Bayesian Persuasion

Emir Kamenica and Matthew Gentzkow (2010)

presented by Johann Caro Burnett and Sabyasachi Das

March 4, 2011

Introduction

- Study of strategic communication between two persons - a Sender and a Receiver.
- The Sender reveals information about the state of the world.
- The Receiver observes the information revealed and then takes an action which affects the Sender.
- The authors ask, can the Sender persuade the Receiver to take an action that benefits him? If yes, when?

Motivational example

- A prosecutor (Sender) wishes to persuade the judge (Receiver) to take an action.
- There are two states of the world, $\omega \in \{guilty, innocent\}$.
- Both players share a prior about the defendant, $Pr(guilty) = 0.3$.

Motivational example

- A prosecutor (Sender) wishes to persuade the judge (Receiver) to take an action.
- There are two states of the world, $\omega \in \{guilty, innocent\}$.
- Both players share a prior about the defendant, $Pr(guilty) = 0.3$.
- The judge get a payoff of one if they take the correct action, and zero otherwise.
- The prosecutor gets a payoff of one if the judge convicts and zero otherwise, regardless of the state.
- The prosecutor conducts an investigation (signal) and is required, by law, to report its full outcome.

Motivational example

The investigation consists of:

- Choosing a realization space S, and conditional distributions $\pi(s|\omega)$.
- And after choosing the signal (S, π), reporting the true realization $s \in S$.

Because of the prior, $Pr(guilty) = 0.3$, the judge would acquit if the prosecutor does not send a signal (a pair (S, π)). So, the payoff of the prosecutor would be zero surely. But we will show that he can get a strictly higher payoff.

Motivational example

Consider, $S \equiv \{i, g\}$ and $\pi(i|innocent) = 4/7$ and $\pi(i|guilty) = 0$. That is, the realization i is totally informative, but the realization g is not.

With this signal (S, π), $\mu(innocent|i) = 1$, and $\mu(innocent|g) = 1/2$. So, if the judge sees i they will acquit. In the case that they see g, they are indifferent; so (we can suggest that) they take the action convict. Moreover, g happens with a probability of $6/10$; which is the expected payoff of the prosecutor.

More impressive is the fact that the judge convicts 60% of the time, while the prior of guilty is only 30%.

The Model

- There are two players, Sender and Receiver.
- There is a finite number of states of the world, $\omega \in \Omega$, and both players share the same prior $\mu_0 \in int\Delta(\Omega)$.
- The Receiver can take an action from a compact set, $a \in A$, which affects both players payoff.
- Preferences are $u(a, \omega)$ for the Receiver and $v(a, \omega)$ for the Sender, and they are both continuous.

Strategies

- Sender only knows the prior and the preferences. He chooses a realization space S, and conditional probabilities $\pi(s|\omega)$. After choosing the signal, he has to report the true realization $s \in S$.
- Receiver observes the signal (S, π) and the realization $s \in S$. She updates her beliefs and takes an action.

Equilibrium Concept

The equilibrium concept is the *Sender-preferred subgame perfect equilibrium*. That is, if the Receiver is indifferent between two actions, she chooses the one the favors the Sender.

When both players hold some belief μ, we denote the (unique) optimal action of Receiver by $\hat{a}(\mu)$, and the Sender's payoff by:

$$\hat{v}(\mu) = E_{\mu} v(\hat{a}(\mu), \omega)$$

Consistency of beliefs

Each realization induces a distribution over states:

$$\mu_s(\omega) = \frac{\pi(s|\omega)\mu_0(\omega)}{\sum_{\omega' \in \Omega} \pi(s|\omega')\mu_0(\omega')} \text{ for all } s \text{ and } \omega$$

Moreover, a signal (S, π) induces a distribution τ over posteriors s.t.

$$Supp\{\tau\} = \{\mu_s : s \in S\} \text{ and}$$

$$\tau(\mu) = \sum_{s:\mu_s=\mu} \sum_{\omega' \in \Omega} \pi(s|\omega')\mu_0(\omega') \text{ for all } \mu$$

Consistency of beliefs

A distribution of posterior is *Bayes-plausible* if the expected posterior equals the prior:

$$\int \mu d\tau(\mu) = \mu_0$$

Proposition 1

A signal is called *straightforward* if $S \subset A$.

Proposition

The following are equivalent:

(i) *There exists a signal that attains value* v^*.

(ii) *There exists a straightforward signal that attains value* v^*.

(iii) *There is a Bayes-plausible distribution of posteriors* τ *such that* $E_\tau \hat{v}(\mu) = v^*$.

