Political Debate on Social Media: Theory and Evidence

Ole Jann* and Christoph Schottmüller**

*CERGE-EI, Charles University and Czech Academy of Sciences
**University of Cologne; TILEC

SAET Paris
July 2023

This paper

- A lot of people discuss politics on social media (Pew 2018: more than half of Americans)
- A lot of people are unhappy with it (find it stressful, find the tone too negative, too offensive etc)

This paper

- A lot of people discuss politics on social media (Pew 2018: more than half of Americans)
- A lot of people are unhappy with it (find it stressful, find the tone too negative, too offensive etc)
- This paper:
  1. A simple model of two people debating on social media
     $\Rightarrow$ some predictions, hypotheses
  2. A dataset of about 150,000 interactions on Twitter
     $\Rightarrow$ we document patterns that are consistent with the model
- Questions:
  1. What kind of debate emerges if people have several, potentially conflicting motivations?
  2. What is the empirical content of theories on communication (cheap talk, signaling, expressive utility)?

How we think about debates

1. People want to win debates (by moving other people's opinion closer to their own)
2. All else equal (i.e. if it did not influence the outcome of debates), people like to inform others
3. People can use sophisticated arguments, statistics, references etc – these take effort but are often not easily verifiable
4. People derive direct payoff from expressing their views (affirm their identity, feel as part of a group, let off steam, ...)

Outline

Model

Analysis

Empirical evidence on what model predicts

Sender and receiver

- We consider the most basic interaction: One sender, one receiver
- S can reply to tweet by R
- State of the world $\theta \in \{0, 1\}$ with equal probability, known to S
- S can communicate θ to R; then R takes an action that S cares about (metaphor for: S cares about R's posterior opinion)
- S and R differ in their ideology (= bias), i.e. some of R's action cannot be changed

The receiver

- The receiver:
  - takes an action (that S cares about)
  - has some ideological difference to S
- Payoff receiver:

$$U_R = -(a - \theta - b)^2$$

where a is action, b is ideological distance between S and R ($b > 0$ wlog)

The sender

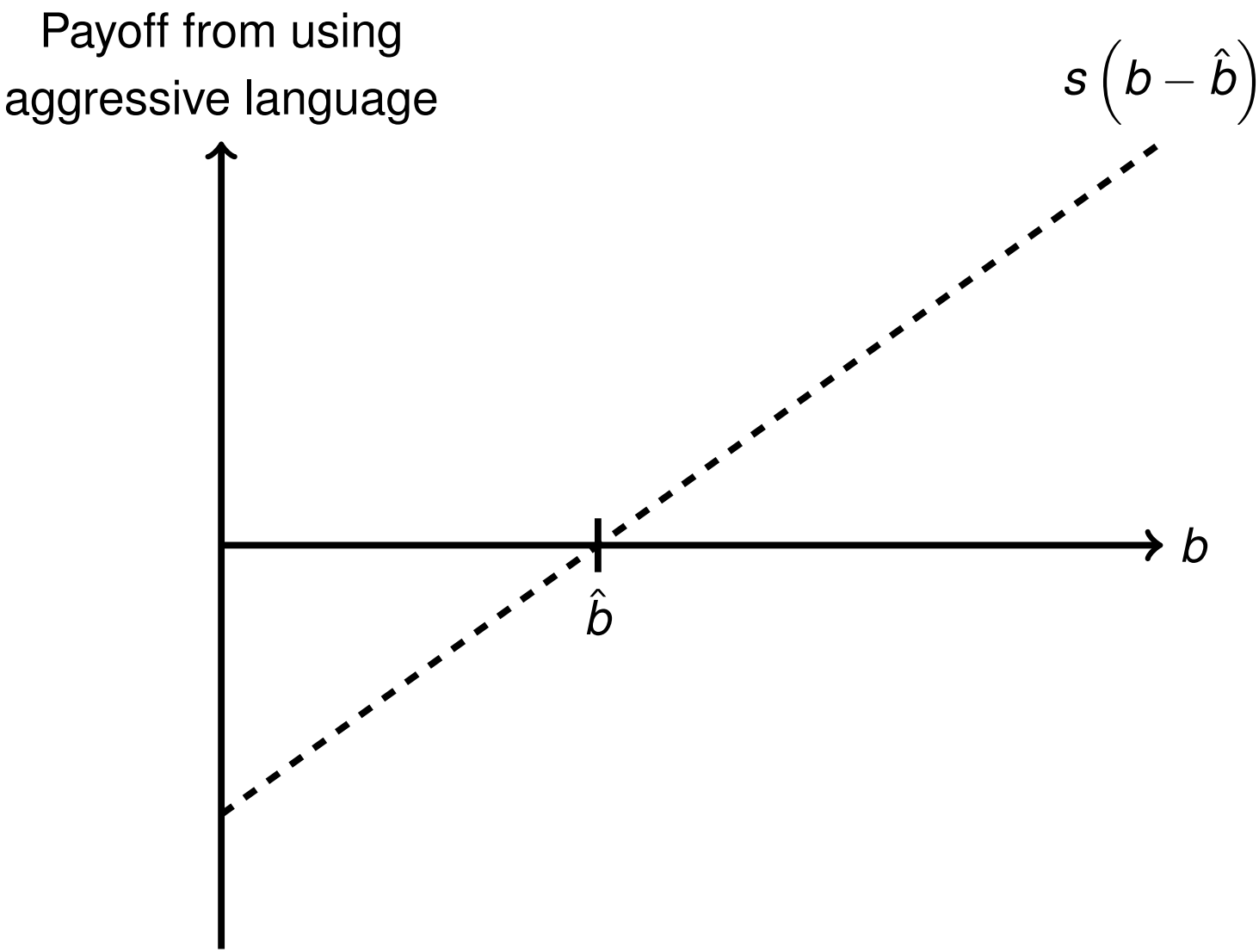
- The sender:
  - sends a message $m(\theta) \in \{0, 1\}$
  - can also provide (non-verifiable) evidence with some effort (e.g. "1_e" is message 1 with evidence)
  - can also choose whether to use aggressive language or not
- Payoff sender:

