NEWS MEDIA AND DELEGATED INFORMATION CHOICE

KRISTOFFER P. NIMARK AND STEFAN PITSCHNER

Abstract. No agent has the resources to monitor all events that are potentially relevant for his decisions. Therefore, many delegate their information choice to specialized news providers that monitor the world on their behalf and report only a curated selection of events. We document empirically that, while different outlets typically emphasize different topics, major events shift the general news focus and make coverage more homogeneous. We propose a theoretical framework that formalizes this type of state-dependent editorial behavior by introducing *news selection functions*. We prove that (i) agents can always reduce the entropy of their posterior beliefs by delegating their information choice, (ii) state-dependent reporting conveys information not only via the contents of a story, but also via the decision of what to report, and (iii) an event that is reported by all news providers is common knowledge among agents only if it is also considered maximally newsworthy by all providers. As an application, we embed delegated news selection into a simple beauty-contest model to demonstrate how it affects actions in a setting with strategic interactions.

1. Introduction

Every day, a large number of events occur, each of them potentially relevant for the decisions of households and firms. However, no individual firm or household has the resources to observe all of these events. Therefore, many delegate their information choice to news media that monitor the world on their behalf and report only a curated selection of events. This editorial aspect of news reporting and information choice is pervasive in reality, but it has not been studied in the existing economics literature. In this paper, we empirically document several salient features of news coverage and develop a theoretical framework that allows us to analyze their economic implications.

Our analysis provides new insights into both the observable reporting behavior of news media, and the implications of this behavior for agents' beliefs and actions. Using a large number of newspaper articles, we show that, while different outlets typically emphasize different topics, important events shift their general focus and make coverage more homogeneous. To understand the implications of this observable behavior, we propose a theoretical

Date: October 16, 2018, Nimark: Economics Department, Cornell University. *e-mail*: pkn8@cornell.edu *webpage*: www.kris-nimark.net. Pitschner: Department of Economics, Uppsala University. *e-mail*:pitschner@gmail.com, *webpage*: www.stefanpitschner.com. The authors thank Mikael Carlsson, Tommaso Denti, Ed Green, Sylverie Herbert, Chris Huckfeldt, Boyan Jovanovich, Alex Kohlhas, Ilse Lindenlaub, Karel Mertens, Laura Veldkamp, Rosen Valchev, Mirko Wiederholt, Nellie Zhao and conference and seminar participants at Penn State, Sveriges Riksbank, Stockholm University, Helsinki University, SED 2015, Cornell University, Federal Reserve Bank of San Francisco, New York University, University of Tokyo, University of Kyoto, Keio University, CERGE-EI, Barcelona GSE Summer Forum 2016 and the *Big Data, Machine Learning and the Macroeconomy* conference at Norges Bank, Indiana University for useful comments and suggestions.

framework that formalizes the editorial decisions of news media and reflects the idea that information choice often involves deciding which information provider to use, rather than what variable or event to get information about.[1]

Using the proposed framework, we prove under general conditions that agents can always reduce the entropy of their posterior beliefs by delegating their information choice to an entity that makes state-dependent reporting decisions, compared to a situation in which they choose for themselves what to get information about. We also show that, when reporting decisions are state dependent, newspapers convey information not only via the contents of their articles, but also via the reporting decision itself. Finally, we show that an event that is reported by all information providers is only common knowledge if all providers also consider it maximally newsworthy.

We begin our analysis by estimating a Latent Dirichlet Allocation (LDA) topic model using texts from almost 15,000 archived newspaper stories. We consider articles from 17 different US newspapers covering 90-day periods around two major events: The 9/11 terrorist attacks in 2001, and the bankruptcy of Lehman Brothers in 2008. Using these texts, we document three stylized facts. First, different outlets generally 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 in our data set. Second, the extent of total news coverage allocated to different topics varies over time and depends on what has happened. The topics associated with, respectively, the 9/11 terrorist attacks, the 2008 political party conventions, the Lehman bankruptcy and the failed financial bailout package 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 front page coverage to them than to any other topic.

Analyzing the consequences of this type of editorial behavior requires a theoretical framework in which agents receive information from news providers that can behave as in the data. Specifically, these news providers must be able to both specialize in different topics and make state-dependent reporting decisions. In this paper, we propose to formalize the notion of the editorial decisions of news media by defining information providers in terms of their *news selection functions*. A news selection function is a provider-specific mapping from a possibly high-dimensional state of the world to a smaller set of reported events. It thus specifies what a given outlet will report in each possible state of the world. Unlike in setups where agents choose their signals ex ante, what our agents receive information about therefore depends on both which news provider they use, and what has happened.

Using this theoretical framework, we are able to prove several results under fairly general conditions. First, we show that if agents are constrained in the number of stories they can read, it is optimal for them to delegate their information choice to news providers that make state-dependent reporting decisions. Formally, by delegating the decision of what to get information about to a news organization, agents can reduce their posterior entropy relative

[1]This idea 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.*"

to any ex-ante choice of which events to observe. We thus provide a novel justification for the existence of news media and the editorial service they perform.

Second, when the information providers make state-dependent decisions about what to report, these decisions are by themselves informative about the state. For example, on a slow news day, outlets may devote most of their coverage to relatively small or mundane events. The readers of these outlets will then of course receive this reported information. However, because they see no stories about more important events, they can also infer that no such more important events have taken place. On the other hand, when an extraordinary event does occur and gets reported, agents cannot rule out the simultaneous occurrence of more mundane events. The more newsworthy a reported event is, the smaller is thus the set of events that can be ruled out.

State-dependent reporting behavior also implies that if two agents receive the same information from different providers they may draw different inferences about the state of the world. For example, consider two information providers that are, respectively, biased for and against a given politician. If neither of these newspapers report negative news about the politician in question, only the reader of the newspaper that would have reported such news can conclude that no negative event for this politician has occurred. Under delegated information choice, news media thus convey information in two distinct ways: via the actual contents of their articles, and via their decisions on what events to cover.

The third general implication from the theoretical framework is that news selection functions and distributions of events jointly determine the degree to which knowledge about an event is common among agents. The concept of common knowledge is important in game theory as well as in strategic settings more broadly.[2] We show that an event that is reported by all news providers is common knowledge among agents only if it is also considered maximally newsworthy by all providers. The form of the news selection functions thus partly determines how agents who receive information from different providers will respond to a given piece of information in a strategic setting.

The proposed framework imposes minimal structure on news selection functions, and the results described above hold regardless of how these functions are determined in practice. However, it is natural to think of news selection functions as equilibrium objects that respond to agents' preferences and their demand for information. We may also ultimately be interested in how they affect agents' actions. To study these questions we embed delegated information choice in a modified version of the simple beauty contest model of Morris and Shin (2002).

The model is a two-agent beauty-contest game in which each agent's pay-off depends on the distance of his action to an agent-specific latent variable, and the distance between his action and that of the other agent. Our agents rely on information providers that monitor the world on their behalf. Agents are heterogeneous in terms of what information they find most useful, and information providers specialize to cater to their different interests. However, because of the strategic motive in their actions, agents also have an indirect interest in events that are only important for predicting the actions of others.

[2]Surveys by Binmore and Brandenburger (1989), Brandenburger and Dekel (1993) and Geanakoplos (1994) provide powerful examples and conceptual discussions of the relevance of common knowledge.

In the model, agents react more strongly to a reported event the more probable they think it is that the other agent also knows about it. This probability depends both on agents' preferences via the news selection functions, and the distribution of events. The degree to which knowledge about an event is common among agents is thus endogenous. This is in contrast to much of the existing literature, where signals are typically either private or common knowledge by assumption, 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 the presence of strategic complementarities, there are some states of the world in which it is optimal for agents to learn about the same events, even if they intrinsically care about different latent variables. We show that under delegated information choice, agents therefore take more correlated actions than they would in a situation in which they must choose ex ante what to get information about. We also show that when states are normally distributed, realized events in the tails of the distribution tend to common knowledge in the limit.

The fact that our agents cannot directly observe the entire state of the economy 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). However, unlike the rational inattention literature, we do not restrict the magnitude of the entropy reduction implied by observing the reported news. Instead, our agents are constrained in the number of news stories they can absorb, leading the information providers in our framework to report a discrete subset of all realized events. This reflects the notion that news stories are to some extent indivisible, and that learning a little about everything is less useful than learning more about fewer events. Our framework also differs from the rational inattention literature, in that it explicitly incorporates information providers that serve as intermediaries between agents and the state of the world.

Throughout the paper, we emphasize the editorial role of information providers. This role is complementary to other aspects of news media such as the non-rivalry of information and the economies of scale in information production studied, for instance, by Veldkamp (2006b). Our setup 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 contrast, in our framework 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. What our agents get information about therefore depends on what has happened.[3]

[3]In the most general formulation of a rational inattention problem, e.g. Sims (2003), the only constraint on agents' beliefs is imposed on the reduction in entropy from agents' priors 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. However, in our framework, what agents update their beliefs about depends on both what is reported and on the news selections functions of the information providers. In our set up, a signal is only completely uninformative about a non-reported event if the decision of what to report is independent of the non-reported event. There is thus no simple mapping between what is reported and what variables agents update their beliefs about.

Like our paper, Gentzkow and Shapiro (2006, 2008) also study the editorial function of news media, but they primarily focus on identifying and analyzing the causes and consequences of ideologically slanted reporting. Perego and Yuksel (2017) study ideological slant in news media markets where agents have heterogenous preferences over both what the political agenda should be, and how issues should be addressed. In their model, increased competition results in news outlets providing more specialized content, making agents more sure of their differences in policy preferences.

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). Like the political economy literature, the political science literature also focuses primarily on documenting and analyzing ideologically or politically biased reporting. While we abstract from such biases, the mechanisms we illustrate also apply to environments in which they are present.

Finally, Nimark (2014) analyzes a set up in which a single information provider is more likely to report unusual events. However, that paper 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.

2. Some Stylized Facts of News Coverage

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

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 newspapers, 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 1. Newspapers in Database

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

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.

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), LDA models are one of the most-widely applied tools in natural language processing. A topic is defined by a frequency distribution of words, and the topics are estimated from the text corpus. LDA models 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.[4] 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 are 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. Unlike approaches that use word counts to measure news coverage, e.g. Baker, Bloom and Davis (2013), LDA models do therefore not require researchers to pre-specify words or topics of interest. Another desirable property of LDA models is that they capture not only changes in the importance of a topic

[4]Thorsrud (2018) uses LDA to classify articles from a large Danish business newspaper and relates the estimated topics to aggregate economic conditions.

Table 2. Sample Text Snippets of Newspaper Articles in the Database

Text Snippet	Newspaper	Publication Date
"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 ..."	The New York Times	2001/10/31
"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 ..."	The Washington Post	2001/09/15
"A day after dividing their votes on a failed proposal for a 700 billion Wall Street bailout, Maines two US House members agreed Tuesday that its vital for lawmakers to pass a relief bill for credit markets."	Portland Press Herald	2008/10/01
"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 commonwealths property-tax system, which raises more ... "	The Philadelphia Inquirer	2008/09/10

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.

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 the topics emerge endogenously from the analysis 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 viewed as describing 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 the document-specific topic distribution generated in (1), with N being the number of words in the document.
(3) Draw one word from each of the N topics generated in (2).

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

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

where β, θ, and z are unobserved parameters. The rows of the $K \times V$ matrix β contain the word distribution β_k for topic k, the columns of the $K \times D$ matrix θ contain the topic proportions θ_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 model 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 topic weights 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 estimate the parameters that govern the topic weights and word distributions from the 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

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

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 β, θ, and z.

2.3. **Text pre-processing and the number of estimated topics.** 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.[5] 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, the vocabulary used in sports-related articles may change little over time and form a topic that is present in news articles from both 2001 and 2008.[6]

Table 3. Estimated LDA Topics: High-Probability Words

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

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.

