Beliefs, coordination and media focus

Kristoffer Nimark[1] Stefan Pitschner[2]

[1]Cornell University

[2]Universitat Pompeu Fabra

March 2, 2016

Beliefs, Coordination and Media Focus

Significant market events generally occur only if there is similar thinking among large groups of people, and the news media are essential vehicles for the spread of ideas.

Robert Shiller, 2002

The man who buys a newspaper does not know beforehand what will be in the news.

Jacob Marschak, 1960

Coordination, Beliefs and Media Focus

Public information can be disproportionately influential in strategic settings

- Public signals are particularly useful for predicting the actions of other agents

Examples:

- Bank runs, currency attacks and political regime change
- Price setting and productions decisions in macroeconomic models with monopolistic competition

But what do we mean when we say that information is *public*?

Public information and common knowledge

In the literature, **public information** means information that is **common knowledge**

- E.g. Morris and Shin (AER 2002), Angeletos and Pavan (Econometrica 2007), Angeletos, Hellwig and Pavan (Econometrica 2007), Amador and Weill (JPE 2010 and Jet 2012), Cespa and Vives (REStud 2012), Edmond (REStud 2013), Hellwig and Veldkamp (REStud 2009).

But that information is common knowledge is a much stronger assumption than the everyday meaning of **publicly available**

The editorial function of news media

Not all publicly available information is common knowledge

- Not all information that is publicly available is observed by everybody
- ...and not all information that is observed by everybody is known to be observed by everybody
- ...and so on.

To understand what determines to what degree information about an event becomes common among agents we need to understand the editorial function of news media.

The plan

I. Stylized facts about news coverage from a statistical topic model

- Different newspapers specialize in different topic
- Major events shift news focus and increase the homogeneity of news across outlets

II. Analyze role of editorial function in a simple beauty contest model

- Heterogenous agents rely on specialized information providers to monitor the world on their behalf
- What agents get information about depends on what has happened and make agents respond stronger to events that they can infer are more widely reported
- Editorial function of news media introduces correlation in actions

Measuring News Coverage

Measuring News Coverage using the LDA

Latent Dirichlet Allocation (LDA) can be used to extract topics from text data

- Originally appeared in Blei, Ng and Jordan (2003)

Statistical topic classification

- A topic is defined by a frequency distribution of words
- Documents probabilistically belong to every topic

Input from researcher:

- Text corpus partitioned into documents
- Number of topics

Advantages:

- Objective and easy to replicate
- Naturally measures importance of topics

Understanding the LDA topic model

A text corpus can be thought of as having been generated by the following steps:

1. For each document, draw topic weights from the distribution over topics.
2. For each word in the document, draw a topic from the document specific distribution of topics.
3. Draw a word from the distribution over the vocabulary defined by the topic drawn in step 2.

The text corpus can be viewed *as if* it was generated by repeating the steps 1-3 for each word in each document.

Estimating the LDA model

The probability of a specific text corpora being generated is described by the distribution

$$p(\beta, \theta, z, w) = \prod_{i=1}^{K} p(\beta_i) \prod_{d=1}^{D} p(\theta_d) \left(\prod_{n=1}^{N} p(z_{d,n} \mid \theta_d)\, p(w_{d,n} \mid \beta_{1:K}, z_{d,n}) \right)$$

where β, θ and z are unobserved parameters and w is a vector space representation of the text corpus.

We want to form a posterior distribution for the latent parameters conditional on the observed text corpus

$$p(\beta, \theta, z \mid w) = \frac{p(\beta, \theta, z, w)}{p(w)}$$

We use Collapsed Gibbs Sampling algorithm of Griffiths and Steyvers (2004)

The news data

The News Data

Use text from Dow Jones Factiva data base

- Contains historical content from more than 30,000 news papers, wire services and online sources beginning in 1970.

Extract text snippets from front page articles

- Focus on events considered to be most newsworthy by individual papers

The sample is two 90-day periods that we knew a priori include major events

- September 11 terrorist attacks
- Lehman Brothers Bankruptcy

Topic classification applied across both periods

- Estimating single model on both periods allows for “timeless” news topics.

