NEWS MEDIA AND DELEGATED INFORMATION CHOICE

KRISTOFFER P. NIMARK AND STEFAN PITSCHNER

Abstract. News media monitor a large number of events on behalf of their audiences but report only those they consider most newsworthy. Analyzing texts from almost 15,000 US newspaper articles, we document that, while different outlets typically emphasize different topics, major events make coverage more homogenous. We analyze the consequences of this type of editorial behavior in a model with delegated information choice that can match the documented facts. In the model, the degree to which information about a given event is common among agents is endogenous. Agents that take strategic decisions react more strongly to information that is closer to common knowledge. The proposed mechanism also induces correlation between different agents’ actions.

1. Introduction

Every day, a very large number of events occur, each of them potentially relevant for the decisions of firms and households. However, no individual firm or household has the resources to observe all of these events. Instead, many rely on news media to monitor the world on their behalf. One important function that news media perform is thus editorial. Among all potential stories that occur, only those that are deemed most newsworthy are reported. In this paper, we analyze how this type of delegated information choice affects the beliefs and strategic interaction of economic agents.

Strategic decisions based on imperfect information are pervasive in economics. For example, producers in oligopolistic markets need to predict the output of their competitors, speculators need to predict whether other speculators plan to attack a currency, and price setters need to predict the pricing decisions of other firms. In such settings, it is well known that public signals are disproportionately influential as they are particularly useful for agents that need to predict the actions of other agents, e.g. Morris and Shin (2002).

Arguably, everything that is reported by news media is public in the sense that it is available to those who care to read it. However, in reality, not all of this information is common knowledge. That is, not all information that is publicly available is also observed by everybody, and not all information that is observed by everybody is also known to be observed by everybody, and so on. In this paper, we argue that understanding the editorial
role played by news media is central to understanding the degree to which knowledge about an event is common among agents.

We begin by estimating a Latent Dirichlet Allocation (LDA) topic model based on texts from almost 15,000 archived newspaper stories from 17 US newspapers. The newspaper stories are from two periods that contained major news events, namely the 90 day period around the 9/11 terrorist attacks on New York and the Pentagon and the 90 day period around the Lehman bankruptcy that marked the start of the more severe phase of the financial crisis. We use the LDA model to document three stylized facts of news coverage. First, different newspapers specialize in different topics. For example, we find that the Wall Street Journal allocated more than twice as much coverage to the financial crisis than the average newspaper. Second, the extent of total news coverage allocated to different topics varies over time and depends on what has happened. The September 11 terrorist attacks, the 2008 political party conventions, the Lehman bankruptcy and the failed bailout package proposed by then Secretary of the Treasury Hank Paulson received a large fraction of the overall news coverage following these events. Third, major events make news coverage more homogenous across different outlets. In the days following the events listed above, a majority of the newspapers in our sample devoted more coverage to them than to any other topic.

Together, these facts suggest that information about major events is closer to common knowledge than information about minor events. In order to analyze how this kind of editorial behaviour of news media affects agents’ beliefs and decisions, we propose a theoretical model with incomplete information that can replicate the stylized facts we document. While our empirical analysis is based on the texts of newspaper articles, the mechanisms that we highlight in the theoretical model apply to all information providers that systematically choose what to report from a large set of events. Our analysis thus encompasses not only print media, but also television and radio news as well as those online information providers that perform an editorial service by presenting a curated selection of events or news stories on their websites.

The model is a two-agent beauty contest game in which each agent’s pay-off depends on two factors: The distance of his action to an agent-specific latent variable, and the distance between his action and that of the other agent. The two agents in our model are thus heterogenous in that their pay-offs depend on different latent variables. Our analysis takes this heterogeneity as given, but one can think of it as reflecting differences in factors such as the agents’ geographical location, their production technologies or their sector membership.

One basic premise of our model is that the dimensionality of the state of the world is too high for individual agents to monitor it on their own. Therefore, they rely on information providers to do so on their behalf. Furthermore, because agents are heterogeneous in terms of what information they find most useful, information providers specialize and cater to their different interests. However, because of a strategic motive, the agents in our model also have an indirect interest in events that are only important for predicting the actions of others. As a result, in some states of the world, all information providers report the same events.

The agents in our model delegate the decision of what to get information about to specialized information providers. These information providers monitor a larger set of events than they eventually end up reporting. Because their reporting decisions depend on the relative newsworthiness of the realized events, what agents get information about depends on what
has happened. The model presented here formalizes this editorial function of news media and thus provides a theory of how and why news media focus changes over time. While the model is abstract, it offers several insights that we argue are general.

One contribution of our paper is to show that a reported news story can be informative about more than the event it actually covers. In fact, we derive formal conditions for when a news report about a specific event also reveals information about events that are not reported. To illustrate this result, consider an example where the heterogeneity in agents’ interests is due to geographical location. What is the inference made by a person who opens a San Francisco newspaper and finds that it only contains stories about New York? If this person knows that the paper always covers all important San Francisco events, the lack of stories on such events reveals to him that none have actually taken place. Therefore, even though he only reads about New York, he also updates his beliefs about San Francisco. Moreover, because this type of information transmission is a direct consequence of the systematic selection of news stories, agents learn something about the unreported event even if the realizations of the reported and unreported events are independent from one another.

The systematic selection of what gets reported also affects the degree to which knowledge about an event is common among agents. In the related existing literature, signals are generally assumed to be either private or common knowledge, e.g. Morris and Shin (2002), Angeletos and Pavan (2007), Angeletos, Hellwig and Pavan (2007), Hellwig and Veldkamp (2009), Amador and Weill (2010, 2012), Cespa and Vives (2012) and Edmond (2013). In our model, information about a particular event is typically neither private nor common knowledge. Instead, the degree to which knowledge about an event is common among agents is endogenous and depends probabilistically on their preferences and the distribution of events. Because news selection is state-dependent, what agents get information about also influences how probable they think it is that other agents read about the same event.

To see how preferences and the distribution of events together determine the degree to which knowledge about an event is common among agents, consider again a person living in San Francisco. If the San Francisco newspaper reports about an event in Manhattan that normally would be of more interest to a reader from New York, the reader in San Francisco can infer that New Yorkers are probably also reading about that event. However, even though both San Franciscans and New Yorkers are reading about the same event, this event may not be common knowledge: While the San Franciscan can be sure that the New Yorker is also reading about the event on Manhattan, the New Yorker cannot draw a corresponding conclusion, because he knows that he has a stronger preference for knowing about events on Manhattan than the person living in San Francisco.

When extreme events such as large terrorist attacks or major financial crises occur, they tend to be reported on the front pages of almost all major newspapers. Our model proposes one mechanism that can explain this behaviour. In the model, individual agents care about the strong actions that other agents will take in response to extreme events even if they do not have a direct interest in the events themselves. Extreme events thus tend to be reported by all information providers.

Information about unlikely but extreme events also tends to be closer to common knowledge. To see the role played by the distributions of events for this result, consider a person
living in New York who reads about a major financial crisis on Wall Street. The San Francisco newspaper will also report the financial crisis on Wall Street unless an almost equally extreme event has also occurred in San Francisco. If such extreme events are very unlikely, the New Yorker can then infer that a person living in San Francisco is almost surely also a reading about the financial crisis.

As an alternative to the motivation we outline here, it is also possible that the increase in homogeneity of news coverage after extreme events is primarily driven by a common desire to know about such large events. However, as long as there is a well-understood underlying reason for why some events are widely reported, such events will be closer to common knowledge among agents. If strategic considerations are important for agents’ optimal decisions, the main results of the paper are then largely unchanged, even if the strategic motive by agents are not what fundamentally motivates the observed editorial behavior.

The fact that the agents in our model cannot directly observe the entire state of world makes them similar to the rationally inattentive agents in Sims (2003), Mackowiak and Wiederholt (2009, 2010), Alvarez, Lippi and Paciello (2011), Matejka (forthcoming), Matejka and McKay (2015) and Stevens (2014). Relative to this literature, the distinctive feature of our set up is that the information providers are constrained to report a discrete subset of all realized events and that the selected events are reported accurately. This reflects a view that news stories are to some degree indivisible and that some facts may be perfectly known by some agents. Agents’ posterior beliefs reflect both the information in the actual reports as well as the information about unreported events implicit in this selection. However, unlike the rational inattention literature, we do restrict how large the entropy reduction implied by observing the signals may be.

Our set-up also differs from the early costly information literature pioneered by Grossman and Stiglitz (1980) and the work following this tradition, e.g. Veldkamp (2006a, 2006b) and Van Nieuwerburgh and Veldkamp (2009, 2010). In this literature, as well as in linearized rational inattention models with quadratic objectives and Gaussian shocks, agents make ex ante information acquisition decisions based on the expected usefulness of a particular signal. In the model presented below, information providers monitor a larger set of events and decide what to report ex post, i.e. after the state of the world has realized. This implies that what has happened, affects what agents get information about.1 Our setup formalizes the editorial function of news media and captures the fact that the decision to

1In the most general formulation of a rational inattention problem, e.g. Sims (2003), the only constraint imposed on agents’ beliefs is on the reduction in entropy implied by going from agents’ prior beliefs to their posterior beliefs. It is thus in principle possible to conceive of a rational inattention model where agents choose to observe signals that are completely uninformative about a subset of latent variables in some states of the world. It could be said that in such a model, too, what agents get information about depends on ex post outcomes. However, the implications of the state dependence of what gets reported in our model is quite different from such a set up: As we show in Proposition 1 below, in our set-up, a signal will only be completely uninformative about the non-reported events if the selection of what is reported is independent of the realized value of the non-reported event. That is, in our set-up, the relationship between what is reported and what agents actually change their beliefs about is more intricate and depends on the news selection functions of the information providers.
acquire information in many instances is a decision about which information provider to use, rather than a decision about what variable or event to get information about.²

Of the existing literature, the papers by Veldkamp (2006b) and Hellwig and Veldkamp (2009) study questions most closely related to those addressed here. Veldkamp (2006b) presents a model in which ex ante identical agents choose asset portfolios and signals simultaneously. If different agents hold different portfolios, they prefer to observe signals about the pay-offs of different assets. However, due to increasing returns to scale in information production, agents tend to purchase similar signals and hold similar portfolios in equilibrium, which can generate *media frenzies* characterized by simultaneous increases in the price and media coverage of a stock. In contrast, our paper studies how widely reported information about an event becomes among agents that have intrinsically different interests.