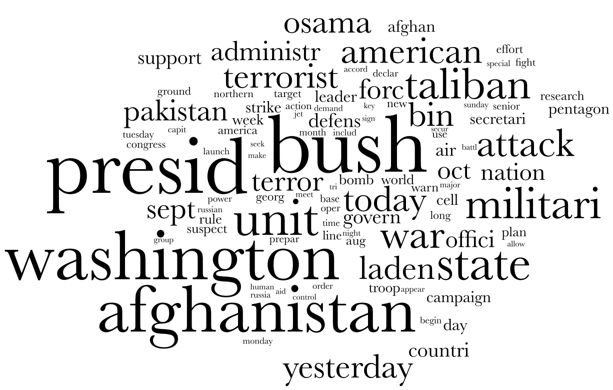
2.4. **Estimated LDA topics.** The topics in the LDA model are estimated endogenously. However, human input is generally required to interpret the resulting topics, and, in our case, associate them with particular events. Table 3 shows the topics identified by our estimated LDA model in terms of their 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 presidential candidates conventions of the 2008 US presidential elections, and Topic 9 covers the September 11 terrorist attacks.

[5]In the Online Appendix, available from the authors' web pages, we also present results from the LDA model using 5, 20, 50 and 100 topics.

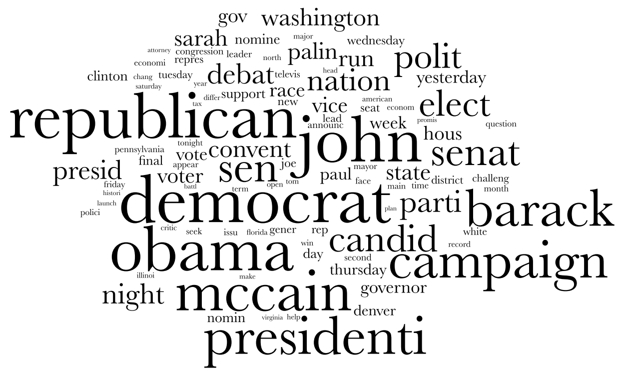
[6]If no topic occurs in both periods and when the number of documents are approximately the same for the two periods, estimating a joint LDA model for both time periods with 10 topics should yield the same topics and assigned topics weights as if we were to estimate two separate models with 5 topics for each period.

Figure 1. Estimated LDA Topics: Word Clouds of Selected Topics

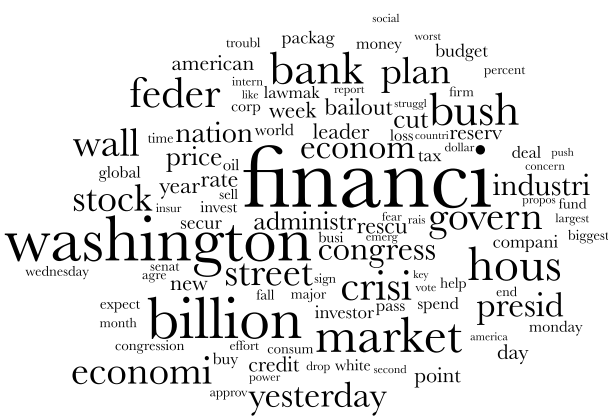
Topic 1: Afghanistan



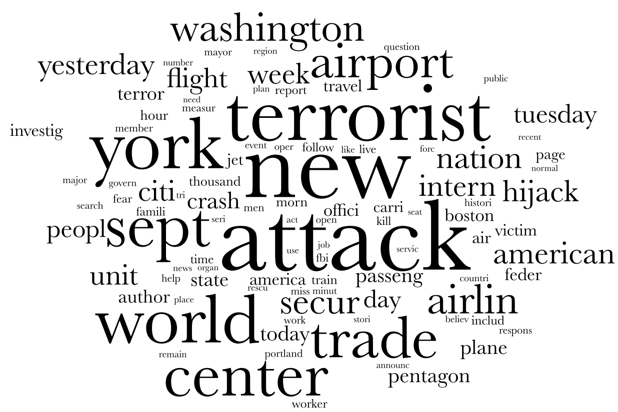
Topic 2: 2008 Presidential Candidates



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.

Furthermore, a relatively clear interpretation can also be attached to Topic 5, which seems to capture both the financial crisis and the corresponding reactions of the US government.[7]

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.

[7]In the Online Appendix, we report sentences of articles that were assigned the highest probability of belonging to each of the topics.

Accordingly, the fraction of difficult-to-interpret topics typically decreases when the number of topics is increased.[8]

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.[9] The interpretations of the four topics that we derive based on the word clouds reinforce the ones obtained from the high-probability words shown in Table 3.

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

$$d_{m,k} = \frac{p_{m,k} - p_k}{p_k} \tag{2.3}$$

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 percent 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 news paper. (A -1 deviation would imply zero coverage of a topic.)

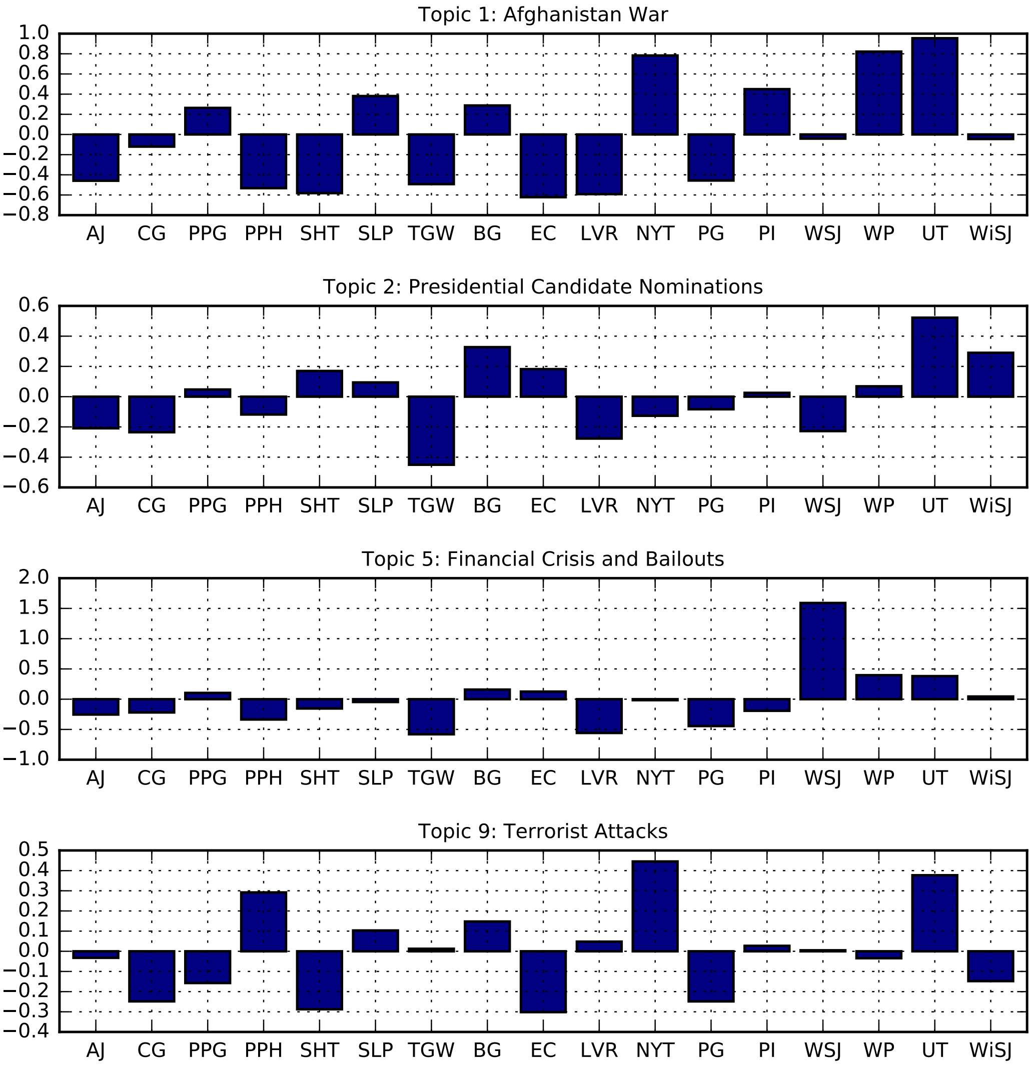
The plots document that there are large amounts of variation in terms of which newspapers tend to cover which topics. For example, the financial crisis as captured by Topic 5 received more than twice as much coverage in the Wall Street Journal than it did in the hypothetical average outlet. Similarly, 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

[8]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 split into several sub-topics, with one topic covering the actual attacks, another topic more closely related to who the suspected perpetrators were, another topic related to the US military response to the attacks, and so on.

[9]Word clouds are not to everyone's liking, see *http://www.wordle.net/show/wrdl/718619/I_hate_word_clouds*.

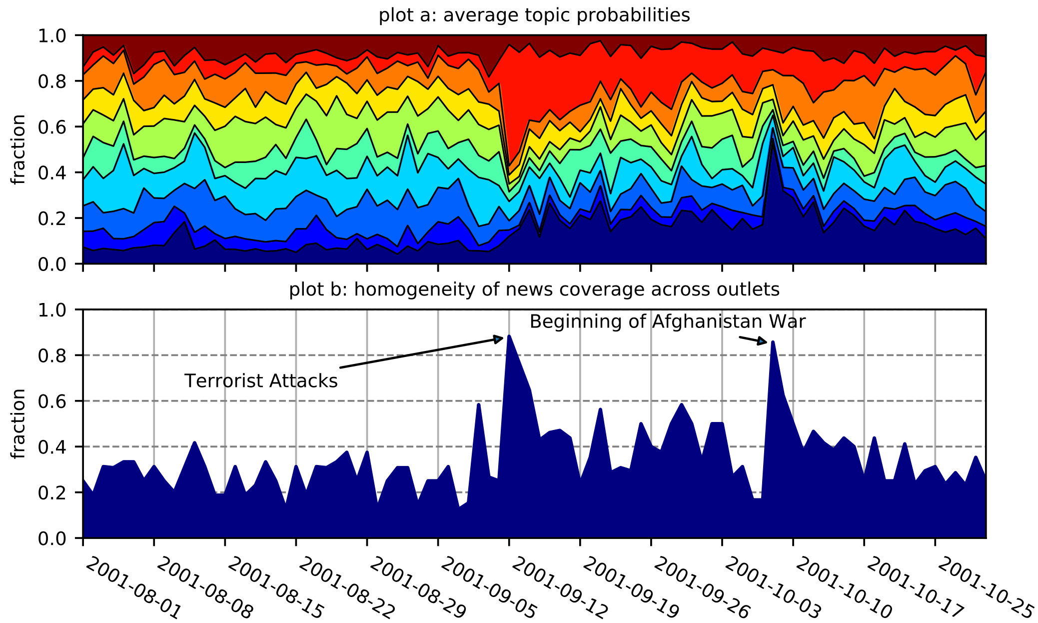
FIGURE 2. Newspaper Specialization: Probabilities of Selected Topics



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 percent 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.

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.

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-period 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.

