

Repeated Communication with Private Lying Cost

HARRY PEI

Department of Economics, Northwestern University

Canadian Economic Theory Conference

May 8th, 2020

Motivation: Sender's Commitment to Disclosure Policies

Bayesian persuasion game (Kamenica and Gentzkow 2011):

- Payoff-relevant state $\omega \in \Omega$, with Ω finite.
- Sender **commits to a disclosure policy** $\{\sigma(\cdot|\omega)\}_{\omega \in \Omega}$, with $\sigma(\cdot|\omega) \in \Delta(M)$.
- Sender honours her commitment after observing ω , and sends $m \in M$.
- Receiver observes message m , and chooses action $a \in A$.

Controversial: Can the sender commit to disclosure policies?

- Sender's optimal disclosure policy requires *nontrivial mixing* & some messages lead to *strictly higher payoffs* compared to others.
- Honouring commitment is against the sender's own interest.

Microfoundation: Repeated Communication Games

A patient sender communicates with a sequence of receivers.

- State of the world $\omega_t \in \Omega$, with $\{\omega_t\}_{t \in \mathbb{N}}$ i.i.d.
- Sender privately observes ω_t and sends a message $m_t \in M$.
- Receiver observes $\{\omega_s, m_s, a_s\}_{s=0}^{t-1}$ and m_t , and then chooses $a_t \in A$.

Folk theorem result of Fudenberg, Kreps and Maskin (1990):

- In many games of interest, patient sender's equilibrium payoff is **strictly bounded away** from her optimal commitment payoff.

Future receivers **cannot perfectly monitor sender's mixed actions**.

Existing solutions: **Allow sender-commitment to kick in**, e.g.,

- sender can commit to truthfully disclose private randomisations;
- with positive prob, sender is a commitment type that mechanically communicates according to her optimal disclosure policy.

My Model: Repeated Communication w/o Commitment

Time: $t = 0, 1, 2, \dots$

- Long-lived sender (discount δ) vs sequence of short-lived receivers.

In period t ,

- State $\omega_t \in \{\omega^h, \omega^l\}$, i.i.d. over time, with $\Pr(\omega_t = \omega^h) = p \in (0, 1/2)$.
- Sender observes ω_t and sends $m_t \in \{m^h, m^l\}$.
- Receiver chooses $a_t \in \{a^h, a^l\}$ after observing $\{\omega_s, m_s, a_s\}_{s=0}^{t-1}$ and m_t .

Receiver's payoff: $\mathbf{1}\{a_t = \omega_t\}$.

Sender's stage-game payoff: $\mathbf{1}\{a_t = a^h\} - C \cdot \mathbf{1}\{m_t \neq \omega_t\}$. [Alternative View](#)

- $C \in \{C_1, \dots, C_m\}$, **perfectly persistent, sender's private info (or type)**.
- Baseline model: $0 \leq C_m < C_{m-1} < \dots < C_1 < 1$.

Every type of sender has a strict incentive to mislead receivers.

- Receivers' full support prior belief $\pi \in \Delta\{C_1, \dots, C_m\}$.

Overview of Three Theorems

1. Characterise every type of patient sender's highest equilibrium payoff.
 - conditions s.t. *all* types' highest equilibrium payoffs converge to their respective optimal commitment payoffs.
2. Distinctions between strategic-type sender and commitment-type.
 - **No type of strategic sender** uses his optimal disclosure policy at every on-path history.
3. Allowing for ethical type senders (i.e., lying cost ≥ 1).
 - Characterise when can **all non-ethical types can attain their optimal commitment payoffs**.
 - The possibility of being ethical and having a high lying cost **can hurt non-ethical type senders**

Related Literature

Repeated communication: Best and Quigley (17), Mathevet et al. (19).

- No type of sender can commit to disclosure policies. Sender cannot commit to truthfully disclose private randomisations.

Static communication with lying cost: Kartik et al. (07), Kartik (09), Guo and Shmaya (19), Nguyen and Tan (19)

- Sender observes the state, cannot perform info control.
- Non-ethical sender can be hurt by ethical types.

Repeated games with persistent private information: Aumann and Maschler (65), Hart (85), Shalev (94), Cripps and Thomas (03), Pęski (14).

- Uninformed player is impatient + non-zero sum.