Hellwig and Veldkamp (2009) analyze a set-up in which agents can choose whether to acquire private or public signals about a single latent variable of common interest. In their set up, public and private signals are equally useful for predicting the latent fundamental. The choice between the two types of signals is then a choice between prediction errors that are correlated with those of other agents and predictions errors that are not. In such a setting, information acquisition inherits the strategic properties of the coordination game, and agents want to observe the signals that other agents observe only if their actions are strategic complements. In our model, the information choice determines what agents get information *about*, and in some states of the world, agents prefer to know what others know regardless of whether their actions are complements or substitutes.

Gentzkow and Shapiro (2006, 2008), like our paper, study the editorial function of news media, but primarily focus on identifying and analyzing the causes and consequences of ideologically slanted reporting. The political science literature has also studied the role of news journalists and newspaper editors as “gatekeepers” that decide what information gets reported, e.g. Soroka (2006, 2012) and Soroka, Stecula, Wlezien (2014). This literature, too, focuses primarily on documenting and analyzing ideologically or politically biased reporting. While we abstract from such biases, the mechanisms we illustrate apply also to environments in which they are present.

Another paper related to the present study is Nimark (2014). That paper analyzes a set up in which a single information provider is more likely to report unusual events. However, it does not model the decision of the information provider explicitly, nor does it incorporate heterogeneity in agents’ interests. In contrast to the model presented below, the set up in Nimark (2014) can thus not be used to study how agents’ preferences and the distribution of events interact to determine the degree to which knowledge about an event is common among agents.

In the next section we document several stylized facts about news coverage using a statistical topic model applied to US newspaper data. Section 3 presents a theoretical model that can replicate the documented facts, and we derive formal results based on discrete distributions of events in Section 4. Section 5 we extend the analysis to continuous distributions and Section 6 concludes.

²The motive of our agents is well-captured by Marschak (1959) who writes that *"The man who buys a newspaper does not know beforehand what will be in the news. He acquires access to potential messages belonging to a set called news."*
2. Three Stylized Facts of News Coverage

In this section, we estimate a Latent Dirichlet Allocation (LDA) topic model based on texts from a large number of archived newspaper articles. We document three stylized facts about news coverage. In particular, we show that different newspapers specialize in different topics, that the total news coverage devoted to different topics depends on what has happened, and that major events make news coverage more homogeneous across papers.

2.1. The News Data. Our empirical analysis focuses on two 3-month periods that contained several major news events. The first period covers the months August to October of 2001 and includes the terrorist attacks on the World Trade center and the Pentagon on September 11. The second period runs from August to October of 2008 and includes the Lehman Brothers bankruptcy that triggered the most severe phase of the financial crisis.

The data we use are parts of news articles obtained from the Dow Jones Factiva database. Factiva contains historical content from more than 30,000 news papers, wire services and online sources from around the world beginning in 1970. We exclude content from wire services since their main audiences are other news organizations. We also limit our data set to articles that appeared either on front pages of US newspapers or on the first pages of their general interest sections.

In total, we obtain data from 14,817 front page articles reported by 17 different US newspapers. The selection of newspapers includes all US newspapers for which we are able to reliably identify the stories that appeared on their front pages or the first pages of their general interest sections. From each of these articles we use a text snippet that typically comprises its first one or two sentences. Table 1 contains an overview of the newspapers in our database as well as corresponding short names that we use in the analysis below. To illustrate the type of information that the text snippets contain, Table 2 shows a number of examples.

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

Notes: The table shows the full names of the newspapers whose front-page articles are in our text corpus. It also shows corresponding short names used in the empirical analysis below. Newspapers that have changed their names over time or have merged are combined into one entry.
Table 2. Sample Text Snippets of Newspaper Articles in the Database

<table>
<thead>
<tr>
<th>Text Snippet</th>
<th>Newspaper</th>
<th>Publication Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>“An 18-year-old student who wounded five people at his suburban San Diego high school earlier this year committed suicide, hanging himself with a sheet in his jail cell. The student, Jason Anthony Hoffman, pleaded guilty last month in the ...”</td>
<td>The New York Times</td>
<td>2001/10/31</td>
</tr>
<tr>
<td>“Passengers returned to US airports in increasing numbers yesterday to find long lines, layers of new security and limited service. But many travelers were able to reach their destinations as more than a third of the usual number of ...”</td>
<td>The Washington Post</td>
<td>2001/09/15</td>
</tr>
<tr>
<td>“A day after dividing their votes on a failed proposal for a 700 billion Wall Street bailout, Maine’s two US House members agreed Tuesday that it’s vital for lawmakers to pass a relief bill for credit markets.”</td>
<td>Portland Press Herald</td>
<td>2008/10/01</td>
</tr>
<tr>
<td>“In a case that could have dramatic consequences for school districts and towns across Pennsylvania, the state Supreme Court will hear arguments today on the constitutionality of the commonwealth’s property-tax system, which raises more ... ”</td>
<td>The Philadelphia Inquirer</td>
<td>2008/09/10</td>
</tr>
</tbody>
</table>

Notes: The table shows examples of the text snippets used to estimate the LDA topic model below. The text snippets were extracted from the Dow-Jones Factiva database. The dates shown are those on which the articles were originally published in the print-editions of the respective newspapers.

2.2. Latent Dirichlet Allocation. To extract topics from our text corpus, we estimate a Latent Dirichlet Allocation (LDA) topic model. Introduced in Blei et al (2003), the LDA model is one of the most-widely applied tools in machine learning and natural language processing. A topic is defined by a frequency distribution of words and the topics are estimated from the text corpus. Variants of the LDA model have been used, for example, to identify scientific topics (Griffiths and Steyvers, 2004) and to classify micro blogs (Ramage et al, 2010). The first application to economics or finance that we are aware of is Mahajan,
Dey and Haque (2008), who used it to classify financial news articles. More recently it has also been used by Bao and Datta (2014) to discover risk-factors disclosed in annual corporate filings. Furthermore, Fligstein, Brundage and Schultz (2014) as well as Hansen, McMahon and Prat (2015) have used LDA models to analyze FOMC transcripts.

Using an LDA model allows us to discover and quantify the topics of a very large number of news texts without relying on manual classifications or pre-defined categories. Moreover, because LDA defines articles as mixtures of different topics, it can accommodate the fact that many news stories talk about more than one specific issue. For example, it can capture that an article about a government bailout package may discuss both politics and financial markets.

The main parameter of choice researchers need to set before estimating an LDA model is the number of topics. Once this number has been set, the actual topics are formed endogenously and are thus outputs of the estimated model. Relative to approaches that use word counts to measure news coverage, e.g. Baker, Bloom and Davis (2013), the LDA does therefore not require researchers to pre-specify words or topics of interest. Another desirable property of LDA is that it captures not only changes in the importance of a topic over time, but also how important that topic is in an absolute sense. Both the limited number of discretionary decisions required for the LDA estimation and the fact that topics emerge from the analysis without having to pre-define them are particularly attractive for our application. These properties allow us to analyze the documents in our database in an objective and replicable manner.

The text data or corpus used for estimating an LDA topic model is described by a vocabulary, which is a list of all words that it contains, and documents, which are partitions of the text corpus. In our case, each text snippet from a news article is one document, and all text snippets together form the corpus. Generally speaking, an LDA topic model can be thought of as a latent structure that could have generated the observed text corpus following probabilistic rules. It is parameterized by (i) a distribution over topics that determine the probability that a document belongs to a topic and (ii) a distribution over the words in the vocabulary that defines each of the topics. In the LDA framework, each document in a corpus can be thought of as having been generated by the following steps:

(1) Draw a set of topic weights from the corpus-specific distribution over topics.

(2) Draw \( N \) topics from this document-specific topic distribution, with \( N \) being the number of words in the document.

(3) Draw one word from each of these \( N \) topics.

To describe the LDA model more formally, we index topics by \( k \in \{1, 2, ..., K\} \), documents by \( d \in \{1, 2, ..., D\} \), words in the vocabulary by \( v \in \{1, 2, ..., V\} \), and words in a document by \( n \in \{1, 2, ..., N\} \). The probability of a specific text corpus being generated is then given by the probability density function

\[
p(\beta, \theta, z, w) = \prod_{i=1}^{K} p(\beta_i) \prod_{d=1}^{D} p(\theta_d) \left( \prod_{n=1}^{N} p(z_{d,n} | \theta_d) p(w_{d,n} | \beta_{1:K}, z_{d,n}) \right) \tag{2.1}
\]

where \( \beta, \theta \) and \( z \) are unobserved parameters. The rows of the \( K \times V \) matrix \( \beta \) contain the word distribution \( \beta_k \) for topic \( k \), the columns of the \( K \times D \) matrix \( \theta \) contain the topic
proportions $\theta_d$ of document $d$ so that $\theta_{k,d}$ is the proportion of words in document $d$ drawn from topic $k$. The topics assignment of document $d$ is $z_d$ so that word $n$ in topic $d$ is drawn from topic $z_{d,n}$. The density (2.1) depends on the text corpus through the matrix $w$, defined so that the words observed in document $d$ is the vector $w_d$ and $w_{d,n}$ is word $n$ in document $d$.

There are two underlying properties that are important for understanding how the LDA is used to extract topics from the text corpus. First, LDA is a mixed membership model. This implies that each document may belong to different topics to different degrees. As discussed above, this is helpful for our application as it allows newspaper articles to be treated as belonging to several topics at the same time. For example, an article could be classified as belonging to the topics financial crisis and congressional politics with topics weight 0.4 and 0.6, respectively. Second, the order and grammatical structure of words within documents is assumed to be irrelevant. This so-called “bag-of-words” assumption simplifies the latent probabilistic structure of the text corpus while retaining the information relevant for discovering the topics that the corpus contains.

In order to apply Latent Dirichlet Allocation to an observed text corpus, the generative process described above needs to be inverted. The posterior distribution for the latent parameters conditional on the text corpus can be formed by dividing the density (2.1) by the probability of observing that corpus

$$p(\beta, \theta, z, w) = \frac{p(\beta, \theta, z, w)}{p(w)}. \tag{2.2}$$

Evaluating the denominator in (2.2) is computationally infeasible as it entails integrating over the distributions of the latent parameters. However, there are several methods that can be used to approximate the posterior distribution, see Asuncion, Welling, Smyth and Teh (2009). Here, we rely on the collapsed Gibbs sampling algorithm of Griffiths and Steyvers (2004) to estimate $\beta, \theta$ and $z$.

