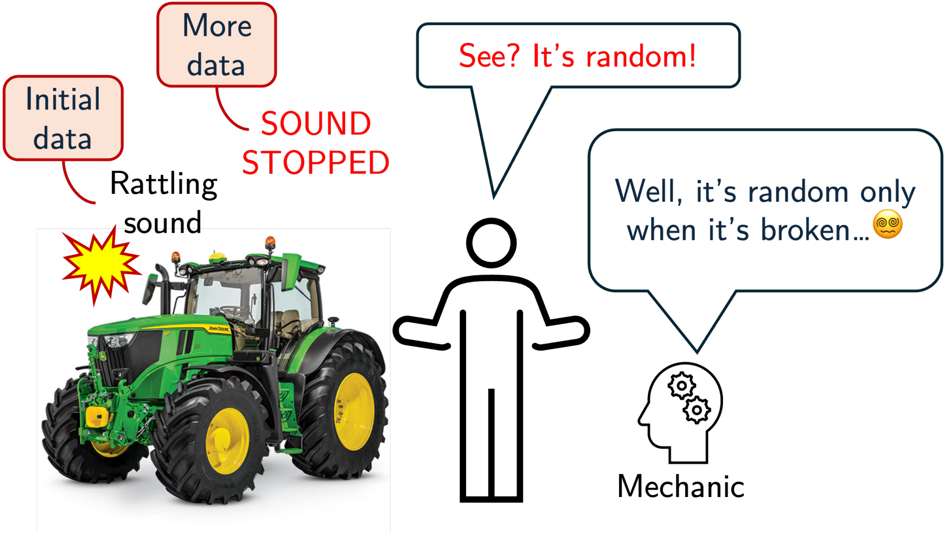
Example

Expert (an auto mechanic) – Client (Garrett's father)



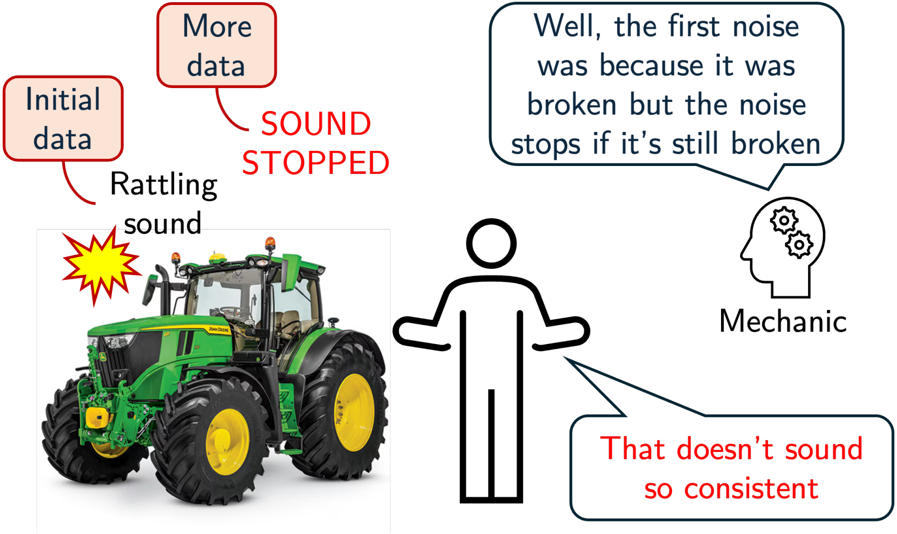
Example

Expert (an auto mechanic) – Client (Garrett's father)

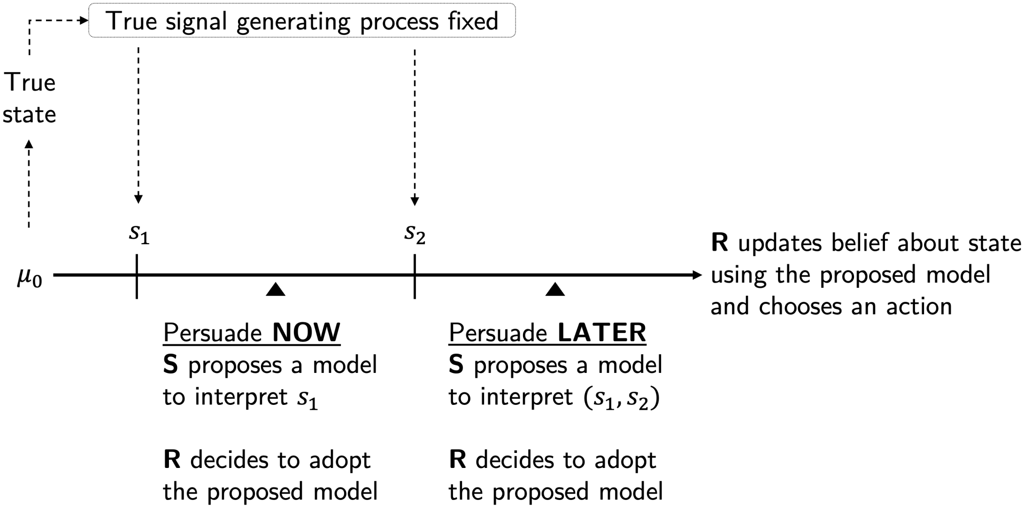


Example

Expert (an auto mechanic) – Client (Garrett's father)