Definition

Sender benefits from persuasion if there exists a signal (S, π) that yields an expected payoff strictly higher than the payoff with no signal.

Corollary

Sender benefits from persuasion if and only if there exists a Bayes-plausible distribution over measures μ on Ω such that:

$$E_\tau \hat{v}(\mu) > \hat{v}(\mu_0)$$

Moreover, the value of an optimal signal is given by:

$$\sup_\tau E_\tau \hat{v}(\mu)$$

$$s.t. \int \mu d\tau(\mu) = \mu_0$$

Value of Optimal Signal

- We define V as the concave closure of $\hat{v}$:

$$V(\mu) \equiv \sup\{z | (\mu, z) \in co(\hat{v})\}$$

where $co(\hat{v})$ denotes the convex hull of the graph of $\hat{v}$.

Value of Optimal Signal

- We define V as the concave closure of $\hat{v}$:

$$V(\mu) \equiv \sup\{z | (\mu, z) \in co(\hat{v})\}$$

  where $co(\hat{v})$ denotes the convex hull of the graph of $\hat{v}$.
- Note that V is the smallest concave function that is everywhere weakly greater than $\hat{v}$.

Value of Optimal Signal

- We define V as the concave closure of $\hat{v}$:

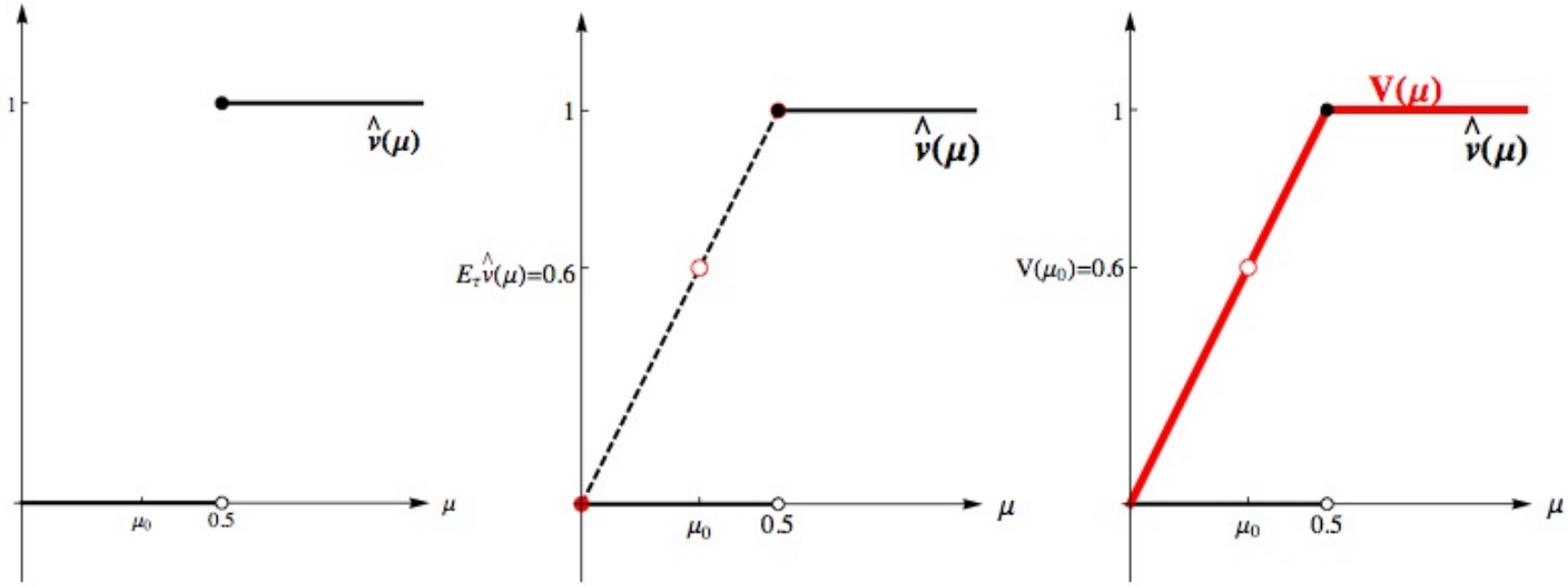
$$V(\mu) \equiv \sup\{z|(\mu, z) \in co(\hat{v})\}$$

  where $co(\hat{v})$ denotes the convex hull of the graph of $\hat{v}$.
- Note that V is the smallest concave function that is everywhere weakly greater than $\hat{v}$.