2.3. Estimation. To be able to estimate the LDA model using the approach described above, we first have to translate the raw newspaper texts into a vector-space representation that captures their word frequencies. For this, we break the text down into single words and remove a number of very common terms that have little informative value in bag-of-words models, see Blei et al (2009). Then, we remove word-suffixes using the Porter (1980) stemming algorithm. This step allows us to group closely related words such as “presidential” and “president” or “worker” and “workers” and thus reduces the size of the resulting vector space. For computational reasons, we also limit our vector-space to words that occur at least 100 times.

The number of topics in the benchmark model is set to 10. While choosing a larger number can generally result in more of the topics having a clear interpretation, it can also yield a classification that is too fine for subsequent analyses.³ We estimate a single LDA model using the texts from both 2001 and 2008 jointly. This allows for the possibility that some topics are recurrent and reported with a terminology that is stable over time. For instance,

³In the Online Appendix, available from the authors’ web pages, we also present results from the LDA model using 5, 20, 50 and 100 topics.
the vocabulary used in sports related articles may change little over time and form a topic that is present in news articles in both 2001 and 2008.\footnote{If no topic occurs in both periods and when the number of documents are approximately the same for the two periods, estimating a joint LDA model for both time periods with 10 topics should yield the same topics and assigned topics weights as if we were to estimate two separate models with 5 topics for each period.}

**Table 3. Estimated LDA Topics: High-Probability Words**

<table>
<thead>
<tr>
<th>Topic</th>
<th>Words with the highest assigned probabilities (in descending order)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>bush presid washington afghanistan unit state militari taliban war attack</td>
</tr>
<tr>
<td>2</td>
<td>democrat john republican obama mccain presidenti campaign barack sen senat</td>
</tr>
<tr>
<td>3</td>
<td>school year student counti high state univers review journal colleg</td>
</tr>
<tr>
<td>4</td>
<td>year old home ago time day just peopl like famili</td>
</tr>
<tr>
<td>5</td>
<td>financi washington billion market hous bush bank feder crisi govern</td>
</tr>
<tr>
<td>6</td>
<td>state million year plan new citi health compani say propos</td>
</tr>
<tr>
<td>7</td>
<td>mail daili staff charleston west counti said virginia st state</td>
</tr>
<tr>
<td>8</td>
<td>yesterday polic said offic anthrax court feder offici investig charg</td>
</tr>
<tr>
<td>9</td>
<td>attack new terrorist york world center sept trade airport airlin</td>
</tr>
<tr>
<td>10</td>
<td>citi new today palestinian aug georgia west day isra south</td>
</tr>
</tbody>
</table>

Notes: For each of the 10 topics estimated using Latent Dirichlet Allocation, the table shows the 10 words with the highest probabilities of occurring in that topic. The order of words is descending in terms of the probabilities assigned to them in the given topic. All words have been stemmed using the Porter (1980) stemmer.

2.4. **Estimated LDA Topics**. The LDA model estimated topics endogenously, but human input is generally required to interpret the resulting topics, and, in our case, associate them with particular events. Table 3 shows the topics identified by our estimated LDA model in terms of their highest-probability words. We find that several of the topics that emerge from our estimation are intuitively meaningful. For example, Topic 1 relates to the war in Afghanistan, Topic 2 relates to the candidates of the 2008 US presidential elections, and Topic 9 covers the September 11 terrorist attacks. Furthermore, a relatively clear interpretation can also be attached to Topic 5, which seems to capture both the financial crisis and the reactions of the US government to it.\footnote{In the Online Appendix, we report the first sentence of each article that were assigned the highest probability of belonging to each of the topics.}

Some topics identified by the LDA model are not associated with easily identifiable real world events, e.g. Topic 4. The presence of such difficult-to-interpret topics is a common feature of LDA models (see Chang, Gerrish, Wang, Boyd-Graber and Blei 2009) and often the result of several less frequent “true” topics being combined into one residual model topic. Accordingly, the fraction of difficult-to-interpret topics typically decreases when the number of topics is increased.\footnote{In the Online Appendix, we list the 10 most frequent words for each topic for an alternative LDA specification with the number of topics set to 100. There, one can see that for instance the 9/11 terrorist attack topic is...}
Figure 1. Estimated LDA Topics: Word Clouds of Selected Topics

Topic 1: Afghanistan

Topic 2: 2008 Presidential Candidates

Topic 5: Financial Crisis and Bailouts

Topic 9: Terror Attacks

Notes: The word-clouds illustrate the probabilities associated with specific words in the topics estimated using Latent Dirichlet Allocation. Words with higher probabilities are shown in a larger size. All words were stemmed using the Porter (1980) stemmer. The topics correspond to those shown in table 3.

Topics 1, 2, 5, and 9 all appear to be associated with separate and well-defined events. To get a more complete understanding of these four topics and their associated word probabilities, we also illustrate them in the form of word clouds (Figure 1). These graphical representations show a larger number of words for each topic, reflecting their probabilities within a given topic in terms of the sizes at which they are displayed. The interpretations of the four topics that we derive based on the word clouds reinforces the ones obtained from the high-probability words shown in Table 3.

7 Word clouds are not to everyone’s liking, see http://www.wordle.net/show/wrdl/718619/I_hate_word_clouds.
Figure 2. Newspaper Specialization: Probabilities of Selected Topics

Notes: The figure illustrates the specialization of newspapers on different topics. The topics correspond to those shown in figure 1 and table 3. The short names of newspapers correspond to those in table 1. The normalized topic-specific deviations of news focus are calculated as \( d_{m,k} = \frac{p_{m,k} - p_k}{p_k} \), with \( p_{m,k} \) denoting the probability that newspaper \( m \) reports on topic \( k \) and \( p_k = \frac{1}{M} \sum_{m=1}^{M} p_{m,k} \) being the corresponding average across all \( I \) newspapers. A positive unit deviation implies that a newspaper devoted 100 per cent more coverage to a topic relative to the average newspaper, a negative deviation implies that the newspaper devoted less than average coverage to a topic.

2.5. Different Newspapers Specialize in Different Topics. The first specific aspect of newspaper coverage that we assess using the estimated LDA model concerns the extent to which newspapers are specialized. In other words, we investigate if and by how much different
newspapers tend to over- or underweight different topics relative to the overall average. For this purpose, Figure 2 plots normalized deviations of newspaper-specific topic probabilities for the same four topics discussed above. We calculate these normalized deviations as

\[ d_{m,k} = \frac{p_{m,k} - p_k}{p_k} \]  

(2.3)

where \( p_{m,k} \) denotes the probability that newspaper \( m \) reports on topic \( k \) and \( p_k = \frac{1}{M} \sum_{m=1}^{M} p_{m,k} \) being the corresponding average across all \( M \) newspapers. A positive unit deviation thus implies that a newspaper devoted 100 per cent more coverage to a topic relative to the average newspaper, a negative deviation implies that the newspaper devoted less coverage to a topic than the average newspaper. (A \(-1\) deviation would imply zero coverage of a topic.)

The plots document that there are large amounts of variation in terms of which newspapers tend to cover which topics. For example, the financial crisis as captured by Topic 5 received more than twice as much coverage in the Wall Street Journal than it did in the hypothetical average outlet. Similarly, both the New York Times and USA Today allocated a larger fraction of their news coverage to the September 11 terror attacks than the average newspaper in our sample. These deviations suggest that newspapers do indeed specialize, resulting in coverage that is heterogeneous in the cross-section of outlets and that the measured specialization conforms to our priors about the target audiences of the most widely read national newspapers.

2.6. Major Events Shift News Focus and Increase the Homogeneity of News. We can now assess how major events affect news coverage along two specific dimensions: the average emphasis specific topics receive and the homogeneity of news coverage in the cross-section of outlets. To do so, we explore time variation in the estimated topic probabilities as well as their distribution across newspapers. If major events do indeed affect the focus of news coverage and its cross-sectional homogeneity, we would expect the September 11 terrorist attacks, the nominations of presidential candidates and the outbreak of the financial crisis to be associated with such a behavior.

To assess if this is the case, we use two different measures. First, we calculate overall topic probabilities at a daily frequency by averaging the estimated topic probabilities of all stories in our database for a given day. The fraction \( F_{t,k} \) of total news devoted to topic \( k \) at date \( t \) is thus given by

\[ F_{t,k} = \frac{\sum_d \theta_{t,d,k}}{D_t} \]  

(2.4)

where \( \theta_{t,d,k} \) is the probability that article \( d \) from date \( t \) belongs to topic \( k \) and \( D_t \) is the total number of articles in the sample from day \( t \).

Second, to assess homogeneity in news-coverage across newspapers, we consider to what extent the outlets agree on which topic is most important on a given day. For this, we first identify the topic that has the highest probability across all articles of a given day. Then, we calculate the fraction of newspapers that assign the highest weight to that same topic,
Figure 3. 2001 Terror Attacks: Time-Variation of Average Topic Probabilities and Homogeneity of Coverage Across Newspapers

Notes: The figure illustrates time-series variation in the probabilities assigned to the estimated topics and the cross-sectional homogeneity in newspaper coverage. Each topic is represented by a separate color. The time-horizon shown is 08/01/2001 to 10/31/2001. Only days with coverage of at least 10 newspapers are shown. The topics correspond to those shown in Table 3 with Topic 1 at the bottom and Topic 10 at the top. The topic probabilities for a specific day shown in plot a are defined as the simple average of the corresponding probabilities of all articles in the database for that day. The homogeneity measure shown in plot b is defined as the fraction of newspapers for which the highest-probability topic is the same one that also carries the highest probability across all articles published on that day.

\[ H_t = \frac{\sum_m I(\arg \max_k F_{t,m,k} = \arg \max_k F_{t,k})}{M} \]  

where \( I \) is an indicator function that takes the value 1 when the equality in brackets holds. \( F_{t,m,k} \) is the fraction of news coverage devoted to topic \( k \) by newspaper \( m \) at time \( t \) and \( M \) is the total number of newspapers. The range of \( H_t \) is thus between 0 and 1, with 1 indicating that all newspapers agree on which topic is the most important one.

Figure 3 shows the evolution of both of these measures for the period August to October 2001. The top panel illustrates the share each topic received on each date in the first sample.
Notes: The figure illustrates time-series variation in the probabilities assigned to the estimated topics and the cross-sectional homogeneity in newspaper coverage. Each topic is represented by a separate color. The time-horizon shown is 08/01/2008 to 10/31/2008. Only days with coverage of at least 10 newspapers are shown. The topics correspond to those shown in Table 3 with Topic 1 at the bottom and Topic 10 at the top. The topic probabilities for a specific day shown in plot a are defined as the simple average of the corresponding probabilities of all articles in the database for that day. The homogeneity measure shown in plot b is defined as the fraction of newspapers for which their highest-probability topic is the same one that also carries the highest probability across all articles published on that day.

