Rules and Commitment in Communication

Guillaume Fréchette Alessandro Lizzeri Jacopo Perego

NYU

Harvard
October 2017

INTRODUCTION

We revisit a classic question in economics from a new perspective:

- How “much” **information** can be shared under direct communication among interested parties?
- How does this depend on rules and protocols governing communication?

This is important for thinking about:

Lobbying, Austen-Smith (1993), Battaglini (2002); **Relation between committees and legislature**, Gilligan-Krehbiel (1987-1989); **Production of evidence to a jury**, Kamenica-Gentzkow (2011), Alonso-Camara (2016), ...

INTRODUCTION

What we do:

- A framework nesting existing models under the same umbrella.
- With this framework, we test comparative statics **across** these models.

We produce comparative statics along two principal dimensions:

1. **Rules**: What can the sender say?
2. **Commitment**: Can sender establish communication protocols?

INTRODUCTION

Focus on a minimal set-up:

- Binary state: Red and Blue.
- Two parties (*sender*, *receiver*) with conflicting interests.
- **Sender** has information, **Receiver** has ability to act.
- Three messages: red, blue and no message.

RULES

Rules: What can the sender say?

We explore two extremes:

- **Unverifiable** messages.
  - There are no rules governing which messages the sender can send.
- **Verifiable** messages.
  - When state **Red**: Sender can send **red** or **no message**.
  - When state **Blue**: Sender can send **blue** or **no message**.

COMMITMENT

Stage 1: Commitment.

- **Sender** selects her *commitment strategy*.
- This strategy will be revealed to the receiver.

Stage 2: Revision.

- **Sender** *learns* color of the ball.
- She can *revise* her previous choice.
- Revision is *not revealed* to the receiver.

With probability ρ

With probability $1 - \rho$

Stage 3: Guess.

- **Receiver** makes decisions as a function of message.
- The message comes from Commitment Stage with probability ρ.

SPECIAL CASES

- **Cheap Talk**. Crawford and Sobel (1982)
  - Unverifiable and no commitment.
- **Disclosure**. Grossman (1981), Milgrom (1981), Okuno-Fujiwara et al (1990)
  - Verifiable and no commitment.
- **Bayesian Persuasion**. Kamenica and Gentzkow (2011)
  - Unverifiable and full commitment.

Variations around a common basic structure, different predictions.

THIS PAPER

Exploit this framework to:

- Provide novel comparative statics: beyond preference alignment.
- Interaction of *Rules* and *Commitment* on strategic information transmission.
- Offer a broader perspective on these communication models.
- Test Bayesian persuasion.

THIS PAPER

Exploit this framework to:

- Provide novel comparative statics: beyond preference alignment.
- Interaction of *Rules* and *Commitment* on strategic information transmission.
- Offer a broader perspective on these communication models.
- Test Bayesian persuasion.

Our **questions**:

1. Are senders able to exploit **commitment**?
2. Do receivers understand information generated by commitment?
3. Do **rules** generate more responsiveness?

FINDINGS

- Subjects understand power of **commitment**: senders figure out how to exploit it and receivers how to react to it.
- Subjects understand the effect of **rules**: senders more informative and receivers more receptive with verifiable information.
- Commitment consistent with Bayesian persuasion. If receiver is more demanding, sender delivers more information.
- Overall informativeness decreases (increases) with commitment under (un)verifiable information.
- Quantitative departures from theory, too much information conveyed in verifiable treatments, too little under unverifiable treatments.

RELATED LITERATURE

- **Cheap talk experiments**: Dickhaut, McCabe, and Mukherji (1995); Blume, De Jong, Kim, and Sprinkle (1998); Cai and Wang (2006); Sanchez-Pages and Vorsatz (2007); Wang, Spezio, Camerer (2010)
- **Disclosure experiments**: Forsythe, Isaac, and Palfrey (1989); King and Wallin (1991); Dickhaut, Ledyard, Mukherji, and Sapra (2003); Forsythe, Lundholm, and Rietz (1999); Benndorf, Kübler, and Normann (2015); Hagenbach and Perez-Richet (2015); Jin, Luca, and Martin (2016)
- **Disclosure field:** Mathios (2000); Jin and Leslie (2003); Dranove and Jin (2010)

GAME

– Binary state $\Theta = \{\theta_L, \theta_H\}$. Common prior μ_0 on θ_H.

– Receiver actions $A = \{a_L, a_H\}$.

– Set of messages $M = \{\theta_L, \theta_H, n\}$.

– Set $M^\theta \subseteq M$: messages that Sender can use in state θ.
  - Information is *unverifiable* if $M^\theta = M$ for all θ.
  - Information is *verifiable* if $M^\theta = \{\theta, n\}$ for all θ.

GAME

- Sender's utility: $v(a) := \mathbf{1}(a = a_H)$.
  - Wins if Receivers chooses a_H.

GAME

- Sender's utility: $v(a) := \mathbf{1}(a = a_H)$.
  - Wins if Receivers chooses a_H.

- Receiver's preferences:
  - $u(a_L, \theta_L) = u(a_H, \theta_H) = 0$.
  - $u(a_L, \theta_H) = -(1-q), u(a_H, \theta_L) = -q$.
  - Choose action a_H if $\mu(\theta_H) \geq q$.
    We call q the *persuasion threshold*.

GAME

Stage 1:

- Sender chooses a **commitment** strategy: $\pi_C : \Theta \to \Delta(M^\theta)$.

Stage 2: With probability $1 - \rho$, she enters an **revision stage**:

- Learns the realization of θ.
- Chooses a **revision** strategy: $\pi_R(\theta) \in \Delta(M^\theta)$ conditional on θ.

Stage 3:

- Receiver guesses. $a : M \times \Pi_c \to A$

Parameter ρ captures the extent of commitment.