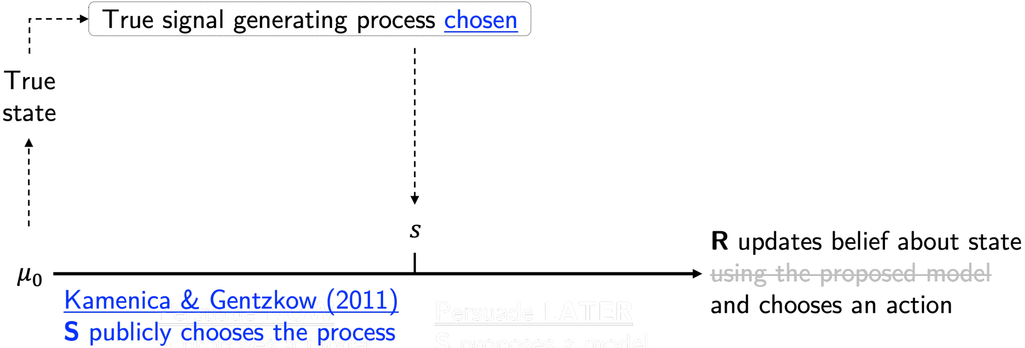


Timeline


True signal generating process fixed
True state
s_1
s_2
μ_0
R updates belief about state using the proposed model and chooses an action
Persuade **NOW**
S proposes a model to interpret s_1
R decides to adopt the proposed model
Persuade **LATER**
S proposes a model to interpret (s_1, s_2)
R decides to adopt the proposed model

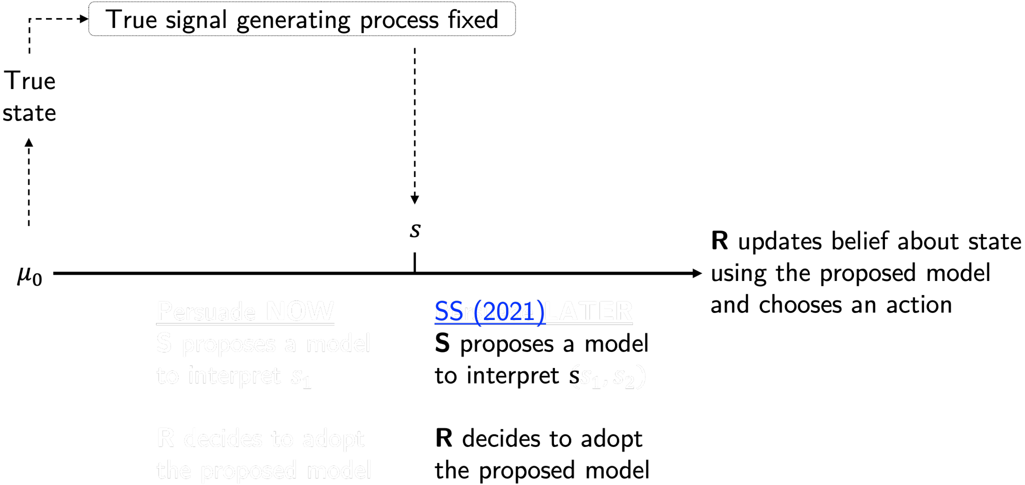
Prior Literature 1

KG (2011) Bayesian Persuasion



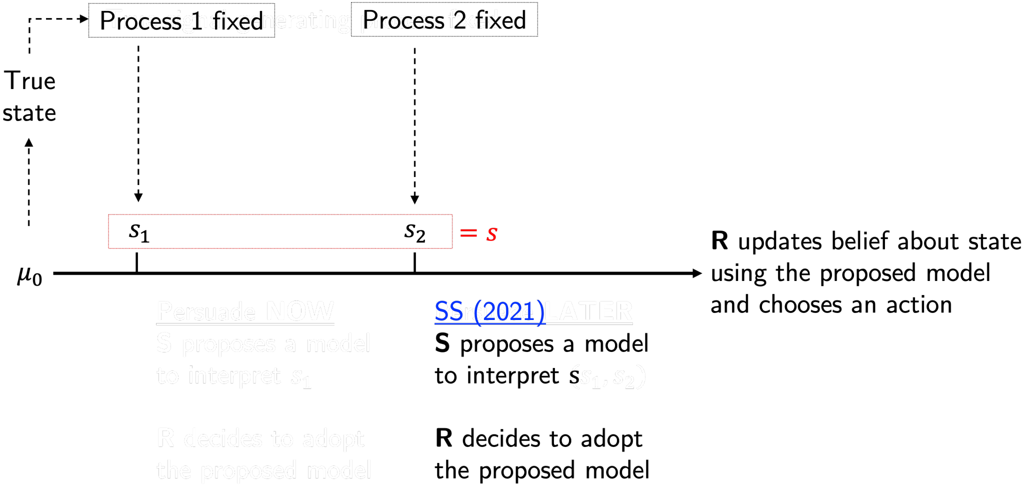
Prior Literature 2

Schwartzstein and Sunderam (2021) Using Models to Persuade



Prior Literature 2

Schwartzstein and Sunderam (2021) Using Models to Persuade



Trade off: Persuading “NOW” vs. “LATER”