$$U_S = -(a - \theta)^2 - \mathbb{1}_e c + \mathbb{1}_a s \left(b - \hat{b}\right)$$

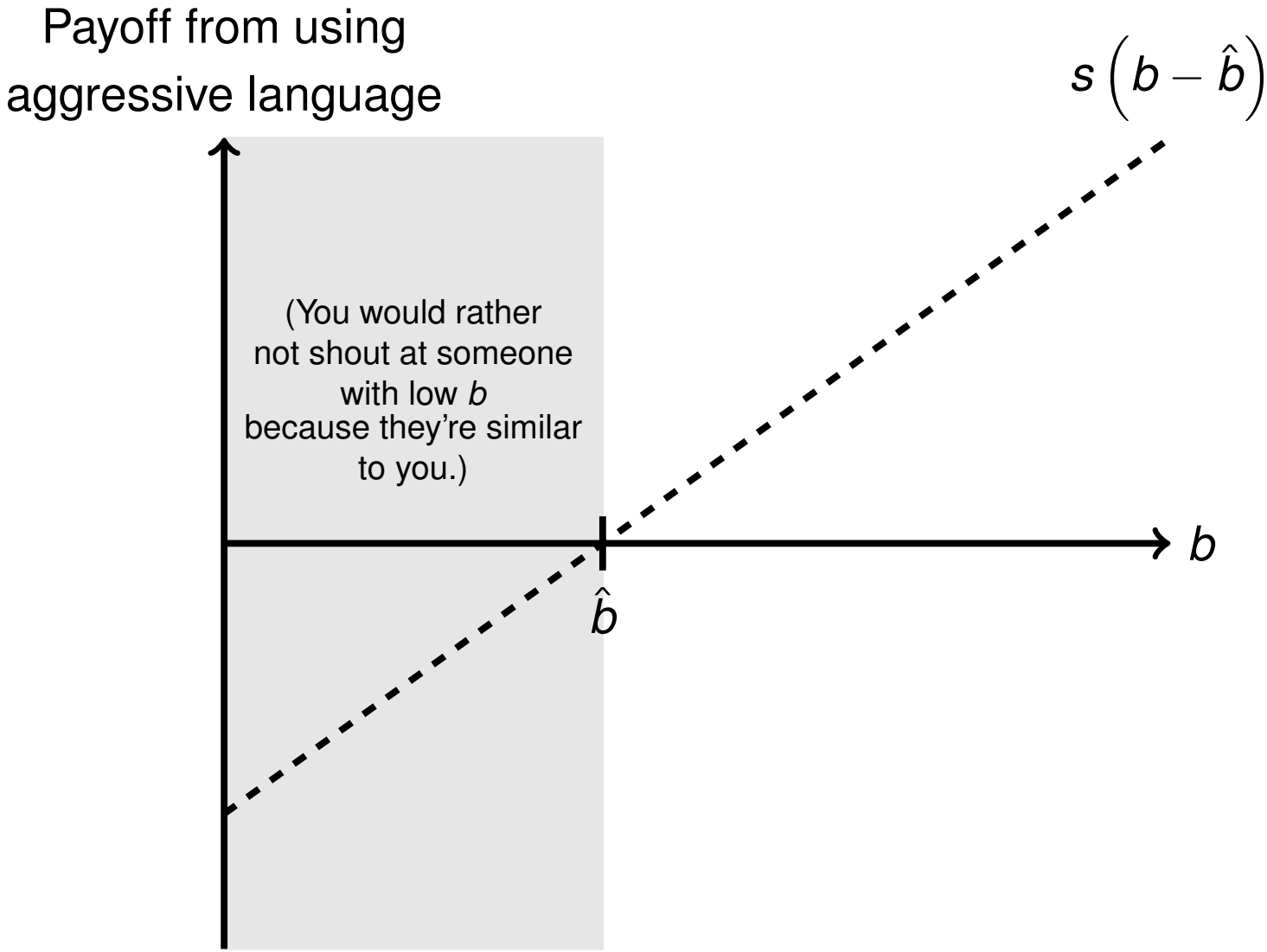
where:

- $a \in \mathbb{R}$ is R's action,
- $\mathbb{1}_e \in \{0, 1\}$ whether S uses evidence,
- $c \in \mathbb{R}_+$ cost of evidence,
- $\mathbb{1}_a \in \{0, 1\}$ whether aggressive language is used,
- $s \in \mathbb{R}_+$ satisfaction from using aggressive language,
- $\hat{b} \in \mathbb{R}_+$ some exogenous threshold

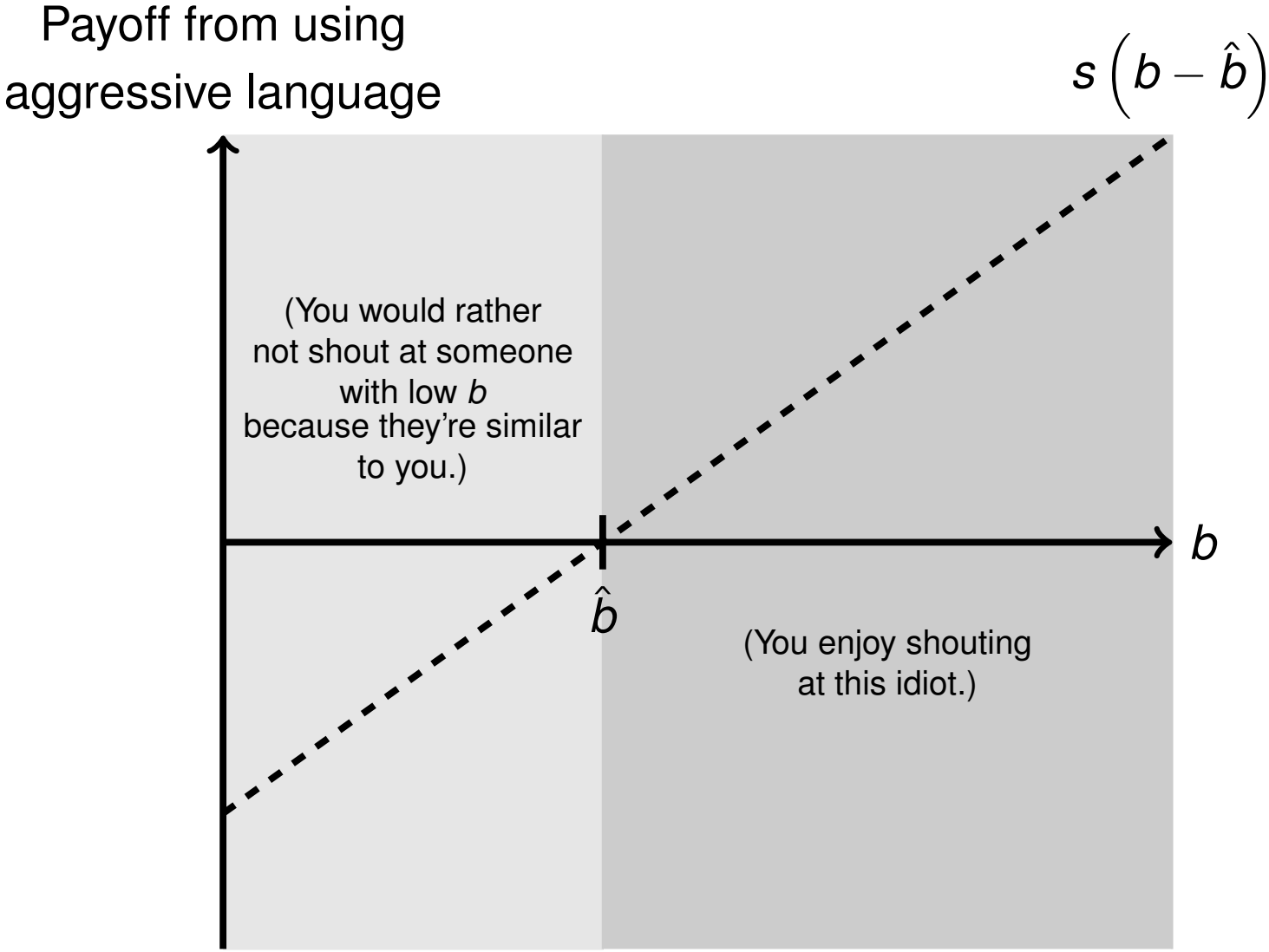
The payoff from using aggressive language



The payoff from using aggressive language



The payoff from using aggressive language



Outline

Model

Analysis

Empirical evidence on what model predicts

Three types of signaling

- We are interested in the most informative PBE
- Besides pure cheap talk, there are three ways for S to signal about θ:

1. Evidence: Making effort on evidence, $m(0) = 0_e$ and $m(1) = 1$

Three types of signaling

- We are interested in the most informative PBE
- Besides pure cheap talk, there are three ways for S to signal about θ:

1. Evidence: Making effort on evidence, $m(0) = 0_e$ and $m(1) = 1$
2. "Biting your tongue": Making an effort to not be aggressive (though you would like to), $m(0) = 0$ and $m(1) = 1_a$ (only possible if $b > \hat{b}$)