THEORY RESULTS/PREDICTIONS

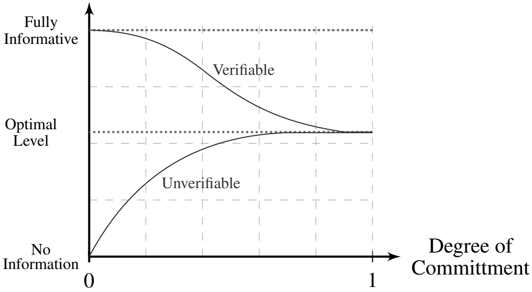
Proposition.

- There is a $\widehat{\rho}$ such that, if $\rho > \widehat{\rho}$:
  1. some information is communicated in U,
  2. less than full information is communicated in V.
- Consider ρ such that $\widehat{\rho} < \rho < 1$. Commitment has opposite effects on the amount of information transmission in V versus U:
  1. under U, less information is transmitted in revision stage than in commitment stage;
  2. under V, more information is transmitted in revision stage than in commitment stage.

THEORY RESULTS/PREDICTIONS

Proposition.

- When messages are *verifiable*, commitment **decreases** informativeness.
- When messages are *unverifiable*, commitment **increases** informativeness.
- For $\rho = 1$, equilibrium outcome is "rule-independent."



THEORY RESULTS/PREDICTIONS

Proposition.
For any $\rho > 0$, for both cases of verifiable and unverifiable messages, as the persuasion threshold q increases, the strategy of the Sender becomes more informative.

SPECIAL CASES

How "much" **information** can be transferred in equilibrium?

1. **Cheap Talk**.
   - No information transmitted: *Babbling*.
2. **Disclosure**.
   - All information transmitted: *Unraveling*.
3. **Bayesian Persuasion**.
   - Some information is transmitted: *Lie, but maintain incentives*.

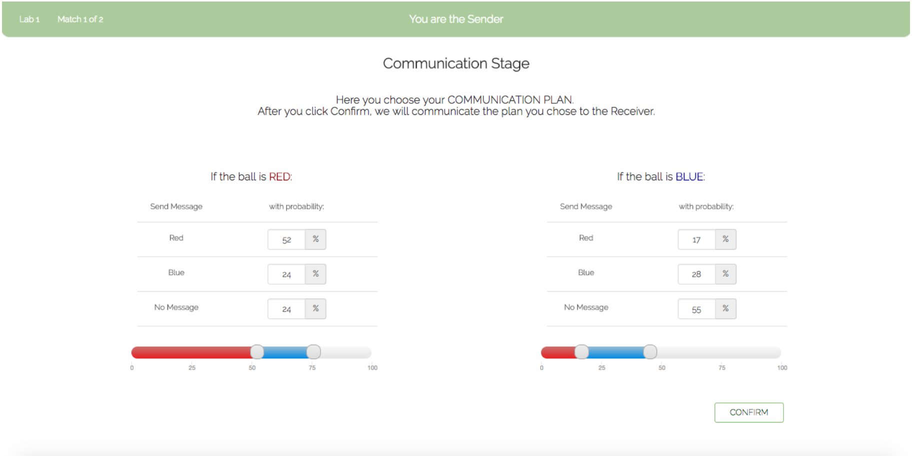
EXPERIMENT

Setup:

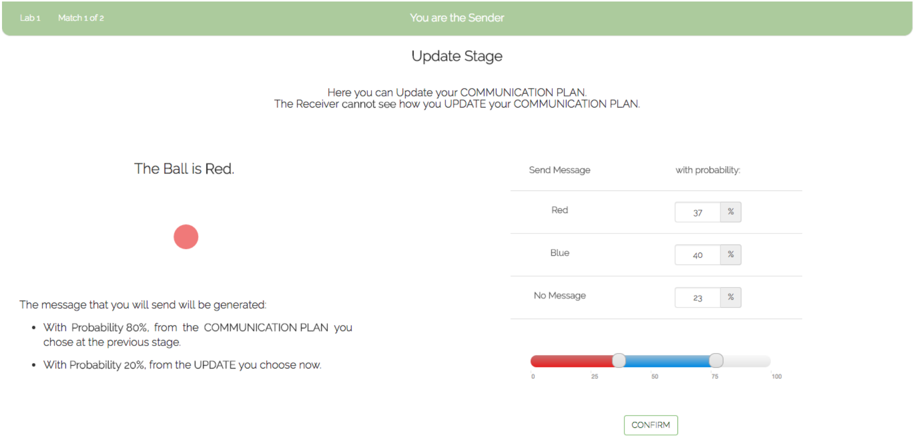
- Urn has three balls: two blue and one red.
- Receiver wins \$2 if guesses correctly.
- Sender wins \$2 if Receiver says Red.
- Up to three messages: **red**, **blue**, **no message**.
- Rules:
  - Verifiable: truth or no message.
  - Unverifiable: no constraints.

Design

Design


Lab 1
Match 1 of 2
You are the Sender
Communication Stage
Here you choose your COMMUNICATION PLAN.
After you click Confirm, we will communicate the plan you chose to the Receiver.
If the ball is RED:
Send Message
with probability:
Red
52 %
Blue
24 %
No Message
24 %
0
25
50
75
100
If the ball is BLUE:
Send Message
with probability:
Red
17 %
Blue
28 %
No Message
55 %
0
25
50
75
100
CONFIRM

DESIGN


Lab 1
Match 1 of 2
You are the Sender
Update Stage
Here you can Update your COMMUNICATION PLAN.
The Receiver cannot see how you UPDATE your COMMUNICATION PLAN.
The Ball is Red.
The message that you will send will be generated:
With Probability 80%, from the COMMUNICATION PLAN you chose at the previous stage.
With Probability 20%, from the UPDATE you choose now.
Send Message
with probability:
Red
37
%
Blue
40
%
No Message
23
%
0
25
50
75
100
CONFIRM