The number of topics is set to 10

Newspaper sources

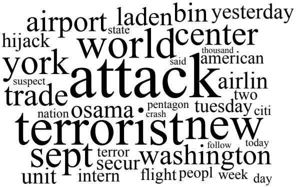
Newspaper Full Name	Short Name	Newspaper Full Name	Short Name
Atlanta Journal	AJ	The Las Vegas Review-Journal	LVR
Charleston Gazette	CG	The New York Times	NYT
Pittsburgh Post-Gazette	PPG	The Pantagraph	PG
Portland Press Herald	PPH	The Philadelphia Inquirer	PI
Sarasota Herald-Tribune	SHT	The Wall Street Journal	WSJ
St. Louis Post-Dispatch	SLP	The Washington Post	WP
Telegram & Gazette Worcester	TGW	USA Today	UT
The Boston Globe	BG	Winston-Salem Journal	WiSJ
The Evansville Courier	EC		

The estimated news topics

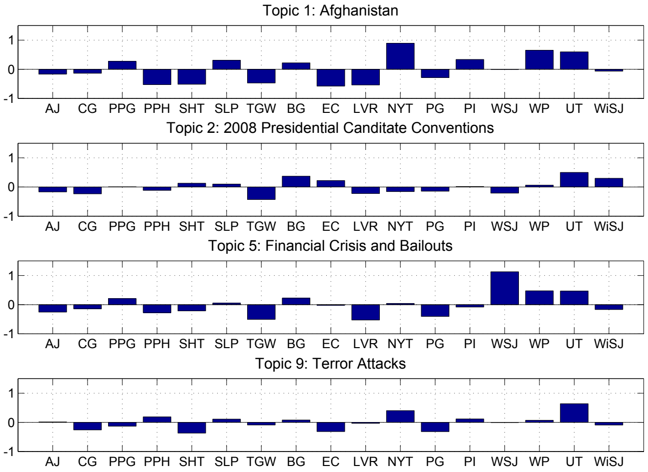
LDA topics classification

Topic	Words with the highest assigned probabilities (in descending order)
1	presid bush afghanistan washington unit today state militari taliban said
2	john democrat obama republican mccain presidenti campaign barack sen candid
3	school citi new counti student high year univers worcest state
4	year two old ago day like today aug just bank
5	financi washington bush billion presid hous plan market bank wall
6	year state million new cut percent month rate price compani
7	mail west state daili virginia staff report get new work
8	state yesterday offici anthrax feder court said offic investig washington
9	attack terrorist new world sept york center trade airport washington
10	year polic old said man review offic counti two journal

Topics 1,2,5 and 9 as word clouds



Specialization of newspapers



Two measures of news coverage over time

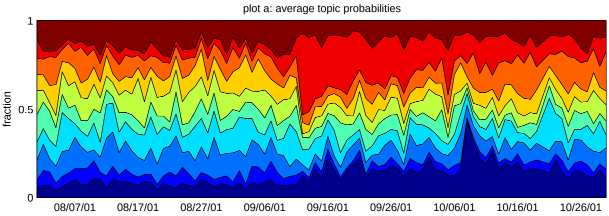
1. **Fraction of total news** devoted to topic k on day t

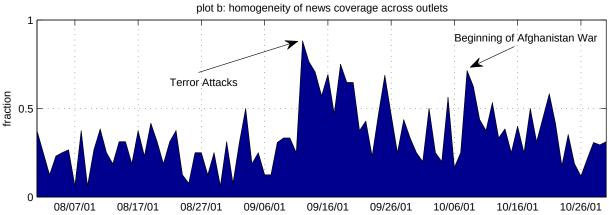
$$F_{t,k} \equiv \frac{\sum_d \theta_{t,d,k}}{D_t}$$

