Role Uncertainty and Punishment Severity in Repeated Cheap Talk: Theory and Experiment

Jinyuqi Huang Wooyoung Lim

Department of Economics, HKUST

July 2026

Motivation

- Talk is cheap: Costless, non-binding, unverifiable.
- Sender talks, Receiver decides: To lie/trust or not depends on whether their incentives are aligned.
- When interests may conflict
  - One-shot: Uninformative communication (Crawford & Sobel, 1982).
  - Repeated: History-based strategies make truth-telling possible.

- We study two realistic features that shape repeated communication:
  1. **Role Uncertainty**: Sender/Receiver roles may vary.
  2. **Punishment Severity**: Payoff distribution in Grim via Nash-reversion.
- **Research question**: Individual effect of each feature on long-run communication, and their interaction effect.

Examples

Role uncertainty arise naturally in professional networks where Sender has **private information advantage** and Receiver has **decision power**:

- **Academic Referral System:** Professors exchange evaluations of students. A letter writer (Sender) today reviews others' letters on a committee (Receiver) tomorrow.
- **Investment Syndication:** Firms share deal flows repeatedly. Today's initiator (Sender) may be tomorrow's participant (Receiver).
- **The "Revolving Door":** Regulatory agencies and the private sector. Industry professionals (Senders) communicate compliance to regulators (Receivers), but frequently switch roles over their careers.

Punishment severity (how bad communication failure is) may depend on exogenous bargaining power etc.

The Model

Infinitely repeated cheap talk (continuation prob. δ) between two players A and B. **Each round:**

1. **Roles assigned**: A is Sender, B is Receiver w.p. γ; opposite w.p. $(1-\gamma)$.
2. **State drawn**: $\theta \in \{L, R\}$, 50–50, only Sender observes.
3. Sender sends message $m \in \{\text{“}L\text{”}, \text{“}R\text{”}\}$; Receiver picks action $a \in \{L, M, R\}$.
4. History is available: Past roles, states, messages, actions.

Payoffs (Sender, Receiver): $b \in [0.5,\ 1)$

	L	M	R
$\theta = L$	$(1,1)$	$(1-b,\ b)$	$(0,0)$
$\theta = R$	$(1,0)$	$(1-b,\ b)$	$(0,1)$

Misaligned incentives:

- Sender always prefers action L; Receiver wants to match the state.

One-shot stage-Nash: Receiver ignores message, chooses M, payoff $(1-b, b)$.
Babbling equilibrium (BE) is inefficient, cannot achieve payoff $(1,1)$ in state L.

Equilibrium in Repeated Cheap Talk

Simple Truth-Telling Equilibrium (STE):

- Always tell the truth as Sender. Always follow message as Receiver.
- Deviation $\rightarrow$ permanent reversion to BE $\rightarrow (1-b,\, b)$, **larger** b **means harsher punishment on Sender.**
- Simple for Receiver; gets full surplus; no incentives to deviate.
- Only Sender is tempted to deviate (lie in state R to get action L).
- STE is efficient, can achieve payoff $(1, 1)$ in state L.
- We use STE as theoretical benchmark due to its **strategic simplicity**.
- Other equilibria (e.g., Wilson & Vespa 2020) discussed in the paper.