Reputation w/o commitment types: Weinstein and Yildiz (16), Pei (19).

Cost of lying: Gneezy (05), Gneezy et al.(18), Sobel (20).

Benchmarks

Recall that in the baseline model, $0 \leq C_m < C_{m-1} < \dots < C_1 < 1$.

1. Type C_j 's optimal commitment payoff: $v_j^{**} \equiv p + p(1 - C_j)$.
2. Repeated communication game when C is common knowledge.
 - Patient sender's highest equilibrium payoff is p ,
i.e., her payoff in a 1-shot interaction by fully disclosing ω .

Why?

- At every history s.t. receiver plays a^h with positive prob, exists message m' s.t. receiver plays a^l for sure, and the sender sends m' when the state is ω^l with positive prob.
- **Sending m' when the state is ω^l** at every such history is sender's best reply, from which her stage-game payoff $\leq p$.

Patient Sender's Highest Equilibrium Payoff

Recall that in the baseline model, $0 \leq C_m < C_{m-1} < \dots < C_1 < 1$.

For every $j \in \{1, 2, \dots, m\}$, let

$$v_j^* \equiv p \left(1 + \frac{C_1 - C_j}{2p + C_1(1 - 2p)} \right)$$

Theorem 1

For every $\varepsilon > 0$, there exists $\underline{\delta} \in (0, 1)$ such that when $\delta > \underline{\delta}$,

1. \nexists BNE such that type C_1 attains payoff strictly more than v_1^* ,
 \nexists BNE such that type C_j attains payoff more than $v_j^* + \varepsilon$, $\forall j \geq 2$.
2. \exists sequential equilibrium s.t. the sender attains payoff within an ε neighbourhood of (v_1^*, \dots, v_m^*) .

Implications of Theorem 1

Formula for type C_j 's *highest limiting equilibrium payoff*:

$$v_j^* \equiv p \left(1 + \frac{C_1 - C_j}{2p + C_1(1 - 2p)} \right)$$

1. Depends only on C_j and C_1 , not on type distribution and other types.

2. $v_1^* = p$, and $p < v_j^* < p + p(1 - C_j)$ for every $j \in \{2, \dots, m\}$.

Every type except for C_1 can *strictly benefit* from incomplete info.

Puzzle: Extract info rent in the long run & preserve info advantage.

3. As $C_1 \rightarrow 1$, $v_j^* \rightarrow p + p(1 - C_j)$ for every $j \in \{1, 2, \dots, m\}$.

Every type can approximately attain her optimal commitment payoff.

Proof of Theorem 1: Overview

Sufficiency: Construct equilibria that approximately attain $v^* \equiv (v_1^*, \dots, v_m^*)$.

- High-cost type mixes to *provide cover* for low-cost types.
- Low-cost types mix except when the posterior prob of high-cost type is sufficiently high, in which they have strict incentive to lie.
- Low-cost types can *rebuild* their reputations after milking them.

According to **every pure best reply of low-cost type**, he needs to tell the truth for a certain number of periods to reach a history in which he has a strict incentive to lie.

Necessity: Type C_j 's payoff cannot significantly exceed v_j^* .

- Relate repeated game outcome to a constrained optimisation problem.
- Value of constrained optimisation problem is v_j^* .
- Establish the necessity of the constraints for BNE in the repeated game.

Rewrite Players' Stage-Game in Normal Form

Sender's stage-game pure action:

- $\mathbf{a} : \Omega \rightarrow M$, with $\mathbf{a} \in \mathbf{A}$.

Receiver's stage-game pure action:

- $\mathbf{b} : M \rightarrow A$, with $\mathbf{b} \in \mathbf{B}$.

Rewrite players' stage-game payoff functions as:

- $u_r(\mathbf{a}, \mathbf{b})$ and $u_s(C, \mathbf{a}, \mathbf{b})$.