Corollary

The value of an optimal signal is $V(\mu_0)$. Sender benefits from persuasion if and only if $V(\mu_0) > \hat{v}(\mu_0))$

Value in Motivating Example



(a) function $\hat{v}$ (b) optimal signal (c) function V

Figure 2: the motivating example

When Does Sender Benefit from Persuasion?

- Or, for what values of prior μ_0 we get $V(\mu_0) > \hat{v}(\mu_0)$?

When Does Sender Benefit from Persuasion?

- Or, for what values of prior μ_0 we get $V(\mu_0) > \hat{v}(\mu_0)$?
- For the motivating example this is the case when $\mu_0 \in (0, \frac{1}{2})$

When Does Sender Benefit from Persuasion?

- Or, for what values of prior μ_0 we get $V(\mu_0) > \hat{v}(\mu_0)$?
- For the motivating example this is the case when $\mu_0 \in (0, \frac{1}{2})$
- In general we require a τ such that $\mathbb{E}_\tau \hat{v}(\mu) > \hat{v}(\mathbb{E}_\tau(\mu))$

When Does Sender Benefit from Persuasion?

- Or, for what values of prior μ_0 we get $V(\mu_0) > \hat{v}(\mu_0)$?
- For the motivating example this is the case when $\mu_0 \in (0, \frac{1}{2})$
- In general we require a τ such that $\mathbb{E}_\tau \hat{v}(\mu) > \hat{v}(\mathbb{E}_\tau(\mu))$
- If $\hat{v}$ is concave, Sender does not benefit from persuasion for any prior. If $\hat{v}$ is convex and not concave, Sender benefits from persuasion for every prior.

When Does Sender Benefit from Persuasion?

- Or, for what values of prior μ_0 we get $V(\mu_0) > \hat{v}(\mu_0)$?
- For the motivating example this is the case when $\mu_0 \in (0, \frac{1}{2})$
- In general we require a τ such that $\mathbb{E}_\tau \hat{v}(\mu) > \hat{v}(\mathbb{E}_\tau(\mu))$
- If $\hat{v}$ is concave, Sender does not benefit from persuasion for any prior. If $\hat{v}$ is convex and not concave, Sender benefits from persuasion for every prior.
- Many times $\hat{v}$ is neither concave nor convex, as in the example. So what do we do?

Conditions for Benefit from Persuasion

- There is *information sender would share* if $\exists\ \mu$ s.t.

$$\hat{v}(\mu) > \mathbb{E}_{\mu} v(\hat{a}(\mu_0), \omega)$$

Conditions for Benefit from Persuasion

- There is *information sender would share* if $\exists\ \mu$ s.t.

$$\hat{v}(\mu) > \mathbb{E}_\mu v(\hat{a}(\mu_0), \omega)$$

- Receiver's *preference is discrete* at belief μ if $\exists\ \epsilon > 0$ s.t.

$$\forall a \neq \hat{a}(\mu), \quad \mathbb{E}_\mu u(\hat{a}(\mu), \omega) > \mathbb{E}_\mu u(a, \omega) + \epsilon$$

Conditions for Benefit from Persuasion

- There is *information sender would share* if $\exists\ \mu$ s.t.

$$\hat{v}(\mu) > \mathbb{E}_\mu v(\hat{a}(\mu_0), \omega)$$

- Receiver's *preference is discrete* at belief μ if $\exists\ \epsilon > 0$ s.t.

$$\forall a \neq \hat{a}(\mu), \quad \mathbb{E}_\mu u(\hat{a}(\mu), \omega) > \mathbb{E}_\mu u(a, \omega) + \epsilon$$

Proposition

If there is no information Sender would share, Sender does not benefit from persuasion. If there is information Sender would share and Receiver's preference is discrete at the prior, Sender benefits from persuasion. If A is finite, Receiver's preference is discrete at the prior generically.

Proof

Claim

If Receiver's belief is discrete at μ then $\hat{a}(\mu)$ doesn't change in the neighborhood of μ.

Proof

Claim

If Receiver's belief is discrete at μ then $\hat{a}(\mu)$ doesn't change in the neighborhood of μ.

Proof.

Suppose this is false.
Let $\mu^{'} \approxeq \mu$ and $\mu^{'} \neq \mu$. Then $\hat{a}(\mu^{'}) \neq \hat{a}(\mu)$.

Proof

Claim

If Receiver's belief is discrete at μ then $\hat{a}(\mu)$ doesn't change in the neighborhood of μ.

