Essays in Political Economy of Media

by

Alessandro Vecchiato

A dissertation submitted in partial fulfillment

of the requirement for the degree of

Doctor of Philosophy

Department of Politics

New York University

September, 2019

Shanker Satyanath

Dedication

In loving memory of my grandmother Anita,
who didn't have the priviledge of my wonderful education and yet
gave me the courage to pursue it, the strenght to endure it, and the depth to desire it.

Acknowledgements

The works included in this dissertation have been produced thanks to the guiding conversations with my doctoral committee. I owe special recognition to my advisor, Shanker Satyanath. He let me delve in any subject matter that I was curious about, and yet he was able to encourage me and point me to the right directions of inquiry. I owe immense gratitude to Cyrus Samii and Pablo Querubin, who helped me develop and implement the research in this work, with attentive suggestions, careful considerations, and honest opinions. The amount of intellectual and personal growth that my committee has allowed me to achieve is the biggest reward I received from my doctoral years.

None of this could have happened without an enduring collaboration with the excellent young academics in New York and abroad. Our discussions challenged my thoughts, posed thorny questions, and allowed me to improve my thinking and sharpen my analysis. I include a special mention to Maria Carreri, Saad Gulzar, Mateo Vasquez, Marlene Guraieb, and Stephanie Zonszein.

A dissertation, like any pursuit of original creative work, is a challenging endeavor that requires motivation and stamina. In my case, as in anyone I know, it was filled with uncertainties, moments of discouragement, and doubt. Learning how to overcome all these difficulties and reach the final line is one of the essential skills that a Ph.D. allows to develop. For this, I have to thank my analyst, Dr. Max Belkin. He unequivocably supported my work and helped me discover within me the ability to explore the world with open eyes and an open mind.

With this dissertation, I reach the end of my education. I need to thank all my inspiring teachers

that since elementary school worked with me and for me to achieve this goal. First among them, my parents, who gave me creativity and intelligence and let me grow up around art and literature. I finally dedicate my work to the memory of my grandmother Anita who did not live to see this moment. She thought me the love for education and for learning without which all I did would have been impossible.

Contents

List of Figures

List of Tables

LIST OF TABLES

Chapter 1

Panem et Circenses: The Persuasive Effect of Soft News in Berlusconi's Italy

"Don't you realize that something doesn't exist—not an idea, a politician, or a product—unless it is on television?"

Silvio Berlusconi

1.1 Introduction

The media are invested with the crucial role of guardians of democracy and accountability. With the day to day relaying of political, economic and social news stories, the media allow citizens of democracies to stay informed on how their representatives are behaving and, thus, hold them accountable for their actions (Przeworski, Stokes and Manin, 1999). Understandably, then, a lot of the scholarship on democracy has investigated ways in which politicians may compromise this process leading, for instance, to slanted news (Groseclose and Milyo, 2005; Gentzkow and Shapiro, 2010) or to under-reporting of malfeasance (Larreguy, Marshall and Snyder, 2018).

In the current media environment, these strategies might be harder to pursue. The expansion of

the internet has made the creation of news outlets much cheaper, with the predictable effect of increasing their number significantly, now even more dramatically with social media. With so many players in the media arena, it seems almost impossible that any politician could exert enough media power so to sway the entire media his way. Yet, for as much as information can make people learn, so distraction can make people forget (Horkheimer and Adorno, 1947). A way to steal attention from a problematic story may be to create many, many more enticing ones.

In this paper, we investigate the extent and effectiveness of a strategy of media bias based on distracting voters from potentially compromising information. In particular, we will look at how flooding the media with soft news, that is information related to gossip, entertainment, and sport (some definitions also include crime), can affect voter political preferences.

Traditionally, there seemed to be little role for soft news in politics (Prior, 2003). Typically, they could foster accidental exposure to salient issues (Baum, 2002) or appeal to inattentive individuals when they were embedding some political message (Baum and Jamison, 2006). In this paper, we argue that soft news may be used to replace politically-relevant information, or to the very least, dilute its impact. This strategy of seeking superficial appeasement to distract the public attention from deeper problems is, in fact, a long-standing form of generation of consent and it is usually defined *panem et circenses* from Juvenal, the Latin poet.