Necessity: v_j^* as a Constrained Optimisation Problem

Proposition 1

v_j^* is the value of the following constrained optimisation problem:

$$\max_{\gamma \in \Delta(\mathbf{A} \times \mathbf{B})} \sum_{(\mathbf{a}, \mathbf{b}) \in \mathbf{A} \times \mathbf{B}} \gamma(\mathbf{a}, \mathbf{b}) u_s(C_j, \mathbf{a}, \mathbf{b}),$$

subject to:

1.

$$\sum_{(\mathbf{a}, \mathbf{b}) \in \mathbf{A} \times \mathbf{B}} \gamma(\mathbf{a}, \mathbf{b}) u_s(C_1, \mathbf{a}, \mathbf{b}) \leq p,$$

2. for every $\mathbf{b} \in \mathbf{B}$ that the marginal distribution of γ on \mathbf{B} attaches positive prob to,

$$\mathbf{b} \in \arg \max_{\mathbf{b}' \in \mathbf{B}} u_r(\gamma(\cdot | \mathbf{b}), \mathbf{b}'),$$

where $\gamma(\cdot | \mathbf{b})$ is the distribution over sender's stage-game action conditional on the receiver's stage-game action being \mathbf{b} .

Relate Maximisation Problem to the Repeated Game

For any strategy profile in the repeated game $((\sigma_{C_j})_{j=1}^m, \sigma_r)$, let

$$\gamma^j(\mathbf{a}, \mathbf{b}) \equiv \mathbb{E}^{(\sigma_{C_j}, \sigma_r)} \left[\sum_{t=0}^{\infty} (1 - \delta) \delta^t \mathbf{1}\{(\mathbf{a}_t, \mathbf{b}_t) = (\mathbf{a}, \mathbf{b})\} \right] \text{ for every } (\mathbf{a}, \mathbf{b}) \in \mathbf{A} \times \mathbf{B}.$$

Type C_j 's discounted average payoff by playing σ_{C_j} in the repeated game *equals* her expected stage-game payoff under $\gamma^j \in \Delta(\mathbf{A} \times \mathbf{B})$.

Objective function in the maximisation problem:

$$\sum_{(\mathbf{a}, \mathbf{b}) \in \mathbf{A} \times \mathbf{B}} \gamma^j(\mathbf{a}, \mathbf{b}) u_s(C_j, \mathbf{a}, \mathbf{b}).$$

Relate Maximisation Problem to the Repeated Game

For any strategy profile in the repeated game $((\sigma_{C_j})_{j=1}^m, \sigma_r)$, let

$$\gamma^j(\mathbf{a}, \mathbf{b}) \equiv \mathbb{E}^{(\sigma_{C_j}, \sigma_r)} \left[\sum_{t=0}^{\infty} (1 - \delta) \delta^t \mathbf{1}\{(\mathbf{a}_t, \mathbf{b}_t) = (\mathbf{a}, \mathbf{b})\} \right] \text{ for every } (\mathbf{a}, \mathbf{b}) \in \mathbf{A} \times \mathbf{B}.$$

Type C_1 's discounted average payoff by playing σ_{C_j} in the repeated game *equals* her expected stage-game payoff under $\gamma^j \in \Delta(\mathbf{A} \times \mathbf{B})$.

First constraint in the maximisation problem:

$$\sum_{(\mathbf{a}, \mathbf{b}) \in \mathbf{A} \times \mathbf{B}} \gamma^j(\mathbf{a}, \mathbf{b}) u_s(C_1, \mathbf{a}, \mathbf{b}) \leq p.$$

This constraint is necessary *in equilibrium* if:

- type C_1 's equilibrium payoff $\leq p$. Proposition 2

Relate Maximisation Problem to the Repeated Game

For any strategy profile in the repeated game $((\sigma_{C_j})_{j=1}^m, \sigma_r)$, let

$$\gamma^j(\mathbf{a}, \mathbf{b}) \equiv \mathbb{E}^{(\sigma_{C_j}, \sigma_r)} \left[\sum_{t=0}^{\infty} (1 - \delta) \delta^t \mathbf{1}\{(\mathbf{a}_t, \mathbf{b}_t) = (\mathbf{a}, \mathbf{b})\} \right] \text{ for every } (\mathbf{a}, \mathbf{b}) \in \mathbf{A} \times \mathbf{B}.$$

Second constraint in the optimisation problem:

- for every $\mathbf{b} \in \mathbf{B}$ that the marginal distribution of γ on \mathbf{B} attaches positive prob to,

$$\mathbf{b} \in \arg \max_{\mathbf{b}' \in \mathbf{B}} u_r(\gamma^j(\cdot | \mathbf{b}), \mathbf{b}').$$

This follows from the learning argument in Gossner (2011):

- Since public signals statistically identify sender's stage-game action.
- Receiver's prediction about sender's action is correct in all except for a bounded number of periods.