If persuade LATER

Unexpected (surprising) data might arise $\Longrightarrow$ wait

- but it might indeed be unexpected $\Longrightarrow$ Say something now to prepare for the expected

Expected data might convince the receiver to take desired action $\Longrightarrow$ wait

- but it might convince him the other way $\Longrightarrow$ Say something now to prevent that

Assumptions– Sender (expert) and Receiver (client)

Sender (expert) and Receiver (client)

- Both do NOT know the true state
- Both do NOT know the true model
- Both know that the two signals are drawn independently from a fixed process

Sender

- Knows the true model "conditional on each state"
- Can only propose a model once
- Can only propose a single model

Assumptions– Receiver (client)

- Has a default model to interpret data
- Does NOT have a prior over models
- Adopts a proposed model if it generates the observed signal with higher likelihood

  —a model that better "fits" the data
- Does NOT consider Sender's incentives
- Limited recall
  - Once the proposed model is adopted, it is never abandoned
  - Once the default model is abandoned, it is never re-adopted

  —the decision to adopt the proposal is taken immediately

Research Question

Timing of persuading a boundedly rational Bayesian's beliefs

by manipulating her interpretation of two signals generated by the same process.

Is it always better to persuade after observing all signals? No.

When?/How?

What matters:

- The prior belief
- Receiver's (default) interpretation of the first signal

⋆ Sender's expectation about the next signal

⋆ What Receiver learns on his own from the first signal

Outline

- Model setup (binary case)
- Period 2: Persuading LATER
- Period 1: Expected payoff of persuading LATER
- Period 1: Persuading NOW
- NOW vs. LATER
  - When Sender knows the next signal with certainty
  - When Sender does NOT know the next signal with certainty

Model Setup: Binary case

- State: $\omega \in \Omega = \{Good, Broken\}$
- Receiver's prior belief $\boldsymbol{\mu}_0 \in \text{int}(\Delta(\Omega))$ $\qquad$ ω_{true} drawn from $\boldsymbol{\mu}_0$
- Signal: $s \in \mathcal{S} = \{Noise, \neg Noise\}$
- Model $\mathcal{M}$: distributions of signals conditional on each state

$$(\pi(s|\omega))_{s\in S,\omega\in\Omega} = (\pi(N|G), \pi(N|B)) \in [\Delta(\mathcal{S})]^{\Omega}$$

- True model $\mathcal{T}$: $(\pi_T(N|G), \pi_T(N|B))$
- Receiver's default model $\mathcal{D}$: $(\pi_d(N|G), \pi_d(N|B))$

- Two signals $(s_1, s_2) \in \mathcal{S}^2$ drawn independently from $\mathcal{T}|\omega_{true}$
- $s_1 = Noise$ is commonly observed $\qquad$ $\cdots$ *Game begins*

Payoffs

Bayesian Receiver who observes s_2 (the second signal)

- updates belief $\boldsymbol{\mu}_0 \to \underbrace{\boldsymbol{\mu}(s_1, s_2; \mathcal{M}, \boldsymbol{\mu}_0)}_{\text{posterior dist. over } \Omega}$ using model $\mathcal{M}$

Posterior belief about $\omega = B$

$$\mu(B|s_1, s_2; \mathcal{M}, \boldsymbol{\mu}_0) = \frac{\mu_0(B)\pi(s_1|B)\pi(s_2|B)}{\sum_{\omega \in \Omega} \mu_0(\omega)\pi(s_1|\omega)\pi(s_2|\omega)} \equiv \Pr(B|s_1, s_2; \mathcal{M}, \boldsymbol{\mu}_0)$$

Payoffs

Bayesian Receiver who observes s_2 (the second signal)

- updates belief $\boldsymbol{\mu}_0 \to \underbrace{\boldsymbol{\mu}(s_1, s_2; \mathcal{M}, \boldsymbol{\mu}_0)}_{\text{posterior dist. over } \Omega}$ using model $\mathcal{M}$
- takes an action $a \in A$ that maximizes his expected utility

$$a^*(s_1, s_2; \mathcal{M}, \boldsymbol{\mu}_0) \in \arg\max_{a \in A} \mathbb{E}_{\mu(\omega|s_1, s_2; \mathcal{M}, \boldsymbol{\mu}_0)} \left[U^R(a, \omega) \right]$$

- Assume $A = [0, 1]$ and $a^*(s_1, s_2; \mathcal{M}, \boldsymbol{\mu}_0) = \Pr(B|s_1, s_2; \mathcal{M}, \boldsymbol{\mu}_0)$

Sender maximizes the receiver's posterior belief about $\omega = B$:

$$\underbrace{\max_{\mathcal{M}} \mathbb{E}\left[\Pr(B|s_1, s_2; \mathcal{M}, \boldsymbol{\mu}_0)\right]}_{\text{“NOW”}} \quad vs. \quad \underbrace{\max_{\mathcal{M}} \Pr(B|s_1, s_2; \mathcal{M}, \boldsymbol{\mu}_0)}_{\text{“LATER’}}$$

(Henceforth, omit $\boldsymbol{\mu}_0$ in every notation.)