Proof.

Suppose this is false.
Let $\mu^{'} \approxeq \mu$ and $\mu^{'} \neq \mu$. Then $\hat{a}(\mu^{'}) \neq \hat{a}(\mu)$.
Now we have $\mathbb{E}_{\mu} u(\hat{a}(\mu), \omega) > \mathbb{E}_{\mu} u(\hat{a}(\mu^{'}), \omega) + \epsilon$.

Proof

Claim

If Receiver's belief is discrete at μ then $\hat{a}(\mu)$ doesn't change in the neighborhood of μ.

Proof.

Suppose this is false.
Let $\mu^{'} \approxeq \mu$ and $\mu^{'} \neq \mu$. Then $\hat{a}(\mu^{'}) \neq \hat{a}(\mu)$.
Now we have $\mathbb{E}_{\mu} u(\hat{a}(\mu), \omega) > \mathbb{E}_{\mu} u(\hat{a}(\mu^{'}), \omega) + \epsilon$.
But $\mathbb{E}_{\mu} u(\hat{a}(\mu), \omega)$ is continuous in μ by the Maximum theorem.

Proof

Claim

If Receiver's belief is discrete at μ then $\hat{a}(\mu)$ doesn't change in the neighborhood of μ.

Proof.

Suppose this is false.
Let $\mu^{'} \approxeq \mu$ and $\mu^{'} \neq \mu$. Then $\hat{a}(\mu^{'}) \neq \hat{a}(\mu)$.
Now we have $\mathbb{E}_{\mu} u(\hat{a}(\mu), \omega) > \mathbb{E}_{\mu} u(\hat{a}(\mu^{'}), \omega) + \epsilon$.
But $\mathbb{E}_{\mu} u(\hat{a}(\mu), \omega)$ is continuous in μ by the Maximum theorem.

Hence $\mathbb{E}_{\mu} u(\hat{a}(\mu), \omega) \approxeq \mathbb{E}_{\mu'} u(\hat{a}(\mu), \omega)$ and $\mathbb{E}_{\mu} u(\hat{a}(\mu^{'}), \omega) \approxeq \mathbb{E}_{\mu'} u(\hat{a}(\mu), \omega)$

Proof

Claim

If Receiver's belief is discrete at μ then $\hat{a}(\mu)$ doesn't change in the neighborhood of μ.

Proof.

Suppose this is false.
Let $\mu^{'} \approxeq \mu$ and $\mu^{'} \neq \mu$. Then $\hat{a}(\mu^{'}) \neq \hat{a}(\mu)$.
Now we have $\mathbb{E}_{\mu} u(\hat{a}(\mu), \omega) > \mathbb{E}_{\mu} u(\hat{a}(\mu^{'}), \omega) + \epsilon$.
But $\mathbb{E}_{\mu} u(\hat{a}(\mu), \omega)$ is continuous in μ by the Maximum theorem.

Hence $\mathbb{E}_{\mu} u(\hat{a}(\mu), \omega) \approxeq \mathbb{E}_{\mu'} u(\hat{a}(\mu), \omega)$ and $\mathbb{E}_{\mu} u(\hat{a}(\mu^{'}), \omega) \approxeq \mathbb{E}_{\mu'} u(\hat{a}(\mu), \omega)$
Hence if $\mu^{'}$ is chosen to be close enough the inequality would hold for $\mu^{'}$ as well.
But that would contradict the definition of $\hat{a}(\mu)$. $\square$

Proof Contd.

Now there exists μ_l such that $\hat{v}(\mu_l) > \mathbb{E}_{\mu_l} v(\hat{a}(\mu_0), \omega)$.

Proof Contd.

Now there exists μ_l such that $\hat{v}(\mu_l) > \mathbb{E}_{\mu_l} v(\hat{a}(\mu_0), \omega)$.

There is μ_h in the neighborhood of μ_0 such that $\hat{a}(\mu_h) = \hat{a}(\mu_0)$.

Proof Contd.

Now there exists μ_l such that $\hat{v}(\mu_l) > \mathbb{E}_{\mu_l} v(\hat{a}(\mu_0), \omega)$.

There is μ_h in the neighborhood of μ_0 such that $\hat{a}(\mu_h) = \hat{a}(\mu_0)$.

Now we can choose μ_h in such a way to make μ_0 a convex combination of the two, i.e. $\exists \gamma \in [0, 1]$ s.t. $\mu_0 = \gamma \mu_h + (1 - \gamma)\mu_l$

Proof Contd.