Distinction between Rational Type and Commitment Type

How will patient sender behave in equilibria in which his payoff $\approx v^*$?

- Will he use his optimal disclosure policy in every period?

Sender's stage-game pure action:

- $\mathbf{a} : \Omega \rightarrow M$, with $\mathbf{a} \in \mathbf{A}$.

Let \mathbf{a}^H be the *honest strategy* and \mathbf{a}^L be the *lying strategy*:

$$\mathbf{a}^H(\omega) = \begin{cases} m^h & \text{if } \omega = \omega^h \\ m^l & \text{if } \omega = \omega^l, \end{cases} \quad \mathbf{a}^L(\omega) = \begin{cases} m^h & \text{if } \omega = \omega^h \\ m^h & \text{if } \omega = \omega^l, \end{cases}$$

Sender's optimal disclosure policy: Mixing between \mathbf{a}^H and \mathbf{a}^L .

Patient Sender's Behaviour in High-Payoff Equilibria

Theorem 2

Suppose $m \geq 2$. For every small $\varepsilon > 0$, there exists $\underline{\delta} \in (0, 1)$ such that when $\delta > \underline{\delta}$, for every BNE that attains payoff within ε of (v_1^, \dots, v_m^*) , no type plays \mathbf{a}^H and \mathbf{a}^L both with positive prob at every on-path history.*

Remarks:

1. Also applies to a type whose cost of lying is 1, i.e., indifferent between lying and not when the state is ω^l .
2. **Contrasts to the commitment type in Mathevet et al. (2019).**

Proof of Theorem 2

Suppose type C_j plays \mathbf{a}^H and \mathbf{a}^L with positive prob at every on-path history.

- Playing \mathbf{a}^H in every period is her best reply against σ_r .
- Playing \mathbf{a}^L in every period is her best reply against σ_r .

For every type C_i that has higher lying cost (i.e., lower index)

- Type C_i plays \mathbf{a}^H with prob 1 at every on-path history.

Consider two cases:

Case 1: This type C_j is not type C_1 .

- Type C_1 plays \mathbf{a}^H with prob 1 at every on-path history.
- Type C_2 separates from type C_1
after the first time she sends m^h when the state is ω^l .
- Type C_2 's payoff is at most $(1 - \delta) + \delta p$,
which is strictly smaller than v_2^* . This leads to a contradiction.

Proof of Theorem 2

Case 2: This type C_j is type C_1 .

- \mathbf{a}^L in every period is type C_1 's best reply against σ_r , from which she attains payoff at least $p - \varepsilon$.
- Type C_1 's stage-game payoff from playing \mathbf{a}^L :
 - * $1 - (1-p)C_1$ if receiver plays a^h following m^h .
 - * $-(1-p)C_1$ otherwise.

This implies a lower bound on the prob that sender gets $1 - (1-p)C_1$:

$$Q\left(1 - (1-p)C_1\right) - (1-Q)(1-p)C_1 \geq p - \varepsilon \Leftrightarrow Q \geq p + (1-p)C_1 - \varepsilon.$$

- Type C_2 's payoff by playing \mathbf{a}^L in every period is at least:

$$Q\left(1 - (1-p)C_2\right) - (1-Q)(1-p)C_2,$$

which is **strictly greater than $v_2^* + \varepsilon$ when ε is small.**

Incorporate Ethical Type Senders

Relax the assumption that $C_1 < 1$. Instead, assume that:

- $C_1 \geq 1$, i.e., exists ethical type
- $C_m < 1$, i.e., exists non-ethical type

Let C^* be the **lowest lying cost among the ethical types**.

Theorem 3

There exist equilibria s.t. all non-ethical types attain their optimal commitment payoffs iff

$$C_1(C^* - 1) < 2.$$

Full Statement

Non-ethical sender can attain commitment payoff:

- *iff* C_1 and C^* are low enough.
i.e., increasing C_1 or C^* hurts non-ethical type sender,
& introducing additional types can hurt non-ethical types.