Topics are ordered from below, with Topic 1 at the bottom and Topic 10 at the top. The most striking episode occurs on September 12 and the following days when more than half of the total news coverage was devoted to the terrorism topic (Topic 9) as displayed in light red.\textsuperscript{8} A second pronounced change occurs on October 8, the day after the war in Afghanistan began (Topic 1) as displayed in dark blue. As can see from the bottom panel, the same two days are also associated with sharp increases in topic homogeneity. That is, both the terror

\textsuperscript{8}This finding is consistent with Eisensee and Strömberg (2007), who by counting broadcast news segments that contain key words, find that major events such as the Olympic Games tend to crowd out other news stories.
attacks and the beginning of the Afghanistan war caused coverage to become more similar across newspapers.

For the second period used in our analysis, i.e. August to October 2008, the same exercise is repeated in Figure 4. Here, too, several events stand out in the sense that they seem to affect both the focus of news coverage and its cross-sectional homogeneity. First, Topic 2, which relates to the presidential nomination conventions, received high levels of media coverage in late August and early September and caused an increase in homogeneity. Then, the Lehman Brothers Bankruptcy on September 15 caused another spike. Finally, a last big spike in the homogeneity measure occurs on September 30, the day after the Emergency Economic Stabilization Act of 2008 failed to pass the US House of Representatives.

The 9/11 terrorist attacks, the war in Afghanistan, the presidential candidate nomination conventions, the Lehman bankruptcy and the failed financial bailout package are all events that robustly and substantially increase the relative level of homogeneity of news coverage across specifications with different number of topics. In the Online Appendix, we plot the graphs for how the homogeneity of news coverage evolves over time when we allow for 5, 10, 20, 50 or 100 topics. Increasing the number of topics affects the topic assignments and our heterogeneity measure in two distinct ways. First, with more topics, individual topics are better defined and may be assigned a higher weight as they may better describe the actual topics of news articles. Second, with many topics, a given event may also give rise to several sub-topics. This effect may decrease our measure of news homogeneity. However, a larger number of topics also reduces the average level of homogeneity: The sample average of our measure homogeneity measure $H_t$ decreases from about 50 per cent with 5 topics to about 20 per cent in the specification with 100 topics. So, while the peaks of the homogeneity measure are somewhat lower with 50 or 100 topics, the relative change in homogeneity after major news events still increases when we allow for a larger number of topics.

3. A Beauty Contest Model with Delegated Information Choice

In the previous section we documented three stylized facts about news coverage that can be attributed to the editorial behaviour of information providers. We showed that, while different newspapers typically emphasize different topics, major events tend to make coverage more homogeneous. In this section we present a theoretical model that allows us to analyze the economic implications of these empirical findings. Specifically, the model can help us understand how the editorial decisions of information providers affect agents’ beliefs and actions in a strategic setting. To have an effect on agents’ beliefs and actions, the reporting behavior must be understood by the agents in the model. That the documented facts are uncontroversial and intuitive, makes this assumption appealing.9

The model is an abstract beauty contest game in the spirit of Morris and Shin (2002) in which agents’ pay-offs depend on the distance of their actions from a latent variable as well as the distance of their action from other agents’ actions. However, we depart from the

9While we believe this to be the first paper that documents how the homogeneity of news coverage increases in response to major events, the fact that large events are widely reported is not entirely surprising. This is consistent with our model, and we view it as a strength of our analysis. In fact, the main economic implications we discuss below arise specifically because our agents are aware of the reporting behavior of news media.
original model in two important ways. First, agents have heterogenous interests in the sense that different agents want their actions to be close to different latent variables. This captures that agent specific factors such as geographical location, sector membership ship and production technologies may affect what type of events agents care the most about. Second, agents are constrained in the number of stories that they can read about and therefore delegate the information choice to specialist information providers that can monitor a large set of events on their behalf.

Each information provider is characterized by a news selection function, which is a mapping from states of the world to a set of reported events. These news selection functions formalize the editorial decisions of news media, and below we will analyze how they affect agents’ beliefs and ability to coordinate their actions. For concreteness, we will refer to the information providers in the model as “newspapers” and say that information consumers “read” about a story. However, the analysis applies equally to TV and radio broadcasters, as well as those online news media that select what news to display on their websites.

In what follows, we assume that information providers are benevolent and choose what to report in order to maximize the utility of their readers. We use this setup as a benchmark for studying how the monitoring and news selection services newspapers provide affect agents’ beliefs and actions. However, while we abstract from other determinants of news reporting decisions, the exact reasons for why a given event is reported are not crucial for the main implications of the behavior we model. In particular, as long as the news selection is systematic and depends on the realized set of events, the forces we study below will be at play.

3.1. Information consumers with heterogeneous interests. Our model is populated by two information consumers, Alice and Bob. They live in a world with two potential stories, $X_a$ and $X_b$. A potential story $X_i : i \in \{a, b\}$ is a random variable that takes values in $\mathcal{X}$ and an event $x_i$ is a particular realization of $X_i$. The state of the world is described by the pair $(x_a, x_b) \in \Omega$ where $\Omega = \mathcal{X} \times \mathcal{X}$ is the set of all (joint) events. We say that an event is of interest to Alice or Bob if their utility increases as a result of knowing about it.\(^{10}\)

3.1.1. Utility and heterogeneous interests. Alice and Bob have different interests, and this heterogeneity is introduced via their utility functions. The basic set-up is a two person beauty contest game in which Alice’s utility depends on the distance between her action $y_a$ and the latent variable $x_a$ as well as the distance between her action and Bob’s action $y_b$. This is formalized by the following utility function for Alice

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

(3.1)

where $\lambda \in (-1, 1)$ is a parameter that governs Alice’s strategic motive. If Bob also wanted to take an action that was close to $x_a$ and close to Alice’s action $y_a$ this setup would be a two-person version of the beauty contest in Morris and Shin (2002). However, we introduce

\(^{10}\)Here we use the word event to mean a specific story that a newspaper might report about, i.e. a realized outcome of $X_i$. This use is more restrictive than how the word is used in the probability theory where “event” describes any set of outcomes that can be assigned a probability.
heterogeneity by making Bob want to take an action that is close to $x_b$ rather than $x_a$. Bob’s utility function $U_b$ is otherwise symmetric to Alice’s and given by

$$U_b = - (1 - \lambda)(y_b - x_b)^2 - \lambda(y_b - y_a)^2$$

(3.2)

where $y_a$ is the action taken by Alice. We say that Alice has a direct interest in $X_a$ because her utility depends directly on the realized value of $X_a$. Symmetrically, Bob has a direct interest in $X_b$. The parameter $\lambda$ governs the strength of the strategic motive. Because of this strategic motive, Alice has an indirect interest in knowing about $X_b$ since that may help her better predict Bob’s action. Symmetrically, Bob has an indirect interest in knowing about $X_a$.

Alice’s optimal action $y_a$ is given by the first order condition

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

(3.3)

where $E_a$ denotes the expectations operator conditional on Alice’s information set, and a symmetric expression describes Bob’s optimal action. If agents could observe both $x_a$ and $x_b$ directly, the equilibrium action would be described by

$$y_i = \frac{1}{1 + \lambda} x_i + \frac{\lambda}{1 + \lambda} x_j : i, j \in \{a, b\}, i \neq j$$

(3.4)

However, Alice and Bob observe neither $x_a$ nor $x_b$ directly and instead have to rely on information providers who monitor the state of the world on their behalf.

3.2. Information providers. There are two information providers, which we call Paper $A$ and Paper $B$. News stories are to some extent indivisible in the sense that reading one word about many different stories is less useful than reading a full paragraph about fewer stories. News media with finite space thus need to select what to report. It is also not feasible for an individual to read all stories that are reported in every newspaper. To capture this constraint, Alice and Bob are restricted to reading only one paper each.\footnote{It would be straightforward to endogenize the decision of how many newspapers each agent chooses to read. A fixed cost of reading a newspaper that is large enough to discourage Alice and Bob from reading both newspapers while not being so large as to make it prohibitively expensive to read one newspaper would result in an outcome identical to the set up posited by assumption here.} Alice reads Paper $A$ because it reports those stories that she finds most interesting. Similarly, Bob reads Paper $B$ because it reports those stories that he finds most interesting. While not modeled explicitly here, this is a simple way of capturing that newspapers compete for readers/subscribers by offering specialized content.

We formalize the editorial decision of a newspaper by defining its news selection function as a mapping from the realized state of the world into a discrete decision of what to report.

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

Depending on the state of the world, Alice observes either $X_a$ or $X_b$. Both newspapers make their editorial decisions in order to maximize the expected utility of their respective
readers. In doing so, they take the news selection function of the other newspaper as given. The news selection functions are thus determined by

$$S_i(x_i, x_j) = \arg\max_{S_i} E[U_i(S_i, S_j)]$$

where the expected utility of an agent depends not only on the news selection of the paper that he or she reads, but also on the news selection function of the paper read by the other agent and his or her utility function.

3.3. News selection and beliefs. Reading a news report about either $X_a$ or $X_b$ is always immediately informative about that specific variable. However, one implication of a systematic news selection is that whether an event is reported or not is by itself informative. We state this result more formally in the following proposition.

**Proposition 1.** Posterior beliefs about the unreported story $x_j$ coincide with $p(x_j | x_i)$ only if the probability of reporting about $x_i$ is conditionally independent of $x_j$, that is

$$p(x_j | S_i = 1, x_i) = p(x_j | x_i)$$

only if

$$p(S_i = 1 | x_i) = p(S_i = 1 | x_j, x_i).$$

**Proof.** By Bayes’ rule we can express the posterior about the unreported variable as

$$p(x_j | S_i = 1, x_i) = \frac{p(S_i = 1 | x_j, x_i)p(x_j | x_i)}{p(S_i = 1 | x_i)}.$$  

It then follows immediately that (3.6) holds only if

$$\frac{p(S_i = 1 | x_j, x_i)}{p(S_i = 1 | x_i)} = 1$$

which completes the proof.

In a standard prediction problem, the joint distribution of the two events would be sufficient to determine what you learn about $x_j$ from observing $x_i$. But Proposition 1 shows that the selection of events by itself is informative when the probability of what you get information about depends on the realized state. This result is general and holds for all distributions of $X_a$ and $X_b$, but a special case is particularly illustrative of its implications.