Comparative Statics: STE

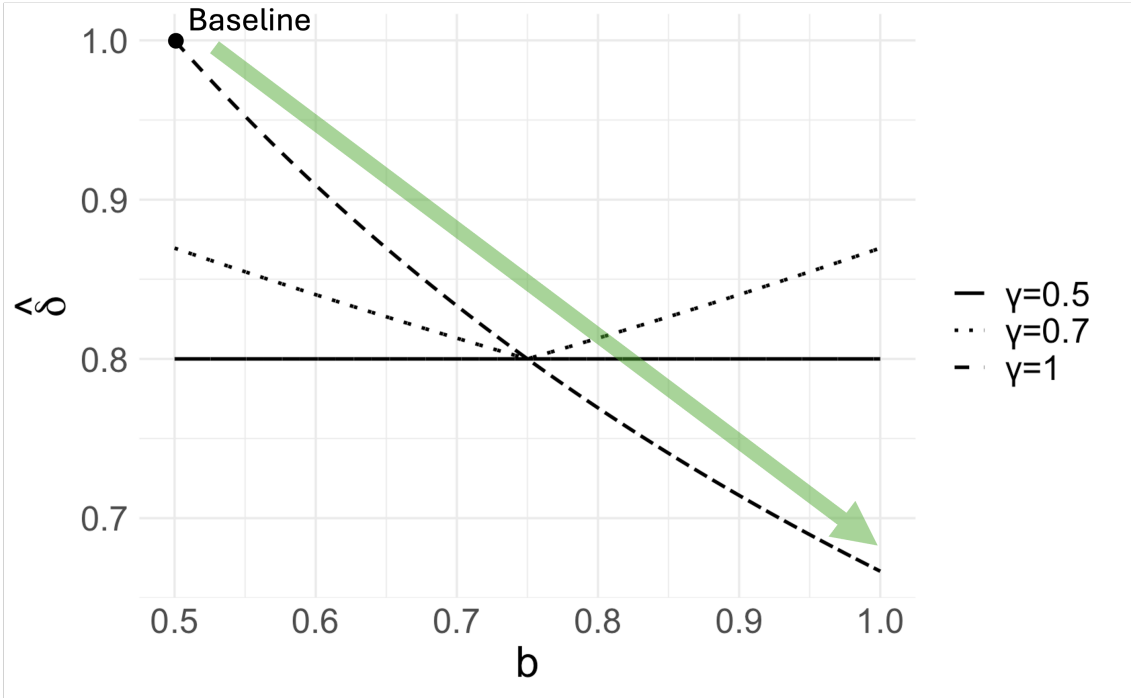
STE is sustainable when future cost $\geq$ current gain from lying.

- Current gain by lying is always 1.
- Future cost is $\frac{\delta}{1-\delta} C(\gamma, b)$.
- $\frac{\delta}{1-\delta} C \geq 1 \Rightarrow \delta \geq \hat{\delta}(\gamma, b)$. The smaller $\hat{\delta}$ is, the easier to support STE.

Comparative Statics w.r.t (γ, b):
Baseline: Fixed roles $(\gamma = 1)$ with mild punishment $(b = 0.5)$.

Comparative Statics: STE

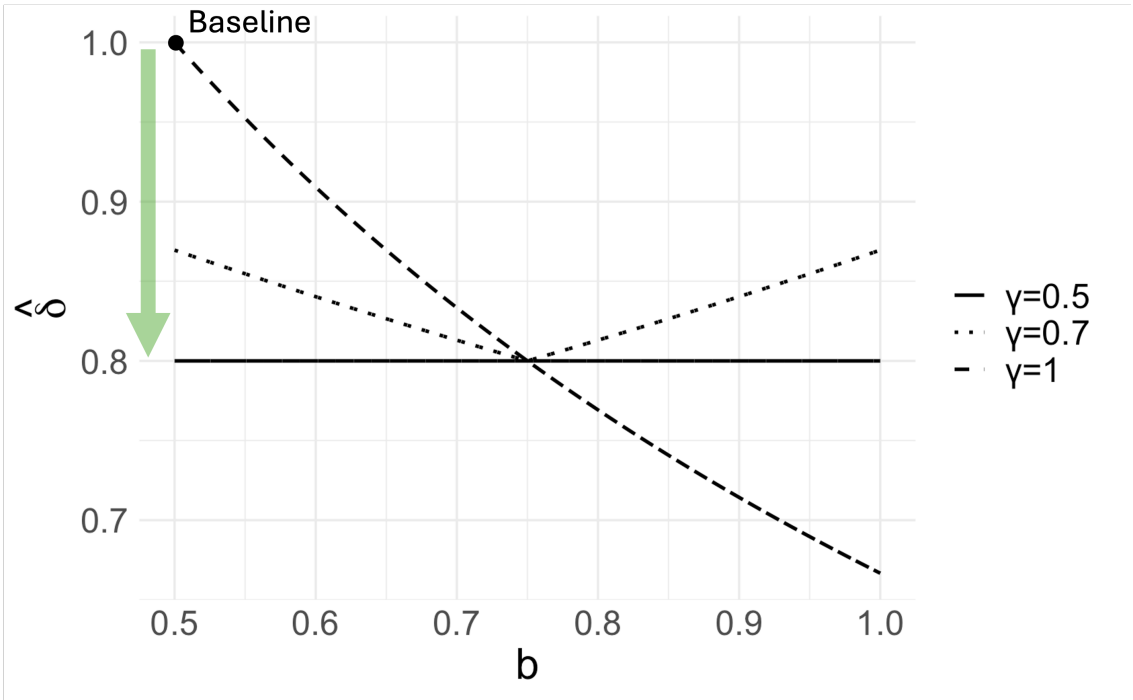
1. **Harsher punishment ($\uparrow b$) helps.**