More Ethical Types Hurt: Outside Option Effect

Illustrative example:

- Two types: C_2, C_3 , with $C_3 \in (0, 1)$ and $C_2 = 1 + \varepsilon$.
Type C_3 can attain his Bayesian persuasion payoff.
- Introduce another type C_1 that is large enough such that:

$$C_1(C_2 - 1) > 2.$$

Type C_3 *cannot* attain his Bayesian persuasion payoff.

Outside Option Effect

If type C_3 attains commitment payoff in equilibrium, then

- Type C_3 lies while does not separate from type C_2 .

Type C_2 needs to lie with some frequency despite his lying cost > 1 :

- **An upper bound on type C_2 's payoff.**

Question: How much is type C_2 willing to lie?

Presence of type C_1 *decreases* type C_2 's willingness to lie.

- Type C_1 can secure payoff 0 by always telling the truth.
- Type C_2 's minimal payoff by imitating type C_1 's equilibrium strategy.

This yields a lower bound on type C_2 's equilibrium payoff.

Concluding Remarks

Repeated communication game *without* any commitment.

- Sender has persistent private info about her psychological cost of lying.

Different from commitment-type approach:

- Strategic-type sender's incentive constraints lead to novel predictions on equilibrium payoffs and behaviours.

Takeaways:

1. Persistent private info (partially) restores sender's commitment.
 - Microfoundation for sender's commitment to disclosure policies.
2. High-cost type's behaviour \neq Stationary commitment behaviour.
3. Possibility of being ethical may hurt non-ethical sender.

Alternative Benchmarks

Sender commits to disclosure policy *after* observing C before observing ω .

- Every type's equilibrium payoff is $p + (1 - p)C$.

One-shot communication w/o commitment:

- Let k be the smallest integer s.t. $\pi(C_1) + \dots + \pi(C_k) \geq \frac{1-2p}{1-p}$.
- For every $i \leq k$, type C_i 's payoff is pC_k .
- For every $j > k$, type C_j 's payoff is $C_k - (1 - p)C_j$.

Not a good benchmark since:

- Short-term info rent does not survive in the long-run.
- Some types of sender receive strictly higher payoff compared to optimal commitment payoff

⇒ Not a satisfactory microfoundation for sender-commitment. [Back](#)

Comment: Two Views on Lying Costs

Sender's stage-game payoff: $\mathbf{1}\{a_t = a^h\} - C\mathbf{1}\{m_t \neq \omega_t\}$

- Non-consequentialism view (Immanuel Kant 1797).
- Sender incurs a psychological cost whenever the face value of her message \neq her knowledge about state.
- Used in economic models of Kartik et al. (07), Kartik (09).

Alternatively, consequentialism view of Martin Luther.

- Sender incurs lying cost *only when* her lie has caused damages.
- Main takeaway messages apply under both views.

1st Constraint: Type C_1 's Highest Equilibrium Payoff

Proposition 2

*For every BNE and for every on-path public history h^t ,
if C_i is the highest-cost type in the support of receiver's belief at h^t ,
then type C_i 's continuation payoff at h^t is at most p .*

Proof is done by **induction on number of types in the support**.

1. Only one type in the support, similar to the argument in FKM.

Caution: Not implied by FKM given the solution concept is BNE.

2. If conclusion applies when there are $\leq n$ types in the support,
then it applies when there are $n + 1$ types in the support.

For simplicity, focus on equilibria s.t. at every on-path history, m^h induces a^h with weakly higher probability than message m^l .

Proof of Proposition 2: Induction Step

Given equilibrium $((\sigma_{C_i})_{i=1}^m, \sigma_r)$. Let C_j be the highest-cost type at h^t .

- I construct a new strategy $\tilde{\sigma}_{C_j}$ based on σ_{C_j} .
- For every $h^s \succeq h^t$,
if σ_{C_j} sends message m^l in state ω^l at h^s with positive prob,
then $\tilde{\sigma}_{C_j}$ sends message m^l with prob 1 when the state is ω^l .
otherwise, $\tilde{\sigma}_{C_j}$ sends the same message as σ_{C_j} .
- By construction, $\tilde{\sigma}_{C_j}$ is type C_j 's best reply against σ_r .

Consider all histories that:

- succeed h^t , and occur with positive prob under $(\tilde{\sigma}_{C_j}, \sigma_r)$.