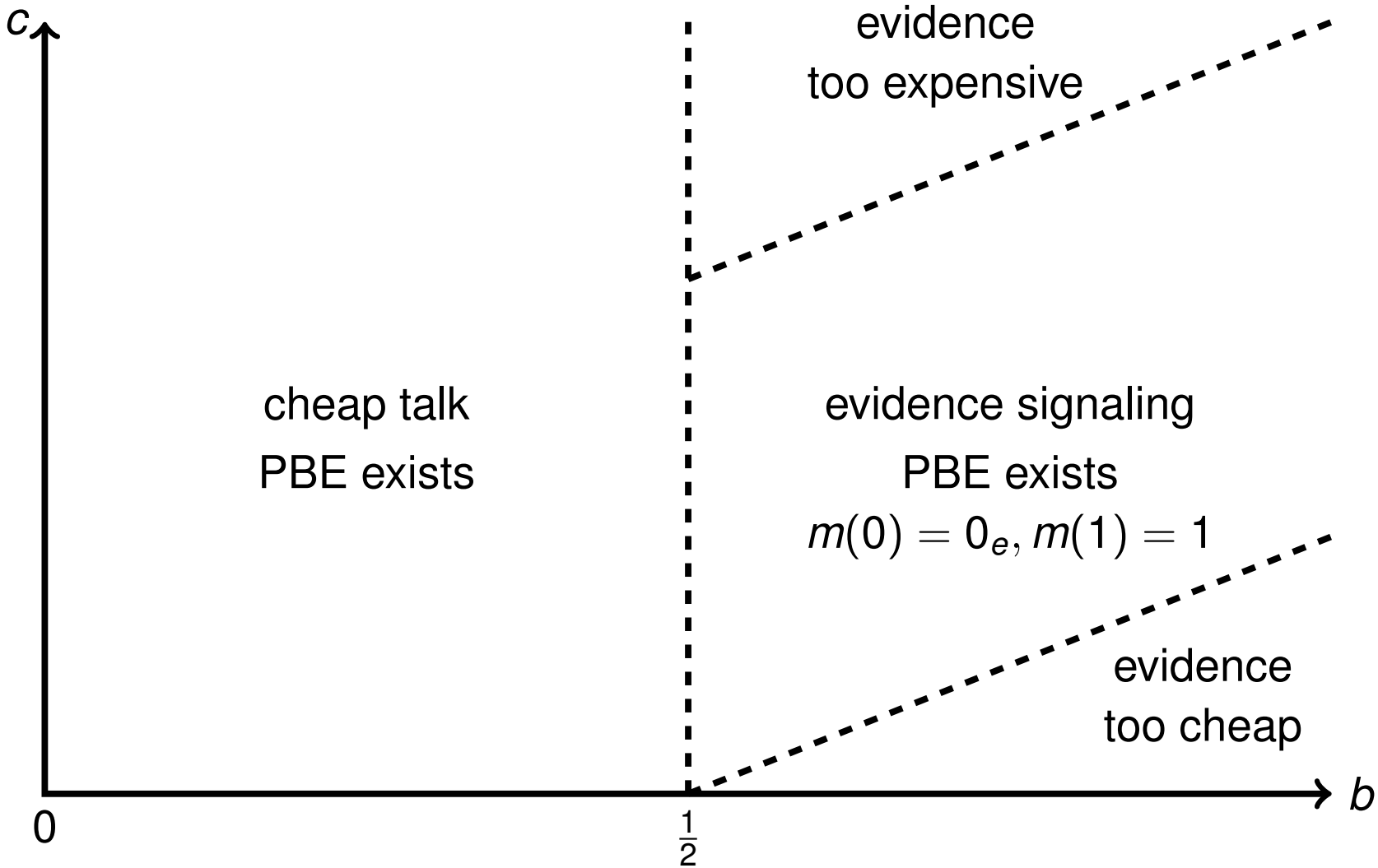
Three types of signaling

- We are interested in the most informative PBE
- Besides pure cheap talk, there are three ways for S to signal about θ:

1. Evidence: Making effort on evidence, $m(0) = 0_e$ and $m(1) = 1$
2. "Biting your tongue": Making an effort to not be aggressive (though you would like to), $m(0) = 0$ and $m(1) = 1_a$
   (only possible if $b > \hat{b}$)
3. "Tough talk among friends": Making an effort to be aggressive towards someone you mostly agree with, $m(0) = 0_a$ and $m(1) = 1$
   (only possible if $b < \hat{b}$)

- Combinations (1+2) or (1+3) are possible

Signaling with evidence



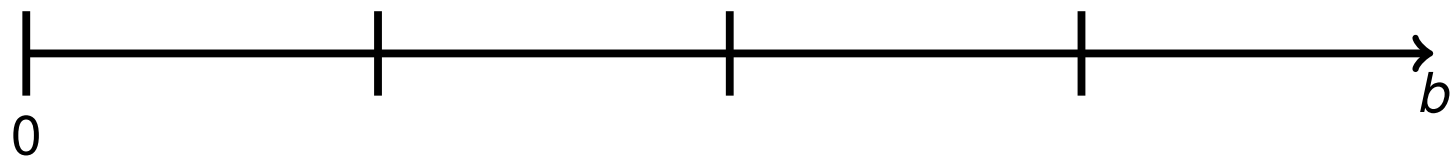
- b: bias; c: cost of evidence

Signaling with evidence and aggressive language

- If for every b, we choose the sender-best among the most informative PBEs, we can get the following (for some parameters):