2. **Homogeneity** of news coverage

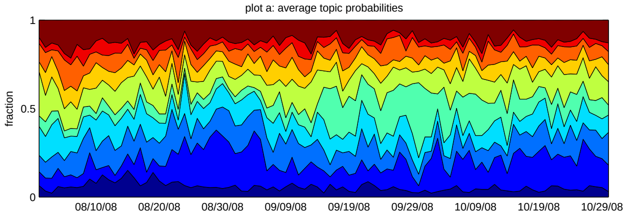
$$H_t \equiv \frac{\sum_m \mathcal{I}\left(\arg\max_k F_{t,m,k} = \arg\max_k F_{t,k}\right)}{M}$$

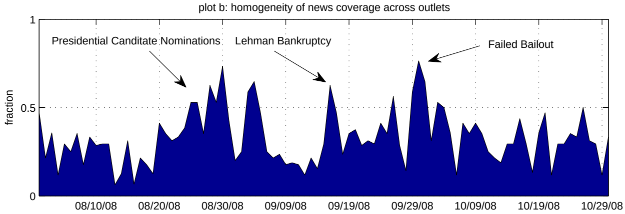
Editorial decisions around 9/11





Editorial decisions around Lehman bankruptcy





Measuring & modeling the editorial function of news media

Newspapers provide specialized content

- Different papers provide specialized content and tend to cover different topics

Major events increase homogeneity of news coverage

- The September 11 terrorist attacks, the 2008 political party conventions, the Lehman bankruptcy and the failed bailout package received large fractions of total news coverage and made news coverage more homogenous across outlets

But what is about an event that makes it *major*?

A beauty contest model with news selection

A beauty contest model with news selection

The model is an abstract coordination game in the spirit of Morris and Shin (2002).

Two essential differences relative to existing models:

1. Agents have heterogeneous interests
2. Agents delegate the information choice to information providers that can monitor more events than they can report

The model will incorporate these features in an as simple set-up as possible

Basic set-up

Two agents (information consumers), Alice and Bob.

Two *potential stories*, $X_a, X_b \in \mathcal{X}$.

- A *potential story* $X_i : i \in \{a, b\}$ is a random variable
- An *event* x_i is a particular realization of X_i.

Two information providers

- Provide specialized content to their respective reader

Information consumers with heterogeneous interests

Heterogeneity is introduced via the agents' utility functions

- Alice's utility function is given by

$$U_a = -(1-\lambda)(y_a - x_a)^2 - \lambda (y_a - y_b)^2$$

  where y_b is the action taken by Bob.

- Alice's optimal action is given by

$$y_a = (1-\lambda) E_a[x_a] + \lambda E_a[y_b]$$

- Symmetric expressions hold for Bob

$$y_b = (1-\lambda) E_b[x_b] + \lambda E_b[y_a]$$

Information providers

Two information providers: Paper A and Paper B

- Alice reads paper A, Bob reads paper B.
- Each paper's news selection maximizes the expected utility of its reader
- Each paper observe the entire realized state of the world but must choose only one story to report
- The chosen story is reported truthfully

Information providers

Information providers are defined by a news selection function

Definition: *The news selection function* $\mathcal{S}_i : \mathcal{X} \times \mathcal{X} \rightarrow \{0, 1\}$ *is an indicator function that takes the value* 1 *when paper* i *reports the realized value of* X_i *and* 0 *otherwise.*

The news selection functions are determined by

$$\mathcal{S}_i\left(x_i, x_j\right) = \arg\max_{\mathcal{S}_i} E\left[U_i\left(\mathcal{S}_i, \mathcal{S}_j, U_j\right)\right]$$

That is, Paper A chooses a selection function in order to maximize the expected utility of Alice, Paper B does the same for Bob.

Simple discrete state space example

Discrete state space

The potential stories X_a and X_b can take the values -1, 0, or 1 with probabilities given by

$$p_i(-1) = \frac{1}{4}, p_i(0) = \frac{1}{2}, p_i(1) = \frac{1}{4} : i \in \{a, b\}$$

The two potential stories are mutually independent

$$p_i(x_i \mid x_j) = p_i(x_i) : i \neq j, \in i, j \{a, b\}$$

Neither symmetry nor the independence of the distributions for X_a and X_b are necessary.