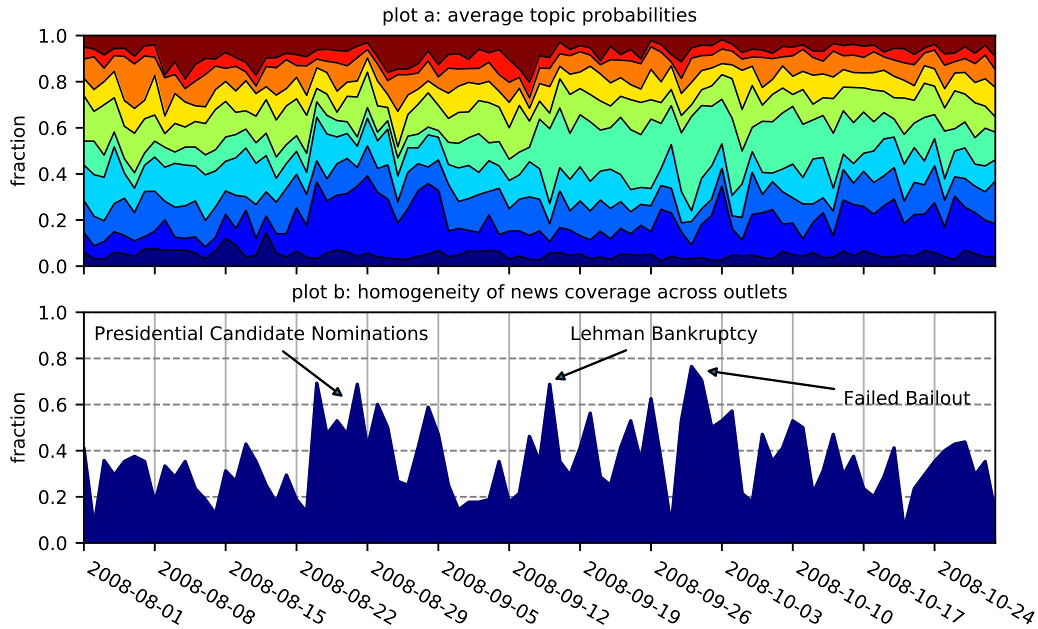
To assess if this is the case, we use two different measures. First, we calculate overall topic probabilities at a daily frequency by averaging the estimated topic probabilities of all stories in our database for a given day. The fraction $F_{t,k}$ of total news devoted to topic k at date t is thus given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

$$F_{t,k} \equiv \frac{\sum_d \theta_{t,d,k}}{D_t} \tag{2.4}$$

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 4. 2008 Financial Crisis: 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-period 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.

Homogeneity H_t of news coverage on day t is thus defined as

$$H_t \equiv \frac{\sum_m \mathcal{I}\left(\arg\max_k F_{t,m,k} = \arg\max_k F_{t,k}\right)}{M} \tag{2.5}$$

where $\mathcal{I}$ is an indicator function that takes the value 1 when the equality in parentheses holds. $F_{t,m,k}$ is the fraction of news coverage devoted to topic k by newspaper m on date t and M is the total number of newspapers. The range of H_t is thus between 0 and 1, with a value of 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. Topics are ordered from below, with Topic 1 at the bottom and Topic 10 at the top. The first striking episode occurs on September 12, when more than 80 percent of the total news coverage was devoted to the terrorism topic (Topic 9, shown in light red).[10] A second pronounced change occurs on October 8, the day after the war in Afghanistan began (Topic 1, shown in dark blue). As can be seen from the bottom panel, the same two days are also associated with sharp increases in topic homogeneity. That is, both the terror 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 affected both the focus of news coverage and its cross-sectional homogeneity. First, the presidential nomination conventions topic (Topic 2, shown in medium blue) received high levels of media coverage and caused an increase in homogeneity in late August and early September. Then, the Financial Crisis topic (Topic 5, shown in bright green) caused another spike with the Lehman Brothers Bankruptcy on September 15. 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 nominations, the Lehman bankruptcy and the failed financial bailout package all are 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 show the homogeneity measure of news coverage for alternative LDA specifications with 5, 10, 20, 50, or 100 topics, respectively. Increasing the number of topics affects the topic assignments and our heterogeneity measure in two distinct ways. First, with more topics, individual topics tend to be 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 percent with 5 topics to about 20 percent in the specification with 100 topics. Thus, while the peaks of the homogeneity measure tend to be somewhat lower in specifications with 50 or 100 topics, the relative changes after major news events remain large.

3. Delegated Information Choice

In the previous section we documented three salient properties of news coverage that reflect the observable reporting behavior of information providers. Specifically, we showed that, while different newspapers typically emphasize different topics, major events shift the general news focus and make coverage more homogeneous across outlets. In this section we develop a general theoretical framework that allows us to analyze the implications of this kind of systematic news selection. We refer to this framework as *delegated information choice*.

[10] 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.

To formalize the editorial behavior of news media, we introduce the concept of a *news selection function*. A news selection function is a mapping from states of the world to sets of reported events. For our purposes, an information provider is completely characterized by its associated news selection function. We prove formally that agents may benefit from delegating their information choice to a mechanism or organization that can make a state-dependent reporting decision. We also show how doing so affects their beliefs, and how news selection functions determine the degree to which knowledge about an event is common among agents that receive information from different news providers.

3.1. **States of the world and news selection.** The state of the world is the realized value of the n-dimensional random vector $\omega \in \Omega_1 \times \Omega_2 \times ... \times \Omega_n \equiv \Omega$. The vector ω is potentially high dimensional, i.e. n could be a large number. An information provider monitors the state of the world and then decides which subset of realized events to report. This monitoring and selection of reported events can then be formalized in terms of a news selection function.

Definition 1. *(News selection functions) A news selection function $\mathcal{S} : \Omega \to s \in \{0,1\}^n$ is a mapping from n-dimensional states of the world into n-dimensional indicator vectors. For $\omega \in \Omega$ and $s \in \{0,1\}^n$, we denote by $\omega^s = (\omega_i : s_i = 1)$ the m-dimensional vector of reported dimensions of the state.*

A news selection function $\mathcal{S}$ associates a pair (ω^s, s) with each state of the world. The vector ω^s contains the values of the reported outcomes. The vector s indicates which dimensions of ω are reported, but it does not contain their actual values. An element of s equal to 1 indicates that the corresponding dimension is reported, and a 0 indicates that the respective dimension is *not* reported. For instance, $s(\omega) = (1, 0 \ldots, 0)$ means that in state $\omega = (\omega_1, \ldots, \omega_n)$ only the first dimension is reported so that $\omega^{s(\omega)} = \omega_1$. Similarly, $s(\tilde{\omega}) = (0, \ldots, 0, 1, 1)$ means that in state $\tilde{\omega} = (\tilde{\omega}_1, \ldots, \tilde{\omega}_n)$ only the last two dimensions are reported so that $\tilde{\omega}^{s(\tilde{\omega})} = (\tilde{\omega}_{n-1}, \tilde{\omega}_n)$. A news selection function thus assigns a 1 to element i of s if the outcome ω_i is sufficiently newsworthy to be reported. Whether the element ω_i is reported or not may generally depend on the entire state vector ω.

The dimension of ω^s and the number of non-zero elements in s is m so that all outcomes are reported if $m = n$. We are mostly interested in non-trivial selections, i.e. cases where s has some zero elements so that ω^s has fewer than n coordinates.

An agent who observes ω^s and knows the function $\mathcal{S}$ has the posterior $p(\omega \mid \omega^s, s)$. By separating the realized values of the reported outcomes from the information about which variables were reported, we can distinguish between information contained in the reported vector ω^s, and information contained in the indicator vector s. The distribution $p(\omega \mid \omega^s)$ is the posterior of an agent who has observed ω^s but does not know how this vector was selected. The distribution $p(\omega \mid s)$ is the posterior of an agent who knows which variables were reported and how they were selected, but who does not know their realized values.

3.2. **Delegated information choice reduces entropy.** Consider an agent who is constrained to observe only $m < n$ possible stories. By delegating the decision of what to get information about to an organization that makes a state-dependent reporting decision, the agent can avoid spending resources on information that ex post turns out not to be useful. Of course, what constitutes useful information depends on the particular setting. However, one can show that, for any ex-ante choice of which m stories to observe, there always exists a news selection function that results in a lower posterior entropy. To make this point formally, we define an ex-ante information choice function as follows.

Definition 2. *(Ex ante information choice) An ex-ante information choice function $\widetilde{\mathcal{S}}$ is defined by an n-dimensional indicator vector $\widetilde{s} \in \{0,1\}^n$ with the $\widetilde{m}$-dimensional random vector $\omega^{\overline{s}}$ of observed outcomes is defined as $\omega^{\overline{s}} \equiv \{\omega_i : \widetilde{s}_i = 1\}$. The indicator vector $\widetilde{s}$ is independent of the state ω.*

As long as the selection is non-trivial, that is, if $\widetilde{m} < n$, it is always possible to find a state-dependent news selection function that reveals strictly more information, without increasing the number of reported outcomes.[11]

Proposition 1. *For any given ex-ante information choice function $\widetilde{\mathcal{S}}$ such that $\widetilde{m} < n$ there exists a news selection function $\mathcal{S}^*$ with $m^* = \widetilde{m}$ that achieves a lower posterior entropy.*

Proof. Start by fixing an indicator vector $\widetilde{s}$ associated with some ex-ante information choice function $\widetilde{\mathcal{S}}$. Define the candidate news selection function $\mathcal{S}^*$ so that $s^*(\omega) = \widetilde{s}$ in every state except ω'. In state ω', for an i and j such that $\widetilde{s}_i = 1$ and $\widetilde{s}_j = 0$, set $s_i^*(\omega') = 0$ and $s_j^*(\omega') = 1$. When $\omega = \omega'$, $\mathcal{S}^*$ then reveals the entire state vector ω, since it is only when $\omega = \omega'$ that $s_j^* = 1$. When $\omega \neq \omega'$, the candidate news selection function $\mathcal{S}^*$ reveals the same vector of outcomes $\omega^{\overline{s}}$ as the ex-ante information choice $\widetilde{\mathcal{S}}$. However, since it is only when $\omega \neq \omega'$ that $s_j^* = 0$, $\mathcal{S}^*$ then also reveals that $\omega \neq \omega'$. The entropy H of ω conditional on $\mathcal{S}^*$ thus satisfies the equality

$$H\left(\omega \mid \omega^{s^*}, s^*\right) = H\left(\omega \mid \omega^{\overline{s}}, s_j^*\right). \tag{3.1}$$

To prove the proposition we need to show that $\mathcal{S}^*$ reveals at least as much information as $\widetilde{\mathcal{S}}$. By the properties of entropy (e.g. Theorem 2.6.5 in Cover and Thomas 2006), we know that conditioning on additional information cannot increase entropy, implying the inequality

$$H\left(\omega \mid \omega^{\overline{s}}, s_j^*\right) \leq H\left(\omega \mid \omega^{\overline{s}}\right). \tag{3.2}$$

(3.2) holds with equality if only if ω is conditionally independent of s_j^* given $\omega^{\overline{s}}$. However, since $p\left(\omega \mid \omega^{\overline{s}}, s_j^*\right) \neq p\left(\omega \mid \omega^{\overline{s}}\right)$ the inequality (3.2) must be strict, which completes the proof. □

Proposition 1 shows that for any ex-ante information choice there exists a state-dependent news selection function $\mathcal{S}^*$ that reveals more information. To achieve this outcome, the state-dependent news selection has to be *delegated*. If agents were to monitor all outcomes on their own and then were to select only m of these outcomes ex post, they would still have to pay the attentional or pecuniary cost of first observing all n outcomes. The editorial service performed by newspapers is useful to agents since it allows them to achieve a lower posterior uncertainty about the state of the world, without increasing the number of stories

[11] $\overline{\mathcal{S}}$ is referred to here as a *function* even though $\overline{s}$ is not state dependent. The reason for this is that it is natural to think of it as an equilibrium object that is determined by preferences and/or technology. Section 4 provides an example where this is the case.

they read. Proposition 1 thus provides a formal justification for the existence of the type of state-dependent news selection that we document in the data.

While Proposition 1 shows that delegating the choice of what to get information about can lead to a lower entropy, the candidate function $\mathcal{S}^*$ constructed in the proposition is not necessarily optimal. In fact, the proposition is silent on how to choose the state ω' in which $\mathcal{S}^*$ deviates from $\widetilde{\mathcal{S}}$, and it does not suggest that deviating in only one state is optimal. However, to say more about optimality, we need additional structure in the form of utility functions and action spaces. In Section 4 we present a simple but complete model that can be used to study optimal news selection functions, and how they depend on agents' preferences.