DESIGN

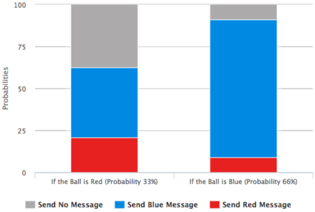
Lab 2 Match 1 of 2 You are the Receiver

Guessing Stage

The message you will receive will come:

- with probability 20%, from the UPDATE, that you can't see.
- with probability 80%, from the COMMUNICATION PLAN you see below:

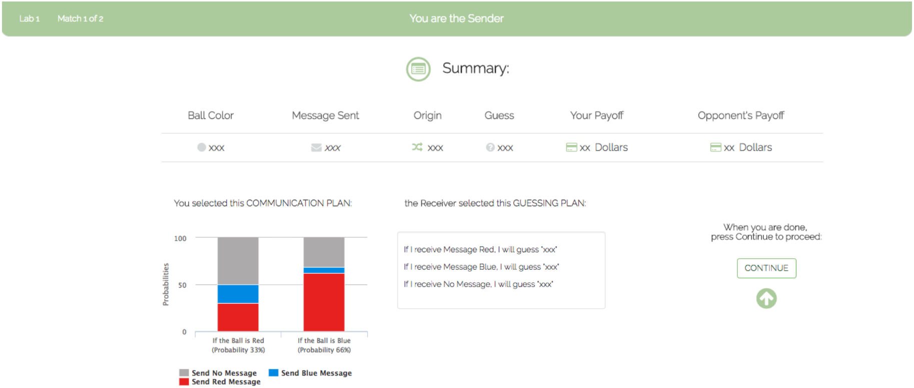
COMMUNICATION PLAN:



Choose your GUESSING PLAN:

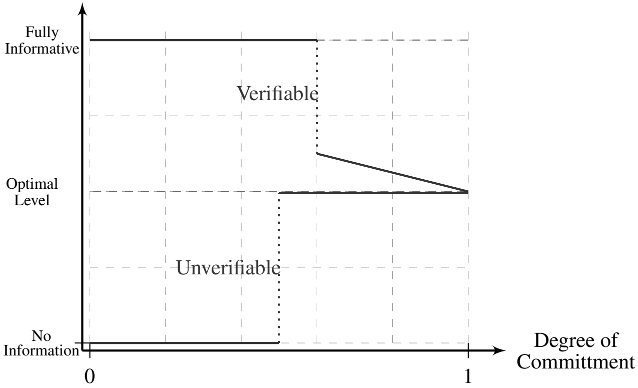
If I Receive Message...	...my guess will be:
The Ball is Red	RED BLUE
The Ball is Blue	RED BLUE
No Message	RED BLUE

DESIGN



Lab 1
Match 1 of 2
You are the Sender
Summary:
Ball Color
Message Sent
Origin
Guess
Your Payoff
Opponent's Payoff
xxx
xxx
xxx
xxx
xx Dollars
xx Dollars
You selected this COMMUNICATION PLAN:
100
50
0
Probabilities
If the Ball is Red (Probability 33%)
If the Ball is Blue (Probability 66%)
Send No Message
Send Blue Message
Send Red Message
the Receiver selected this GUESSING PLAN:
If I receive Message Red, I will guess 'xxx'
If I receive Message Blue, I will guess 'xxx'
If I receive No Message, I will guess 'xxx'
When you are done, press Continue to proceed:
CONTINUE

PREDICTION (REVISITED)



TREATMENTS

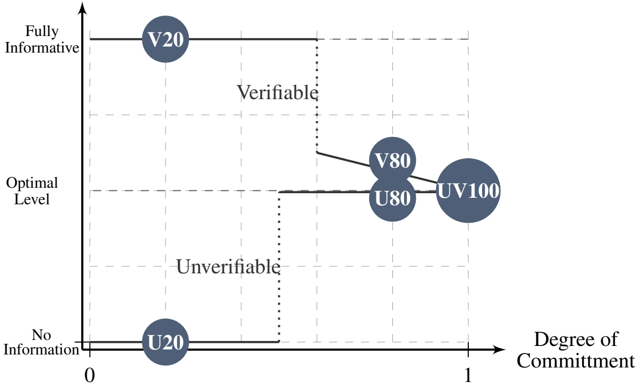
Treatments (2x3):

Rules: Verifiable *vs* Unverifiable.
Commitment: $\rho = \{20, 80, 100\}$.

Labeling:

	Commitment		
Rules	V20	V80	V100
	U20	U80	U100

TREATMENTS


Fully Informative
Optimal Level
No Information
V20
Verifiable
V80
U80
UV100
Unverifiable
U20
0
1
Degree of Committment

EQUILIBRIUM

Sender's **equilibrium behavior** in two extreme cases:

U100

		messages		
		r	b	n
Ball	**R**	100%	0	0
	B	50%	50%	0

V100

		messages		
		r	b	n
Ball	**R**	0	0	100%
	B	0	50%	50%

Intuition and main *tensions*:

– **U100**. Lie as much as you can, but preserve incentives.

– **V100**. Never release good news: "No news, good news."

