Commitment and Communication in Bayesian Persuasion: Theory and Experiment

Yan Xu
University of Vienna
yan.xu@univie.ac.at

(with) Yun Wang
Renmin University of China

November, 2022

Motivation

- A principal attempts to influence a DM's action due to conflict interests:
  — A pharm. company wants the FDA to approve a drug, effective or not
  — The FDA only wants to approve effective drugs

- **information design**: communication **or** Bayesian persuasion
  - communication games: private info. + send messages
    — The pharm. company has news about the drug and talks to the FDA.
  - Bayesian persuasion: no private info. + commit to an info. environment
    — The pharm. company pre-registers trials to test the drug. The FDA learns the trial result and decides to approve or reject the drug.

This paper: a principal uses both communication **and** Bayesian persuasion?

Scenario: the company still commits to a trial, but learns trial results (private info.) and send messages. The FDA acts upon messages (not signal).

- How will the principal design info structure and communication strategy?
- For the principal: effectiveness of persuasion?
- For a policy maker: information transmission?

An example of Bayesian Persuasion between a pharmaceutical company and FDA

- Common prior of the drug being effective $P(\omega = L) = 0.3$

- The company picks and commits to an info. structure $\pi \equiv (p, q)$
  — $p \equiv \mathbb{P}(s = l \mid \omega = L)$ and $q \equiv \mathbb{P}(s = l \mid \omega = R)$
  — e.g., $(p, q) = (1, 0)$; $(1, 1/2)$; $(1, 3/7)$; $(1/2, 1/2)$

1. State ω realizes, unknown to the receiver
   — BP benchmark: sender knows ω or not is irrelevant due to commitment

2. Signal s realizes (mechanically) from the chosen $\pi(\cdot \mid \omega)$
   — BP benchmark: sender learns signal and truthfully reports to receiver
   — **BP with communication: sender learns signal and sends message**

3. The FDA observes π and takes action $a \in \{Approve, Reject\}$ upon signal s or message $m(s)$. Payoffs are materialized.

   — company gets a prize if $a = Approve$, 0 otherwise
   — FDA gets a prize if action matches the state, 0 otherwise

- Measure persuasion effectiveness: persuasion rate $\phi(a_L)$
- Measure info transmission: State-Action match rate $\phi(a_\omega)$

An example of Bayesian Persuasion between a pharmaceutical company and FDA

The company commits to the optimal information structure (π^*) to maximize the FDA's chance to choose the action $a = Approve$.

- Set-up: binary states $\{L, R\}$, signals $\{l, r\}$, and actions $\{Approve, Reject\}$
- With prior $P(\omega = L) = 0.3$, the FDA always Reject under no persuasion.
  — persuasion rate: the probability of choosing $a = Approval$ is 0
  — match rate: the probability of the action matches the state $=$ $0.3 * 0 + 0.7 * 1 = 0.7$
- Upon observing the signals realized from the chosen π, the FDA updates his posterior belief via Bayes' Rule and takes action.

$$\mu(L \mid l) = \frac{0.3 \times p}{0.7 \times q + 0.3 \times p} > \frac{1}{2}, \Rightarrow \ a = Approve$$

$$\mu(L \mid r) = \frac{0.3 \times (1-p)}{0.7 \times (1-q) + 0.3 \times (1-p)} < \frac{1}{2}, \Rightarrow \ a = Reject$$

- Note that $\mu(L \mid l) > \mu(L \mid r)$ always hold iff $p > q$ and can never be both $>1/2$.
- **The optimal information structure is** $\pi^* = (1, 3/7)$.
  — The induced posteriors are $\mu(L \mid l) = 1/2$ and $\mu(L \mid r) = 0$.
  — The induced persuasion rate $= Prob(l) = 0.3 \times 1 + 0.7 \times 3/7 = 0.6$
- Whereas the induced State-Action match rate $= Prob(l)\mu(L \mid l) + Prob(r)\mu(R \mid r) = 0.6 \times 1/2 + 0.4 \times 1 = 0.7$

Research questions: Commitment and Communication in Bayesian Persuasion

How does sender's private information and communication affect persuasion and information transmission in the Bayesian persuasion game?