3.3. **Beliefs and news selection functions.** The proposition above demonstrates that delegated news selection can reduce the entropy of an agent's posterior beliefs. Implicit in the proof is that a provider's decision of what to report is by itself informative about the state of the world. We now derive general conditions under which that is the case.

Reading a news report ω^s is clearly informative about the outcomes in the vector ω^s itself. However, when the news provider's reporting decision depends on the state ω, agents may receive additional information. This information is conveyed by the reporting decision itself, and it allows them to also update their beliefs about those dimensions of ω that are not reported. To see how, define the vector of realized but unreported outcomes as $\omega^{\not s} \equiv \{\omega_i : s_i = 0\}$. We can then state the following proposition.

Proposition 2. *Posterior beliefs about the unreported stories $\omega^{\not s}$ coincide with $p\left(\omega^{\not s} \mid \omega^s\right)$ if and only if the probability of reporting about $\omega^{\not s}$ is conditionally independent of ω^s. That is*

$$p\left(\omega^{\not s} \mid \omega^s, s\right) = p\left(\omega^{\not s} \mid \omega^s\right) \tag{3.3}$$

if and only if

$$p\left(s \mid \omega^s\right) = p\left(s \mid \omega^s, \omega^{\not s}\right). \tag{3.4}$$

Proof. By Bayes' rule the posterior beliefs about the unreported variables $\omega^{\not s}$ is given by

$$p\left(\omega^{\not s} \mid \omega^s, s\right) = \frac{p\left(s \mid \omega^s, \omega^{\not s}\right)}{p\left(s \mid \omega^s\right)} p\left(\omega^{\not s} \mid \omega^s\right). \tag{3.5}$$

It then follows immediately that (3.3) holds if and only if

$$\frac{p\left(s \mid \omega^s, \omega^{\not s}\right)}{p\left(s \mid \omega^s\right)} = 1 \tag{3.6}$$

which completes the proof. □

In a standard prediction problem, the joint distribution of the two variables is sufficient to determine what can be learned about $\omega^{\not s}$ from observing ω^s. But the state-dependent news selection performed by the information providers makes agents update their beliefs about the unreported outcomes beyond what is implied by the joint distribution of ω^s and $\omega^{\not s}$.

Proposition 2 is general and holds for all distributions of ω, but a special case is particularly illustrative. Consider a setup where all the elements in ω are independent so that $p\left(\omega^{\not s} \mid \omega^{s}\right) = p\left(\omega^{\not s}\right)$. Observing ω^{s} is then by itself uninformative about $\omega^{\not s}$. But Proposition 2 states that if the probability of reporting the outcomes in ω^{s} depends on the realized outcomes of events not included in ω^{s}, the fact that ω^{s} was reported *is* informative about the outcomes in $\omega^{\not s}$.

The implications of Proposition 2 are starkest if there are states of the world in which some outcomes not currently in ω^{s} would have been reported had they occurred. Since these outcomes were not reported, they can then be ruled out. For example, readers of the Wall Street Journal or the Financial Times know that when stock market crashes occur, these outlets will always report them. Therefore, the absence of such reporting allows these readers to conclude that no crash has taken place.

The fact that certain outcomes can be ruled out if they are not reported has broader implications. To see this, define the following two sets of states that contain, respectively, more newsworthy and less newsworthy outcomes than the outcome ω_j.

Definition 3. *(Set of more newsworthy states)* $\mathcal{M}\left(\mathcal{S}, \omega_j\right) \equiv \{\omega : s_j = 0\}$ *is the set of states that, according to* $\mathcal{S}$*, contain at least* m *outcomes that are more newsworthy than* $\omega_j \in \Omega_j$.

The set $\mathcal{M}\left(\mathcal{S}, \omega_j\right)$ contains all states such that if any state in that set occurs, ω_j is not reported by an information provider characterized by $\mathcal{S}$. Therefore, the larger this set is, the less newsworthy is ω_j considered by this provider. Similarly, the set of less (or equally) newsworthy states can be defined as follows.

Definition 4. *(Set of less or equally newsworthy states)* $\mathcal{L}\left(\mathcal{S}, \omega_j\right) \equiv \{\omega : s_j = 1\}$ *is the set of all states that, according to* $\mathcal{S}$*, do not contain* m *elements that are more newsworthy than* ω_j.

The set $\mathcal{L}\left(\mathcal{S}, \omega_j\right)$ is the set of all states that are consistent with ω_j being reported by $\mathcal{S}$. We say that, the larger the set $\mathcal{L}\left(\mathcal{S}, \omega_j\right)$ is, the more newsworthy is ω_j considered, since there are then many states that could occur without changing what $\mathcal{S}$ reports.

Corollary 1. *Reporting of the least newsworthy outcomes leads to the largest number of states that can be ruled out, and reporting of the most newsworthy outcomes leads to the smallest number of states that can be ruled out.*

Proof. By Proposition 2, all states in $\mathcal{M}\left(\mathcal{S}, \omega_j\right)$ can be ruled out if $s_j = 1$. The larger the set $\mathcal{M}\left(\mathcal{S}, \omega_j\right)$ is, the larger is thus the number of states that can be ruled out. A minimally newsworthy outcome $\underline{\omega}_j$ is associated with the set $\mathcal{L}\left(\mathcal{S}, \underline{\omega}_j\right)$ being a singleton. The set $\mathcal{M}\left(\mathcal{S}, \underline{\omega}_j\right)$ is then maximally large, implying that all states except the one consistent with $\underline{\omega}_j$ and $s_j = 1$ can be ruled out. There is then no posterior uncertainty about the state ω. In the opposite direction, a maximally newsworthy outcome $\overline{\omega}_j$ is always reported whenever it occurs. The set $\mathcal{L}\left(\mathcal{S}, \overline{\omega}_j\right)$ is then maximally large and equal to all states in Ω consistent with $\omega_j = \overline{\omega}_j$. The set of more newsworthy states $\mathcal{M}\left(\mathcal{S}, \overline{\omega}_j\right)$ is then empty, so that no states can be ruled out. $\square$

Reporting of less newsworthy events thus leads to larger reductions in uncertainty than reporting of more newsworthy events. This is intuitive: When only mundane events appear on the front page of a newspaper, readers can rule out all other events that would have been more newsworthy. On the other hand, when an extreme event such as a major terrorist attack is reported, readers understand that almost anything else could have also occurred without replacing the terrorist attack on the front page of the paper.

Another corollary of Proposition 2 is that under state-dependent news selection, how agents update their beliefs depends not only on what information they receive, but also on who they receive it from.

Corollary 2. *Individuals who observe the same reported events but from different information providers may draw different inference about* ω*.*

Proof. Define the two news selection functions $\mathcal{S}$ and $\mathcal{S}'$ so that $\mathcal{M}(\mathcal{S}, \omega_j) = \varnothing$ and $\mathcal{M}(\mathcal{S}', \omega_j) \neq \varnothing$ for some $\omega_j \in \Omega_j$. An agent who gets the report ω_j from the provider characterized by $\mathcal{S}'$ can then rule out more states of the world than an agent who receives the same report from the information provider characterized by $\mathcal{S}$. □

The proof uses a simple abstract example to prove the corollary by construction. A more concrete, and perhaps more interesting, example is if two newspapers have news selection functions that are, respectively, biased for and against a given politician. If neither of the newspapers report negative news about the politician in question, only the reader of the newspaper that would have reported such news can conclude that there has been no negative events to report about the politician.

3.4. **Newsworthiness and common knowledge.** So far, we have only analyzed how an agent's beliefs about the state ω depend on the news selection function. But news selection also affects agents' higher-order beliefs, i.e. what agents believe about the beliefs of other agents and the degree to which knowledge about an event is common among agents who receive information from different providers.

An event is common knowledge if all agents know that the event has occurred, that all agents know that all agents know that the event has occurred, that all agents know that all agents know that all agents know, and so on, ad infinitum. Higher-order beliefs and common knowledge are central concepts in strategic games. The notion of common knowledge was alluded to already by Schelling (1960), but it was first formalized by Aumann (1976). Since then, it has received a lot of attention in the theory literature. One reason for this is that common knowledge is a strong assumption, and that relaxing it in seemingly innocuous ways can in fact lead to very different outcomes, e.g. the electronic mail game of Rubinstein (1989). Surveys by Binmore and Brandenburger (1989), Brandenburger and Dekel (1993) and Geanakoplos (1994) present canonical examples of settings where common knowledge is central to agents' behavior. Given the centrality of common knowledge in strategic settings, one may ask under what circumstances common knowledge about an event can be achieved among agents that receive information from different providers.

We follow Brandenburger and Dekel (1987) and use a Bayesian notion of knowledge. In the iterative definition of common knowledge stated above, this implies replacing each occurrence of "know" with "believe with probability 1". However, an equivalent definition that is more practical for our purposes is known as the fixed-point characterization of common knowledge.[12]

Lemma 1. *An outcome ω_j' is common knowledge whenever it occurs if and only if all agents assign probability 1 to $\omega_j = \omega_j'$ whenever ω_j' occurs.*

Proof. If all agents assign probability one to $\omega_j = \omega_j'$ whenever ω_j' occurs, then ω_j' is evident knowledge for all agents. For a proof that this implies that the outcome is also common knowledge, see p.174 of Monderer and Samet (1989). $\square$

We need to expand our notation to allow for multiple agents that receive information from different providers. Denote the set of agents as $\mathcal{A}$. Analogously to the notation used for the single agent case, denote the news selection function that determines what agent a observes as $\mathcal{S}^a$, and the associated vector of observed outcomes as $\omega^{s,a}$. The indicator vector s^a is defined so that $s_i^a = 1$ if $\omega_i \in \omega^{s,a}$. With this notation in place, the next proposition derives conditions under which an outcome that is reported by all providers is not merely mutual knowledge among agents, but also common knowledge.

Proposition 3. *The outcome ω_j' is common knowledge when reported by all providers if and only if $\mathcal{M}\left(\mathcal{S}^a, \omega_j'\right) = \varnothing$ for all $a \in \mathcal{A}$.*

Proof. The condition $\mathcal{M}\left(\mathcal{S}^a, \omega_j'\right) = \varnothing$ for all $a \in \mathcal{A}$ is a necessary and sufficient condition for the outcome $\omega_j = \omega_j'$ to be reported by all providers whenever ω_j' occurs. Because ω_j' is always reported by all providers, $\omega_j = \omega_j'$ is an evident knowledge event. The result then follows from Lemma 1. $\square$

The condition in Proposition 3 is quite stringent, but it does capture one reason why we may think of events such as the 9/11 terrorist attacks as being close to common knowledge. As we documented above, the 9/11 terrorist attacks were widely reported and generated a large spike in our measure of news coverage homogeneity. However, that an event is reported by all information providers does not necessarily imply that it is also common knowledge. Instead, what is required is that it is inconceivable that any news outlet would not report the event in question. Most people would arguably be surprised to learn that a friend or a colleague was unaware of the 9/11 terrorist attacks. This sense of surprise would arise from a belief that an event such as these attacks must be considered maximally newsworthy by all news outlets.[13] It is this common understanding of how newsworthy such an event is that makes it common knowledge, not the fact that it happens to be reported by all newspapers (though that is one consequence).

Common knowledge of a particular outcome thus relies on an underlying assumption that the news selection functions themselves are also common knowledge. This has echoes of the discussion regarding the implicit self-reference in definitions of common knowledge that

[12]The equivalence of the fixed-point and iterative definitions of common knowledge is one of the central results in Aumann (1976). However, Aumann did not use these terms. As far as we can tell, the terminology originated with Barwise (1987), reprinted as Barwise (2016).