Objective of Analysis: NOW vs. LATER

When persuading "NOW", Receiver adopts $\mathcal{M}$ if

Model $\mathcal{M}$'s likelihood of s_1 is higher than that of the default model $\mathcal{D}$

$$\Pr(s_1|\mathcal{M}) \equiv \sum_{\omega\in\Omega} \mu_0(\omega)\pi(s_1|\omega) \geq \sum_{\omega\in\Omega} \mu_0(\omega)\pi_d(s_1|\omega) \equiv \Pr(s_1|\mathcal{D})$$

When persuading "LATER", Receiver adopts $\mathcal{M}$ if

Model $\mathcal{M}$'s likelihood of (s_1, s_2) is higher than that of the default model $\mathcal{D}$

$$\Pr(s_1, s_2|\mathcal{M}) \equiv \sum_{\omega\in\Omega} \mu_0(\omega)\pi(s_1|\omega)\pi(s_2|\omega) \geq \sum_{\omega\in\Omega} \mu_0(\omega)\pi_d(s_1|\omega)\pi_d(s_2|\omega) \equiv \Pr(s_1, s_2|\mathcal{D})$$

In both cases, Sender is proposing a model that better fits the *observed* data

LATER problem

$$V(s_1, s_2) := \max_{\mathcal{M}} \Pr(B|s_1, s_2, \mathcal{M}) \quad \text{s.t.} \quad \Pr(s_1, s_2|\mathcal{M}) \geq \Pr(s_1, s_2|\mathcal{D})$$

$$= \max_{\mathcal{M}} \frac{\mu_0(B)\pi(s_1|B)\pi(s_2|B)}{\sum_{\omega\in\Omega} \mu_0(\omega)\pi(s_1|\omega)\pi(s_2|\omega)}$$

$$\text{s.t.} \quad \sum_{\omega\in\Omega} \mu_0(\omega)\pi(s_1|\omega)\pi(s_2|\omega) \geq \sum_{\omega\in\Omega} \mu_0(\omega)\pi_d(s_1|\omega)\pi_d(s_2|\omega)$$

$$\text{Solution} = \arg\max_{\mathcal{M}} \left(\frac{\pi(s_1|B)\pi(s_2|B)}{\pi(s_1|G)\pi(s_2|G)} \right) \quad \text{s.t.} \quad \Pr(s_1, s_2|\mathcal{M}) \geq \Pr(s_1, s_2|\mathcal{D})$$

LATER problem: Solution

Schwartzstein and Sunderam (2021)

If Sender can be inconsistent (propose a model for each period separately)

$$i.e. \quad \underset{\mathcal{M}^{t_1},\mathcal{M}^{t_2}}{\arg\max} \left(\frac{\pi^{t_1}(s_1|B)\pi^{t_2}(s_2|B)}{\pi^{t_1}(s_1|G)\pi^{t_2}(s_2|G)} \right) \quad \text{s.t.} \quad \Pr(s_1, s_2|\mathcal{M}^{t_1}, \mathcal{M}^{t_2}) \geq \Pr(s_1, s_2|\mathcal{D})$$

- $\pi^{t_1*}(s_1|B)\pi^{t_2*}(s_2|B) = 100\% \quad \forall (s_1, s_2) \in \mathcal{S}^2$
- $\pi^{t_1}(s_1|G)\pi^{t_2}(s_2|G)$ is as small as possible—binds the constraint
- $V(s_1, s_2) = \min\left\{ 100\%, \ \dfrac{\mu_0(B)}{\Pr(s_1, s_2|\mathcal{D})} \right\}$

Receiver finds the signals (s_1, s_2) more surprising $\Longrightarrow$ better persuasion

LATER problem: The cost of consistency

Mixed signals limit Sender's ability to confidently link the data to the desired state

- The best Sender can do is "In bad state, the signals are purely random"

Corollary 1 (The cost of consistency)

If Sender has to be consistent (propose a single model),

- $\pi^*(s_1|B)\pi^*(s_2|B) = \pi^*(s_1|B)(1 - \pi^*(s_1|B)) = \sqrt{50\%} \quad \text{if } s_1 \neq s_2$
- $$V(s_1, s_2) = \begin{cases} \min\left\{100\%, \dfrac{\mu_0(B)}{\Pr(s_1, s_2|\mathcal{D})}\right\} & \text{if } s_1 = s_2 \\ \min\left\{100\%, \dfrac{1}{4}\left(\dfrac{\mu_0(B)}{\Pr(s_1, s_2|\mathcal{D})}\right)\right\} & \text{if } s_1 \neq s_2 \end{cases}$$

LATER problem: Key Points

- Receiver's surprise $\uparrow$ (poor fit) $\Longrightarrow$ persuasion $\uparrow$
- The cost of being consistent
  - mixed signals $\Longrightarrow$ persuasion $\downarrow$ by $1/4$
- True model does not matter at all
- What Receiver learns on his own from the first signal does not matter
  - What matters is the "fit"