EQUILIBRIUM BEHAVIOR

	Sender								**Receiver**	
	Commitment				*Revision*				*Guessing*	
Treat.	Ball	Message			Ball	Message			Mes.	Guess
		red	blue	no		red	blue	no		
V20	R	1		0	R	1		0	red	*red*
	B		x	$1-x$	B		x	$1-x$	blue	*blue*
									no	*blue*
V80	R	0		1	R	1		0	red	*red*
	B		$\frac{3}{4}$	$\frac{1}{4}$	B		0	1	blue	*blue*
									no	*red*
V100	R	0		1					red	*red*
	B		$\frac{1}{2}$	$\frac{1}{2}$					blue	*blue*
									no	*red*
U20	R	x	y	$1-x-y$	R	1	0	0	red	*blue*
	B	x	y	$1-x-y$	B	1	0	0	blue	*blue*
									no	*blue*
U80	R	1	0	0	R	1	0	0	red	*red*
	B	$\frac{3}{8}$	$\frac{5}{8}$	0	B	1	0	0	blue	*blue*
									no	*blue*
U100	R	1	0	0					red	*red*
	B	$\frac{1}{2}$	$\frac{1}{2}$	0					blue	*blue*
									no	*blue*

EXPERIMENTAL DETAILS

Implementation:

- Two unpaid practice rounds.
- 25 periods played for money in **fixed roles**.
- Random rematching between periods.

General Information:

- Six treatments, four sessions per treatment.
- 384 subjects ($\approx$ 16 per session; between 12 and 24).
- Average earnings: \$24 (including \$10 show up fee).
- Average duration: 100 minutes.

RESULTS

INFORMATIVENESS: CORRELATION

How to measure equilibrium **informativeness**?

- Pearson **correlation index** ϕ between Ball and Guess.
  (Definition ▷)

Intuition:

- If no information, $\phi = 0$. Receiver always says blue.
- If full information, $\phi = 1$. Receiver perfectly matches the state.

We focus attention on data from last 15 rounds.

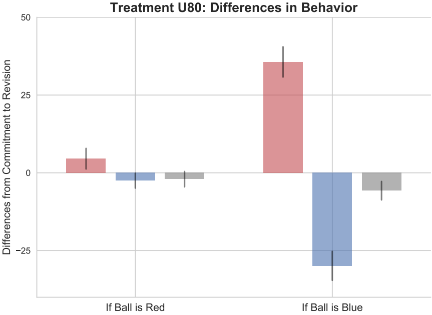
TO FOCUS ON SENDERS

Assume Bayesian receiver:

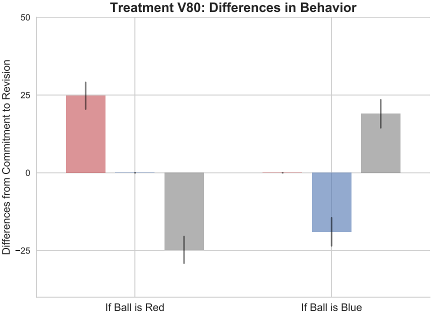
1. Receives **message** m.
2. Computes **posterior** belief $\mu(R|m) \in [0, 1]$.
3. Guesses Red if and only if $\mu(R|m) \geq \frac{1}{2}$.

SENDER: COMMITMENT VS. REVISION, $\rho = 0.8$

Informativeness in U80:
Commitment stage: 0.43
Revision stage: 0.0

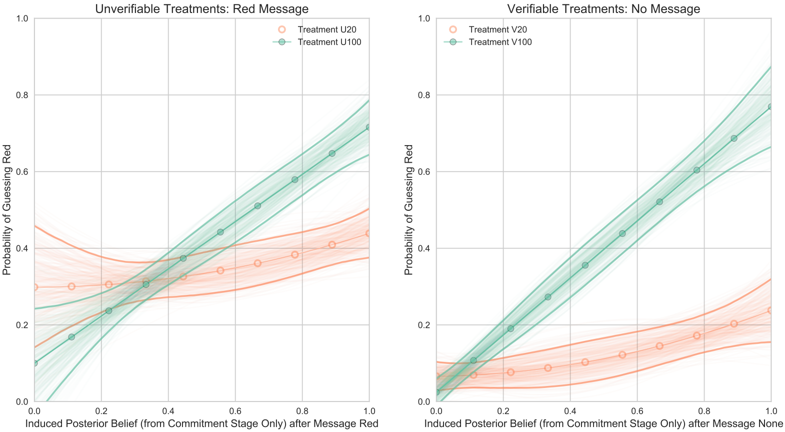


Informativeness in V80:
Commitment stage: 0.83
Revision stage: 0.98



RECEIVER'S RESPONSE TO PERSUASIVE MESSAGES

$\rho = 0.2$ vs. $\rho = 1$



DO SUBJECTS REACT TO RULES?

THE CASE OF $\rho = 0.2$

- Senders send more information in V20 than U20:
  - $\phi^B = 0.89$ vs 0.00.
- Receivers' probability of guessing red is higher in V20 than U20:
  - 97% vs 37%.

TREATMENT U100H

New payoffs:

– Receiver wins if correctly guesses the color of the ball:
  - 2 if ball is Blue.
  - $\frac{2}{3}$ if ball is Red.

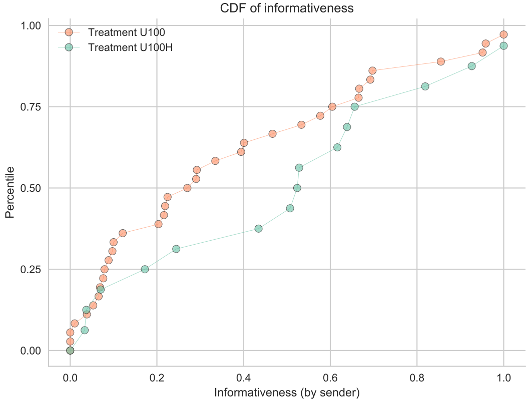
– Sender wins 3 if Receiver guesses Red.

Bayesian Receiver guesses Red iff $\mu(R) \geq 0.75$.