Intuition: Sender's BE payoff is $1-b$.

Comparative Statics: STE

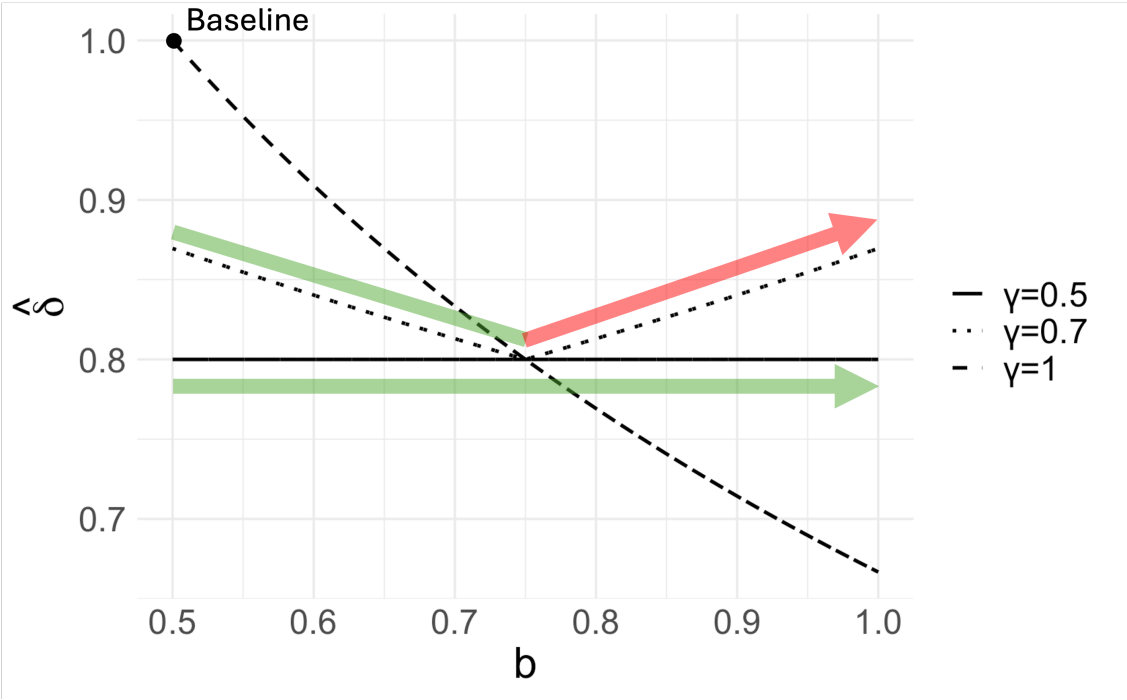
2. **More role uncertainty ($\downarrow \gamma$) helps.**



Intuition: Both players get to enjoy Receiver's information surplus, balancing incentives to cooperate.