NOW problem

- Expected payoff of waiting (i.e., entering the LATER problem)

$$\mathbb{E}_{\mathcal{T}}\Big[V(s_1, s_2)\Big]$$

where Sender's (correct) expectation about s_2 conditional on s_1 based on $\mathcal{T}$:

$$\Pr(s_2|s_1, \mathcal{T}) = \frac{\Pr(s_1, s_2|\mathcal{T})}{\Pr(s_1|\mathcal{T})} = \frac{\sum_{\omega\in\Omega}\mu_0(\omega)\pi_T(s_1|\omega)\pi_T(s_2|\omega)}{\sum_{\omega\in\Omega}\mu_0(\omega)\pi_T(s_1|\omega)}$$

Detail structure of the True model does not matter

Denote the expectation of the next signal:

$$\rho^{\mathcal{T}}(s_1) \equiv \Pr(s_2 = s_1|s_1, \mathcal{T}) \quad \text{and} \quad 1 - \rho^{\mathcal{T}}(s_1) \equiv \Pr(s_2 \neq s_1|s_1, \mathcal{T})$$

NOW problem

- Payoff of persuading "NOW"

$$V(s_1) := \max_{\mathcal{M}} \mathbb{E}_{\rho^{\mathcal{T}}} \Big[\Pr(B|s_1, s_2, \mathcal{M}) \Big] \quad \text{s.t.} \quad \Pr(s_1|\mathcal{M}) \geq \Pr(s_1|\mathcal{D})$$

$$= \max_{\mathcal{M}} \sum_{s_2 \in S} \rho^{\mathcal{T}}(s_2) \left[\frac{\mu_0(B)\pi(s_1|B)\pi(s_2|B)}{\sum_{\omega\in\Omega} \mu_0(\omega)\pi(s_1|\omega)\pi(s_2|\omega)} \right]$$

$$\text{s.t.} \quad \sum_{\omega\in\Omega} \mu_0(\omega)\pi(s_1|\omega) \geq \sum_{\omega\in\Omega} \mu_0(\omega)\pi_d(s_1|\omega)$$

Maximal “NOW” Persuasion: convincing $\omega = B$ with 100%

Proposition 1 (Maximal “NOW” Persuasion)

Sender can achieve maximal persuasion “NOW” if either

$$\text{(a) } \mu_0(B) > \Pr(s_1|\mathcal{D}) \quad \text{or} \quad \text{(b) } \rho^{\mathcal{T}}(s_1) = 0$$

(a) The prior is very favorable compared to how much the default model $\mathcal{D}$ fits s_1

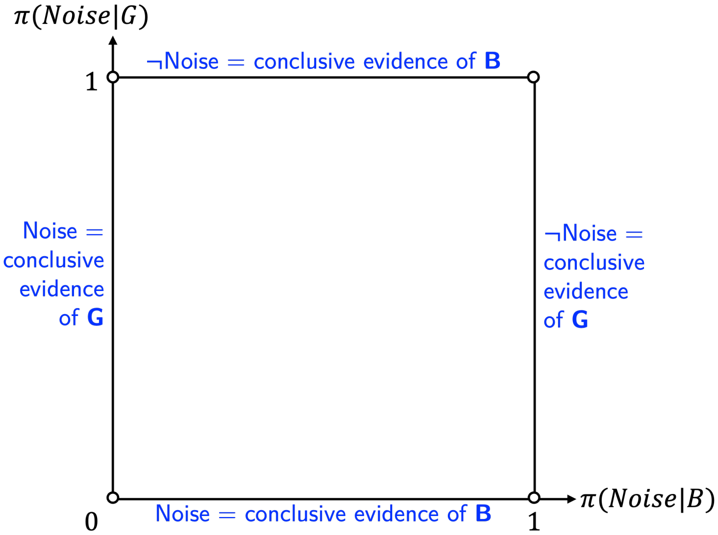
(b) Sender expects mixed signals ($s_1 \neq s_2$) for certainty

When to stop acquiring more data

(a) also applies to “LATER” problem

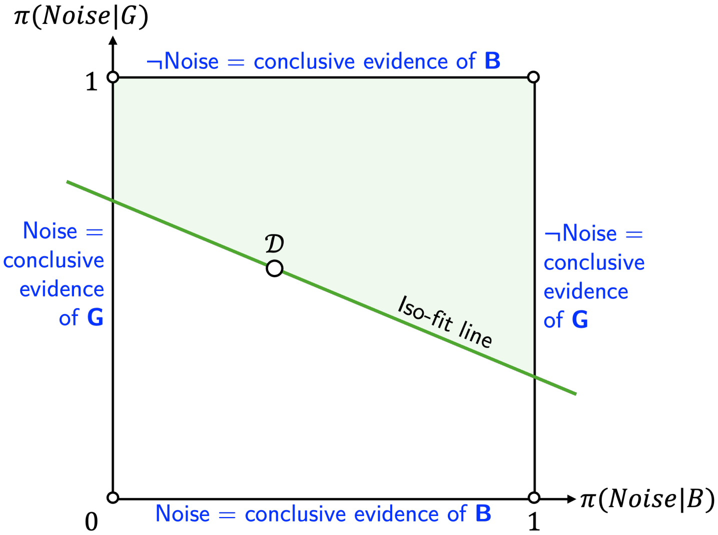
- For any s at any point, $\mu_0(B) > \Pr(s|\mathcal{D})$ is enough to stop acquiring more data