$m(0)$:

$m(1)$:

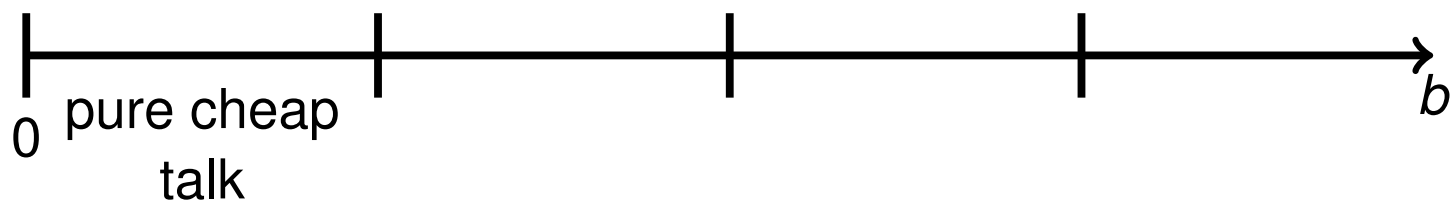


Signaling with evidence and aggressive language

- If for every b, we choose the sender-best among the most informative PBEs, we can get the following (for some parameters):

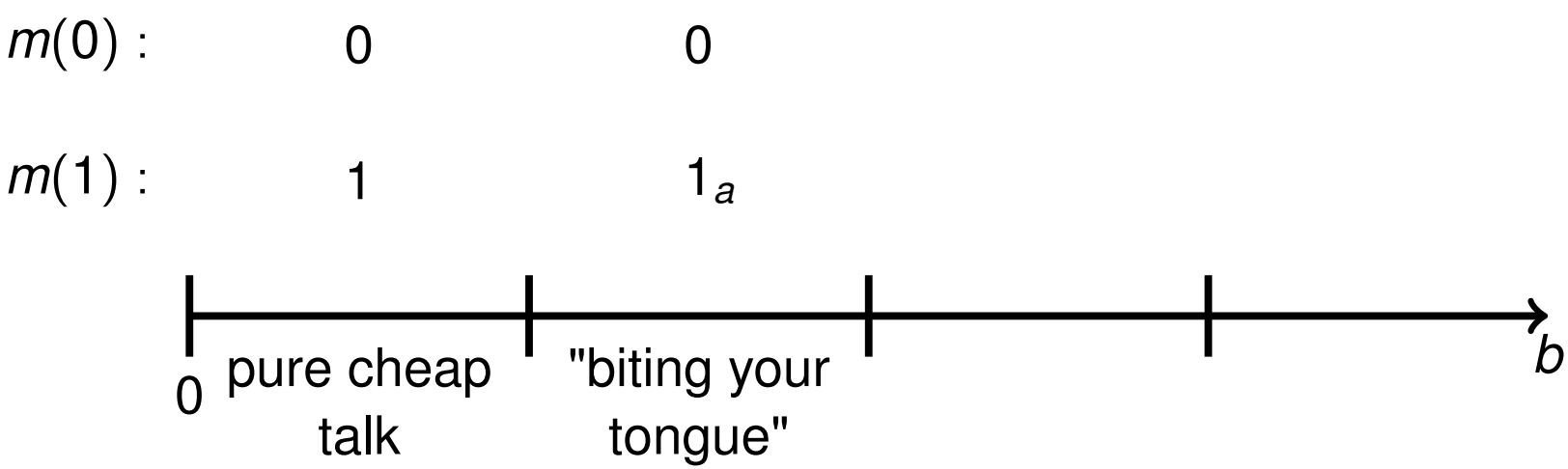
$m(0)$: 0

$m(1)$: 1



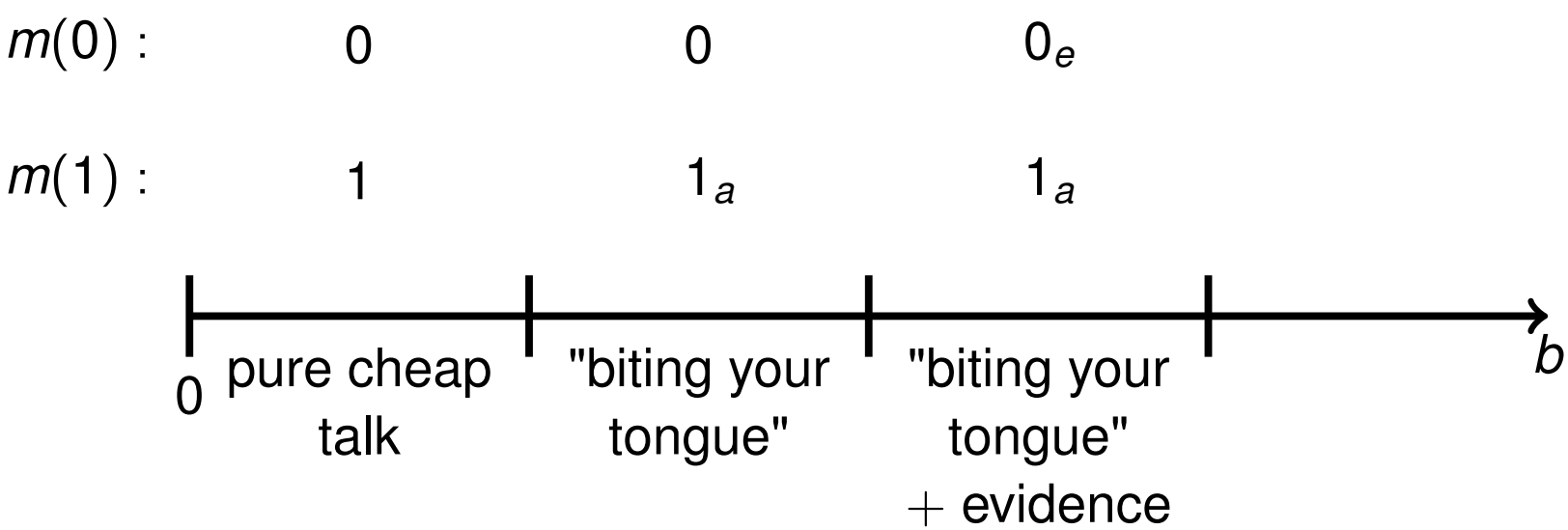
Signaling with evidence and aggressive language