- Benchmark: Kamenica and Gentzkow (2011)'s Bayesian Persuasion
  - sender commits to the chosen information structure after state is realized.
  - **sender commits to truthful reporting the realized signals.**
- Does the type of private information matter?
  - the sender knows the state: Frechette, Lizzeri, and Perego (FLP, 2022).
  - **the sender knows the signal**: this paper
- Does the communication protocols matter?
  - lying: cheap talk message (Crawford and Sobel (1982))
  - concealment: partially verifiable messages (Dye (1985), Verrecchia (1983)) — requires an additional message "I don't know"

Empirical relevance: regulations, monitoring, and disclosure policies in persuasion activities

Behavioral relevance: educative nudging in strategic reasoning in Bayesian persuasion.

Experimental design: treatment

Table: Variants of Bayesian persuasion games with communication

Communication protocol	Treatments and subjects: three-message-framing	Treatments and subjects: two-message-framing
No message	**BP3** (64)	BP2 (32)
Cheap talk message	**CBP3** (68)	CBP2 (32)
Verifiable message	**VBP** (64)	—
Mix message	MBP3 (100)	MBP2 (52)

- Main treatments: BP3, CBP3, and VBP
- BP2 and CBP2 controls framing effect: does the strategically irrelevant message "I don't know" matter for BP and CBP?
- MBP2 and MBP3 are robustness treatments: truthful and cheap talk messages are equally likely to be implemented

Bayesian persuasion in the lab

Experimental task: the receiver bets the state to be L or R

- The sender gets 100 (tokens) if the receiver bets L; 20 otherwise.
- The receiver gets 100 (tokens) if betting correctly; 20 otherwise.

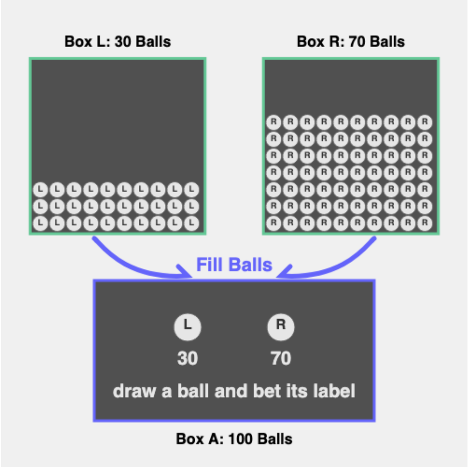
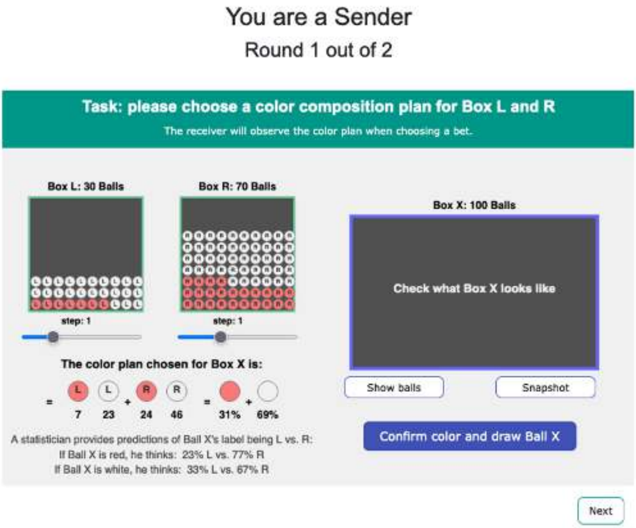


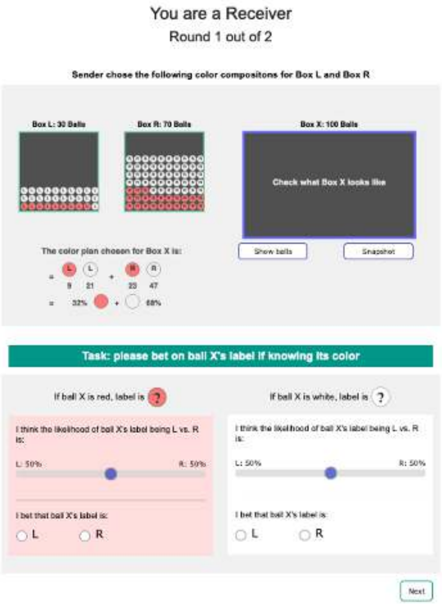
Figure: Draw a ball from Box A. The receiver bets its label to be either L or R. The bet choice determines payoffs for the pair.