Graphical Illustration: the set of all models



- The set of all models
- Models on edges

Graphical Illustration: the constraint



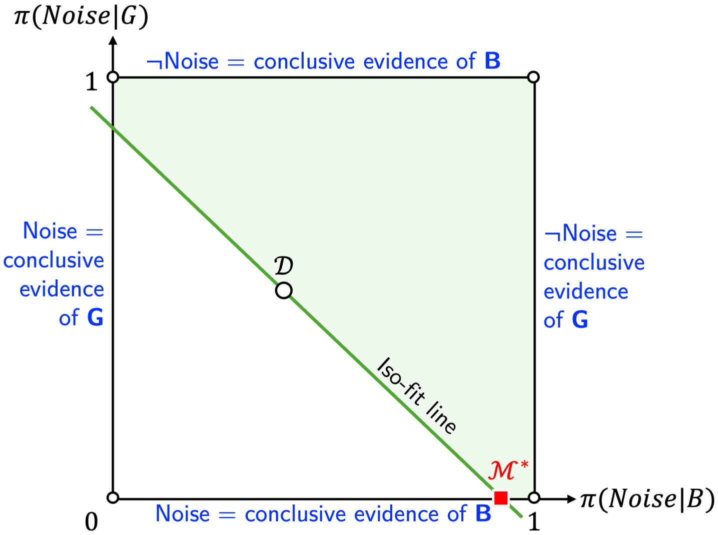
Iso-fit line

$\Pr(Noise|\mathcal{M}) = \Pr(Noise|\mathcal{D})$

- Slope $= -\left(\frac{1-\mu_0(B)}{\mu_0(B)}\right)$

- Default model $\mathcal{D}$

Graphical Illustration: Maximal Persuasion "NOW" (a)



Favorable prior

$\mu_0(B) > \Pr(Noise|\mathcal{D})$

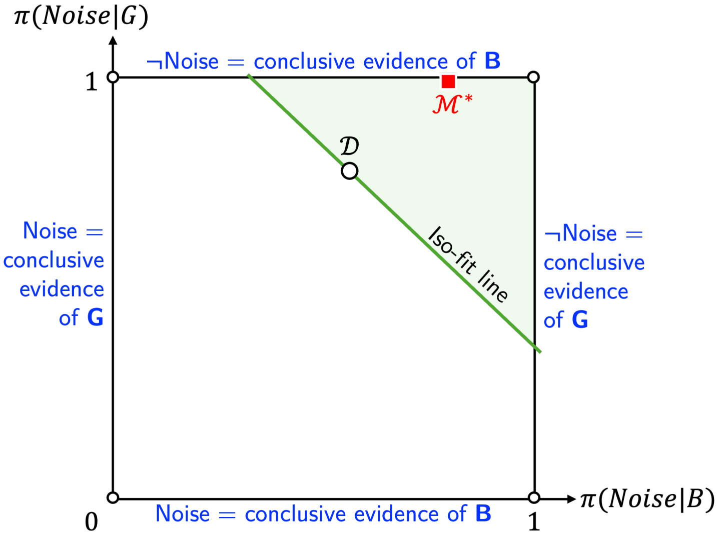
- The optimal model $\mathcal{M}^*$

  $\pi^*(Noise|B) < 1$

  $\pi^*(Noise|G) = 0$

"Noise means red light! Just as you suspected, this tractor needs a repair for sure."

Graphical Illustration: Maximal Persuasion “NOW” (b)



Sender knows $s_2 = \neg Noise$

$\rho^{\mathcal{T}}(Noise) = 0$

- The optimal model $\mathcal{M}^*$
  $\pi^*(Noise|B) < 1$
  $\pi^*(Noise|G) = 1$

“Noise is probably nothing serious. But if the noise ever stops, that’s a flashing red light.”

When the sender knows $s_2 = s_1$

Lemma 1 (Optimal model for "NOW" when Sender knows $s_2 = s_1$)

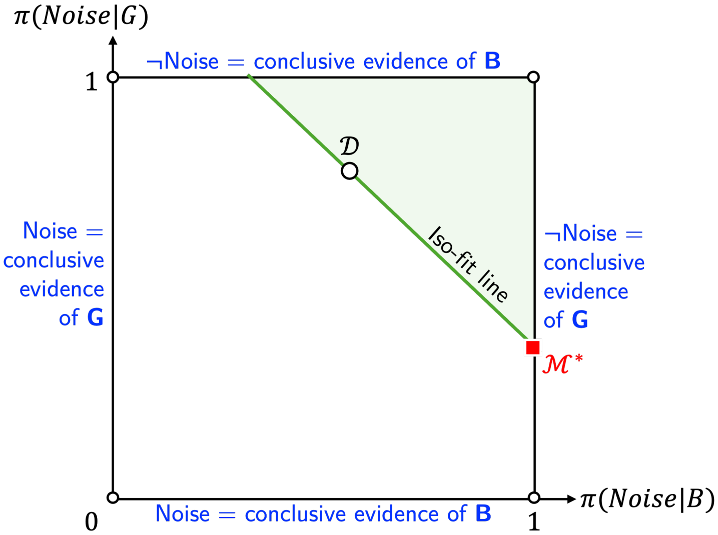
Suppose $\rho^{\mathcal{T}}(s_1) = 1$. Then,