Comparative Statics: STE

3. **Interaction: with $\gamma = 0.5$, no effect; with $\gamma > 0.5$, can backfire.**



Intuition: $\uparrow b$ creates two opposing forces.

- It raises Sender-self's cost but reduces Receiver-self's (BE payoff $b \uparrow$).
- Equally likely ($\gamma = 0.5$), effects on two selves cancel out.
- Unequally likely ($\gamma > 0.5$), Receiver-self dominates for player B who becomes harder to regulate.

Experimental Design & Predictions

Six Treatments: Role Uncertainty (γ) $\times$ Punishment (b)

	Fixed ($\gamma=1$)	Unequal ($\gamma=0.7$)	Equal ($\gamma=0.5$)
Mild ($b=0.5$)	**Baseline**	UR	ER
Harsh ($b=0.9$)	H	HUR	HER

- Implemented $\delta = 0.8$.
- Information transmission **efficiency** $\mathbb{P}(a = L|\theta = L)$.
- Truth-telling eqm. $\Rightarrow$ states revealed and followed, efficiency $= 1$.

Predicted efficiency relative to Baseline:

	H	UR	ER	HUR	HER
Prediction	$\uparrow$	$\approx$	$\uparrow$	$\downarrow$	$\uparrow$

Remark: Efficiency is *not* an identifying statistic for any TE. High efficiency can arise despite low truth if Receiver unconditionally trusts, and high honesty needs not translate into high efficiency if Receiver does not follow message "L". We complement efficiency with Sender and Receiver behaviors.

Implementation

- Online English experiment via oTree at HKUST.
- 658 undergraduates, 7 sessions/treatment, 7 supergames/session.
- Average payment $\approx$ US \$20.
- Caveat: We identified a procedural issue in the data collection. We plan to re-run sessions under the correct protocol to replace the data, but believe the qualitative patterns are robust.

Results

	Mild ($b=0.5$)			**Harsh** ($b=0.9$)		
	Baseline	UR	ER	H	HUR	HER
Efficiency $\mathbb{P}(a = L \mid \theta = L)$	0.48	**0.58****	**0.55**	0.47	**0.31*****	**0.38***
Truth $\mathbb{P}(m = \text{“}R\text{”} \mid \theta = R)$	0.34	0.36	0.34	**0.55*****	0.41	0.42
Trust $\mathbb{P}(a = L \mid m = \text{“}L\text{”})$	0.45	**0.56****	**0.52**	0.44	**0.30*****	**0.38***

Significance vs. Baseline using two-way clustered std. err. Improvement. Decline.

Three striking findings:

1. Harsh punishment disciplines Sender but fails to elicit Receiver trust.
2. Role uncertainty triggers unconditional trust despite unchanged truth.
3. Their interaction leads to a collapse in trust despite higher honesty.

They challenge two implicit assumptions in STE-based predictions.

The Safety Trap: Harsh Punishment Treatments

Expected: ↑ punishment → ↑ truth → ↑ trust → ↑ efficiency.
Actual: ↑ punishment → ↑ truth → ↓ trust → ↓ efficiency.

Why? In H, HUR and HER, "safe" action M guarantees Receiver **0.9** regardless of Sender's honesty.

- Receivers default to M instead of trying to trust.
- Role uncertainty makes it worse: Selfish choice feel less unfair. Future consequence of distrust is ignored.

Evidence: $\mathbb{P}(a = M \mid m = \text{“}L\text{”})$ is much higher in harsh treatments:

	Baseline	UR	ER	H	HUR	HER
$\mathbb{P}(a = M \mid m = \text{“}L\text{”})$	0.36	0.32	0.38	**0.54*****	**0.69*****	**0.60*****

In HUR and HER, first M-to-"L" actions occur in round 1 by 10% more than other treatments. Trapped by safety-seeking in the beginning, not punitive.