- If for every b, we choose the sender-best among the most informative PBEs, we can get the following (for some parameters):



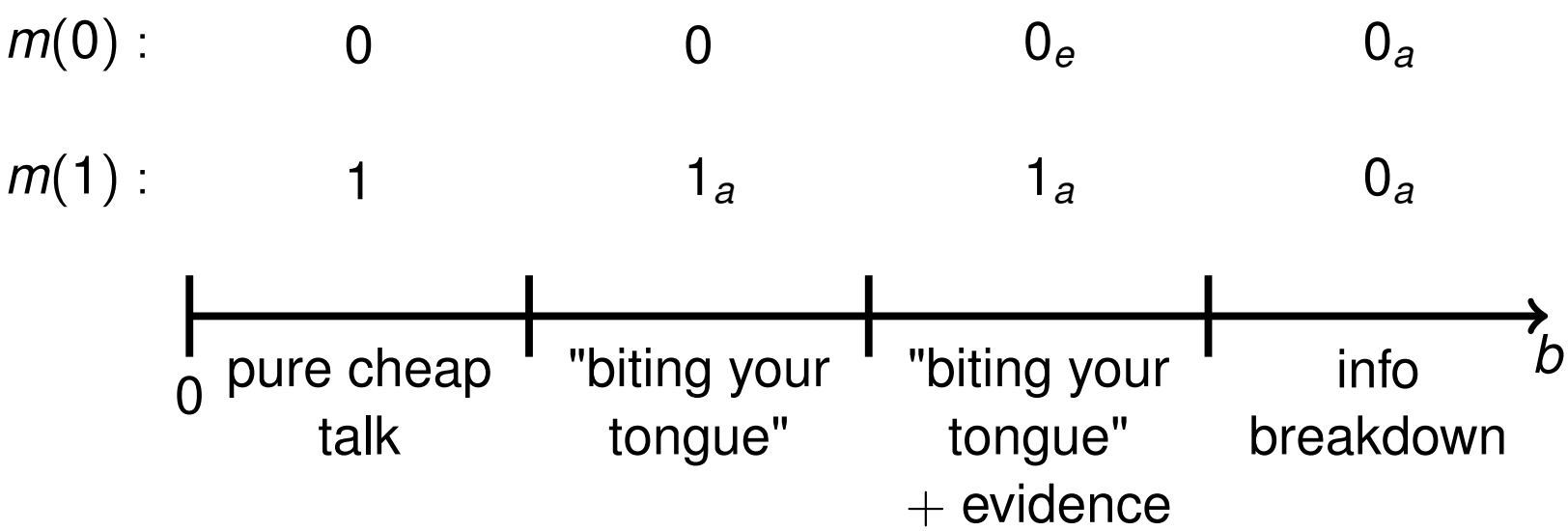
Signaling with evidence and aggressive language

- If for every b, we choose the sender-best among the most informative PBEs, we can get the following (for some parameters):