Equilibrium news selection

Strategic motive in Alice and Bob's actions $\Rightarrow$ news selection of Paper A depends on news selection of Paper B (and vice versa)

- Nash equilibrium news selection

Equilibrium news selection is a fixed point:

- Optimal selection depends on actions and actions depend on news selection functions

Strategy:

- Posit candidate selection functions
- Derive optimal actions given news selections functions
- Verify that news selection functions are optimal given Alice and Bob's actions

Equilibrium news selection functions

Table 1: News selection functions

No strategic motive

Paper A

	$X_a = -1$	$X_a = 0$	$X_a = 1$
$X_b = -1$	A	A	A
$X_b = 0$	A	A	A
$X_b = 1$	A	A	A

Paper B

	$X_a = -1$	$X_a = 0$	$X_a = 1$
$X_b = -1$	B	B	B
$X_b = 0$	B	B	B
$X_b = 1$	B	B	B

Strategic complementarities

Paper A

	$X_a = -1$	$X_a = 0$	$X_a = 1$
$X_b = -1$	A	B	A
$X_b = 0$	A	A	A
$X_b = 1$	A	B	A

Paper B

	$X_a = -1$	$X_a = 0$	$X_a = 1$
$X_b = -1$	B	B	B
$X_b = 0$	A	B	A
$X_b = 1$	B	B	B

Equilibrium news selection functions

Table 1: News selection functions

	Paper A				**Paper B**		
	No strategic motive						
	$X_a = -1$	$X_a = 0$	$X_a = 1$		$X_a = -1$	$X_a = 0$	$X_a = 1$
$X_b = -1$	A	A	A	$X_b = -1$	B	B	B
$X_b = 0$	A	A	A	$X_b = 0$	B	B	B
$X_b = 1$	A	A	A	$X_b = 1$	B	B	B
	Strategic complementarities						
	$X_a = -1$	$X_a = 0$	$X_a = 1$		$X_a = -1$	$X_a = 0$	$X_a = 1$
$X_b = -1$	A	B	A	$X_b = -1$	B	B	B
$X_b = 0$	A	A	A	$X_b = 0$	A	B	A
$X_b = 1$	A	B	A	$X_b = 1$	B	B	B

News selection functions and beliefs

The news selection functions affect agents' beliefs also about the story that is not reported

Consider the state $(0, 1)$

- Alice knows that $X_b = 1$
- But she also knows that $X_a = 0$ since in the states $(1, 1)$ and $(-1, 1)$ Alice observes X_a

Table 1: News selection functions

Strategic complementarities

Paper A

	$X_a = -1$	$X_a = 0$	$X_a = 1$
$X_b = -1$	A	B	A
$X_b = 0$	A	A	A
$X_b = 1$	A	B	A

Paper B

	$X_a = -1$	$X_a = 0$	$X_a = 1$
$X_b = -1$	B	B	B
$X_b = 0$	A	B	A
$X_b = 1$	B	B	B

News selection functions and beliefs

The news selection functions affect agents' beliefs also about the story that is not reported

Consider the state $(0, 1)$

- Alice knows that $X_b = 1$
- But she also knows that $X_a = 0$ since in the states $(1, 1)$ and $(-1, 1)$ Alice observes X_a

Table 1: News selection functions

Strategic complementarities

Paper A

	$X_a = -1$	$X_a = 0$	$X_a = 1$
$X_b = -1$	A	B	A
$X_b = 0$	A	A	A
$X_b = 1$	A	B	A

Paper B

	$X_a = -1$	$X_a = 0$	$X_a = 1$
$X_b = -1$	B	B	B
$X_b = 0$	A	B	A
$X_b = 1$	B	B	B

News selection functions and beliefs

Proposition: Posterior beliefs about the unreported story X_j coincides with the prior distribution $p\left(x_j\right)$, i.e.

$$p\left(x_j \mid \mathcal{S}_i = 1, x_i\right) = p\left(x_j\right) \tag{1}$$

only if the probability of reporting x_i is conditionally independent of x_j

$$p\left(\mathcal{S}_i = 1 \mid x_i\right) = p\left(\mathcal{S}_i = 1 \mid x_j, x_i\right).$$