Solution is to provide more information:

$\pi_1^\star$:		Message		
		r	b	n
Ball	**R**	1	0	0
	B	1/6	5/6	0

CDF of ϕ^B for treatments U100 and U100H



CORRELATIONS BY TREATMENT

Theory:

	Commitment (ρ)			
	20%	**80%**	**100%**	**100% H.**
Verifiable	1	0.57	0.50	
Unverifiable	0	0.50	0.50	0.79

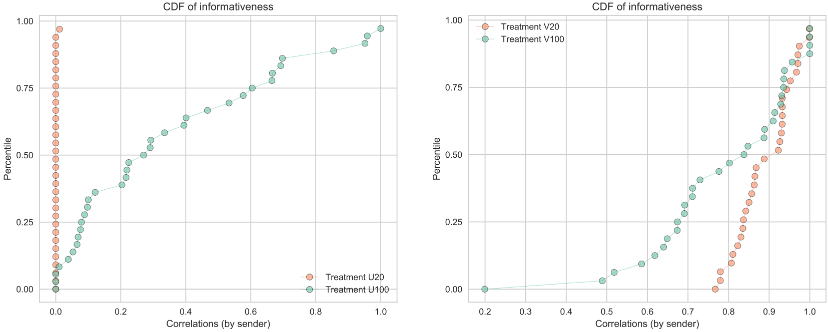
Data:

	Commitment (ρ)						
	20%		**80%**		**100%**		**100% H.**
Verifiable	0.80	$\approx$	0.78	$>$	0.67		
	$\vee$		$\vee$		$\vee$		
Unverifiable	0.09	$<$	0.21	$\approx$	0.21	$\approx$	0.20

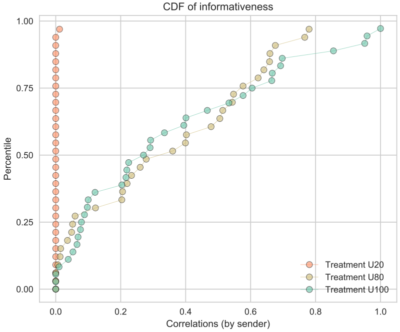
Data + Bayesian Rec:

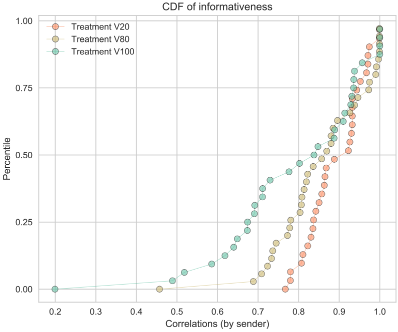
	Commitment (ρ)						
	20%		**80%**		**100%**		**100% H.**
Verifiable	0.89	$\approx$	0.85	$>$	0.78		
	$\vee$		$\vee$		$\vee$		
Unverifiable	0.00	$<$	0.33	$\approx$	0.34	$\approx$	0.45

CDF OF ϕ^B: $\rho = 0.2$ VS $\rho = 1$



CDF OF ϕ^B: ρ 0.2, 0.8, AND 1.





Informativeness: Correlation

Verifiable:

- Commitment decreases correlation, although much less than it should.

Unverifiable:

- Commitment increases correlation, although much less than it should.

INFORMATIVENESS: CORRELATION

Verifiable:

- Commitment decreases correlation, although much less than it should.

Unverifiable:

- Commitment increases correlation, although much less than it should.

This measure takes into account at the same time:

1. Senders' behavior.
2. Receivers' behavior.

Cumulates mistakes from all sides.

- Who is getting it wrong and why?

CORRELATION WITH BAYESIAN RECEIVERS

Point predictions on informativeness increase in all treatments.

Observation 1.

Informativeness reacts to commitment in a manner consistent with the theory. When receivers are Bayesian, predictions close to theory for unverifiable case, mixed for unverifiable case.

Most interesting deviation:

- Even with rational receivers: $U100 \ll V100$

INFORMATIVENESS: RANDOM POSTERIORS

What posteriors do senders attempt to induce?

Chain of events: $\theta \quad \Rightarrow \quad m \quad \Rightarrow \quad \mu(R|r)$

Goal:

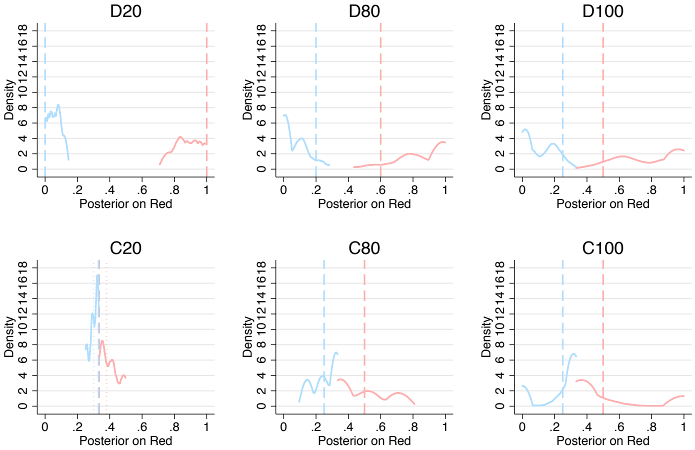
- Extracting informativeness from induced posteriors.

We use:

- Variation in conditional **posterior** beliefs.

A richer measure than correlation.

INFORMATIVENESS: RANDOM POSTERIORS



Posteriors on the ball being RED.
The color of the line indicates the state.
Vertical lines indicate the equilibrium predictions.