Now there exists μ_l such that $\hat{v}(\mu_l) > \mathbb{E}_{\mu_l} v(\hat{a}(\mu_0), \omega)$.

There is μ_h in the neighborhood of μ_0 such that $\hat{a}(\mu_h) = \hat{a}(\mu_0)$.

Now we can choose μ_h in such a way to make μ_0 a convex combination of the two, i.e. $\exists \gamma \in [0,1]$ s.t. $\mu_0 = \gamma \mu_h + (1-\gamma)\mu_l$

$$\begin{aligned} V(\mu_0) &\geq \gamma \mathbb{E}_{\mu_h} v(\hat{a}(\mu_h), \omega) + (1-\gamma)\mathbb{E}_{\mu_l} v(\hat{a}(\mu_l), \omega) \\ &> \gamma \mathbb{E}_{\mu_h} v(\hat{a}(\mu_0), \omega) + (1-\gamma)\mathbb{E}_{\mu_l} v(\hat{a}(\mu_0), \omega) \\ &= \textstyle\sum_{\omega \in \Omega} v(\hat{a}(\mu_0), \omega)[\gamma \mu_h(\omega) + (1-\gamma)\mu_l(\omega)] \\ &= \hat{v}(\mu_0) \end{aligned}$$

Optimal Signal

- We say *no disclosure* is optimal if $\mu_s = \mu_0$ for all s realized with positive probability under the optimal signal.
- If μ_s is degenerate for all s realized with positive probability under the optimal signal, we say *full disclosure* is optimal under the optimal signal.
- If $\hat{v}$ is (strictly) concave, no disclosure is (uniquely) optimal, and if $\hat{v}$ is (strictly) convex full disclosure is (uniquely) optimal.

Lobbying

- We consider a setting where a lobbying group commissions a study with the goal of influencing a benevolent, but nonetheless rational, politician.

Lobbying

- We consider a setting where a lobbying group commissions a study with the goal of influencing a benevolent, but nonetheless rational, politician.
- The politician (Receiver) chooses a unidimensional policy $a \in [0, 1]$
- The state $\omega \in [0, 1]$ is the socially optimal policy.
- The lobbyist (Sender) is employed by the interest group whose preferred action is $a^* = \alpha\omega + (1 - \alpha)\omega^*$, with $\alpha \in [0, 1]$ and $\omega^* > 1$.
- Politician's payoff $u = -(a - \omega)^2$
- Lobbyist's payoff $v = -(a - a^*)^2$
- Therefore we know that $\hat{a}(\mu) = \mathbb{E}_\mu[\omega]$

Lobbying Cntd.

- Simple algebra reveals that

$$\hat{v}(\mu) = \underbrace{-(1-\alpha)^2\omega^{*2} + 2(1-\alpha)^2\omega^*\mathbb{E}_\mu[\omega] - \alpha^2\mathbb{E}_\mu[\omega^2]}_{\text{linear in } \mu} + (2\alpha - 1)\underbrace{(\mathbb{E}_\mu[\omega])^2}_{\text{convex in } \mu}$$

Lobbying Cntd.

- Simple algebra reveals that

$$\hat{v}(\mu) = \underbrace{-(1-\alpha)^2\omega^{*2} + 2(1-\alpha)^2\omega^*\mathbb{E}_\mu[\omega] - \alpha^2\mathbb{E}_\mu[\omega^2]}_{\text{linear in } \mu} + (2\alpha - 1)\underbrace{(\mathbb{E}_\mu[\omega])^2}_{\text{convex in } \mu}$$

- $\hat{v}$ is linear in μ if $\alpha = \frac{1}{2}$, strictly convex if $\alpha > \frac{1}{2}$, and strictly concave when $\alpha < \frac{1}{2}$.

Lobbying Cntd.

- Simple algebra reveals that

$$\hat{v}(\mu) = \underbrace{-(1-\alpha)^2\omega^{*2} + 2(1-\alpha)^2\omega^*\mathbb{E}_\mu[\omega] - \alpha^2\mathbb{E}_\mu[\omega^2]}_{\text{linear in } \mu} + (2\alpha - 1)\underbrace{(\mathbb{E}_\mu[\omega])^2}_{\text{convex in } \mu}$$

- $\hat{v}$ is linear in μ if $\alpha = \frac{1}{2}$, strictly convex if $\alpha > \frac{1}{2}$, and strictly concave when $\alpha < \frac{1}{2}$.
- Therefore we have full disclosure if $\alpha > \frac{1}{2}$ and no disclosure if $\alpha < \frac{1}{2}$.