Consider a set up where $X_a$ and $X_b$ are independent so that $p(x_j | x_i) = p(x_j)$. Knowing about $x_i$ is then by itself uninformative about $x_j$. But Proposition 1 states that if the probability of reporting $x_i$ depends on the realized value of $X_j$, the fact that $x_i$ was reported is informative about $x_j$. The systematic news selection by the information providers thus makes agents update their beliefs about both reported and unreported events, even when the two events by themselves are uninformative about each other.

The implications of Proposition 1 are starkest if there are states of the world where paper $i$ would report $X_j$ instead of $X_i$. Since paper $i$ did not report $X_j$, these states can then be ruled out, i.e. these states are associated with a zero probability conditional on $S_i = 1$ and $x_i$. As an example, consider somebody reading the Wall Street Journal. If there is no report about a stock market crash, the reader can infer that no stock market crash has occurred since the Wall Street Journal would for sure have reported such an event, had it occurred.
4. **News selection, commonality of information and correlated actions**

We first study the implications of delegated information choice for agents’ beliefs and actions in a simple setting where the state space of the random variables $X_a$ and $X_b$ is low-dimensional and discrete. While we relax these assumptions in Section 5, the simple set up analyzed here allows us to derive explicit expressions for agents’ optimal actions and the degree to which information is common across agents. It also enables us to establish analytically how delegated information choice affects the correlation between agents’ actions.

4.1. **Discrete states of the world.** Consider a world where the potential stories $X_a$ and $X_b$ are discrete random variables that can take the values $-1, 0, 1$. The different states occur with probabilities given by

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

where $p_i(x_i)$ is the pmf of $x_i$. The potential stories $X_a$ and $X_b$ are thus identically and symmetrically distributed, zero mean random variables. We also assume that $X_a$ and $X_b$ are independent of one another so that

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

Neither the symmetry nor the independence of $X_a$ and $X_b$ are necessary for what follows, but help simplify the presentation.

4.2. **Optimal news selection functions.** Each information provider chooses which story to report in order to maximize the expected utility of its respective reader. Because of the strategic motive in agents’ utility, what information will be most useful to Alice depends on Bob’s action. Since Bob’s action in turn depends on what information he receives, the news selection function of Paper $A$ depends on the news selection function of Paper $B$. A Nash equilibrium in the news-selection game is a fixed point at which neither newspaper wants to change its selection function, taking the other paper’s selection function as given. To find such an equilibrium, we first conjecture explicit news selection functions for Paper $A$ and Paper $B$ (without proof of optimality) and derive Alice and Bob’s implied optimal actions. Then, we verify that the postulated news-selection functions do indeed constitute a Nash equilibrium. In the following, we go through these steps separately for the case with no strategic motive, the case with strategic complementarity and the case with strategic substitutability.

4.2.1. **No strategic motive.** As a benchmark, consider first the case in which $\lambda = 0$ and where agents’ thus do not have an incentive to coordinate. In this case, it is optimal for Paper $A$ to always report $X_a$ since Alice’s utility then depends neither directly nor indirectly on $X_b$. Symmetrically, it is optimal for Paper $B$ to always report $X_b$. In the absence of a strategic motive in actions, the news selection functions are thus simply described by

$$S_i = 1 \ \forall \ \{x_i, x_j\} \in \Omega.$$  

The news selection functions for Paper $A$ and Paper $B$ when $\lambda = 0$ are also given in tabular form in the top row of Table 1.
4.2.2. **Strategic complementarity.** When agents have an incentive to take an action that is close to the action of the other agent, i.e. when $\lambda > 0$, the equilibrium news selection function is described by

$$S_i = \begin{cases} 0 & \text{if } x_i = 0 \text{ and } x_j \in \{-1, 1\} \\ 1 & \text{otherwise} \end{cases}. \quad (4.4)$$

The corresponding reporting decisions of this case are illustrated in the lower panel of Table 1. As in the case without strategic motives, Paper A will report about $X_a$ most of the time. However, now that Alice wants to take an action close to that of Bob, it is optimal for Paper A to report about $X_b$ whenever $x_a = 0$ and $x_b \neq 0$. The intuition of this is simple. When the realized value of $X_a$ is zero, it is more important for Alice to know whether Bob will take a positive or negative action. Knowing the realized value of $X_b$ is then more useful to her than knowing about $X_a$ since this allows her to better predict Bob’s action.\(^{12}\)

4.2.3. **Strategic substitutability.** Changing the sign of $\lambda$ so that agents want to take actions that are far from the actions of the other agent leaves the equilibrium news selection functions unchanged. When $x_a = 0$, it is still more useful to Alice to observe $X_b$ so that she knows whether Bob took a positive or negative action. Having this information, she can then take an action in the opposite direction. Thus, regardless of whether $\lambda$ is positive or negative, there are states of the world in which Alice and Bob want to know what the other agent knows.

This contrasts with the result in the coordination game in Hellwig and Veldkamp (2009). There, ex ante identical agents can chose to observe different combinations of private and public signals about a single latent variable of common interest. In such a setting, information acquisition inherits the strategic properties of the coordination game. Thus, if there is a strategic complementarity in actions, agents also want to buy the same signals as other agents. The different implications arise from the fact that in Hellwig and Veldkamp’s model, agents do not choose what to get information about, but rather if the noise in their signal is common to all agents or idiosyncratic. Clearly, public and private signals of the same precision are equally useful for predicting the latent fundamental. The choice between private and public signals faced by an agent in the model of Hellwig and Veldkamp is thus a choice about having prediction errors that are positively correlated or uncorrelated with the prediction errors of other agents. With strategic complementarities, the former is preferred, with strategic substitutes, the latter. In our model, the information choice determines whether an agent makes an error in predicting the action of the other agent or not. When actions are either strategic complements or substitutes, not knowing the action of the other agent is costly.\(^{13}\)

\(^{12}\)In fact, given the news selection function (4.4), Alice can infer that if she reads about $X_b$, then $x_a = 0$ with probability 1. However, that Alice can infer the realized value of the unreported value with certainty is to some degree an artefact of the low dimensional state space. (Proposition 1 above provided a more general characterization of the posterior beliefs about the unreported event, conditional on what was reported.)

\(^{13}\)Sufficiently strong complementarities result in multiple equilibria in news selection strategies. This case is discussed in the Online Appendix.
### Table 1: News selection functions

<table>
<thead>
<tr>
<th>Paper A</th>
<th>Paper B</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\omega$</td>
<td>$\omega$</td>
</tr>
<tr>
<td>$x_b = -1$</td>
<td>$x_a = -1$</td>
</tr>
<tr>
<td>$x_b = 0$</td>
<td>$A$</td>
</tr>
<tr>
<td>$x_b = 1$</td>
<td>$A$</td>
</tr>
</tbody>
</table>

| $x_b = -1$ | $x_a = -1$ | $x_a = 0$ | $x_a = 1$ |
| $x_b = 0$ | $B$ | $B$ | $B$ |
| $x_b = 1$ | $B$ | $B$ | $B$ |

### 4.3. News selection, higher-order beliefs and common knowledge.

Public signals that are commonly known to be observed by all agents are particularly influential when privately informed agents interact strategically (e.g. Morris and Shin 2002). Arguably, everything that is reported by newspapers is public in the sense that it is available for those who care to look for it. However, not all information that is printed in a newspaper is necessarily observed by everybody, and even when an event is widely reported, it may not be known to readers of all newspapers how widely reported it is. In the model above, when $\lambda \neq 0$, there are some states of the world in which Alice and Bob read about the same event and yet this event is not common knowledge.

To see this, consider first the case when Paper A reports about $X_b$. This only happens in the states $(0, 1)$ and $(0, -1)$ i.e. only in states of the world where Paper B also reports about $X_b$. This is natural since Alice has no direct interest in $X_b$ and finds it useful to know about $X_b$ only to the extent that it helps her predict the action of Bob. Because Alice understands that Bob will read about $X_b$ for sure whenever she does, she knows not only that $X_b = x_b$ but also that Bob knows this as well. Yet, this fact is not common knowledge. Bob knows that he observes $X_b$ in the states $(-1, 1), (1, 1), (-1, -1), (1, 1), (0, -1)$ and $(0, 1)$. But since Alice observes $X_b$ only in the latter two states and because in these states, Bob attaches positive probability also to states where Alice does not observe $X_b$, the fact that $X_b = x_b$ is not common knowledge even though both Alice and Bob knows this to be true.\(^{14}\)

The degree to which knowledge about an event is common among agents matters for their optimal decisions. As we will now demonstrate, the probability with which Bob believes that Alice observes $X_b$, conditional on Bob observing $X_b$, affects not only how strongly Bob responds to the realized value $x_b$, but also how strongly Alice responds in those states of the world where she also observes $X_b$.

---

\(^{14}\)In fact, in the simple discrete example here, the only state in which any event is common knowledge is $(0, 0)$ since it is only in this state that Alice or Bob reads a report stating that the variable they have a direct interest in equals zero.
4.4. Equilibrium actions. Alice and Bob’s equilibrium actions depend on the degree of strategic complementarities both through the first order condition (3.3) and through the effect that the strategic motive has on the equilibrium news selection functions. We now derive the optimal actions, taking the news selection functions described by (4.3) and (4.4) as given.

4.4.1. No strategic motive. In the absence of a strategic motive, Alice always observes $X_a$ and Bob always observes $X_b$. Alice and Bob’s equilibrium actions are then trivially given by $y_a = x_a$ and $y_b = x_b$. Since $X_a$ and $X_b$ are independent random variables, Alice’s and Bob’s actions are then also independent.

4.4.2. Strategic motives in actions. With a strategic motive, Alice sometimes observes $X_b$ and Bob sometimes observes $X_a$. The optimal response to a news report depends on whether it is about an event that the agents have a direct or indirect interest in. We start by deriving the optimal action of Bob when he reads about $X_a$, i.e. the story that he only has an indirect interest in.