[13]Hence the expression, "Have you been living under a rock?", implying that only if you have received *no news at all*, could you possibly be unaware of the event in question.

require all agent's information partitions to be common knowledge, e.g. Brandenburger and Dekel (1993).[14]

Checking whether an outcome is considered maximally newsworthy for a given news selection function is straightforward in a setting where Ω is finite. However, if the state space is continuous, there may be no outcomes that strictly satisfy this condition. Yet, there may still be some outcomes that are almost common knowledge. In order to make the notion of "almost common knowledge" more specific, we define the following two concepts.

Definition 5. *(Common p-belief) It is a common p-belief that $\omega_j = \omega_j'$, if all agents believe with probability at least p that $\omega_j = \omega_j'$, that all agents believe with at least probability p that all agents believe with probability at least p that $\omega_j = \omega_j'$, and so on.*

Definition 6. *(Approximate common knowledge) It is approximate common knowledge that $\omega_j = \omega_j'$ if it is common p-belief that $\omega_j = \omega_j'$, with p arbitrarily close to 1.*

Common p-beliefs were introduced by Monderer and Samet (1989). Using these definitions, we can now make the statement that some outcomes are "almost" to common knowledge formally.

Proposition 4. *The outcome ω_j tends to approximate common knowledge almost surely as $\omega_j' \to c$ if*

$$\lim_{\omega_j' \to c} p\left(\omega \in \mathcal{M}\left(\mathcal{S}^a, \omega_j\right) \mid \omega_j\right) = 0 \tag{3.7}$$

for all $a \in \mathcal{A}$.

Proof. We need to show that as $\omega_j' \to c$, it is almost surely approximate common knowledge that $\omega_j = \omega_j'$. The condition in the proposition directly implies that for ω_j' close enough to c, all agents almost surely know that $\omega_j = \omega_j'$. If all agents almost surely know that $\omega_j = \omega_j'$ and they know that for ω_j' close enough to c, all agents almost surely know that $\omega_j = \omega_j'$, then all agents almost surely know that all agents almost surely know that $\omega_j = \omega_j'$. By induction, this argument can be extended to any order of p-beliefs. Since an event that occurs almost surely occurs with probability one, the desired result follows. □

The proposition states that an outcome ω_j tends to common knowledge in the limit c, if the probability of a more newsworthy state occurring vanishes as $\omega_j' \to c$. One natural way to think of the limit c is in a setting where outcomes that are more extreme in some direction are considered more newsworthy. For instance, if Ω_j is the percentage point change in the Dow Jones Industrial Average stock price index, then one limit value of ω_j may be $c = -100\%$. In other words, a sufficiently large stock market crash would be reported by all newspapers and understood by all agents to be considered maximally newsworthy by all news outlets.

Monderer and Samet (1989), Sonsino (1995) and Monderer and Samet (1995) showed that common p-beliefs are a natural notion of "almost common knowledge" in the sense that economic outcomes are in many settings continuous in the limit of common p-beliefs with $p = 1$. This is also the case in the model we present in the next section.

[14]The conditions in Proposition 3 are sufficient but not necessary for an outcome to be common knowledge. The model presented in Section 4 includes an example of a state that is not reported by all providers and yet common knowledge.

4. A Beauty Contest Model with Delegated Information Choice

In this section, we embed delegated information choice in a two-agent beauty contest model in the spirit of Morris and Shin (2002). The purpose of the model is to demonstrate how systematic news selection can affect agents' beliefs and actions in a strategic setting. Our agents are heterogeneous in terms of what information they find most useful, and information providers specialize to cater to their different interests. However, because of a strategic motive in their utility functions, agents also have an indirect interest in events that are only important for predicting the actions of others. We assume that news providers are benevolent and maximize the utility of their readers. Unlike Crawford and Sobel (1982) and the literature following in that tradition, we thus abstract from any strategic considerations between the sender and the receiver of information. We also abstract from potential reporting biases such as political slant or a focus on negative events. Here, we demonstrate that even in the absence of such strategic considerations or biases, agents' beliefs and actions may still be affected by providers' systematic selection of what to report.

4.1. **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, ω_a and ω_b. A potential story $\omega_i : i \in \{a, b\}$ is a random variable that takes values in Ω_i. The state of the world is described by the pair $(\omega_a, \omega_b) = \omega \in \Omega_a \times \Omega_b = \Omega$. We say that an outcome ω_i is of interest to Alice or Bob if knowing about it allows them to take an action that increases their expected utility.

4.1.1. *Utility and heterogenous interests.* Alice and Bob find different information interesting, and this heterogeneity is introduced via their utility functions. Alice's utility depends on the distance between her action y_a and the latent variable ω_a, as well as on the distance between her action and Bob's action y_b. This is formalized by the following utility function for Alice

$$U_a = -(1-\lambda)(y_a-\omega_a)^2 - \lambda(y_a-y_b)^2. \tag{4.1}$$

In the original beauty contest model of Morris and Shin (2002), the payoffs of all agents depend on the same latent fundamental. We deviate from this setup and introduce heterogeneity with respect to the variables Alice and Bob are fundamentally interested in. Specifically, while Alice wants to take an action that is close to the latent variable ω_a, Bob wants to taken an action that is close to the latent variable ω_b. His preferences are thus given by

$$U_b = -(1-\lambda)(y_b-\omega_b)^2 - \lambda(y_b-y_a)^2. \tag{4.2}$$

We say that Alice has a direct interest in ω_a because her utility depends directly on the realized value of ω_a. Symmetrically, Bob has a direct interest in ω_b. The parameter $\lambda \in (-1, 1)$ governs the strength of the strategic motive. When $\lambda \neq 0$, Alice has an indirect interest in knowing about ω_b since that may help her better predict Bob's action. Symmetrically, Bob then has an indirect interest in knowing about ω_a.

The optimal action y_i is given by the first order condition

$$y_i = (1-\lambda)E_i[\omega_i] + \lambda E_i[y_j] : i, j \in \{a, b\}, i \neq j \tag{4.3}$$

where E_i denotes the expectations operator conditional on an agent's information set (which we define below). If the agents could observe both ω_a and ω_b directly, the equilibrium action would be described by

$$y_i = \frac{1}{1+\lambda}\omega_i + \frac{\lambda}{1+\lambda}\omega_j : i, j \in \{a, b\}, i \neq j. \tag{4.4}$$

However, neither ω_a nor ω_b are directly observable. Instead, Alice and Bob rely on information providers that monitor the state of the world on their behalf and report only the most interesting outcomes.

4.2. **Information providers.** There are two information providers, Paper A and Paper B. Both newspapers decide what to report in order to maximize the expected utility of their respective readers. By assumption, 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.[15]

To make the selection non-trivial, we assume that each newspaper must choose only one dimension of the state to report. In the notation of Section 3, we thus have $n = 2$ and $m = 1$. When choosing what to report, each newspaper takes the news selection function of the other newspaper as given. The news selection functions given by

$$\mathcal{S}_i = \arg\max_{\mathcal{S}_i} E\left[U_i\left(\omega, \mathcal{S}_j\right)\right] : i, j \in \{a, b\}, i \neq j. \tag{4.5}$$

thus determine what each agent observes in each state of the world.

4.3. **Discrete states of the world.** We first study the implications of delegated information choice for agents' beliefs and actions in a simple setting where the state space Ω is discrete. While we relax this assumption later, the simple setup 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.

To this end, for $\omega_a, \omega_b \in \{-1, 0, 1\}$ let the different outcomes occur with probabilities given by

$$p_i(\omega_i = -1) = \frac{1}{4}, \quad p_i(\omega_i = 0) = \frac{1}{2}, \quad p_i(\omega_i = 1) = \frac{1}{4} : i \in \{a, b\}. \tag{4.6}$$

The potential stories ω_a and ω_b are thus identically distributed zero mean random variables. We also assume that ω_a and ω_b are independent of one another so that

$$p_i(\omega_i \mid \omega_j) = p_i(\omega_i) : i, j \in \{a, b\}, i \neq j. \tag{4.7}$$

Neither of these assumptions are necessary for what follows, but they help simplify the presentation.

[15]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.

4.4. **Equilibrium news selection functions and actions.** Equilibrium is determined in two stages. First, the newspapers choose news selection functions $\mathcal{S}_i$ in order to maximize the expected utility of their respective readers. Then, the state ω is realized, and the relevant elements of ω as determined by $\mathcal{S}_a$ and $\mathcal{S}_b$ are reported to Alice and Bob, respectively, who then choose their actions y_a and y_b.

When $\lambda \neq 0$, what information will maximize Alice's utility 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, and vice versa. Equilibrium in the news selection game is a fixed point at which neither newspaper wants to change its selection function, taking the conditional actions of the agents and the other paper's selection function as given. Due to this fixed-point nature of the equilibrium, we here first conjecture explicit news selection functions for Paper A and Paper B, without providing a proof of optimality. We then derive Alice and Bob's implied optimal actions conditional on these news selection functions. In the appendix, we describe how to verify that the postulated news selection functions indeed constitute a Nash equilibrium.

4.4.1. *No strategic motive.* We discuss equilibrium outcomes with and without a strategic motive separately. As a benchmark, consider first the case in which $\lambda = 0$ so that agents' do not have an incentive to coordinate. In this case, it is optimal for Paper A to always report ω_a since Alice's utility then depends neither directly nor indirectly on ω_b. Symmetrically, it is optimal for Paper B to always report ω_b. In the absence of a strategic motive in actions, the news selection functions are thus given by $s^i_j = 1$ for $i = j$ and $s^i_j = 0$ for $i \neq j$. Alice and Bob's equilibrium actions are then trivially given by $y_a = \omega_a$ and $y_b = \omega_b$. Since ω_a and ω_b are independent random variables, so are Alice's and Bob's actions.

4.4.2. *Strategic motives.* 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 functions are given by

$$s^i_i = \begin{cases} 0 & \text{if } \omega_i = 0 \text{ and } \omega_j \in \{-1, 1\}, i \neq j \\ 1 & \text{otherwise} \end{cases} . \tag{4.8}$$

Given this news selection function, Paper A will report about ω_a most of the time. However, if Alice wants to take an action close to that of Bob, it is optimal for Paper A to report about ω_b whenever $\omega_a = 0$ and $\omega_b \neq 0$. Since Paper A only reports ω_b when $\omega_a = 0$, the entire state vector is then revealed to Alice. Because she also knows that she only reads about ω_b when Bob does so as well, delegated and state-dependent information choice implies that in some states of the world, Alice faces no uncertainty about either the state or Bob's action.

Changing the sign of λ, so that each agent wants to take an action far from that of the other agent, leaves the equilibrium news selection functions unchanged. When $\omega_a = 0$ and $\omega_b \neq 0$, it is still more useful to Alice to observe ω_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 λ is positive or negative, there are states of the world in which Alice and Bob prefer to know what the other agent knows, rather than the value of the variable they have a direct interest in.

Table 4 shows the news selection functions derived above and thus indicates what each newspaper reports in each state of the world (ω_a, ω_b). The top panel describes what is reported in the absence of a strategic motive, the bottom panel describes what is reported when a strategic motive is present.

TABLE 4. News selection functions

No strategic motive ($\lambda = 0$)

Paper A	$\omega_a = -1$	$\omega_a = 0$	$\omega_a = 1$	Paper B	$\omega_a = -1$	$\omega_a = 0$	$\omega_a = 1$
$\omega_b = -1$	A	A	A	$\omega_b = -1$	B	B	B
$\omega_b = 0$	A	A	A	$\omega_b = 0$	B	B	B
$\omega_b = 1$	A	A	A	$\omega_b = 1$	B	B	B

Strategic motive ($\lambda \neq 0$)

Paper A	$\omega_a = -1$	$\omega_a = 0$	$\omega_a = 1$	Paper B	$\omega_a = -1$	$\omega_a = 0$	$\omega_a = 1$
$\omega_b = -1$	A	**B**	A	$\omega_b = -1$	B	B	B
$\omega_b = 0$	A	A	A	$\omega_b = 0$	**A**	B	**A**
$\omega_b = 1$	A	**B**	A	$\omega_b = 1$	B	B	B

Notes: News selection of Paper A and Paper B functions of ω_a (columns) and ω_b (rows). The letter in a cell corresponds to the reported story in the relevant state. The top panel describes the news selection functions when there is no strategic motive in agents' actions. The bottom panel describes the news selection functions when agents have a strategic motive.