Lobbying Cntd.

- Simple algebra reveals that

$$\hat{v}(\mu) = \underbrace{-(1-\alpha)^2\omega^{*2} + 2(1-\alpha)^2\omega^*\mathbb{E}_\mu[\omega] - \alpha^2\mathbb{E}_\mu[\omega^2]}_{\text{linear in } \mu} + (2\alpha - 1)\underbrace{(\mathbb{E}_\mu[\omega])^2}_{\text{convex in } \mu}$$

- $\hat{v}$ is linear in μ if $\alpha = \frac{1}{2}$, strictly convex if $\alpha > \frac{1}{2}$, and strictly concave when $\alpha < \frac{1}{2}$.
- Therefore we have full disclosure if $\alpha > \frac{1}{2}$ and no disclosure if $\alpha < \frac{1}{2}$.
- Optimal signal is independent of ω^*.

Lobbying Cntd.

- Simple algebra reveals that

$$\hat{v}(\mu) = \underbrace{-(1-\alpha)^2\omega^{*2} + 2(1-\alpha)^2\omega^*\mathbb{E}_\mu[\omega] - \alpha^2\mathbb{E}_\mu[\omega^2]}_{\text{linear in } \mu} + (2\alpha - 1)\underbrace{(\mathbb{E}_\mu[\omega])^2}_{\text{convex in } \mu}$$

- $\hat{v}$ is linear in μ if $\alpha = \frac{1}{2}$, strictly convex if $\alpha > \frac{1}{2}$, and strictly concave when $\alpha < \frac{1}{2}$.
- Therefore we have full disclosure if $\alpha > \frac{1}{2}$ and no disclosure if $\alpha < \frac{1}{2}$.
- Optimal signal is independent of ω^*.
- Thus theory suggests that lobbyist either commissions a fully revealing study or no study at all.

Lobbying Cntd.

- Simple algebra reveals that

$$\hat{v}(\mu) = \underbrace{-(1-\alpha)^2\omega^{*2} + 2(1-\alpha)^2\omega^*\mathbb{E}_\mu[\omega] - \alpha^2\mathbb{E}_\mu[\omega^2]}_{\text{linear in } \mu} + (2\alpha - 1)\underbrace{(\mathbb{E}_\mu[\omega])^2}_{\text{convex in } \mu}$$

- $\hat{v}$ is linear in μ if $\alpha = \frac{1}{2}$, strictly convex if $\alpha > \frac{1}{2}$, and strictly concave when $\alpha < \frac{1}{2}$.
- Therefore we have full disclosure if $\alpha > \frac{1}{2}$ and no disclosure if $\alpha < \frac{1}{2}$.
- Optimal signal is independent of ω^*.
- Thus theory suggests that lobbyist either commissions a fully revealing study or no study at all.
- This contrasts with the real life observation.

Lobbying Cntd.

- Simple algebra reveals that

$$\hat{v}(\mu) = \underbrace{-(1-\alpha)^2\omega^{*2} + 2(1-\alpha)^2\omega^*\mathbb{E}_\mu[\omega] - \alpha^2\mathbb{E}_\mu[\omega^2]}_{\text{linear in } \mu} + (2\alpha - 1)\underbrace{(\mathbb{E}_\mu[\omega])^2}_{\text{convex in } \mu}$$

- $\hat{v}$ is linear in μ if $\alpha = \frac{1}{2}$, strictly convex if $\alpha > \frac{1}{2}$, and strictly concave when $\alpha < \frac{1}{2}$.
- Therefore we have full disclosure if $\alpha > \frac{1}{2}$ and no disclosure if $\alpha < \frac{1}{2}$.
- Optimal signal is independent of ω^*.
- Thus theory suggests that lobbyist either commissions a fully revealing study or no study at all.
- This contrasts with the real life observation.
- May be that benevolent politician is not that rational! Or, there may be a commitment problem on the part of the lobbyist not to distort the study results *ex post*.

Conclusion

- The paper studies a scenario of strategic communication where one agent reveals information to another in a strategic way in order to persuade him to take a preferred action.
- It gives us necessary and sufficient conditions for the Persuader to benefit from persuasion.
- As it turns out, the scope for persuasion is, surprisingly, quite large.
- The model can be easily extended to Receiver having private information about the state.
- Also can be extended to the case of multiple receivers.
- How about multiple Persuaders? See Gentzkow and Kamenica (Working Paper, 2011)