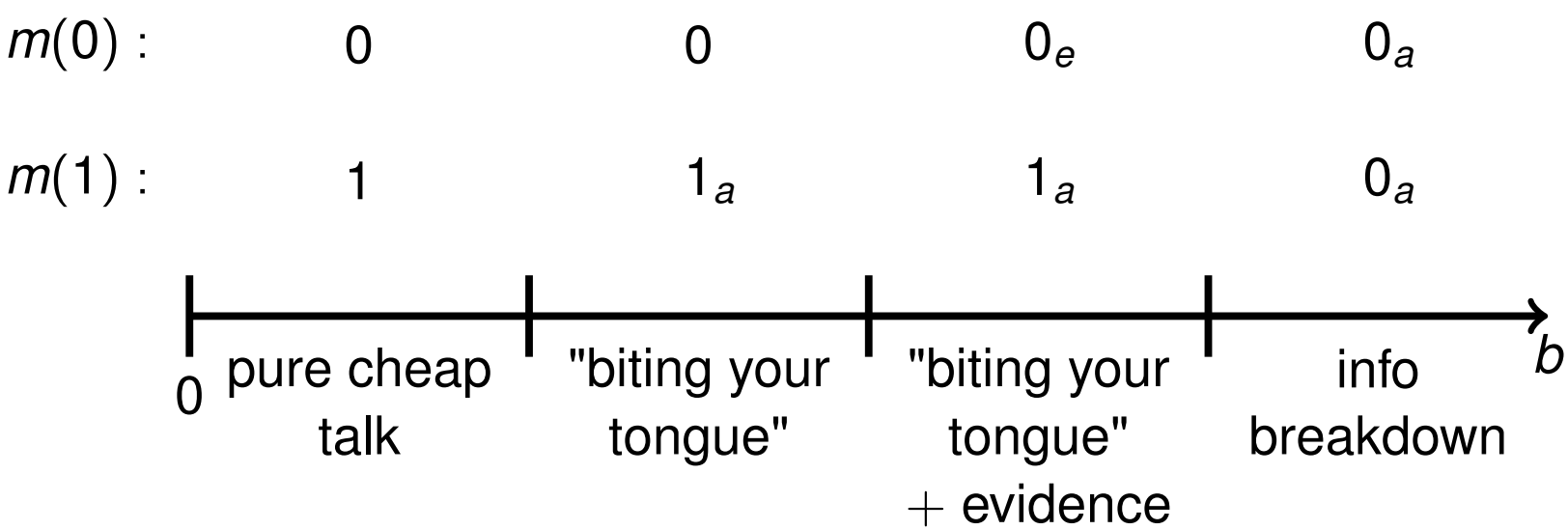
Signaling with evidence and aggressive language

- If for every b, we choose the sender-best among the most informative PBEs, we can get the following (for some parameters):



Signaling with evidence and aggressive language

- If for every b, we choose the sender-best among the most informative PBEs, we can get the following (for some parameters):



- As b increases, more aggressive language and more evidence (but usually not at the same time)

Beliefs

- Which beliefs support these PBE?
- Consider the equilibrium "biting your tongue + evidence", i.e. $m(0) = 0_e$ and $m(1) = 1_a$
- Equilibrium beliefs are (write $\mu(m)$ for posterior belief that $\theta = 1$):
  - $\boldsymbol{\mu(0_e) = 0}$
  - $\mu(0) = \mu(0_a) = \mu(0_{ea}) = \mu(1) = \mu(1_e) = \boldsymbol{\mu(1_a)} = \mu(1_{ea}) = 1$

Beliefs

- Which beliefs support these PBE?
- Consider the equilibrium "biting your tongue + evidence", i.e. $m(0) = 0_e$ and $m(1) = 1_a$
- Equilibrium beliefs are (write $\mu(m)$ for posterior belief that $\theta = 1$):
  - $\boldsymbol{\mu(0_e) = 0}$
  - $\mu(0) = \mu(0_a) = \mu(0_{ea}) = \mu(1) = \mu(1_e) = \boldsymbol{\mu(1_a)} = \mu(1_{ea}) = 1$
- Most profitable sender deviations (that give us the "band" on the previous slide):
  - $m(0) = 0$ instead of $m(0) = 0_e$ ($\Rightarrow$ b has to be large enough)
  - $m(1) = 0_e$ instead of $m(1) = 1_a$ ($\Rightarrow$ b has to be small enough)