INFORMATIVENESS: RANDOM POSTERIORS

	Commitment (ρ)					
	20%		**80%**		**100%**	
Verifiable	0.86	**(1.00)**	0.78	**(0.40)**	$\boxed{0.69}$	$\boxed{(0.25)}$
	B	**R**	**B**	**R**	**B**	**R**
	0.05	0.91	0.07	0.85	0.10	0.80
Unverifiable	0.11	**(0.00)**	0.23	**(0.25)**	0.30	**(0.25)**
	B	**R**	**B**	**R**	**B**	**R**
	0.30	0.40	0.26	0.49	0.23	0.53

Informativeness: Random Posteriors

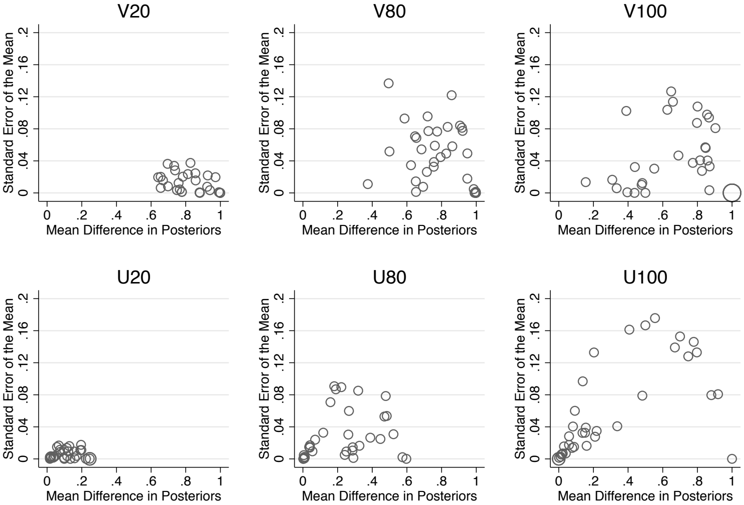
This confirms that senders react to commitment and, to some extent, know how to exploit it.

Also, this shows under a different light that:

Observation 2.

Point prediction of V100 is further off than U100.

SENDERS' HETEROGENEITY



Size of circle proportional to number of observations.

FULL COMMITMENT (THEORY)

Let's review equilibrium behavior in **U100** and **V100**.

U100

		messages		
		r	b	n
States	R	100%	0	0
	B	50%	50%	0

V100

		messages		
		r	b	n
States	R	0	0	100%
	B	0	50%	50%

FULL COMMITMENT (DATA)

What is going on in V100?

- Full commitment, no lies.

Let's see the aggregate data in **U100** and **V100**.

U100

		messages		
		r	b	n
States	R	74%	12%	14%
	B	44%	39%	17%

V100

		messages		
		r	b	n
States	R	**51%**	0	**49%**
	B	0	58%	42%

FULL COMMITMENT (DATA)

Unpacking Senders's heterogeneity in V100:

We compute the *most representative* strategies for Senders in V100.

49% of data points

		messages		
		r	b	n
States	R	16%	0	84%
	B	0	72%	28%

33% of data points

		messages		
		r	b	n
States	R	95%	0	5%
	B	0	20%	80%

18% of data points

		messages		
		r	b	n
States	R	96%	0	4%
	B	0	95%	5%

QUANTAL RESPONSE EQUILIBRIUM

To understand who is mostly responsible for these documented deviations, we estimate a QRE model with heterogeneous λ's.

We use the empirical method in Bajari and Hortacsu (2005). (link)

Challenges: dynamic game with a continuum of actions.

Denoting $EU_i(a_i)$ the expected utility of action a_i for player i:

$$\mathbb{P}(a_i) = \frac{e^{\lambda_i EU_i(a_i)}}{\sum_{a_i' \in A_i} e^{\lambda_i EU_i(a_i')}}$$

- When $\lambda_i = \infty$, the player is perfectly rational.
- When $\lambda_i = 0$, the player is perfectly naive.

QUANTAL RESPONSE EQUILIBRIUM

Our results: (Preliminary)

Treatment V100: $\lambda_S = 0.17$ and $\lambda_R = 1.73$.

Treatment U100: $\lambda_S = 0.99$ and $\lambda_R = 1.28$

The comparison among treatments is legitimate because:

(a) Binary actions.

(b) Same "transformed" strategy spaces.

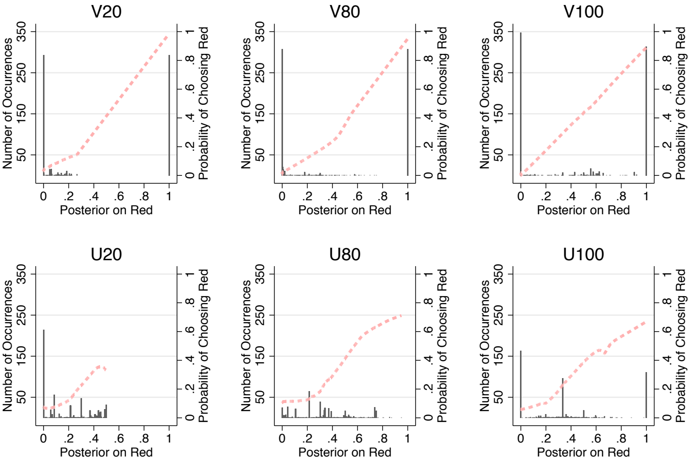
FOCUS ON RECEIVERS

How to establish **rationality** of a receiver?

A weak requirement of rationality:

- The likelihood of guessing red is **increasing** $\mu(R|m)$.
- Conditional on posterior, message should not matter.

FOCUS ON RECEIVERS



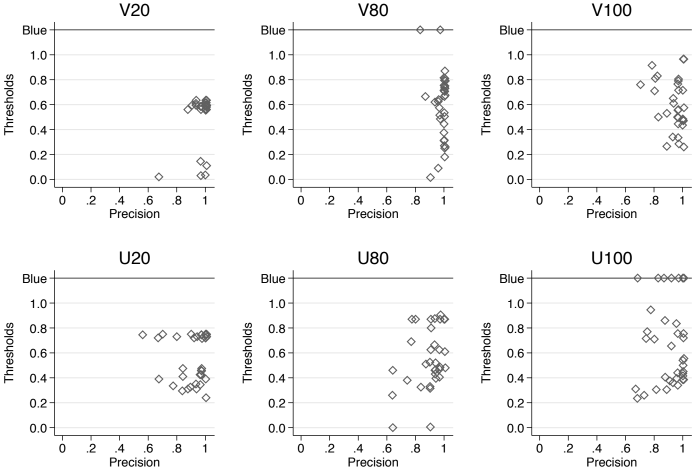
Bars indicate the number of messages inducing this posteriors on the ball being RED (left axis).
The red line indicates the probability that such a message yields a red guess (right axis).