- $V(s_1)$ increases in $\dfrac{\mu_0(B)}{\Pr(s_1|\mathcal{D})}$
- The optimal model for "NOW" is

$$\pi^*(s_1|B) = 100\% \quad \text{and} \quad \pi^*(s_1|G) = \max\left\{0, \frac{\Pr(s_1|\mathcal{D}) - \mu_0(B)}{1 - \mu_0(B)}\right\}$$

- $\pi^*(s_1|B)$ is as small as possible—binds the constraint

When the sender knows $s_2 = s_1$



Sender knows $s_2 = Noise$

$\rho^{\mathcal{T}}(Noise) = 100\%$

- The optimal model $\mathcal{M}^*$
  $\pi^*(Noise|B) = 1$
  with binding constraint

"If it's in good shape, you should probably hear nothing"

NOW vs. LATER: When the sender knows $s_2 = s_1$

Assumption 1 (Maximal "NOW" persuasion is not feasible)

$$\mu_0(B) < \Pr(s_1|\mathcal{D})$$

Proposition 2 (NOW vs. LATER)

Suppose $\rho^{\mathcal{T}}(s_1) = 1$. Then, the followings are equivalent:

(a) "LATER" is better than "NOW"

$$\text{i.e., } V(s_1, s_1) \geq V(s_1)\Big|_{s_2=s_1}$$

(b) The optimal model for "NOW" is also feasible "LATER"

(c) $\Pr(s_1|\mathcal{D}) \leq \dfrac{1 + \pi_d(s_1|B)}{2}$

NOW vs. LATER: When the sender knows $s_2 = s_1$

Corollary 2

"LATER" is better if either

1) $\mu_0(B) < \Pr(s_1|\mathcal{D}) \leq 0.5$

unfavorable prior but surprised $\implies$ to-be very surprised

2) $0.5 \leq \mu_0(B) < \Pr(s_1|\mathcal{D})$

favorable prior and not surprised $\implies$ to-be not surprised (convinced about B)

"NOW" is better if either

1) $\Pr(s_1|\mathcal{D}) \leq \mu_0(B)$

Proposition 1 (Maximal "NOW" persuasion)

2) $\mu_0(B) < 0.5 < \Pr(s_1|\mathcal{D})$ with high enough $\pi_d(s_1|G)$

unfavorable prior and not surprised $\implies$ to-be not surprised (convinced about G)

NOW vs. LATER: When the sender knows $s_2 = s_1$

- Before Sender's proposal, Receiver learns from s_1 that the probability of ω is

$$\Pr(\omega|s_1;\mathcal{D}) \equiv \frac{\mu_0(\omega)\pi_d(s_1|\omega)}{\Pr(s_1|\mathcal{D})}$$

Remark 1 (What Receiver learns on his own from the first signal matters)

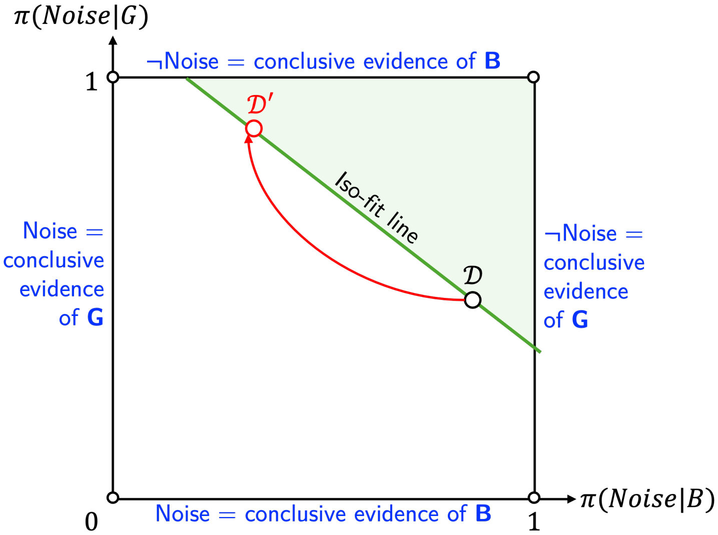
Fix $\Pr(s_1|\mathcal{D}) > 0.5 > \mu_0(B)$. Then, "NOW" is better than "LATER" for any other default model $\mathcal{D}'$ such that $\Pr(s_1|\mathcal{D}') = \Pr(s_1|\mathcal{D})$ and

$$\Pr(G|s_1;\mathcal{D}') \geq 1 - \mu_0(B)\left[\frac{2\Pr(s_1|\mathcal{D}) - 1}{\Pr(s_1|\mathcal{D})}\right]$$

Model-based persuasion timing problem is influenced by both

- the fit of Receiver's initial interpretation of the first signal s_1
- what Receiver initially learns from it