4.4.3. *Conditional actions.* Given the news selection functions (4.8), Bob only observes ω_a in states when Alice does so as well. Symmetrically, if Alice observes ω_b, then so does Bob. Alice's expectation of Bob's action conditional on Bob observing a different variable is zero, and vice versa. Combined with the first order condition (4.3), this implies that the optimal conditional actions can be described as

$$y_i\left(\omega_i, s_i^i = 1\right) = \frac{(1-\lambda)}{1-\lambda^2 p\left(s_i^j = 1 \mid \omega_i, s_i^i\right)}\omega_i \tag{4.9}$$

and

$$y_j\left(\omega_i, s_j^j = 0\right) = \lambda\frac{(1-\lambda)}{1-\lambda^2 p\left(s_i^j = 1 \mid \omega_i, s_i^i\right)}\omega_i. \tag{4.10}$$

The first of these equations describes the action of an agent who observes the variable he or she has a direct interest in. The second equation describes the action of an agent who observes the variable he or she has only an indirect interest in. The probability in the denominator of both equations is the probability that the agent without a direct interest in ω_i nevertheless observes it, conditional on the information available to the agent that does have a direct interest in ω_i. The responses of both agents are thus increasing in this probability. Given the distributional assumptions (4.6) and the news selection function (4.8), this probability is equal to $\frac{1}{2}$.

4.5. **Strategic motives and knowing what others know.** That our agents prefer to know what the other agent knows even in the presence of strategic substitutability contrasts with the result in the coordination game in Hellwig and Veldkamp (2009). There, ex-ante identical agents can choose 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. This difference arises 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 (2009) 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.[16]

4.6. **News selection functions and common knowledge.** The conditional actions (4.9) and (4.10) illustrate that knowing what agents observe is not sufficient for us to be able to predict their actions. In a strategic setting, agents respond more strongly to a given outcome if they believe it is likely that the other agent observes the same outcome as them. This belief is determined jointly by the observed outcomes, the distribution of the state, and the agents' knowledge of the news selection functions.

To see how the news selection functions affect actions, consider the state the state $(1, 0)$, in which both newspapers report $\omega_a = 1$ so that the value of ω_a is mutual knowledge. Yet, ω_a is not common knowledge. While Bob knows that he reads about ω_a only in those states where Alice does so as well, Alice cannot draw the same inference. She knows that she would observe $\omega_a = 1$ also in the states $(1, -1)$ and $(1, 1)$. Since these states occur with probability $\frac{1}{2}$, Alice's and Bob's responses are then given by the expressions above, where the probability in the denominator is $\frac{1}{2}$.

For a different set of news selections functions, Alice and Bob may have been able to infer that whenever $\omega_a = 1$, both newspapers always report the value of ω_a. The outcome $\omega_a = 1$ is then evident knowledge and thus common knowledge whenever it occurs. Ceteris paribus, Alice and Bob's respective responses, which we denote $\widehat{y}_a$ and $\widehat{y}_b$, would then be given by

$$\widehat{y}_a = \frac{(1-\lambda)}{1-\lambda^2}\omega_a, \quad \widehat{y}_b = \lambda\frac{(1-\lambda)}{1-\lambda^2}\omega_a. \tag{4.11}$$

In this alternative scenario, the agents observe the same outcomes in the state $(1, 0)$ as in the model above, yet the responses $\widehat{y}_a$ and $\widehat{y}_b$ to what they observe are stronger than those implied by the news selection functions (4.8). When agents delegate their information choice

[16]Sufficiently strong complementarities result in multiple equilibria in news selection strategies. This case is discussed in the Online Appendix.

to different providers, the form of the news selection functions are central for understanding how agents respond to a given set of reported outcomes.[17]

4.7. **Correlation of actions with and without delegated news selection.** In some states of the world, the two newspapers report the same outcomes. Unlike in a situation where agents cannot delegate the choice of which variable to get information about, delegation introduces correlation in agents' actions of the same sign as λ. In the Appendix, we show that for values of λ such that

$$\left(1-\lambda^{2}\right)^{2}+\lambda>0 \tag{4.12}$$

Alice would choose to always observe ω_a if she had to decide ex ante which variable to observe.[18] Symmetrically, Bob would then choose to always observe ω_b. Since ω_a and ω_b are independent, observing ω_a is then uninformative about ω_b and vice versa. The conditional expectation of the unobserved variable is then equal to its unconditional mean, and the optimal action $\widetilde{y}_i$ with ex-ante information choice is given by

$$\widetilde{y}_i=(1-\lambda)\,\omega_i : i\in a,b. \tag{4.13}$$

If ω_a and ω_b are independent, then so are Alice's and Bob's actions. However, this is not the case with delegated information choice.

Proposition 5. *Delegated news selection introduces a correlation between Alice's and Bob's actions of the same sign as* λ*.*

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

$$\frac{\sum_{\omega\in\Omega}p\left(\omega\right)y_a(\omega)y_b(\omega)}{\sqrt{var\left(y_a\right)}\sqrt{var\left(y_b\right)}}=\lambda\frac{2\left(1-\lambda\right)^{2}}{var\left(y_i\right)\times\left(2-\lambda^{2}\right)^{2}} \tag{4.14}$$

where $y_i(\omega)$ is agent i's action in state ω. The result then follows from the fact that the ratio on the right hand side of (4.14) is positive for all values of λ. □

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 (4.14) associated with states where $\omega_a=\omega_b$ cancel against the terms associated with the equally probable states where $\omega_a=-\omega_b$. The correlation in actions is thus driven by those states in which both agents read about the same event. That is, in the states $(1,0)$ and $(-1,0)$ both Alice and Bob read about ω_a, and in the states $(0,1)$ and $(0,-1)$ they both read about ω_b. The products of Alice's 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.[19]

[17]In the simple discrete example here, the only state in which any outcome is common knowledge is $(0,0)$ since it is only in this state that Alice and Bob reads a report stating that the variable they have a direct interest in equals zero.

[18]The condition (4.12) holds for all positive values of λ and for negative values of λ in the interval $(-0.53,0)$.

[19]In the model, the correlation due to the state-dependent delegated information choice arises between readers of different news outlets. In reality, correlation of actions may of course also arise within the readership of a given news provider.

4.8. **Extreme events and common knowledge.** The simple set up with discrete states that we have studied so far allows for closed-form solutions, but it is not suitable for studying how news coverage, beliefs and actions are affected by the magnitude of events. In this section, we therefore extend the model to allow for a continuous distribution of the state. This extended version of the model can be used to study how the magnitude of outcomes matters in a setting where more extreme events are considered more newsworthy.

4.8.1. *Optimal simple news selection functions.* With continuous distributions of the potential stories ω_a and ω_b, the optimal news selection functions are infinite-dimensional objects with unknown functional forms. We therefore approximate the optimal news selection functions using a simple but flexible parametric class of threshold functions in the absolute values of ω_a and ω_b of the form

$$s_i^i = \begin{cases} 1 \text{ if } |\omega_i| \geq \alpha_i \left|\omega_j\right|^{\beta_i} \\ 0 \text{ otherwise} \end{cases} . \tag{4.15}$$

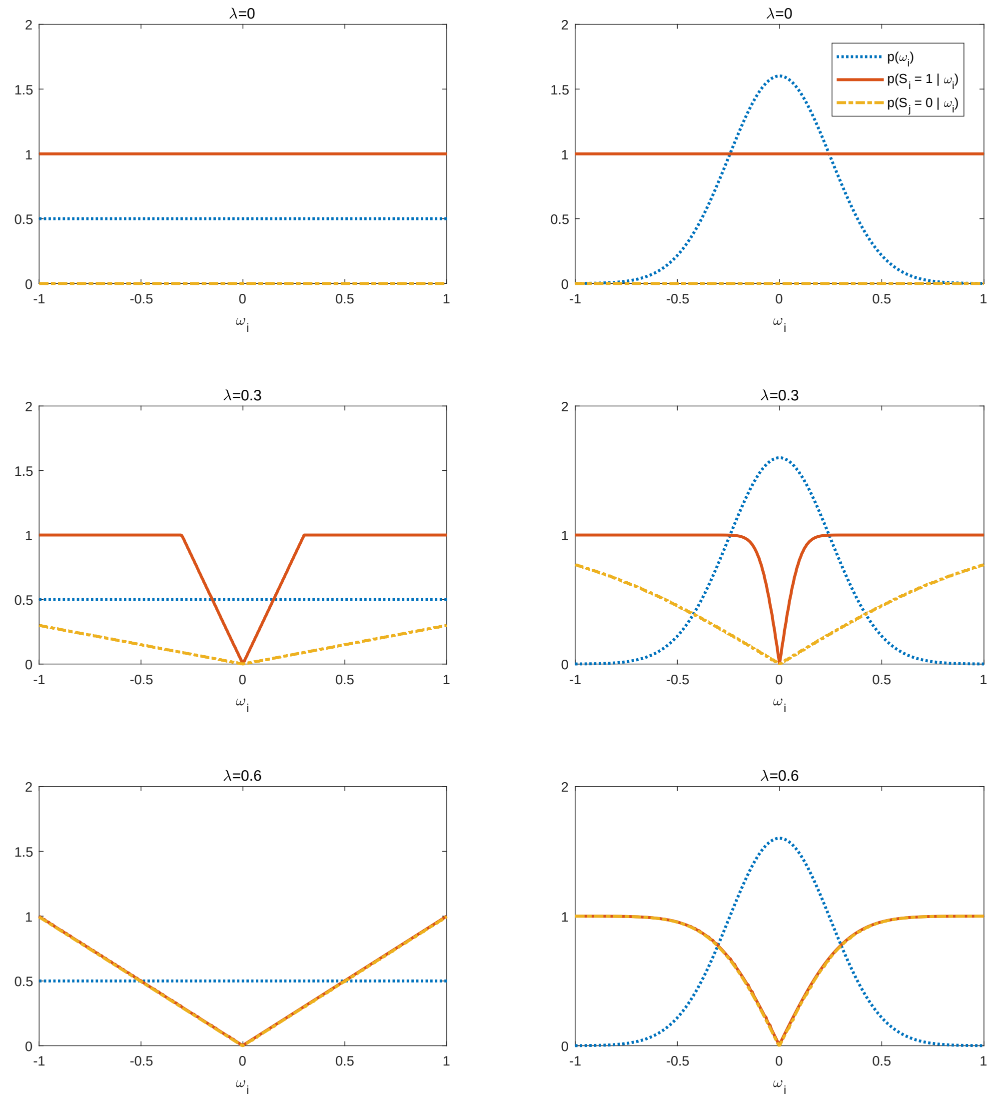
This functional form implies that the relative newsworthiness of different events depends only on their relative magnitudes. While this is a simplification, the effects we study below should be robust to richer functional forms. Given the constraint that the news selection functions must be of the form (4.15), Paper A now chooses α_a and β_a, in order to maximize the expected utility of Alice, and Paper B chooses α_b and β_b, to do the same for Bob. The optimal values of α_i and β_i for each provider are found by numerically maximizing the expected utility of the relevant agent. In general, the optimal values for α_i and β_i are a function of λ.

4.8.2. *Conditional actions.* In equilibrium, α_i is always smaller than or equal to 1 and $\beta = 1$. This means that Alice only observes ω_b in states where Bob does so as well, and vice versa. The optimal conditional actions are thus again described by (4.9) and (4.10). The main difference here, relative to the discrete state space case, is that the probabilities in the denominator of the conditional actions now vary smoothly with the realized value of ω_i. The exact mapping from realized value of ω_i to this probability is determined by α_i and β_i in the news selection functions (4.15).