Bob knows that he only observes $X_a$ when Alice does so as well. Bob can thus infer Alice’s action with certainty when he observes $X_a$. Since Bob only observes $X_a$ when $x_b = 0$, Bob’s optimal action when he observes $X_a$ is simply given by

$$y_b(x_a, S_b = 0) = \lambda y_a(x_a, S_a = 1).$$ (4.5)

When Alice observes $X_a$ she does not know with certainty whether Bob does so as well. Her optimal action can then be expressed as function of $x_a$ and a weighted average of Bob’s actions when he observes $X_a$ and his action when he does not

$$y_a(x_a, S_a = 1) = (1 - \lambda)x_a + p(S_b = 0 | x_a, S_a = 1) \lambda y_b(x_a, S_b = 0)$$
$$+ p(S_b = 1 | x_a, S_a = 1) \lambda E[y_b(x_b, S_b = 1) | x_a, S_a = 1, S_b = 1].$$ (4.6)

Note that Bob will also observe $x_a \in \{-1, 1\}$ if $x_b = 0$, which happens with probability $\frac{1}{2}$, and he then takes the action (4.5). Because $X_b$ is symmetrically distributed around zero, Alice’s expectation of Bob’s action when he observes $X_b$ equals zero. Taken together, these two facts means that Alice’s optimal action when $x_a \in \{-1, 1\}$ simplifies to

$$y_a(x_a, S_a = 1) = (1 - \lambda)x_a + \frac{1}{2}y_b(x_a, S_b = 0).$$ (4.7)

Substituting (4.5) into (4.7) gives the optimal actions as functions of the observed variables

$$y_i(x_i, S_i = 1) = \frac{(1 - \lambda)}{1 - \frac{1}{2}\lambda^2}x_i$$ (4.8)

and

$$y_j(x_i, S_j = 0) = \lambda \frac{(1 - \lambda)}{1 - \frac{1}{2}\lambda^2}x_i$$ (4.9)

where the symmetry of Alice’s and Bob’s decision problem allowed us to switch to general indices. We can see from (4.8) - (4.9) that the magnitude of both Alice and Bob’s responses to $x_a$ depends on the probability $p(S_b = 0 | x_a, S_a = 1)$. When Alice observes $x_a$, this is the probability she attaches to the event that Bob also observes $x_a$. When Bob observes $x_a$, this
is the probability that Bob believes Alice attaches to the event that he observes $x_a$. The higher this probability is, the stronger is the response of both agents and this effect will be stronger when strategic motives are more important.

The degree to which information about an event is common among agents depends on the news selection functions $S_a$ and $S_b$, even when the event in question is mutual knowledge. The news selection functions thus directly affect how strongly agents respond to a given event, even after conditioning on what events they have observed. As a result, measuring the coverage a specific event receives is neither sufficient for determining the extent to which the event is common knowledge, nor for predicting how strongly agents will react to it.

Incidentally, the expression (4.8) also describes the optimal action when agents observe that the variable they have a direct interest in equals zero. Since the agents only observe a zero realization in the state $(0,0)$, and because this state is common knowledge, it is then optimal for both agents to take a zero action.

4.5. Verifying the optimality of the conjectured news selection functions. Given the optimal actions derived above, it is straightforward to verify by direct computation that neither Paper $A$ nor Paper $B$ has an incentive to deviate from the conjectured news selection functions described by (4.3). The Appendix describes an algorithm for doing so.

4.6. Correlation of actions with and without delegated news selection. To isolate the implications of the editorial function of the newspapers for agents’ actions, we now compare the predictions of the model to a natural alternative. In the alternative model, Alice and Bob are, as in the benchmark model, restricted to observing only one out of the two realized events. However, instead of delegating the news selection to an information provider that can condition on ex post outcomes, Alice and Bob have to make a decision ex ante about which variable to observe.

When actions are strategic complements, that is when $\lambda > 0$, Alice will choose to always observe $X_a$ and Bob will choose to always observe $X_b$. As long as the strategic motive is not too strong, the same is also true when actions are strategic substitutes.\textsuperscript{15} Since $X_a$ and $X_b$ are independent, observing $x_a$ is then uninformative about $x_b$ and vice versa. The conditional expectation of the unobserved variable is then equal to its unconditional mean and the optimal action $\tilde{y}_i$ with ex ante information choice is given by

$$\tilde{y}_i = (1 - \lambda) x_i : i \in a, b$$

(4.10)

If $X_a$ and $X_b$ are independent, Alice and Bob’s actions $\tilde{y}_a$ and $\tilde{y}_b$ are thus uncorrelated in this alternative model. However, as we now show, this is not the case with delegated information choice.

**Proposition 2.** Delegated news selection introduces correlation between Alice and Bob’s actions of the same sign as $\lambda$.

\textsuperscript{15}The formal condition for Alice and Bob to choose to always observe the variable they have a direct interest in is that $(1 - \lambda^2)^2 + \lambda > 0$. The model with ex ante signal choice is derived and solved in the Appendix.
Proof. Direct computation of the correlation of Alice and Bob’s actions gives
\[
\sum_{\omega \in \Omega} p(\omega) y_a(\omega) y_b(\omega) \sqrt{\text{var}(y_a)} \sqrt{\text{var}(y_b)} = \lambda^2 (1 - \lambda^2)^2 \text{var}(y_i)^{-1}
\]
where \(y_i(\omega)\) is agent \(i\)'s action in state \(\omega\).
\[\square\]

To see why the delegated news selection introduces correlation in the actions of the agents, first note that the terms in the sum of the left hand side of (4.11) associated with states where \(x_a = x_b \in \{-1, 1\}\) cancel against the terms associated with the equally probable states where \(x_a = -x_b \in \{-1, 1\}\). (The term associated with the state \((0, 0)\) is zero.) The correlation in actions is thus driven by those states where both agents read about the same event. That is, in the states \((1, 0)\) and \((-1, 0)\) both Alice and Bob read about \(X_a\), and in the states \((0, 1)\) and \((0, -1)\) they both read about \(X_b\). The products of Alice and Bob’s actions in these states are then always either positive (if \(\lambda > 0\)) or negative (if \(\lambda < 0\)). The editorial function of newspapers thus introduces a correlation in agents’ actions that is absent if agents choose ex ante what variable to get information about. This correlation is positive if actions are strategic complements and negative if they are strategic substitutes.\(^{16}\)

5. Extreme events and common knowledge

The model presented in the previous section can reproduce the main features of the news coverage data that we documented above using actual newspaper texts. First, the two newspapers specialize in the sense that they tend cover different topics to different degrees. Second, in some states of the world, they report on the same topic and thus cause coverage to become more homogenous. However, the simple set up with discrete states does not allow us to capture how large-magnitude events, such as the 9/11 terrorist attacks or the Lehman Brothers bankruptcy, affect news coverage. Arguably, what made these events special and so widely reported were their magnitudes, as both terrorist attacks and bank failures occur frequently on a smaller scale. In this section we therefore extend the model to allow for continuous distribution of events. This extended version of the model can be used to study how the magnitude of events affects how widely reported they are and, in extension, how it affects agents’ beliefs and actions. We again begin by deriving the optimal news-selection functions as well as the agents’ conditional actions.

5.1. Optimal simple news selection functions. With continuous distributions of the potential stories \(X_a\) and \(X_b\), the optimal news selection functions are infinite-dimensional objects with unknown functional forms. We therefore approximate the true news selection functions using a simple but flexible parametric class of threshold functions in the absolute values of \(x_a\) and \(x_b\) of the form
\[
S_i = \begin{cases} 
1 & \text{if } |x_i| \geq \alpha_i |x_j|^\beta_i \\
0 & \text{otherwise} 
\end{cases}
\]

\(^{16}\)The coordinating effect of news media identified in Proposition 2 works by affecting the actions of audiences across media outlets. In reality, most media outlets reach more than one agent. The correlation in actions introduced by the delegated news selection thus goes beyond any coordination of agents’ actions that occurs simply because a single news outlet may reach a large audience.
This functional form makes our analysis more tractable and captures a wide variety of news selection functions. Given the constraint that the news selection functions must be of the form (5.1), Paper A now chooses $\alpha_a$ and $\beta_a$ in order to maximize the expected utility of Alice, and Paper B chooses $\alpha_b$ and $\beta_b$ to do the same for Bob.

5.2. Conditional actions. When $X_a$ and $X_b$ are continuously distributed, the conditional expectations in the first order condition (3.3) can be expressed as

$$y_i(x_k, S_i) = (1 - \lambda) \int \int x_i \frac{p(x_i, x_j, S_i)}{p(x_k, S_i)} dx_i dx_j + \lambda \int \int y_j(x_i, x_j) \frac{p(x_i, x_j, S_i)}{p(x_k, S_i)} dx_i dx_j$$

where: $i, j, k \in \{a, b\}$ and $i \neq j$. Now, first note that for independent zero mean distributions of $X_a$ and $X_b$, the news selection function (5.1) implies that the expected value of the unreported variable is the same as its unconditional mean.

Second, and as in the previous section, it is again optimal for Paper B to report about $X_a$ only when Paper A does so as well. That is, if $S_j = 0$, then $S_i = 1$. Taken together, these two facts allow us to simplify the optimal action (5.2) to

$$y_i(x_i, S_i = 1) = \frac{(1 - \lambda)}{1 - p(S_j = 0 \mid x_i, S_i = 1)} \lambda^2 x_i$$

and

$$y_i(x_j, S_i = 0) = \lambda \frac{(1 - \lambda)}{1 - p(S_i = 0 \mid x_j, S_j = 1)} \lambda^2 x_j$$

As in the discrete states model above, the strength of agents’ responses to news stories depends on the degree to which the observed event is common knowledge, i.e. the probability in the denominators of (5.3) and (5.4). Here, this probability varies smoothly with the realized value of the state and depend on the values of $\alpha_j$ and $\beta_j$. Because the optimal news selection functions depend on the strength of the strategic motive, the degree to which knowledge about an event becomes common among agents depends both on the realized value of $X_a$ and $X_b$ as well as on the preferences of Alice and Bob.

5.3. Solving the model. With continuous distributions, we need to solve the model numerically. A solution can be found by letting Paper A choose $\alpha_a$ and $\beta_a$ in order to maximize Alice’s utility, taking Bob’s actions as given. Paper B then chooses $\alpha_b$ and $\beta_b$ in order to maximize Bob’s expected utility, taking Alice’s actions from the previous step as given. Iterating between these two steps until convergence yields a solution.

5.4. Preferences, distributions and common knowledge. From the expressions (5.3) - (5.4) of the optimal action, we know that the strength of both Alice and Bob’s responses to observing $x_a$ depend on how likely Alice thinks it is that Bob also observes $x_a$. We therefore first derive these probabilities and how they depend on the realized value of $x_a$.

---

17 The main implicit assumption here is that the relative news worthiness of different events depends only on their relative magnitudes. Nevertheless, the types of effects we are able to identify using this more simple set up should be robust to richer functional forms.