Sender's choice: a color composition of two boxes = an information structure



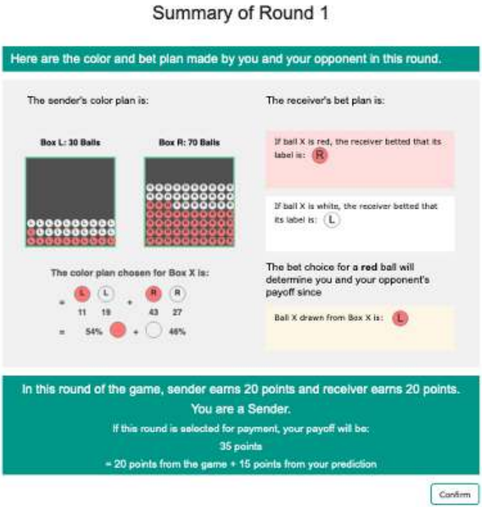
- State $\omega \in \{L, R\}$; Signal $s \in \{red, white\}$; Action $a \in \{bet\ L, bet\ R\}$
- Each color composition is an info. structure: $(p, q) = (7/30,\ 24/70)$

Receiver's strategy: a contingent plan for each signal realization

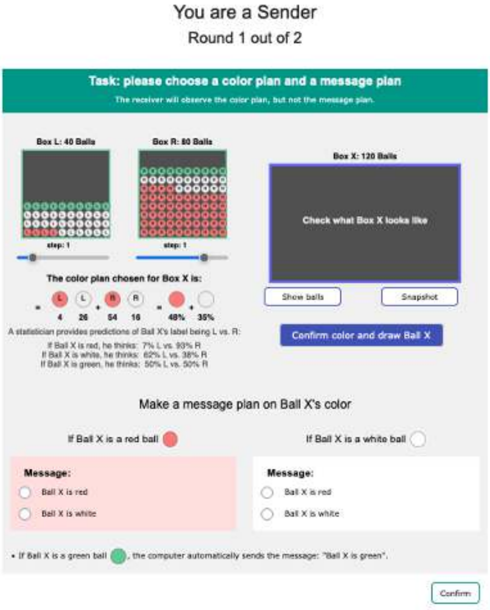


- Receiver's strategy profile: $(a_{red}, a_{white}) = (bet\ R, bet\ R)$

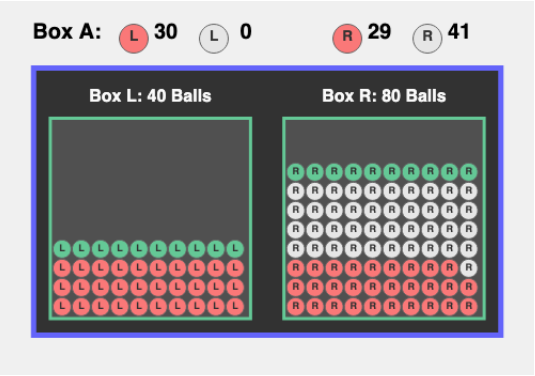
Feedback after each round



Interfaces for CBP and VBP treatment: sender



Re-framing with the message “I don’t know” for VBP: add an ex-post useless color



- BP2, CBP2 treatment: no green balls
- BP3, CBP3, VBP treatments: fixed number of *green* balls in Box L and R
- Choice of info structure π is the same for all treatments:
  — senders always choose the color composition of *red* and *white* balls

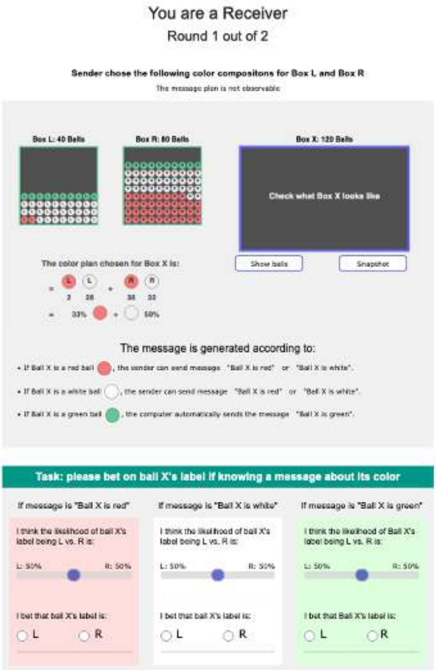
Re-framing with the message "I don't know" for VBP: add an ex-post useless color

- If a green ball is realized, computer automatically sends a message "green"
  $\Rightarrow$ Pure message strategy is $(m(red), m(white))$ for both CBP and VBP.