PRECISION OF RECEIVERS' RESPONSE TO POSTERIORS

- The choices of a majority of subjects in each treatment is consistent with a threshold strategy at least 90% of the time.
- A large fraction of subjects in every treatment have a precision of at least 80%:

	Commitment (ρ)		
	20%	**80%**	**100%**
Verifiable	0.80	0.96	0.96
Unverifiable	0.92	0.85	0.75

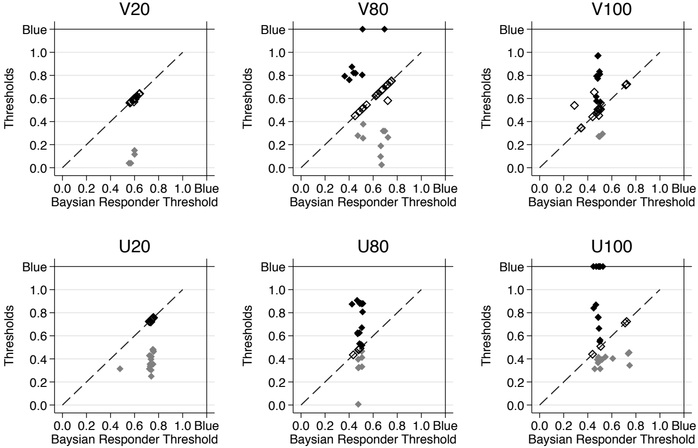
RECEIVERS' THRESHOLDS



Values are jittered slightly to make multiple overlapping thresholds distinguishable.

In more details

RECEIVERS' THRESHOLDS



Baysian Responder Threshold are the thresholds that would be estimated if the responders were Baysian given the posteriors in the data.
Black for subjects harder to convince than a Bayesian, gray for subjects easier to convince than a Baysian.
Values are jittered slightly to make multiple overlapping thresholds distinguishable.

REACTION TO IRRELEVANT INFORMATION

Unverifiable Treatments

	Commitment (ρ)		
	20%	**80%**	**100%**
Posterior	0.49**	0.45***	0.55***
Blue Message	-0.11**	-0.18***	-0.15**
No Message	-0.03	-0.16***	-0.05

Marginal effects on receiver's guess of red.

FOCUS ON RECEIVERS (SUMMARY)

Overall, receivers respond to communication protocol.

Observation 4.

- Most Receivers use threshold strategy most of the time
- Posterior beliefs not sufficient statistic, actions not sensitive enough to posteriors.
- Significant fraction indistinguishable from Bayesian
- Significant fraction too skeptical in high commitment treatments.
- Skepticism reduced by rules (Pareto improvement)

CONCLUSIONS

CONCLUSIONS

We study the role of *rules* and *commitment* on informativeness.

- Present a simple framework nesting known models as special cases.
- We perform comparative statics **across** models.
- Look at communication models from a different perspective.

CONCLUSIONS

RESULTS

- Many ways in which behavior responds to rules and commitment in line with (complex) theory.
- In aggregate data: in V receivers are close to optimal, not so in U.
- Senders' behavior heterogeneous.
- Some senders more likely to play close to equilibrium in V, but some senders also more likely to be "noisy" in V. Partly explains why, as ρ increases, informativeness decreases in V.

APPENDIX

Quantal Response Equilibrium

As in Bajari and Hortacsu (2005), we estimate H-QRE using a two-step procedure:

1. For every binned Sender's strategy $\tilde{\pi}_C \in \Pi$, we estimate the expected utility $EU_S(\pi_C)$–an equilibrium object–with $\hat{EU}_S(\pi_C)$, its empirical mean.
2. Then we use $\hat{EU}_S(\pi_C)$ to compute the Likelihood function as a function of the parameters λ_S and λ_R.

This procedure eliminates the need to compute the equilibrium, as in McKelvey and Palfrey (1995).

This greatly reduces the computational complexity of estimating the model.

(back)

Informativeness: Correlation

Pearson Correlation index btw Ball and Guess.
$\phi := \dfrac{n_{Rr}n_{Bb} - n_{Rb}n_{Br}}{\sqrt{n_R n_B n_r n_b}}$.

	$a = r$	$a = b$	$\theta = R$
$\theta = R$	n_{Rr}	n_{Rb}	n_R
$\theta = B$	n_{Br}	n_{Bb}	n_B
	n_r	n_b	

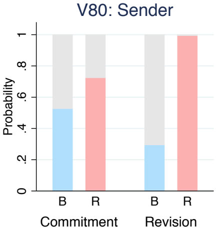
where

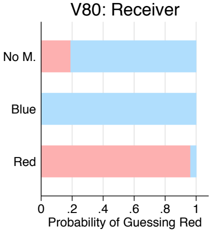
$$n_{\theta,a} = \sum_{m \in M} \hat{\pi}(m|\theta)\sigma(a|m)$$