Partition all those histories into two subsets:

Subset 1: h^s is such that from h^t to h^s , \nexists outcome (ω^l, a^h) .

Subset 2: h^s is such that from h^t to h^s , \exists outcome (ω^l, a^h) .

Proof of Proposition 2: Induction Step

Consider type C_j 's continuation payoff at h^t by adopting $\tilde{\sigma}_{C_j}$:

- In subset 1, type C_j has never received positive payoff in state ω^l .

If h^s belongs to subset 2 *but* h^{s-1} belongs to subset 1, then:

- σ_{C_j} sends message m^l at h^{s-1} with prob 0,
- ⇒ some type C_k sends m^l at h^{s-1} with positive prob.
- After observing m^l at h^{s-1} , **number of types in the support of receiver's posterior belief is reduced by at least one.**
 - C_k : highest-cost type in the support of receiver's belief at (h^{s-1}, m^l) .

By definition, $C_j > C_k$.

Induction hypothesis: C_k 's continuation payoff in period t is $\leq p$.

- **After observing ω^l at h^{s-1} ,**
type C_k 's continuation payoff $\leq \delta p$ if she sends message m^h .
- Type C_j 's continuation payoff is also less than δp .

2nd Constraint is Approximately Satisfied

Proposition 3

For every $\varepsilon > 0$ and j , there exists $\underline{\delta} \in (0, 1)$ such that, for every BNE when $\delta > \underline{\delta}$, if γ^j attaches probability more than ε to \mathbf{b} , then \mathbf{b} is an ε -best reply against $\gamma^j(\cdot|\mathbf{b}) \in \Delta(\mathbf{A})$.

Intuition: Learning argument in Gossner (11).

- Receiver's prediction about sender's action is ε -close to sender's action (based on her true type) in *all except for a bounded number of periods*.
- If δ is close to 1, then this bounded number of periods have negligible impact on $\gamma^j(\cdot|\mathbf{b})$ if the prob of \mathbf{b} is not too small.
- $\gamma^j(\cdot|\mathbf{b})$ is *close* to the subset of $\Delta(\mathbf{A})$ that \mathbf{b} best replies against.

Convergence of ε -relaxed problem

Let v_j^ε be the value of the optimisation problem that satisfies:

- the first constraint, i.e., type C_1 's payoff $\leq p$.
- the ε -relaxed version of second constraint.

Proposition 4

$\lim_{\varepsilon \downarrow 0} v_j^\varepsilon = v_j^*$ for every $j \in \{1, 2, \dots, m\}$.

[Back](#)

Sufficiency Part of Theorem 1

Let

$$v^N \equiv \left(-C_1(1-p), -C_2(1-p), \dots, -C_m(1-p) \right),$$

$$v^L \equiv \left(p + (1-C_1)(1-p), p + (1-C_2)(1-p), \dots, p + (1-C_m)(1-p) \right),$$

and $v^H \equiv (p, p, \dots, p)$.

For every $\rho \in [0, \frac{p}{1-p}]$, let

$$v(\rho) \equiv \frac{(1-\rho)C_1}{\rho(1-C_1)+C_1} v^H + \frac{\rho C_1}{\rho(1-C_1)+C_1} v^L + \frac{\rho(1-C_1)}{\rho(1-C_1)+C_1} v^N.$$

Proposition 5

For every $\underline{\pi} > 0$ and $\rho \in [0, \rho^)$, there exists $\underline{\delta} \in (0, 1)$ such that for every π with $\pi(C_1) \geq \underline{\pi}$ and $\delta > \underline{\delta}$, there exists an equilibrium that attains $v(\rho)$.*

Characterization Result

Theorem 3

1. If

$$C_1(C^* - 1) - 2 < 0,$$

then for every $\varepsilon > 0$, there exists $\underline{\delta} \in (0, 1)$, such that when $\delta > \underline{\delta}$, there exists a sequential equilibrium in which every non-ethical type C_j obtains payoff at least $v_j^{**} - \varepsilon$.

2. If

$$C_1(C^* - 1) - 2 > 0,$$

then there exist $\eta > 0$ and $\underline{\delta} \in (0, 1)$, such that in every Bayes Nash equilibrium when $\delta > \underline{\delta}$, every non-ethical type C_j obtains payoff no more than $v_j^{**} - \eta$.