	Pure message strategies $(m(red), m(white))$			
Treatment	truth-telling	full lie	partial lie: s_{red}	partial lie: s_{white}
BP2 & BP3	(red, white)	—	—	—
CBP2 & CBP3	(red, white)	(white, red)	(white, white)	(red, red)
VBP	(red, white)	(green, green)	(green, white)	(red, green)

- In BP3 and CBP3, green ball and green message play no strategic role
  — BP3: two states, two signals, two actions, no message
  — CBP3: two states, two signals, two actions, two messages
- In VBP, sender can delegate the message task to the computer, who always reports the message "green".
  — VBP: two states, two signals, two actions, three messages
  $m_s \in \{red, white, green\}$

Interfaces for CBP3 and VBP treatment: receiver



Experimental procedures

- Auxiliary tasks (incentivized):
  - Receiver: predicts the sender's color plan and message plan
  - Sender: predicts the receiver's bet plan (for each signal or message)
- Experiment was run in the lab at the Renmin University of China during Dec 2021 – March 2022.
- Two games: $(30L, 70R)$ and $(40L, 60R)$, each has 10 rounds
- Matching: fixed role, random matching, summary after each round
- Payment: random pay for one round for each game
- Duration: 90 – 120 minutes

Theoretical analyses

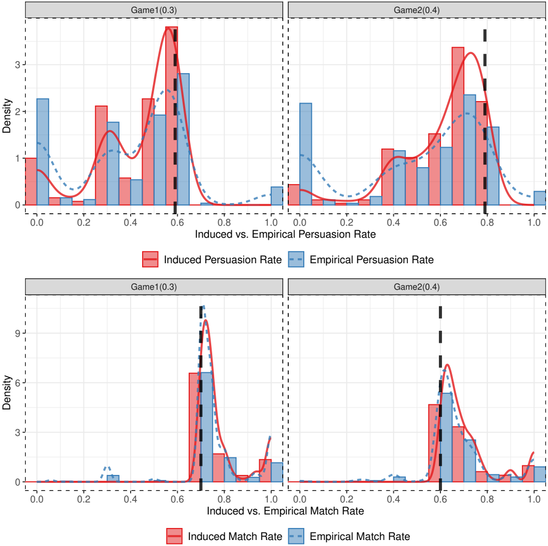
Strategies and equilibrium concept

- only consider the pure strategy Sender-preferred PBE $(\pi(p,q), \{m_s\}_{s\in S}; \{a_m\}_{m\in M})$
    - Information structure $\pi(\cdot)$: what information to generate
    - Message strategy m_s: what information to reveal
    - pure message strategy and pure bet strategy
    - belief systems: $\mu_\pi(\omega \mid s)$ and $\rho_{\pi,m_s}(\omega \mid m)$

- Equilibrium Predictions:
    - BP3: π_{BP}, $(a_{red}, a_{white}, a_{green}) = (L, R, L)$
    - CBP3: any π, babbling message $(m_{sred}, m_{swhite}) = (red, red)$, $(a_{mred}, a_{mwhite}, a_{mgreen}) = (R, R, L)$
    - VBP: π_{BP}, Partially unravelling messages $(m_{sred}, m_{swhite}) = (red, green)$, $(a_{mred}, a_{mwhite}, a_{mgreen}) = (L, R, R)$
    - MBP3: π_{BP}, babbling message $(m_{sred}, m_{swhite}) = (red, red)$,
    - Induced Persuasion rate: BP3 $(\simeq 2\mu_0) \simeq$ VBP $>$ CBP3 $(\simeq 0)$
    - State-Action Match rate: BP3 $(\simeq 1-\mu_0) \simeq$ VBP $\simeq$ CBP3