Proof: By Bayes' rule

$$p\left(x_j \mid \mathcal{S}_i = 1, x_i\right) = \frac{p\left(\mathcal{S}_i = 1 \mid x_j, x_i\right)}{p\left(\mathcal{S}_i = 1 \mid x_i\right)} p\left(x_j\right)$$

so that (1) holds only if

$$\frac{p\left(\mathcal{S}_i = 1 \mid x_j, x_i\right)}{p\left(\mathcal{S}_i = 1 \mid x_i\right)} = 1.$$

News selection functions and higher order beliefs

Some events are observed by both Alice and Bob, yet the event may not be common knowledge

Consider the state $(1, 0)$: Both Alice and Bob knows that $X_a = 1$.

- Bob can infer with certainty that Alice knows this too.
- Alice infer that Bob knows that $X_a = 1$ with probability $\frac{1}{2}$.

Table 1: News selection functions

Strategic complementarities

Paper A

	$X_a = -1$	$X_a = 0$	$X_a = 1$
$X_b = -1$	A	B	A
$X_b = 0$	A	A	A
$X_b = 1$	A	B	A

Paper B

	$X_a = -1$	$X_a = 0$	$X_a = 1$
$X_b = -1$	B	B	B
$X_b = 0$	A	B	A
$X_b = 1$	B	B	B

News selection functions and higher order beliefs

Some events are observed by both Alice and Bob, yet the event may not be common knowledge

Consider the state $(1, 0)$: Both Alice and Bob knows that $X_a = 1$.

- Bob can infer with certainty that Alice knows this too.
- Alice infer that Bob knows that $X_a = 1$ with probability $\frac{1}{2}$.

Table 1: News selection functions

Strategic complementarities

Paper A	$X_a = -1$	$X_a = 0$	$X_a = 1$	Paper B	$X_a = -1$	$X_a = 0$	$X_a = 1$
$X_b = -1$	A	B	A	$X_b = -1$	B	B	B
$X_b = 0$	A	A	A	$X_b = 0$	A	B	A
$X_b = 1$	A	B	A	$X_b = 1$	B	B	B

Equilibrium actions

Alice's action when she observes X_a

$$y_a\left(x_a\right) = \left(1-\lambda\right)x_a + \lambda p\left(S_b = 0 \mid S_a = 1, x_a\right) y_b\left(x_a\right)$$

Bob's action when he observes X_a

$$y_b\left(x_a\right) = \lambda y_a\left(x_a\right)$$

Using symmetry between agents and simplifying we get

$$y_a\left(x_a\right) = \frac{\left(1-\lambda\right)}{1-\frac{1}{2}\lambda^2}x_a, \qquad y_b\left(x_a\right) = \lambda\frac{\left(1-\lambda\right)}{1-\frac{1}{2}\lambda^2}x_a$$

The strength of the response of agents depends on $p\left(S_b = 0 \mid S_a = 1, x_a\right)$.

Delegated news selection and correlated actions

Alternative benchmark model: Optimal actions with ex ante signal choice

Agents subject to same constraint on number of stories but must choose *ex ante* which story to read about

- Alice will choose to always observe X_a
- Bob will choose to always observe X_b.

Since

$$E\left[x_i \mid x_j\right] = 0 : i \neq j$$

the optimal action is given by

$$y_i = (1 - \lambda)\, x_i : i \in a, b$$

Alice and Bob's actions are uncorrelated if X_a and X_b are independent

News selection and correlation of actions

Direct computation of the correlation of Alice and Bob's actions gives

$$\begin{aligned}\frac{\sum p\left(x_a, x_b\right) y_a(x_a, x_b) y_b(x_a, x_b)}{\sqrt{var\left(y_a\right)}\sqrt{var\left(y_b\right)}} &= 2\lambda\frac{\left(1-\lambda\right)^2}{\left(2-\lambda^2\right)^2} var\left(y_i\right)^{-1} \\ &> 0\end{aligned}$$

- The terms in the sum associated with the states $\left(0, 1\right), \left(0, -1\right), \left(1, 0\right)$ and $\left(-1, 0\right)$ are all positive with delegated news selection
- The same terms are zero with ex ante information choice