4.8.3. *Equilibrium probability of reporting.* Figure 5 illustrates the probability that ω_i is reported conditional on its realized value for different strengths of the strategic motive. The solid red lines are the probabilities that paper i reports ω_i. The dashed yellow lines are the probabilities that paper j reports ω_i. The left column corresponds to $\omega_i \sim U(-1, 1)$, and the right column corresponds to $\omega_i \sim N(0, \frac{1}{3})$. The pdfs of the distributions are plotted using dotted blue lines. The rows correspond to, from top to bottom, values of λ equal to 0, 0.3, and 0.6.

When $\lambda = 0$, the optimal values of α_i and β_i are 0 and 1, respectively. Paper i then always reports ω_i, but Paper j never does. With $\lambda = 0.3$, the optimal values of α_i and β_i are 0.3 and 1, respectively. Finally, with $\lambda = 0.6$, the optimal values of α_i and β_i are both equal to 1. A stronger strategic motive in actions thus decreases the probability that Paper i reports ω_i, but it increases the probability that Paper j does. The reason is that with a with a stronger strategic motive and for a given realized state, it is more costly for Alice not to know what action Bob takes. There are then more values of ω_b that imply a strong enough response by

FIGURE 5. Conditional probabilities of each paper reporting ω_i.



Notes: The figure illustrates the probability that ω_i is reported by Paper i (solid red) and Paper j (yellow dashed) conditional on the realized value of ω_i. The distribution of ω_i and ω_j is $U(-1, 1)$ (left column) and $N(0, 1/3)$ (right column). The top row corresponds to no strategic motive ($\lambda = 0$), the middle row to a moderate strategic motive ($\lambda = 0.3$) and the bottom row to a strong strategic motive ($\lambda = 0.6$).

Bob for Alice to prefer to know about ω_b and indirectly, Bob's response to it, rather than to know about ω_a. For $\lambda = 0.6$, both newspapers always report the same outcome, since values of $\alpha_i = \alpha_j = 1$ imply that both newspapers simply report the outcome that had the larger realized value in absolute terms. Effectively, it is then as if there were only one single information provider.

4.8.4. *The probability of an outcome being common knowledge.* From Bayes Rule, the probability that Bob observes ω_a conditional on Alice doing so is given by

$$p\left(s_a^b = 1 \mid s_a^a = 1, \omega_a\right) = \frac{p\left(s_a^b = 1 \mid \omega_a\right)}{p\left(s_a^a = 1 \mid \omega_a\right)} p\left(s_a^a = 1 \mid s_a^b = 1, \omega_a\right). \tag{4.16}$$

Since $p\left(s_a^a = 1 \mid s_a^b = 1, \omega_a\right) = 1$, the probability that Bob observes ω_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 ω_a as the absolute value of ω_a increases, so larger magnitude events tend to be closer to common knowledge. With uniform distributions, the probability that Bob observes ω_a as ω_a tends to the boundary of its support is simply given by the value of α_b.[20] Alice thus attaches approximately a 30 percent probability to that Bob observes ω_a as when ω_a is close to either -1 or 1.

With normally distributed variables, the corresponding probability at $|\omega_a| = 1$ is about 75 percent. The difference compared to the uniform distributions is explained by the fact that with normally distributed variables, more probability mass is concentrated around the (zero) mean. Conditional on the realized value of ω_a, it is then less likely that the realized (absolute) value of ω_b is large enough to make Paper B report ω_b instead of ω_a. With normally distributed variables and for a large enough absolute realization of ω_a, both Paper A and Paper B report ω_a almost surely. In the limit, information about extreme realizations of ω_a or ω_b thus approaches common knowledge as the probability that ω_b has a realization that is considered more newsworthy by Paper B then tends to zero.

4.8.5. *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, they determine how likely it is that an 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, news selection functions also affect how likely it is that the other agent is reading about the same event. When 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 it.

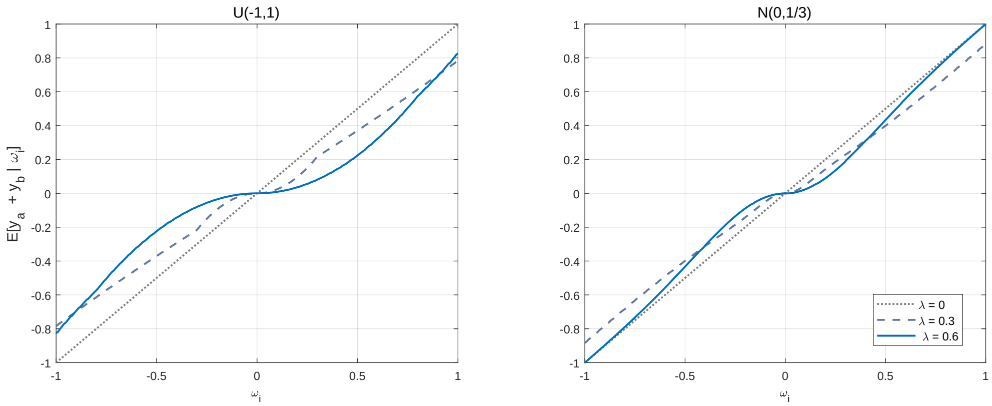
As we saw in Figure 5 above, the probability that the two agents observe an event ω_i is increasing in its absolute realized value. Figure 6 illustrates how these probabilities translate into expected aggregate actions for different values of λ and for different distributions of events.

When $\lambda = 0$, Alice always observes ω_a, but Bob never does. The probability in the denominator of (4.9) and (4.10) is then zero, and Alice's response is linear in ω_a. Because Bob never observes ω_a, the conditional expectation of his action is zero for all values of ω_a. The conditional expectation of the aggregate action illustrated by the dotted grey line is thus simply Alice's conditional action.

When $\lambda > 0$, the probability that Alice observes ω_a increases in the absolute value of ω_a, and so does the probability that Bob observes ω_a conditional on Alice doing so. These

[20]This is so since $p\left(\alpha_b \left|x_a\right| > \left|x_b\right|\right) = \alpha_b$ when the distributions of x_a and x_b are both uniform and with identical support.

FIGURE 6. Expected aggregate action conditional on ω_i.



Notes: The figure illustrates the expected aggregate action conditional on ω_i with $\omega_i, \omega_j \sim U(-1, 1)$ (left panel) and $\omega_i, \omega_j \sim N(0, \frac{1}{3})$ (right panel). When $\lambda \neq 0$, the slope of the expected aggregate action is zero around the point where $\omega_i = 0$. This reflects that the probability that ω_i is reported by any paper is then also zero (see Figure 5).

effects introduce a nonlinearity in the expected aggregate response, as illustrated by the dashed grey and solid blue lines in Figure 6.

The source of this non-linearity are the probabilities in the denominator of the conditional actions. To see how, consider first realizations of ω_a that are close to zero. At the zero limit, the probability that Alice or Bob observe ω_a is also zero, and then so is the conditional expectation of the sum of their actions. The expected response curve in Figure 6 is therefore flat around the point where $\omega_a = 0$. As the absolute value of ω_a increases, the probability that Alice and Bob read about it also increases. Thus, both the probability that the agents read about ω_a and the degree to which this fact is common knowledge are increasing in $|\omega_a|$. The magnitude of the expected aggregate response to ω_a is thus increasing more than proportionally in $|\omega_a|$. Delegated information choice with benevolent information providers and quadratic loss functions therefore generates weak responses to small events, and strong responses to large events. The reason for this is that large events are not only widely reported, but that their magnitude also conveys information about the degree to which they are common knowledge. When one agent receives information about a very large event, he can therefore also infer that the other agent has probably received the same information. This then allows the two agents to coordinate more effectively.

5. CONCLUSIONS

Economic agents live in a complex reality they cannot fully monitor on their own. Therefore, many delegate their information choice to specialized news providers that report only a curated selection of events. We have shown that the reporting behavior of such news providers is state dependent, and that it exhibits a number of intuitive properties. Specifically, while different outlets tend to emphasize different topics, large events shift the overall news focus and make coverage more homogeneous. The key message of our paper is that under this type of systematic news selection, information is conveyed not only via the actual content of a news story, but also via the observable reporting decision itself.

For our theoretical analysis, we have developed a general framework for studying the consequences of information providers' editorial decisions. We refer to the framework as *delegated information choice* and the main underlying idea of this framework is that individuals often decide whom to get information from, not what to get information about. This is in contrast to much of the incomplete information literature, which typically assumes that agents make ex-ante choices about what variables to observe. Our framework also differs from rational inattention literature that has followed Sims (1998, 2003) in that it explicitly incorporates information providers that serve as intermediaries between agents and the state of the world.

Our main results are proven in a general but abstract setting. However, we also study the implications of delegated information choice in a more explicit strategic setting by embedding the mechanism in a modified version of Morris and Shin's (2002) beauty-contest model. As in the original version of the model, the actions an agent takes depend on what he observes. However, under delegated information choice they also depend on the news selection functions of both his own news provider and that of the other agent. This suggests that, in order to fully understand how agents respond to news, one should study not only what information they receive, but also how their news providers decide what to report.

The information providers in our beauty-contest model maximize the utility of their respective readers, and they do not exhibit reporting biases such as political slant or a focus on negative news. However, neither the specific set of micro foundations used to generate systematic news selection nor the absence of reporting biases are important for our results. Instead, what matters is that news selection exhibits some form of state dependence, and that agents understand this. If they do, their beliefs and actions will be affected by both the reported information a story contains and the reporting decision itself.

Finally, one may ask to what extent our results apply to a world in which agents move away from traditional news media and increasingly consume information online. If agents who consume news online do not visit web sites that make state-dependent reporting decisions, our findings may be less relevant now than they were before the advent of the internet. Recent empirical evidence suggests, however, that this is not the case. Using browser history data of 50,000 US households, Flaxman et al (2016) find that "*the vast majority of online news consumption is accounted for by individuals simply visiting the home pages of their favorite, typically mainstream, news outlets*". This suggests that most people still tend to return to specific news providers selected according to what kinds of stories they typically report. While the internet does provide easy access to vast amounts of information, it therefore does not appear to have necessarily diminished the importance of the mechanisms we analyze in this paper.

References

[1] Alvarez, F., F. Lippi and L. Paciello, 2011, "Optimal Price Setting With Observation and Menu Costs", *Quarterly Journal of Economics* 126, pp1909-1960.

[2] Amador, M. and Weill, P.O., 2010. Learning from prices: Public communication and welfare. *Journal of Political Economy,* 118(5), pp.866-907.

[3] Amador, M. and Weill, P.O., 2012. Learning from private and public observations of others' actions. *Journal of Economic Theory*, 147(3), pp.910-940.

[4] Angeletos, G.M., Hellwig, C. and Pavan, A., 2007, "Dynamic global games of regime change: Learning, multiplicity, and the timing of attacks", *Econometrica*, 75(3), pp.711-756.

[5] Angeletos, G.M. and A. Pavan, 2007, "Efficient use of information and social value of information", *Econometrica*, 75(4), pp1103-1142.

[6] Asuncion, A., M. Welling, P. Smyth and Y.W. Teh, 2009, "On smoothing and inference for topic models", *Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence*, pp.27-34. AUAI Press.

[7] Aumann, R.J., 1976. Agreeing to disagree. *The annals of statistics*, pp.1236-1239.

[8] Baker, S.R., N. Bloom and S.J. Davis, 2013, "Measuring economic policy uncertainty", Chicago Booth research paper 13-02.

[9] Bao, Y. and A. Datta, 2014, "Simultaneously Discovering and Quantifying Risk Types from Textual Risk Disclosures", *Management Science* vol 60, pp1371-1391.

[10] Barwise, Jon. "Three views of common knowledge." Readings in Formal Epistemology. Springer, Cham, 2016. 759-772.