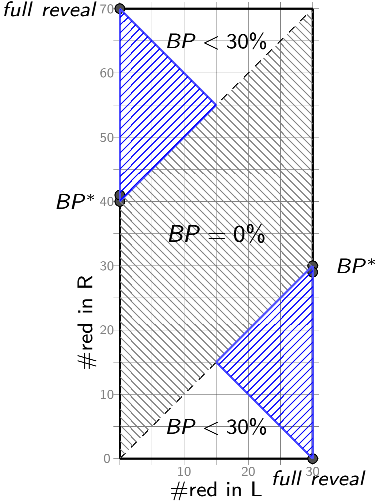
Experimental results

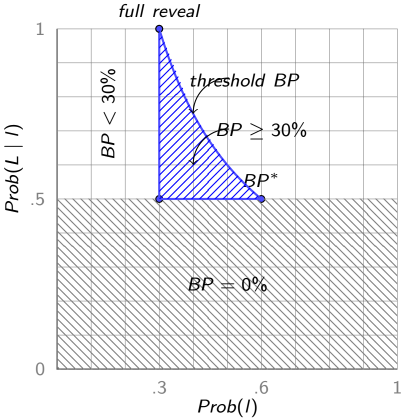
BP3 Benchmark: persuasion $\phi(a_L)$ and information transmission $\phi(a_\omega)$



- Induced vs. Theoretical: sender's deviations from the equilibrium $\pi_{BP}(\cdot)$
- Empirical vs. Induced: receiver's deviations from the optimal $(a_l^*(\pi), a_r^*(\pi))$

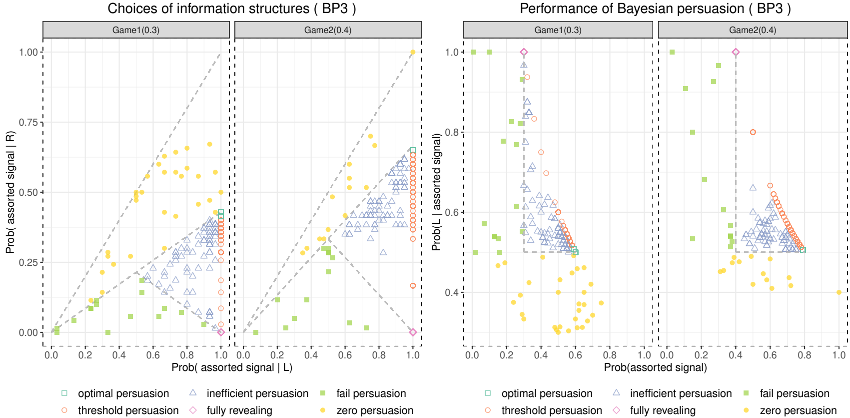
Analyzing choices of $\pi(\cdot)$: BP partitions and clusters





- optimal persuasion: *BP**
- threshold persuasion: hyperbola frontier
- inefficient persuasion: blue region
- fully revealing: most informative
- fail persuasion: $BP < \mu_0$
- zero persuasion

Can sender perform Bayesian persuade? BP3 Partitions and Clusters



	Successful persuasion $BP_{\geq \mu_0}$				Failed persuasion	
	optimal	**threshold**	inefficient	fully revealing	$BP_{<\mu_0}$	$BP_{=0\%}$
Game1	17%	33%	24%	12%	5%	9%
Game2	8%	52%	23%	9%	4%	4%

Will receivers be persuaded? Bet Strategies

Table: Average persuasion rates by clusters (Induced vs. Empirical)

	Successful persuasion $BP_{\geq \mu_0}$				Failed persuasion	
	optimal	threshold	inefficient	fully reveal	$BP_{<\mu_0}$	$BP_{=0\%}$
Game1: Induced	**.59**	.54	.44	.30	.18	.00
Game1: Empirical	**.43**	.49	.38	.38	.15	*.07*
Game2: Induced	.79	.70	.61	.40	.26	.00
Game2: Empirical	.56	.55	.49	.43	.26	.05

- Empirical persuasion rates are close to the induced optimal ones.
- Their differences are explained by either non-optimal bet a_l or a_r.