18 The conditional distribution $p(x_i \mid x_j, S_i = 0)$ is simply $p(x_i)$ with symmetrically truncated tails.
By Bayes’ rule, this probability is given by

\[ p(S_b = 0 \mid S_a = 1, x_a) = \frac{p(S_a = 1 \mid S_b = 0, x_a) p(S_b = 0 \mid x_a)}{p(S_a = 1 \mid x_a)} \] (5.5)

Now, since Bob only observes \( X_a \) when Alice does so as well, we have that \( p(S_a = 1 \mid S_b = 0, x_a) = 1 \), implying that (5.5) simplifies to

\[ p(S_b = 0 \mid S_a = 1, x_a) = \frac{p(S_b = 0 \mid x_a)}{p(S_a = 1 \mid x_a)} . \] (5.6)

The probabilities in the ratio on the right hand side of (5.6) can be computed from the news selection functions (5.1) as

\[ p(S_b = 0 \mid S_a = 1, x_a) = \frac{p(|x_b| < \alpha_b |x_a|^{\beta_b} |x_a|)}{p(|x_a| \geq \alpha_a |x_b|^{\beta_a} |x_a|)} . \] (5.7)

Since the parameters \( \alpha \) and \( \beta \) in (5.1) depend on the preferences of the agents, the degree to which information about a realized event is common among agents depends on agents’ preferences as well as on the distributions of events. To illustrate how these interact, we solve the model for different values of \( \lambda \) and for different distributions of \( X_a \) and \( X_b \). For each set up we compute the conditional probability that Alice and/or Bob observes \( X_a \) as a function of its realized value.

Figure 5 illustrates the probabilities that Alice and/or Bob observe \( X_i \) or \( X_j \) as a function of the realized value of \( X_i \) and the strength of the strategic motive. The left column corresponds to \( X_i \sim U(-1,1) \) and the right column to \( X_i \sim N(0, \frac{1}{3}) \), illustrated by the dotted blue lines.\(^{19}\)

5.4.1. **No strategic motive.** The first row of Figure 5 illustrates the probability that Alice (solid red) and Bob (dashed yellow) observe \( X_a \) conditional on the realized value \( x_a \) when \( \lambda = 0 \). Without a strategic motive, Alice always observes \( X_a \) but Bob never observes \( X_a \). For all values of \( x_a \) the associated conditional probabilities are then either 1 or 0. The optimal news selection functions are given by \( \alpha_i = 0 \) (and an indeterminate \( \beta_i \)) for \( i \in \{a,b\} \).

5.4.2. **Moderate strategic motive.** With a strategic motive, agents may value being informed about the action of the other agent more than they value being informed about the latent variable they have a direct interest in. When \( \lambda = 0.3 \), this is reflected in the optimal news selection functions, which are again symmetric and characterized by \( \alpha_i \) approximately equal to 0.3 and \( \beta_i \) equal to 1. The unconditional probability that a newspaper report about the event that its reader has a direct interest in is given by

\[ p(S_i = 1) = p(|x_i| > 0.3 \times |x_j|) \] (5.8)

\[ > \frac{1}{2} \] (5.9)

where the inequality follows from the fact that \( X_a \) and \( X_b \) have identical distributions. The papers thus specialize their coverage to cater to their respective readers in the sense that

\(^{19}\)To facilitate the comparison, the variance of the Gaussian distribution is chosen so that most of its probability mass lies within the support of the uniform distribution.
Paper A will report about \( X_a \) most of the time and Paper B will report about \( X_b \) most of the time. However, when \(|x_a|\) is sufficiently large relative to \(|x_b|\), that is when \(0.3 \times |x_a| > |x_b|\), both papers will report about \( X_a \). The reason for this is that Alice will then take an action that is sufficiently large (in absolute terms) to make it more important for Bob to know about Alice’s action than about the small (in absolute terms) realized value of \( X_b \).

The middle row of Figure 5 shows that the probability that Alice observes \( X_a \) increases rapidly as the absolute value of \( x_a \) increases for both the uniform and the normal distribution. The probability that Bob observes \( X_a \) is also increasing in \(|x_a|\). Since the states in which Bob observes \( X_a \) is a strict subset of the states where Alice observes \( X_a \), Bob’s probability of observing \( X_a \) is lower than the probability that Alice does so for every value of \( x_a \).

The probability \((5.6)\) that Bob observes \( x_a \) conditional on Alice doing so corresponds to the ratio of the (dashed) yellow and (solid) red lines in the graph. Alice knows that Bob is more likely to observe \( X_a \) as the absolute value of \( x_a \) increases, so larger magnitude events tend to be closer to common knowledge. With uniform distributions, the probability that Bob observes \( X_a \) as \( x_a \) tends to the boundary of its support is simply given by the value of \( \alpha_b \).\(^{20}\) Alice thus attaches about a 30 per cent probability to the event that Bob observes \( X_a \) as \( X_a \) tends to \(-1\) or \(1\).

If events are normally distributed, the corresponding probability at \(|x_a| = 1\) is about 75 per cent. The difference compared to the uniform distributions is explained by the fact that with normally distributed variables, more probability mass is concentrated around the mean. Conditional on the realized value of \( X_a \), it is then less likely that the realized (absolute) value of \( X_b \) is large enough to make Paper B report \( x_b \) instead of \( x_a \). With normally distributed variables and for a large enough absolute realization of \( X_a \), both Paper A and Paper B report \( x_a \) almost surely. In the limit, information about extreme realizations of \( x_a \) or \( x_b \) thus approaches common knowledge. The conditional correlation of Alice and Bob’s actions then tends to 1 as \(|x_i| \to \infty\).

5.4.3. Strong strategic motive. As the strategic motive is strengthened, the cost of not observing the event that the other agent is responding to increases. The news selection functions of the two papers then become more similar in the sense that the the two papers are more likely to report about the same event. When \( \lambda = 0.6 \) (bottom row of Figure 5) both Paper A and Paper B will simply report the variable that has had the largest absolute realization and \( \alpha_i = \beta_i = 1 \) for both Paper A and Paper B. Since the news selection functions are known to both agents, Alice can then infer that if she observes \( X_a \), \(|x_a|\) is larger than \(|x_b|\) so that Paper B will also report \( X_a \). With sufficiently strong complementarities in actions, both papers will always report the same event, and that event will be common knowledge. Effectively, it is then as if there was a single information provider. The realized values \( X_a \) and \( X_b \) are then either common knowledge or not known to any agent.

5.5. News selection and aggregate actions. Ultimately, we are interested in how delegated information choice affects agents’ actions. The news selection functions do so through two distinct channels. First, the news selection functions determine how likely it is that an

\(^{20}\)This is so since \( p(\alpha_b |x_a| > |x_b|) = \alpha_b \) when the distributions of \( x_a \) and \( x_b \) are both uniform and with identical support.
agent knows about an event. If an agent does not know that an event has occurred, he or she cannot respond to it. Second, conditional on reading about an event, the news selection functions affect how likely it is that the other agent is reading about the same event. When

**Figure 5.** The probability that the agent i (solid red) and agent j (yellow dashed) observe \( x_i \) conditional on \( x_i \) with \( X_i \) and \( X_j \) distributed as \( U(-1, 1) \) (left column) and \( N(0, 1/3) \) (right column). Top row is model without strategic motive (\( \lambda = 0 \)), middle row with moderate strategic motive (\( \lambda = 0.3 \)) and bottom row with strong strategic motive (\( \lambda = 0.6 \)).
actions are strategic complements, an agent will respond more strongly to an event if he thinks it is more likely that the other agent also observes the same event.

As we saw in Figure 5 above, the probability that the two agents observe an event $x_i$ is increasing in its absolute realized value. Figure 6 illustrates how these probabilities translate into expected aggregate actions for different values of $\lambda$ and for different distributions of events.

![Graphs showing expected aggregate response](image)

**Figure 6.** Expected aggregate response conditional on $x_i$ with $x_i, x_j \sim U(-1,1)$ (left panel) and $x_i, x_j \sim N(0, \frac{1}{3})$ (right panel).

When $\lambda = 0$, Alice always observes $X_a$, but Bob never does. The probability in the denominator of (5.3) and (5.4) is then zero, and Alice’s response is linear in $x_a$. Because Bob never observes $X_a$, the conditional expectation of his action is zero for all values of $x_a$. The conditional expectation of the aggregate action illustrated by the dotted grey line is thus simply Alice’s expected action. When $\lambda > 0$, the probability that Bob and/or Alice observe $X_a$ increases in the absolute value of $x_a$, and so does the probability that Bob observes $x_a$ conditional on Alice doing so. These effects introduce a nonlinearity in the expected aggregate response, as illustrated by the dashed grey and solid blue lines in Figure 6.

To understand the source of this non-linearity, consider first realizations of $X_a$ that are close to zero. In the zero limit, the probability that Alice or Bob observe $X_a$ is also zero, and then so is the conditional expectation of the sum of their actions. The expected response curve in the Figure 6 is therefore flat around the point where $x_a = 0$. As the absolute value of $x_a$ increases, the probability that Alice and Bob reads about it increases. Now, both the probability that the agents read about $X_a$ and the degree to which this fact is common knowledge is increasing in $|x_a|$. The magnitude of the expected aggregate response to $x_a$ is thus increasing more than proportionally in $|x_a|$.

The mechanism thus generates weak responses to small magnitude events, and comparably strong responses to events of large magnitude. The model here thus captures how events compete for media coverage. Small magnitude events only become widely known if no competing more extreme event has occurred. But because even when all agents observe a small event, the agent with a direct interest in the event will generally not be able to infer
that it is in fact widely reported. The resulting responses of both agents are then weaker. More extreme events, on the other hand, are not only more likely to be widely reported, but they also tend to be closer to common knowledge. Agents’ responses to such events are therefore stronger.\footnote{In this section, we analyzed the effect of news selection when $\lambda \geq 0$. Results are qualitatively unchanged if we allow for actions to be strategic substitutes. Additional results pertaining to the cases when $\lambda < 0$ are reported in the Online Appendix.}

6. Conclusions

News media are an important source of information for a large part of society. In this paper we have argued that, in order to understand how news media affect decisions and beliefs, we first need to understand how they select what stories to report. To do this, we obtained text fragments from a large number of news stories published in US newspapers during the months around the September 11 terrorist attacks and the Lehman bankruptcy in 2008. We then estimated a Latent Dirichlet Allocation topic model to document three stylized facts about news coverage. First, different newspapers provide specialized content and tend to cover different topics to different degrees. Second, major events such as terrorist attacks or financial crises result in a large fraction of news content being devoted to the topics associated with these events. Third, major events make news coverage more homogenous across newspapers.