and

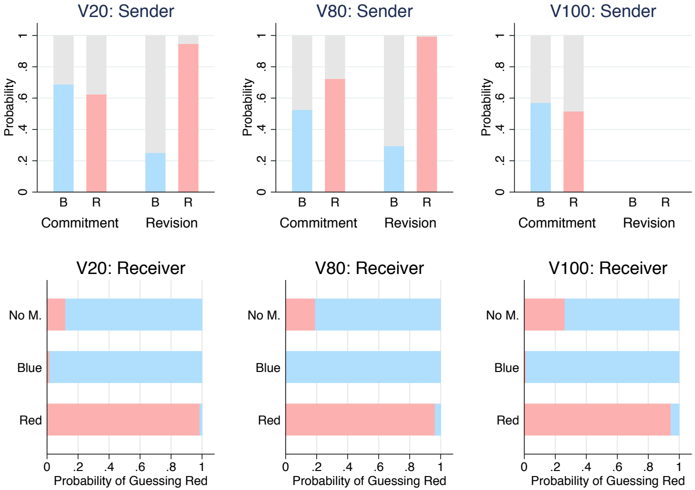
$$\hat{\pi}(m|\theta) := \rho\pi_C(m|\theta) + (1-\rho)\pi_U(m|\theta)$$

BEHAVIOR UNDER VERIFIABLE MESSAGES

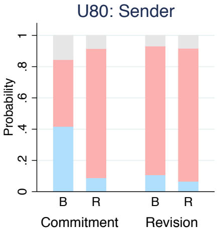


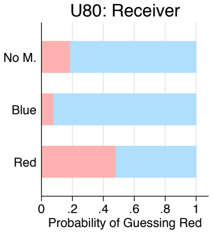


Behavior Under Verifiable Messages

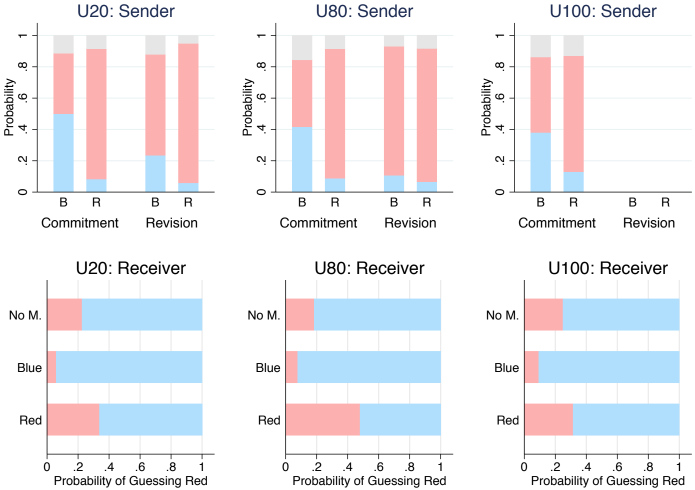


Behavior Under Unverifiable Messages

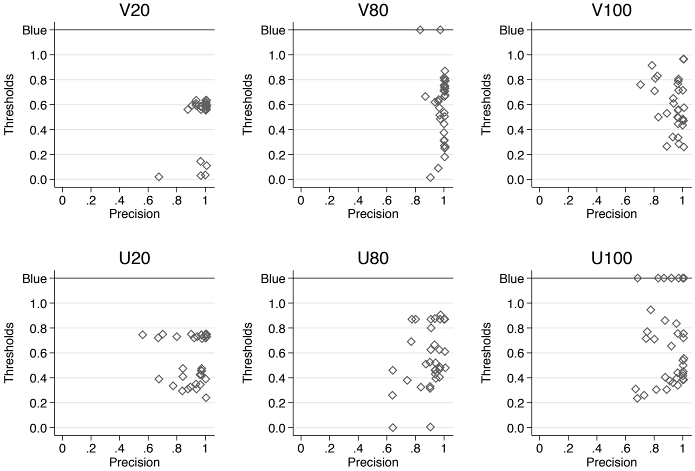




Behavior Under Unerifiable Messages



RECEIVERS' THRESHOLDS



Values are jittered slightly to make multiple overlapping thresholds distinguishable.

Go Back

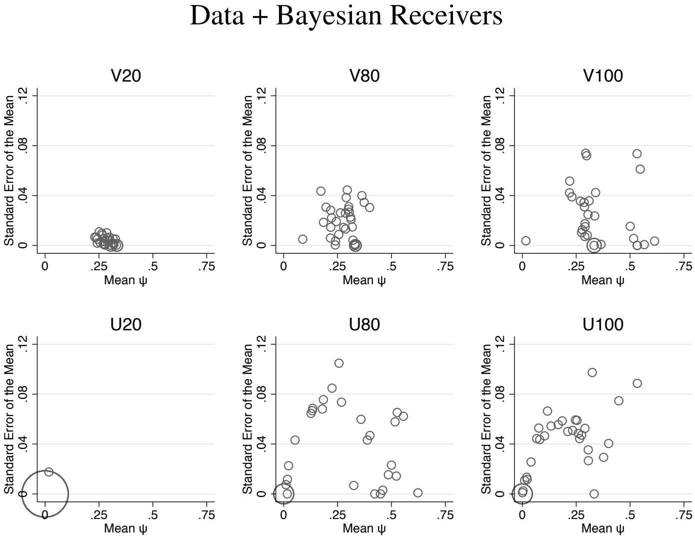
SENDERS' PAYOFFS

Data + Bayesian Receivers

	Commitment (ρ)							
	20%		**80%**		**100%**		**100% High**	
Verifiable	0.29		0.28		0.35			
	Theory	Simulated	Theory	Simulated	Theory	Simulated		
	0.33	0.33	0.60	0.30	0.67	0.30		
Unverifiable	0.00		0.25		0.17		0.29	
	Theory	Simulated	Theory	Simulated	Theory	Simulated	Theory	Simulated
	0.00	0.27	0.67	0.32	0.67	0.26	0.44	0.19

Table: Expected Payoffs (Normalized for maximal win)

SENDERS' HETEROGENEITY IN PAYOFFS



Size of circle proportional to number of observations.