	Bet L after signal l		>	Bet L after signal r	
	Game 1	Game 2		Game 1	Game 2
BP3: Optimal	91%	96%		0%	0%
BP3: Predicted	91%	97%		2%	2%
BP3: Observed	77%	78%	>	5%	4%

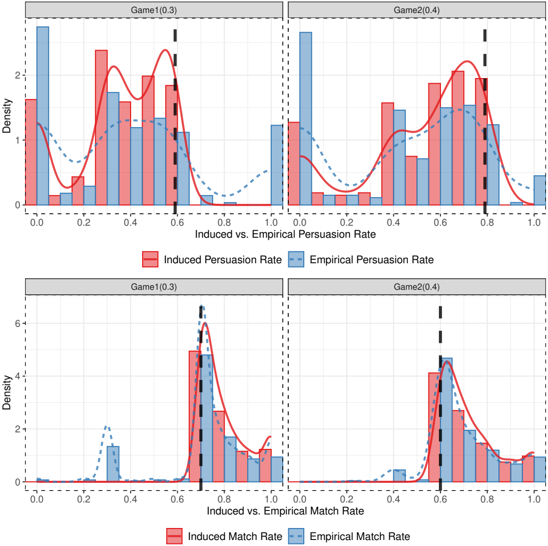
- 15% - 30% receivers cannot be persuaded due to resistance and suspicion.

Summary of BP Benchmark

- Senders choose information structures that induce (quite good) Bayesian persuade and also anticipate the effectiveness of BP.
- Most receivers act optimally with respect to sender's info structure. The non-optimal acts are mainly due to resistance and suspicion.
  - Do receivers understand the mechanism of BP?
    — 65% predict info structures in $BP^* + BP_{>\mu_0}$
  - Do receivers have correct posteriors?
    — (unincentivized) likelihood estimates deviate from the correct Bayesian ones, but they are correct in updating direction.
- Receivers' non-optimal acts are costly for persuasion rate, but not for match rate.
- Overall, BP2 $\approx$ BP3
  — No framing effects with strategically irrelevant green balls
  — Observed and predicted bets after green ball: 50% L and R
- No learning within 10 rounds and between Game 1 and Game 2

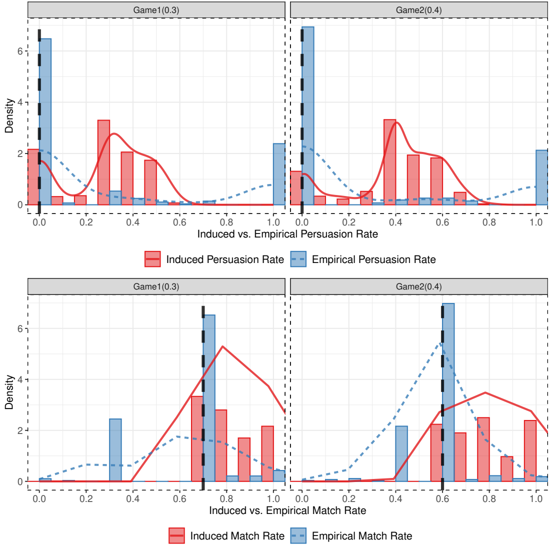
With private information and communication...

VBP: persuasion $\phi(a_L)$ and information transmission $\phi(a_\omega)$



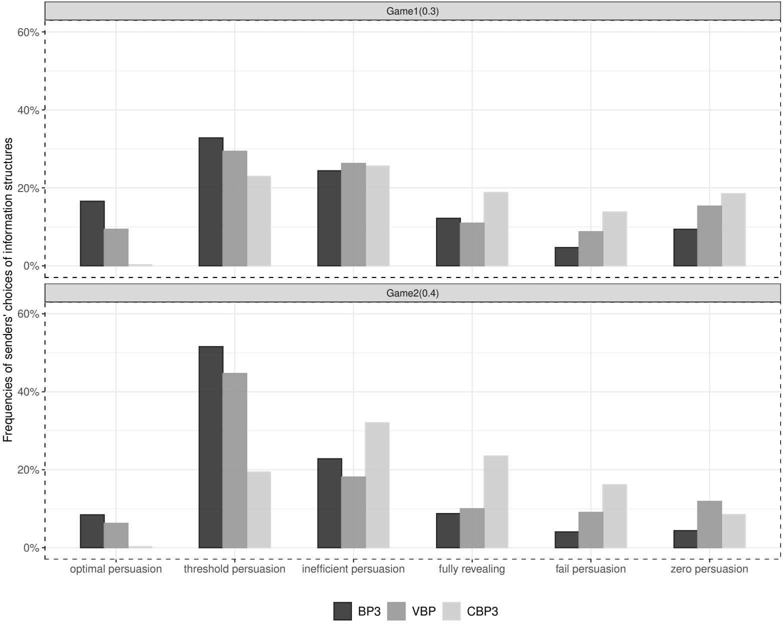
- Induced vs. Theoretical: sender's deviations from equilibrium $\pi(\cdot) + (m_l, m_r)$
- Empirical vs. Induced: receiver's dev. from optimal (a_l, a_r) + unknown (m_l, m_r)