Blind Trust: Random Role Treatments

Expected: Role uncertainty $\rightarrow \uparrow$ truth $\rightarrow \uparrow$ trust $\rightarrow \uparrow$ efficiency.
Actual: In UR and ER, truth is unchanged but $\uparrow$ trust $\rightarrow \uparrow$ efficiency.

Mechanism — Blind Trust Equilibrium (BTE):

- Sender *always* sends "*L*"; Receiver *always* plays *L*.
- Efficiency $\mathbb{P}(a = L \mid \theta = L) = 1$, but **no info. is actually transmitted**.
- Sender captures all surplus; roles alternate $\Rightarrow$ both benefit over time.

Why does BTE emerge in UR/ER but not elsewhere?

- **Trivial** $\hat{\delta} = 0$ for UR and ER.
  - Harsh + role uncertainty: $\hat{\delta} = 0.62$ in HER; impossible in HUR.
  - Fixed roles: BTE is undefined (Baseline, H).
- **Strategically simple**: Sender needn't resist lying; Receiver needn't assess credibility.

Supergame-Level Evidence: Safety Trap in HUR/HER

	Truth ≤ 0.5		Truth > 0.5	
	Trust $\leq .5$	Trust $> .5$	Trust $\leq .5$	Trust $> .5$
UR	57%	32%	2%	9%
ER	53%	29%	5%	12%
HUR	**72%**	12%	**9%**	8%
HER	**64%**	16%	**11%**	10%

Truth: $\mathbb{P}(m = \text{“}R\text{”} | \theta = R)$; Trust: $\mathbb{P}(a = L | m = \text{“}L\text{”})$.
Percentages are based on retained supergames that include valid obs. $\theta = R$ and $m = \text{“}L\text{”}$..

- **Safety trap**: HUR/HER cluster in low-trust regardless of truth.

Supergame-Level Evidence: Blind Trust in UR/ER

	Truth ≤ 0.5		Truth > 0.5	
	Trust $\leq .5$	Trust $> .5$	Trust $\leq .5$	Trust $> .5$
UR	57%	**32%**	2%	9%
ER	53%	**29%**	5%	12%
HUR	72%	12%	9%	8%
HER	64%	16%	11%	10%

Truth: $\mathbb{P}(m = \text{“}R\text{”} | \theta = R)$; Trust: $\mathbb{P}(a = L | m = \text{“}L\text{”})$.
Percentages are based on retained supergames that include valid obs. $\theta = R$ and $m = \text{“}L\text{”}$..

- **Blind trust**: UR/ER cluster in high-trust even when truth is low.

Takeaways & Next Steps

We study how role uncertainty and punishment severity jointly shape communication in repeated cheap talk.

- Theory (STE): Each feature alone helps, while interaction may hurt.
- Experiment data reveal different patterns:
  - **Safety trap:** Harsh punishment improves truth but fails to elicit trust.
  - **Blind trust:** Role uncertainty leads to efficient but untruthful outcome.

Next steps:

1. **Break the Safety Trap**: Lower the Sender's stage-Nash payoff while holding the Receiver's stage-Nash payoff fixed.
2. **Penalize uninformed decisions**: BTE may be less favorable as STE if information has intrinsic value. We can reduce Sender's payoff when Receiver acts L in state R, capturing the welfare loss of misallocating the resource.

Literature

- **Facilitating communication in cheap talk:** Crawford and Sobel (1982); Blume, Lai and Lim (2020); Golosov et al. (2014); **Wilson and Vespa (2020)**; Kolotilin and Li (2021); Kuvalekar, Lipnowski and Ramos (2022); Best and Quigley (2024).
- **Role switching in games:** Classen (2005); Iriberri and Rey-Biel (2011); Ben-Ner et al. (2004); Herne, Lappalainen and Kestila-Kekkonen (2013); Burks, Carpenter and Verhoogen (2003); Johnson and Mislin (2011).

Thank You!

Questions or comments?

jhuangde@connect.ust.hk
wooyoung@ust.hk