Extreme events

Extreme events and approximate common knowledge

The discrete, low dimensional set up does not lend itself to study large magnitude, or extreme, events

Continuous distributions of events allow us to think of how the magnitude of an event affect beliefs and actions

- $X_i \sim N(0, \frac{1}{3})$
- News selection parameterized a

$$\mathcal{S}_i = \begin{cases} 1 \text{ if } \ |x_i| \geq \alpha + \beta\,|x_j|^{\gamma} \\ 0 \text{ otherwise} \end{cases}$$

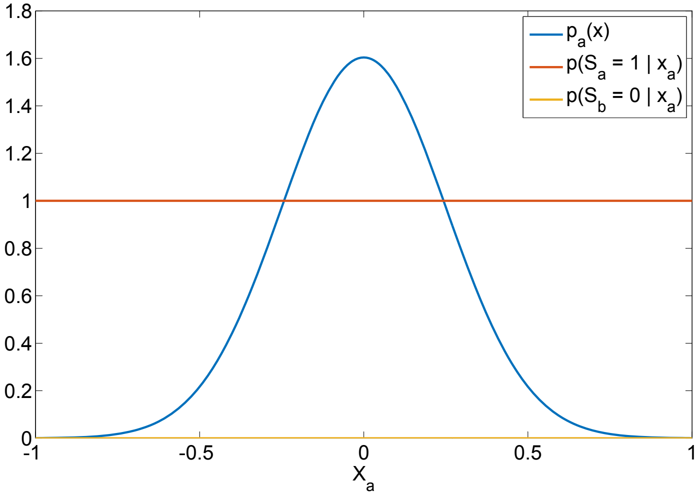
- Optimal actions

$$y_i\,(x_i) = \frac{(1-\lambda)}{1-\lambda^2 p\,(\mathcal{S}_j = 0 \mid x_i, \mathcal{S}_i = 1)} x_i$$

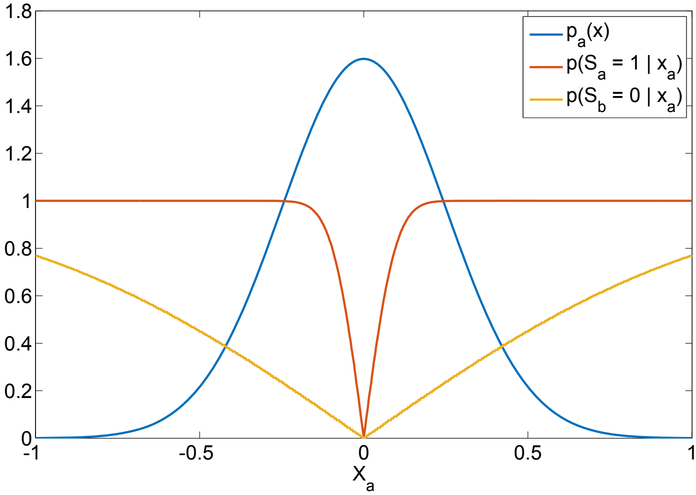
and

$$y_i\,(x_j) = \lambda \frac{(1-\lambda)}{1-\lambda^2 p\,(\mathcal{S}_i = 0 \mid x_j, \mathcal{S}_j = 1)} x_j$$

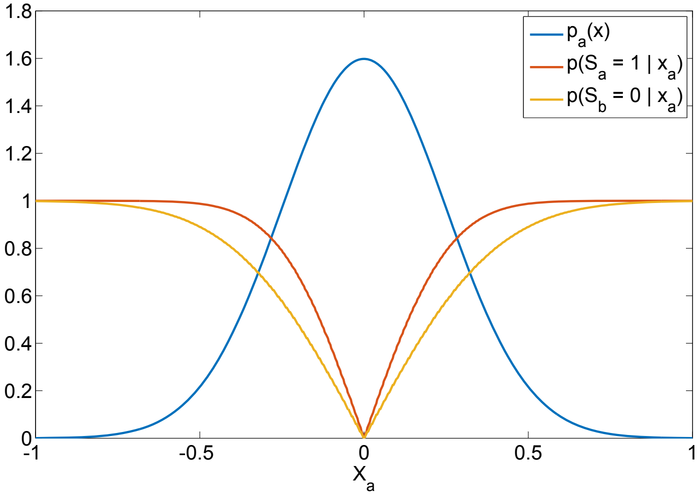
Extreme events and common knowledge
$\lambda = 0$