CBP3: persuasion $\phi(a_L)$ and information transmission $\phi(a_\omega)$

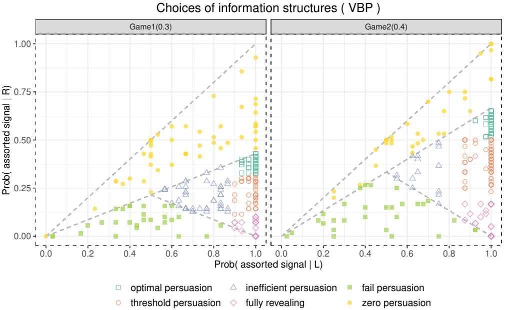


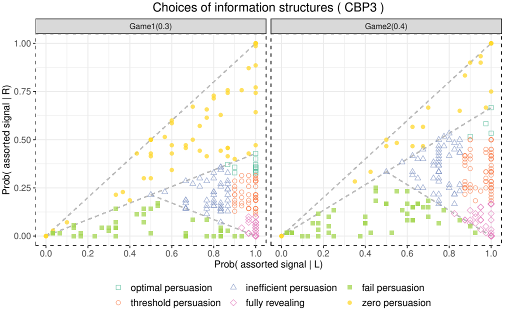
- Induced and Empirical persuasion rate: $BP3 > VBP >> CBP3$
- Induced match rate: $BP3 < VBP < CBP3$
- Empirical match rate: $BP3 > VBP > CBP3$

Senders' choice of $\pi(\cdot)$: exact partitions and clusters

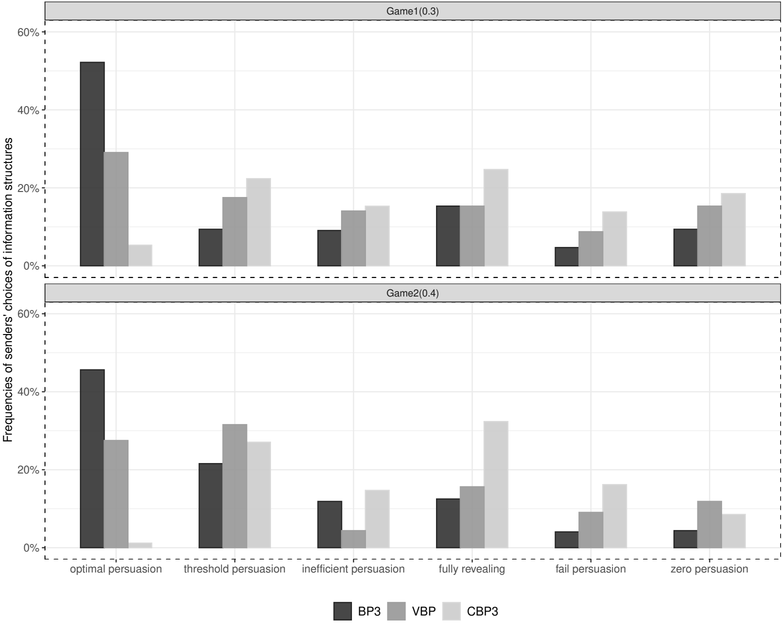


Senders' choice of $\pi(\cdot)$: coarse clusters based on the minimum distance





Senders' choice of $\pi(\cdot)$: coarse partitions and clusters



CBP and VBP: sender's message strategies

	m_l after signal l		m_r after signal r	
	Game 1	Game 2	Game 1	Game 2
CBP3&VBP: Theoretical	1	1	.0	.0
CBP3: Sender	.90	.90	.24	.25
CBP3: Receiver	.84	.93	.41	.41
VBP: Sender	.86	.93	.35	.37
VBP: Receiver	.79	.94	.49	.45

- CBP3: (all lie, partial lie, full truthful) = (6%, 73%, 20%)
  — receiver's prediction: (5%, 65%, 30%)
- VBP: (all lie, partial lie, full truthful) = (5%, 64%, 30%)
  — receiver's prediction: (6%, 57%, 37%)
- Fully revealing information structures are associated with low truthful rate.