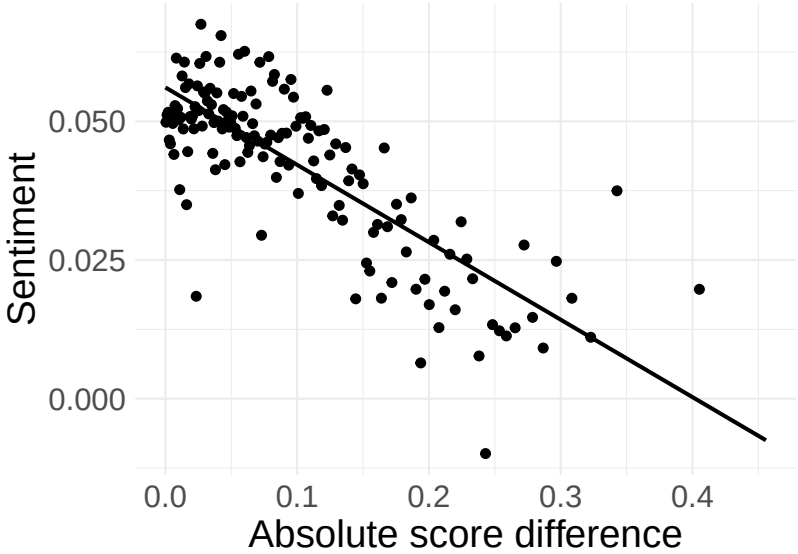
Outline

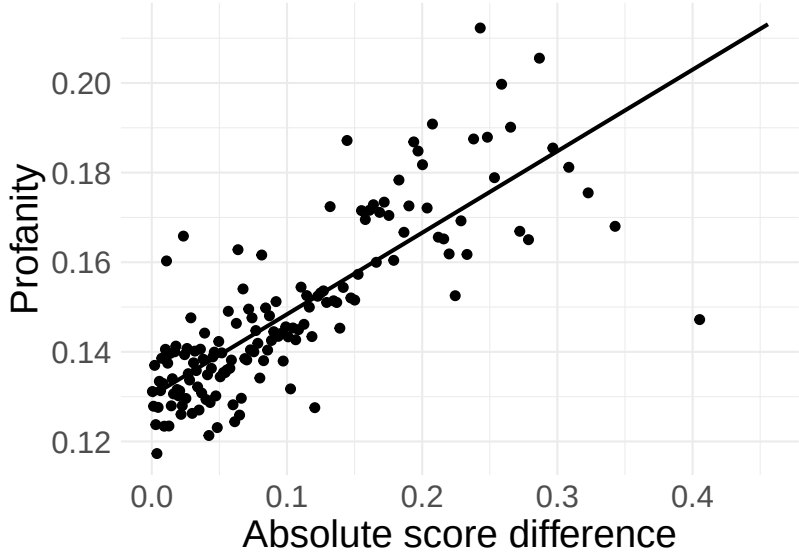
Model

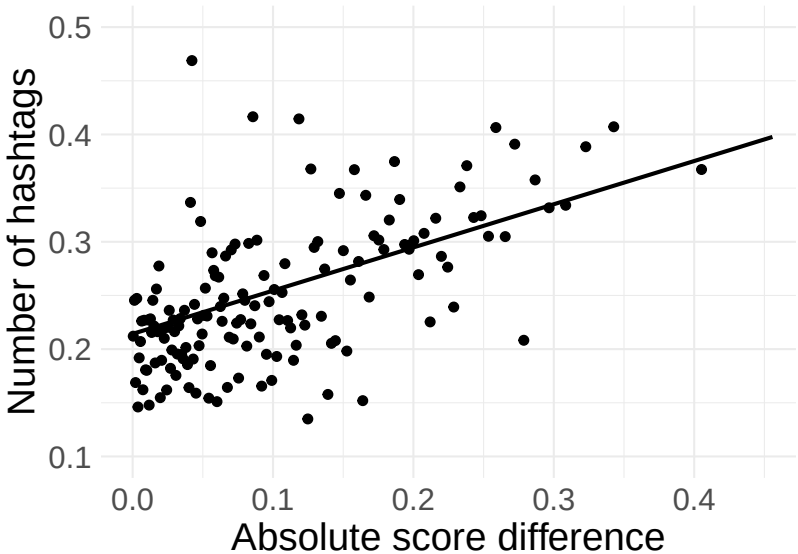
Analysis

Empirical evidence on what model predicts

Larger ideological distance ⇒ Negative language, more profanity, more hashtags







(One dot = 1000 interactions)

Larger ideological distance $\Rightarrow$ More complex language, longer tweets, more links, more pictures

- Fixed-effects OLS:

$$\text{property}_i = \beta \left| \text{score}_{S(i)} - \text{score}_{R(i)} \right| + \text{FE}_{S(i)} + \varepsilon_i$$

	nLinks	linkDummy	tweet length	word length	media
	(1)	(2)	(3)	(4)	(5)
absolute score difference	0.021*	0.022**	10.489*	0.220***	0.119***
	(0.008)	(0.008)	(4.743)	(0.030)	(0.018)
sender fixed effects	Yes	Yes	Yes	Yes	Yes
Estimator	OLS	OLS	OLS	OLS	OLS
N	147,634	147,634	147,634	143,595	147,634
R^2	0.408	0.305	0.275	0.148	0.362

No increase in aggressive language in tweets with links

- For those tweets that contain links we see no increase in profanity or hashtag use (and smallery change in emotional tone):

	profanity	sentiment	hashtags
	(1)	(2)	(3)
absolute score difference	0.018	-0.084**	-0.369
	(0.054)	(0.031)	(0.377)
sender fixed effects	Yes	Yes	Yes
Estimator	OLS	OLS	OLS
N	5,307	5,307	5,307
R^2	0.189	0.269	0.690

Conclusion

- A model in which people (i) want to win arguments, (ii) can use costly, non-verifiable evidence, (iii) have direct expressive utility
- Evidence and aggressive language are used as costly signals to transmit information
- Main predictions are consistent with data from Twitter
- Implications:
  - Increasing the cost of using references/arguments/evidence could make more communication possible
  - But effort spent on evidence should be easily observable
  - Censoring aggressive language could make less communication possible