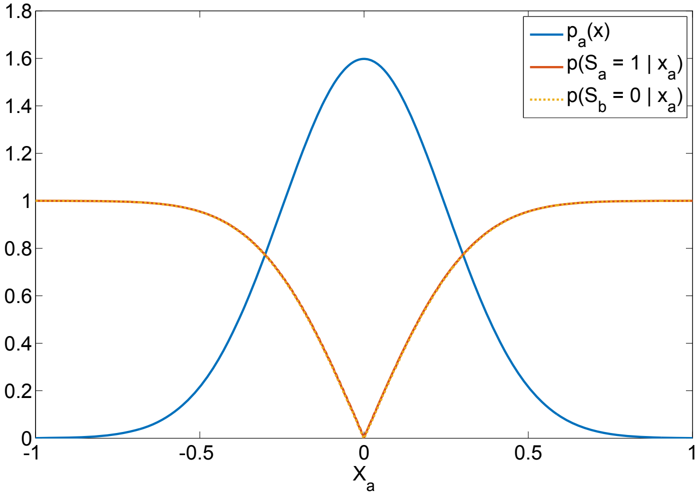
Extreme events and common knowledge
$\lambda = 0.3$



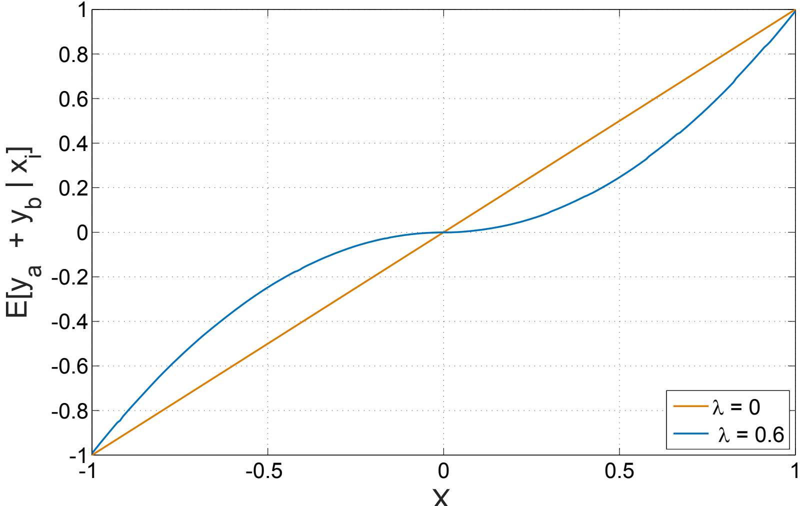
Extreme events and common knowledge
$\lambda = 0.45$



Extreme events and common knowledge
$\lambda = 0.6$



Expected aggregate action



Conclusions

Documented stylized facts about news coverage

- Different newspapers provide specialized content and tend to cover different topics to different degrees
- Major events increase homogeneity of news coverage.

Formalized the editorial service provided by news media

- Differ from ex ante information choice literature, e.g. Sims (2003), Mackowiak and Wiederholt (2009, 2010), Alvarez, Lippi and Paciello (2011), Matejka (forthcoming), Matejka and McKay (2015), Stevens (2014),Grossman and Stiglitz (1980), Veldkamp (2006a,2006b), Van Nieuwerburgh and Veldkamp (2009, 2010)

Implications

- Editorial function induces correlation in agents' actions
- The strength of agents' responses depends on the degree to which knowledge about the event is common

More Conclusions!

Strong assumptions regarding benevolence of news media

- Reports events with perfect accuracy
- Select stories to maximize utility of reader

As long as bias is systematic and understood by the agents, the mechanism applies.

Change in news focus at lower frequencies