CBP and VBP: bet choices

	Bet L after signal l		>	Bet L after signal r	
	Game 1	Game 2		Game 1	Game 2
CBP3: Optimal	81%	91%		0%	0%
CBP3: Predicted	58%	57%		20%	21%
CBP3: Empirical	33%	28%	≈	26%	24%
VBP: Optimal	85%	88%		0%	0%
VBP: Predicted	86%	88%		12%	9%
VBP: Empirical	69%	73%	>	16%	6%

- Optimal vs. Predicted in CBP3: knowing $\pi(\cdot)$ and (m_l, m_r), senders anticipated that many receivers will not be persuaded after l and falsely persuaded after r.
- Optimal vs. Empirical in CBP3: with $\pi(\cdot)$ and unknown (m_l, m_r), receivers in CBP3 act as ignoring info structure and signals
- Fully revealing $\pi(\cdot)$s in CBP3 do not induce higher frequencies of betting L than $\pi(\cdot)$s in threshold and inefficient persuasion cluster.

Summary of CBP and VBP

- BP + partially verifiable communication slightly decreases the empirical persuasion rate and match rate.
    - Many senders choose info structures and message strategies close to the theoretical predictions.
    - Most receivers understand the mechanism of persuasion via π and unravelling message, and act optimally wrt. them.
    - Receivers' uncertainties about message strategies are not costly for persuasion and info transmission.

- BP + cheap talk communication significantly decreases the empirical persuasion rate and slightly decreases the match rate.
    - More senders choose the fully revealing information structure and babbling message to influence the receiver's decisions.
    - Receivers understand the mechanism of informative π and untruthful message, and thus are not persuaded.
    - cheap talk communication increases the interim information transmission, but decreases the empirical one.

- The sender's private information about signal realizations harms the persuasion and information transmission under cheap-talk protocol but not much for partially verifiable protocol.

Contributions and Discussions

Contribution to the literature

This project provides:

- A simple experiment on Bayesian Persuasion
  - Frechette, Lizzeri, and Perego (2022); Aristidou, Coricelli, Vostroknutov, et al. (2019); Nguyen (2017); Au and Li (2018); Kwon (2020); Zhou (2022)
  - limited choice of (p, q); restricted equilibrium strategies; partition method

- An umbrella nesting variants of BP with communication
  - BP + state: Perez-Richet (2014); Galperti (2015); Hedlund (2017); Alonso and Camara (2018); Degan and Li (2021).
  - BP + signal: Pei (2015); Gentzkow and Kamenica (2017); Nguyen and Tan (2019); Lyu and Suen (2022)

- An empirical investigation on the interaction between commitment and communication in Bayesian persuasion
  - Frechette, Lizzeri, and Perego (2022): limited commitment on π
  - This project: full commitment on π, limited credibility on s

Discussions

- Robustness treatments: CBP2; MBP (50% BP + 50% CBP);
- Measuring interim information transmission
  - three measures of informativeness of π: entropy reduction; $|p - q|$; $|\mu_l - \mu_r|$
    — BP2 $\approx$ BP3 $\approx$ VBP $<$ CBP
    — negative correlation b/w informativeness of information structure and truthfulness of message strategies.
- The information structure serves both persuasion and **signaling role**?
  - Senders signal their truthfulness in message strategies by committing to an informative $\pi(\cdot)$.
  - Receivers infer the sender's "type" through the informativeness of $\pi(\cdot)$.
  - e.g. Degan and Li (2021): Persuasion with costly precision.
- Comparative statics: when lying is costly (receiver types)
  - truth-telling might be supported for some lying costs $k \sim F(\cdot)$
  - $E(kv) \leq E(kc)$, we will have more truth-telling in VBP than in CBP

Thanks for your patience!

Contact: yan.xu@univie.ac.at