To understand the implications of the documented reporting behavior for agents beliefs and actions, we proposed a simple beauty-contest model that can replicate the documented facts. The model is populated by agents that have heterogeneous interests and delegate their decisions of what to get information about to specialized news providers that cater to their specific interests. These news providers perform an editorial service for the agents by monitoring the entire state of world and by choosing ex post which of the realized events to report. This mechanism of delegated information choice generates new predictions relative to models in which agents choose what to get information about based on the ex ante expected usefulness of different signals. In our set up, what has happened affects what agents get information about. Because of this, the model can match the stylized facts about news coverage documented in the first part of the paper.

We formalized the editorial decision of news media as a mapping between states of the world and a selection of reported events. Because this selection is systematic, and understood by the agents, the selection of reported events is by itself potentially informative also about non-reported events. One contribution of the paper is to provide formal conditions for when this is the case. The systematic selection of reported events also makes it possible for agents to infer how likely it is that other agents read about the same events as them. When agents have a strategic motive, they respond stronger to events that they can infer that other agents are more likely to know about as well. Events can be mutually known and yet fail to be common knowledge. To understand how agents will respond to an event in a strategic setting, it is thus not sufficient to know how widely reported the event is, but one must also know the degree to which information about the event is common among agents.
In the model, the degree to which information about an event is common among agents is endogenous and depends on agents’ preferences, the event itself and the distribution of possible events. Information about most events is then neither perfectly public nor perfectly private. This contrasts with the large literature that has studied the role of private and public information in strategic settings, e.g. Morris and Shin (2002) and Angeletos and Pavan (2007). In this literature, signals are either perfectly private or perfectly public, and this structure is unaffected by preferences and the realized state of world. Here, we break this clean dichotomy and make positive predictions about the types of environments and events to which the results of this existing literature are more likely to apply.

We also demonstrated that delegated and state-dependent information choice can introduce correlation between agents’ actions compared to a setting in which agents choose ex-ante what to get information about. That the editorial role of information providers facilitates coordination in some states of the world has implications for the large existing literature proposing that business cycles are (at least partly) caused by agents coordinating on either pure sun-spot shocks, e.g. Cass and Shell (1983), on noisy public signals e.g. Lorenzoni (2009) and Nimark (2014), or on “sentiment” shocks e.g. Angeletos and La’O (2013). One feature these papers have in common is that the coordination of actions cannot rely solely on the information that is transmitted through prices. The argument we make in this paper is that, to the extent that coordination works via news media, it will be facilitated in those states of the world in which news coverage is understood to be more uniform across different news providers.

Finally, in the theoretical model above, we take a very benevolent view of news media and assume that what is reported is selected to maximize the utility of the reader and reported without bias. This allows us to abstract from other motives of how news stories are selected and reported. However, the mechanisms that we study here do not rely on these particular assumptions. As long as agents understand the mapping between states of the world and what gets reported, agents beliefs and actions will be affected by the news selection functions. For instance, if all newspapers are more likely to report bad news events relative to good ones, events that are considered bad will be closer to common knowledge and provoke stronger responses than good events of similar magnitude.

References


Hansen, S., M. McMahon and A. Prat, 2015, ”Transparency and Deliberation within the FOMC: a Computational Linguistics Approach”.


Marschak, Jacob, 1960. ”Remarks on the Economics of Information”, Cowles Foundation Paper 146, Reprinted from Scientific Research in Management, University of California Printing, UCLA


Soroka, S.N., D.A. Stercula and C. Wlezien, 2015, ”It’s (Change in) the (Future) Economy, Stupid: Economic Indicators, the Media, and Public Opinion”, *American Journal of Political Science* 59, pp457-474.
Appendix A. Verifying equilibrium in news selection functions

Define $y_i(\omega)$ as the optimal action associated with state $\omega \in \Omega$ for agent $i$. A news selection function $S_i(\omega)$ determines whether agent $i$ observes $x_a$ or $x_b$. The optimal action when agent $i$ observes $x_k$ and $S_i$ can be expressed as

$$y_i(x_k, S_i) = (1 - \lambda) \sum x_i p(x_i, x_k, S_i) + \lambda \sum y_j p(y_j, x_k, S_i)$$

(A.1)

The news selection function is defined by a binary choice in each of the $3 \times 3 = 9$ states of the world, implying that there are $2^9 = 512$ different news selection functions for each information provider. The conjectured news selection functions in Section 3 can be verified to be a Nash equilibrium as follows.

1. For each possible news selection function for $S_a$
   a. Find Alice’s optimal actions in each state of the world as described by (A.1) taking Bob’s action in each state as given.
   b. Given Alice’s actions computed in Step (1a) and the conjectured news selection function $S_b$, compute Bob’s optimal action as described by (A.1).
   c. Iterate on steps (1a) and (1b) until both Alice and Bob’s actions have converged.
   d. Compute expected utility of Alice and save.

If Alice’s maximum expected utility in Step 1d coincides with the expected utility in the conjectured equilibrium, Paper A has no incentive to deviate. Because of symmetry, Paper B then also do not have an incentive to deviate, and the conjectured news selections functions are then an equilibrium.

Appendix B. Alternative model with ex ante information choice

Here we derive the solution to the alternative model discussed in Section 4 of the paper. The set up is identical to the benchmark model except that agents choose ex ante which story to get information about.

B.1. Optimal action. As in the benchmark model, the optimal action $\tilde{y}_i$ of agent $i$ is described by the first order condition

$$\tilde{y}_i = (1 - \lambda) E_i [x_i] + \lambda E_i [y_j] : i \neq j$$

(B.1)
B.2. Information choice. Define the utility of agent $i$ when she observes $X_j$ as $U_i(X_j)$. Agents choose ex ante whether to observe $X_a$ or $X_b$ and agent $i$ will choose to observe $X_i$ when the expected utility of doing so is higher than the expected utility of observing $X_j$. To solve for the information choice, we thus need to find expressions for the expected utility under the two choices.

B.2.1. Alice and Bob observe different stories. If agent $i$ observes $x_i$ and agent $j$ observes $x_j$ their respective actions are

$$\tilde{y}_i = (1 - \lambda) x_i, \quad \tilde{y}_j = (1 - \lambda) x_j.$$ 

Agent $i$’s expected utility when she observes $X_i$ is then given by

$$EU_i(X_i) = - (1 - \lambda) E [(1 - \lambda) x_i - x_i]^2 - \lambda E [(1 - \lambda) (x_i - x_j)]^2$$  \hspace{1cm} (B.2)

or

$$EU_i(X_i) = - (1 - \lambda) \lambda^2 E x_i^2 - \lambda \{(1 - \lambda)^2 E x_i^2\} = - (1 - \lambda) \lambda^2 E x_i^2$$ \hspace{1cm} (B.3)

B.2.2. Alice and Bob observe the same story. When both agents choose to observe $X_j$ the actions are given by

$$\tilde{y}_i = \frac{1 - \lambda}{1 - \lambda^2} x_j, \quad \tilde{y}_j = \frac{1 - \lambda}{1 - \lambda^2} x_j$$

The expected utility of agent $i$ then is

$$EU_i(X_j) = - (1 - \lambda) E \left[\frac{1 - \lambda}{1 - \lambda^2} x_j - x_i\right]^2 - \lambda E \left[\frac{1 - \lambda}{1 - \lambda^2} - \frac{1 - \lambda}{1 - \lambda^2} x_j\right]^2$$ \hspace{1cm} (B.4)

which can be rearranged to

$$EU_i(X_j) = - (1 - \lambda) E \left[\frac{1 - \lambda}{1 - \lambda^2} x_j - x_i\right]^2 - \lambda E \left[\frac{(1 - \lambda) (\lambda - 1)}{1 - \lambda^2} x_j\right]^2$$ \hspace{1cm} (B.5)

and simplified to

$$EU_i(X_j) = - (1 - \lambda) \lambda^2 \frac{(1 - \lambda)^2}{(1 - \lambda^2)^2} E x_j^2 - (1 - \lambda) E x_i^2 - \lambda \left\{\frac{(1 - \lambda) (\lambda - 1)}{1 - \lambda^2}\right\}^2 E x_j^2$$ \hspace{1cm} (B.6)

B.2.3. Solving for the information choice. Without loss of generality, we can normalize the variances of $X_i$ and $X_j$ to 1. The expected utilities can then be written as

$$EU_i(X_i) = - (1 - \lambda) \lambda^2 - \lambda^2 (1 - \lambda)^2$$ \hspace{1cm} (B.7)

and

$$EU_i(X_j) = - (1 - \lambda) \lambda^2 \frac{(1 - \lambda)^2}{(1 - \lambda^2)^2} - (1 - \lambda) - \lambda \frac{(1 - \lambda)^2 (\lambda - 1)^2}{(1 - \lambda^2)^2}$$ \hspace{1cm} (B.8)

Agent $i$ will choose to observe $X_i$ when $EU_i(X_i) > EU_i(X_j)$, that is, when the inequality

$$- (1 - \lambda) \lambda^2 - \lambda^2 (1 - \lambda)^2 > - (1 - \lambda) \lambda^2 \frac{(1 - \lambda)^2}{(1 - \lambda^2)^2} - (1 - \lambda) - \lambda (\lambda - 1)^2 \frac{(1 - \lambda)^2}{(1 - \lambda^2)^2}$$ \hspace{1cm} (B.9)
holds. Move all terms in (B.9) to the left hand side and divide by $(1 - \lambda)$ to get

$$ -\lambda^2 - 2(1 - \lambda) + 1 + \lambda^2 \frac{(1 - \lambda)^2}{(1 - \lambda^2)^2} + \lambda(\lambda - 1)^2 \frac{(1 - \lambda)}{(1 - \lambda^2)^2} > 0. $$  \hspace{1cm} (B.10)

The resulting inequality can then be simplified to

$$ (1 - \lambda^2)^2 + \lambda > 0 $$  \hspace{1cm} (B.11)

The inequality (B.11) holds for all $\lambda > 0$. Alice will thus choose to always observe $X_a$ when actions are strategic complements, and Bob will then also choose to always observe $X_b$. When actions are strong enough strategic substitutes, agents will choose to coordinate so that they both always observe either $X_a$ or $X_b$. While there is no simple analytical solution to (B.11), solving (B.11) numerically shows that agents will choose to observe the same variable when $\lambda \in (-1, -0.53)$. 