When the sender knows $s_2 = s_1$



Sender knows $s_2 = Noise$

$\rho^{\mathcal{T}}(Noise) = 100\%$

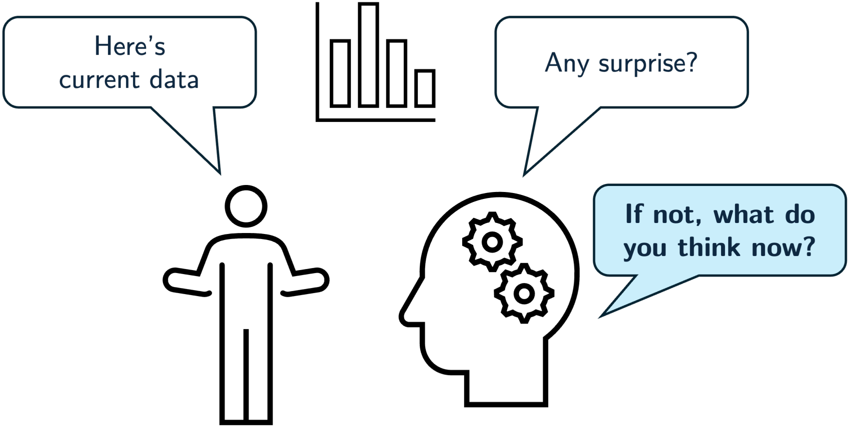
- $\mathcal{D}$ and $\mathcal{D}'$ have same fit

  $\Pr(N|\mathcal{D}) = \Pr(N|\mathcal{D}')$

"LATER" is better for $\mathcal{D}$

"NOW" is better for $\mathcal{D}'$

The impact of future data



In short

"NOW" is better

- when the prior belief is very favorable (Proposition 1a)

  Consistent with Schwartzstein and Sunderam (2021)

- when Sender knows $s_2 \neq s_1$ (Proposition 1b)

  Diverse dataset benefits present persuasion but hurts future persuasion

- when Sender knows $s_2 = s_1$, and (Proposition 2)
  s_1 initially moves Receiver's belief toward the unfavorable state ($\omega = Good$)

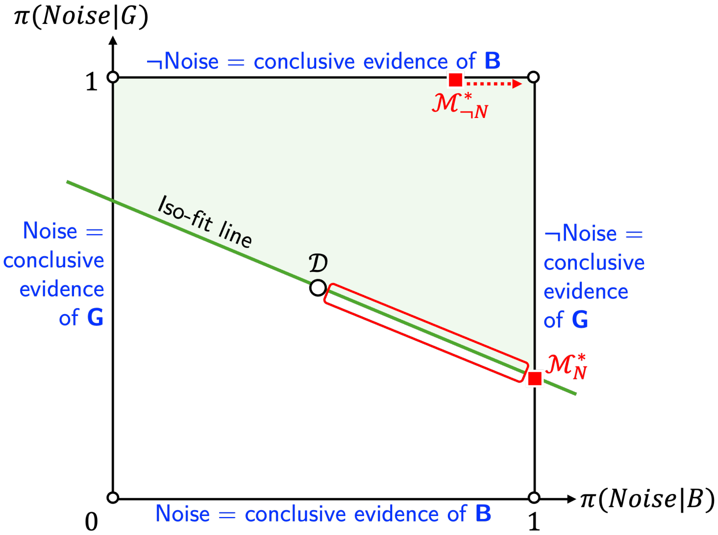
  Sender prevents s_2 from further convincing Receiver unfavorably

Next Step

- Model setup (binary case)
- Period 2: Persuading LATER
- Period 1: Expected payoff of persuading LATER
- Period 1: Persuading NOW
- NOW vs. LATER
  - When Sender knows the next signal with certainty
  - **When Sender does NOT know the next signal with certainty**

- **Period 0** related to Aina (2023) "Tailored Stories"
- **Receiver sophistication: relaxing limited recall**
- **Finite states/signals**

Appendix: When the sender doesn't know the next signal for sure



$\rho^{\mathcal{T}}(Noise) \in (0, 1)$

- $\mathcal{M}^*_{\neg N}$

  $\pi^*(Noise|B) = 1 - \varepsilon$

  $\pi^*(Noise|G) = 1$

- Or

  $\pi^*(Noise|B) > \Pr(s_1|\mathcal{D})$

  with binding constraint

Appendix: When the sender doesn't know the next signal for sure

Example

- $\mu_0(B) = 20\%$
- True model = Receiver's default model: purely random signals

$$\pi_T(Noise|B) = \pi_d(Noise|B) = \pi_T(Noise|G) = \pi_d(Noise|G) = 0.5$$

- Expected payoff of persuading LATER:

$$\mathbb{E}[V(\text{“LATER”})] = \frac{1}{2}\Big(V(N,N) + V(N,\neg N)\Big) = \frac{1}{2}\left(80\% + 20\%\right) = 50\%$$

- Proposing “NOW”

$$\pi^*(Noise|B) = 1 - \varepsilon \quad \text{and} \quad \pi^*(Noise|G) = 1$$

yields the payoff

$$\rho^{\mathcal{T}}(N) \cdot \mu + (1 - \rho^{\mathcal{T}}(N)) \cdot 1 = 50\% \cdot 20\% + 50\% \cdot 100\% = 60\%$$