[11] Binmore, K. and Brandenburger, A., 1988. Common Knowledge and Game Theory (No. 167). Suntory and Toyota International Centres for Economics and Related Disciplines, LSE.

[12] Blei, D.M., A. Ng, and M. Jordan. Latent Dirichlet allocation. Journal of Machine Learning Research, 3:993–1022, January 2003

[13] Blei, D.M. and J.D. Lafferty, 2009, "Topic Models", in A. Srivastava and M. Sahami, editors, *Text Mining: Classification, Clustering, and Applications*, Chapman and Hall CRC Data Mining and Knowledge Discovery Series.

[14] Brandenburger, A. and Dekel, E., 1987. Common knowledge with probability 1. *Journal of Mathematical Economics*, 16(3), pp.237-245.

[15] Brandenburger, A. and Dekel, E., 1993. "Hierarchies of beliefs and common knowledge", *Journal of Economic Theory*, 59(1), pp.189-198.

[16] Cespa, G. and Vives, X., 2011. Dynamic trading and asset prices: Keynes vs. Hayek. The *Review of Economic Studies*, 79(2), pp.539-580.

[17] Chang, J., Gerrish, S., Wang, C., Boyd-Graber, J.L. and Blei, D.M., 2009. Reading tea leaves: How humans interpret topic models. In *Advances in neural information processing systems* (pp. 288-296).

[18] Crawford, V.P. and Sobel, J., 1982, "Strategic information transmission", *Econometrica*, pp.1431-1451.

[19] Edmond, C., 2013. Information manipulation, coordination, and regime change. *Review of Economic Studies*, 80(4), pp.1422-1458.

[20] Eisensee, T. and Strömberg, D., 2007, "News droughts, news floods, and US disaster relief", *Quarterly Journal of Economics*, pp.693-728.

[21] Fogarty, B.J., 2005, "Determining Economic News Coverage", *International Journal of Public Opinion Research*, vol 17, pp.149-172.

[22] Flaxman, S., Goel, S. and Rao, J.M., 2016. Filter bubbles, echo chambers, and online news consumption, *Public Opinion Quarterly*, 80(S1), pp.298-320.

[23] Fligstein, N., Brundage, J.S. and Schultz, M., 2014. Why the Federal Reserve failed to see the financial crisis of 2008: the role of "macroeconomics" as a sense making and cultural frame, mimeo, UC Berkeley.

[24] Geanakoplos, J., 1994. Common knowledge. Handbook of game theory with economic applications, 2, pp.1437-1496.

[25] Gentzkow, M. and J. Shapiro, 2006, "Media Bias and Reputation", *Journal of Political Economy,* vol 14, pp280-316.

[26] Gentzkow, M. and J. Shapiro, 2008, "Competition and Truth in the Market for News, *Journal of Economic Perspectives*, pp133-154.

[27] Griffiths, T. and M. Steyvers, 2004, "Finding scientific topics", *Proceedings of the National Academy of Sciences*, vol 101, pp5228-5235

[28] Hansen, S., M. McMahon and A. Prat, 2015, "Transparency and Deliberation within the FOMC: a Computational Linguistics Approach".
[29] Hellwig, C. and L. Veldkamp, 2009, "Knowing what others know", *Review of Economic Studies*, pp223-251.
[30] Mackowiak, B. and M. Wiederholt, 2009, "Optimal Sticky Prices under Rational Inattention", *American Economic Review*, vol. 99(3), pages 769-803.
[31] Mackowiak, B. and M. Wiederholt, 2010, "Business Cycle Dynamics under Rational Inattention", *Review of Economic Studies*, pp1502-1532.
[32] Mahajan, A., L. Dey, and S. M. Haque, 2008, "Mining financial news for major events and their impacts on the market", in *Web Intelligence and Intelligent Agent Technology*, vol. 1, pp.423-426, IEEE.
[33] Marschak, Jacob, 1959, "Remarks on the Economics of Information", Cowles Foundation Paper 146, Reprinted from Scientific Research in Management, University of California Printing, UCLA.
[34] Matějka, F., 2015. Rationally inattentive seller: Sales and discrete pricing. *Review of Economic Studies*, 83(3), pp.1125-1155.
[35] Matějka, F. and A. McKay, 2015, "Rational inattention to discrete choices: A new foundation for the multinomial logit model", *American Economic Review* 105, pp272-98.
[36] Marschak, Jacob, 1960, "Remarks on the Economics of Information", Cowles Foundation Paper 146, Reprinted from Scientific Research in Management, University of California Printing, UCLA.
[37] Monderer, D. and Samet, D., 1989. Approximating common knowledge with common beliefs. *Games and Economic Behavior*, 1(2), pp.170-190.
[38] Monderer, Dov, and Dov Samet, 1995, "Stochastic common learning." *Games and Economic Behavior,* pp161-171.
[39] Morris, S. and H.S. Shin, 2002, "The social value of public information", *American Economic Review* 92, pp1521-1534.
[40] Nimark, K.P., 2014, "Man-bites-dog business cycles", *American Economic Review* 104, pp2320–2367.
[41] Perego, J. and Yuksel, S., 2017, "Media competition and the source of disagreement", mimeo, Yale University.
[42] Porter, M.F., 1980, "An algorithm for suffix stripping", *Program* Vol. 14, pp. 130-137.
[43] Ramage, D., S. Dumais and D. Liebling, 2010, "Characterizing Microblogs with Topic Models", *Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media.*
[44] Rubinstein, A., 1989, "The Electronic Mail Game: Strategic Behavior Under Almost Common Knowledge", *American Economic Review*, pp.385-391.
[45] Schelling, T.C., 1960. The strategy of conflict. Harvard University Press.
[46] Sims, C.A., 1998, Stickiness, *Carnegie-Rochester Conference Series on Public Policy* (Vol. 49, pp. 317-356). North-Holland.
[47] Sims, C.A., 2003. Implications of rational inattention. *Journal of Monetary Economics*, 50(3), pp.665-690.
[48] Sonsino, Doron. "'Impossibility of speculation' theorems with noisy information." *Games and Economic Behavior* 8.2 (1995): 406-423.
[49] Soroka, S.N., 2006, "Good news and bad news: Asymmetric responses to economic information", *Journal of Politics* 68, pp372-385.
[50] Soroka, S.N., 2012, "The gatekeeping function: Distributions of information in media and the real world." *Journal of Politics* 74, pp514-528.
[51] Soroka, S.N., D.A. Stecula 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.
[52] Stevens, L., 2014, "Coarse Pricing Policies", working paper, University of Maryland.
[53] Thorsrud, L.A., 2018, "Words are the New Numbers: A Newsy Coincident Index of the Business Cycle", *Journal of Business & Economic Statistics*, pp.1-35.
[54] Van Nieuwerburgh, S. and L. Veldkamp, 2009, "Information immobility and the home bias puzzle", *Journal of Finance* 64, pp1187-1215.

[55] Van Nieuwerburgh, S. and L. Veldkamp, 2010, "Information acquisition and under-diversification", *Review of Economic Studies* 77, pp779-805.
[56] Veldkamp, L., 2006a, "Information markets and the comovement of asset prices", *Review of Economic Studies* 73, pp823-845.
[57] Veldkamp, L., 2006b, "Media Frenzies in Markets for Financial Information", *American Economic Review*, Vol. 96, pp. 577-601.

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 $\mathcal{S}_i(\omega)$ determines whether agent i observes ω_a or ω_b. The optimal action when agent i observes ω_k and $\mathcal{S}_i$ can be expressed as

$$y_i(\omega_k, \mathcal{S}_i) = (1-\lambda)\frac{\sum \omega_i p(\omega_i, \omega_k, \mathcal{S}_i)}{p(\omega_k, \mathcal{S}_i)} + \lambda \frac{\sum y_j p(y_j, \omega_k, \mathcal{S}_i)}{p(\omega_k, \mathcal{S}_i)} \tag{A.1}$$

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 $\mathcal{S}_a$
   (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 $\mathcal{S}_b$, 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 $\widetilde{y}_i$ of agent i is described by the first order condition

$$\widetilde{y}_i = (1-\lambda) E_i[\omega_i] + \lambda E_i[y_j] : i \neq j \tag{B.1}$$

where E_i is the expectations operator conditional on agent i's information set.

B.2. **Information choice.** Define the utility of agent i when she observes ω_j as $U_i(\omega_j)$. Agents choose ex ante whether to observe ω_a or ω_b and agent i will choose to observe ω_i when the expected utility of doing so is higher than the expected utility of observing ω_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 ω_i and agent j observes ω_j their respective actions are

$$\widetilde{y}_i = (1-\lambda)\,\omega_i, \quad \widetilde{y}_j = (1-\lambda)\,\omega_j.$$

Agent i's expected utility when she observes ω_i is then given by

$$EU_i(\omega_i) = -(1-\lambda)\,E\left[(1-\lambda)\,\omega_i - \omega_i\right]^2 - \lambda E\left[(1-\lambda)\,(\omega_i-\omega_j)\right]^2 \tag{B.2}$$

or

$$EU_i(\omega_i) = -(1-\lambda)\,\lambda^2 E\omega_i^2 - \lambda\,(1-\lambda)^2\,E\,(\omega_i-\omega_j)^2 \tag{B.3}$$

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

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

The expected utility of agent i then is

$$EU_i(\omega_j) = -(1-\lambda)\,E\left[\lambda\frac{1-\lambda}{1-\lambda^2}\omega_j - \omega_i\right]^2 - \lambda E\left[\lambda\frac{1-\lambda}{1-\lambda^2}\omega_j - \frac{1-\lambda}{1-\lambda^2}\omega_j\right]^2 \tag{B.4}$$

which can be rearranged to

$$EU_i(\omega_j) = -(1-\lambda)\,E\left[\lambda\frac{1-\lambda}{1-\lambda^2}\omega_j - \omega_i\right]^2 - \lambda E\left[\frac{(1-\lambda)\,(\lambda-1)}{1-\lambda^2}\omega_j\right]^2 \tag{B.5}$$

and simplified to

$$EU_i(\omega_j) = -(1-\lambda)\,\lambda^2\frac{(1-\lambda)^2}{(1-\lambda^2)^2}E\omega_j^2 - (1-\lambda)\,E\omega_i^2 - \lambda\left(\frac{(1-\lambda)\,(\lambda-1)}{1-\lambda^2}\right)^2 E\omega_j^2 \tag{B.6}$$

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

$$EU_i(\omega_i) = -(1-\lambda)\,\lambda^2 - \lambda 2\,(1-\lambda)^2 \tag{B.7}$$

and

$$EU_i(\omega_j) = -(1-\lambda)\,\lambda^2\frac{(1-\lambda)^2}{(1-\lambda^2)^2} - (1-\lambda) - \lambda\frac{(1-\lambda)^2\,(\lambda-1)^2}{(1-\lambda^2)^2} \tag{B.8}$$

Agent i will choose to observe ω_i when $EU_i(\omega_i) > EU_i(\omega_j)$, that is, when the inequality

$$-(1-\lambda)\,\lambda^2 - \lambda 2\,(1-\lambda)^2 > -(1-\lambda)\,\lambda^2\frac{(1-\lambda)^2}{(1-\lambda^2)^2} - (1-\lambda) - \lambda\,(\lambda-1)^2\,\frac{(1-\lambda)^2}{(1-\lambda^2)^2} \tag{B.9}$$

holds. Move all terms in (B.9) to the left hand side and divide by $(1-\lambda)$ to get

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

The resulting inequality can then be simplified to

$$\left(1-\lambda^2\right)^2 + \lambda > 0 \tag{B.11}$$

The inequality (B.11) holds for all $\lambda > 0$. Alice will thus choose to always observe ω_a when actions are strategic complements, and Bob will then also choose to always observe ω_b. When actions are strong enough strategic substitutes, agents will choose to coordinate so that they both always observe either ω_a or ω_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 $\lambda \in (-1, -0.53)$.