We empirically test these theories, and investigate the effect of changes in news content, by exploiting a natural experiment occurring in Italy as a consequence of Mr. Berlusconi's resignations as Prime Minister on November 12, 2011. The government turnover was reportedly due to the dramatic financial crisis that invested Italy for the large part of 2011. Right before his resignation, the interest spread between Italian and German bonds reached record highs. This case lends itself to the test of our theory because as a result of this resignation, Mr. Berlusconi lost his hold on public television, while he maintained control of his private Mediaset group. This peculiar arrangement allows us to detect Mr. Berlusconi media strategy directly when damaging information is not just possible, but likely. We exploit this quasi-exogenous variation in media control and unique fine-grained text data to detect changes in television news content. We measure news content by

applying the unsupervised learning algorithm WORDFISH to scale news segments on a hard-soft news dimension (Slapin and Proksch, 2008). Then, we define our treatment group as the public TV channels, and the control group as the private ones owned by Berlusconi. By comparing TV news segments on the national broadcast in the private and the public channels before and after Berlusconi's resignation, we can detect changes in the type of news that were covered as a result of his tenure. We find that Berlusconi manipulated the public media by increasing the concentration of their soft news content, thus making public channels overall more similar to his own Mediaset channels. This strategy is unlikely to be a commercial tactic since evidence points to declining profitability of public television during Berlusconi's tenure (DellaVigna et al., 2016). Thus, these results suggest changing news type can become an instrument impacting on public attention and favoring incumbent parties influencing the media.

The second set of empirical tests tackles this conclusion directly: we exploit plausibly exogenous variation in news exposure to evaluate the effect of changes in coverage on voters political preferences. Since television viewers have no control over broadcasted content beyond choosing which channel to watch, we develop a novel empirical strategy relying on content scores and panel data, to isolates within-channel content variation to detect its impact on voters preferences. We find that a standard deviation increase in coverage of soft news led to about 3% increase in the propensity to vote for Berlusconi's party. Moreover, we document that greater soft news also significantly impacted Berlusconi's rating as a leader and voters' left-right self-placement. To provide additional validity to our results and further explore the mechanism through which soft news can influence political attitudes, we reassuringly find that heavy news watchers are particularly affected by changes in news content. Our results are consistent with recent evidence that entertainment based television has political consequences, in particular in fostering the electoral success of populist leaders (Durante, Pinotti and Tesei, Forthcoming).

Italy is the ideal setting to comprehend the origins and outcomes of political influence on media. Its television market is characterized by a relatively uncompetitive ownership structure and a visible political parallelism (Hallin and Mancini, 2004). Moreover, the appointment of public networks'

directors is controlled by the Ministry of Economy and Finance. This setting facilitates the task of attributing changes in news' content to political influence. Our results provide troubling conclusions on the power of media to shape the electorate and better clarify an underappreciated mode of political communication.

The remainder of the paper is structured as follows: section 2 reviews the literature addressing the political influence on media, including the forms of media capture and bias; section 3 provides information regarding the Italian political and television system, before presenting the research design; section 4 documents our analyses and reports the main empirical findings; section 5 concludes.

1.2 Assessing Political Influence on Media: Challenges and Opportunities

The relationship between media and politics marks the nature and quality of democratic regimes (Dahl, 2000, p. 97). Democratic deliberation requires that voters receive truthful and accurate information. When media are politically influenced, however, news stories may be exaggerated, suppressed or generally distorted, and the democratic process compromised.

According to agency theory (Hölmstrom, 1979; Prat, 2006), news' distortion compromises the relationship between the agents (politicians) and the principal (voters). In absence of perfect monitoring ability by the principal, agents can engage in moral hazard and deliver suboptimal outcomes to the principal. For instance, this may take place, if news are tailored to deviate public attention from public mismanagement[1] of elected officials. In this case, voters cannot enforce accountability and are will likely support suboptimal political choices (Besley and Prat, 2006; Stanig, 2015).

The normative ideal of media political independence can be lessened in a various ways (Enikolopov and Petrova, 2015). Direct forms of political control, such as directly muting the opposition, appear to be widespread. According to Freedom House (2017), only 31% of the world's population

[1]This may include forms of corruption, misuse of public funds and pork barrel politics

lives in a country with free press. But even in the absence of direct control from the government, political and interests groups can capture the media, exerting pressure through more indirect forms of control and influence on the information flow (Acemoglu, Egorov and Sonin, 2013). For instance, in countries with weak constitutional checks and balances, media capture typically results from execrable practices such as targeting of non-compliant journalists with judiciary prosecution, indiscriminate arrests, intimidating or defamatory acts, and bribes. McMillan and Zoido (2004) document a process of media capture based on bribing practices occurring in Peru in the 1990s. Their estimate of the bribing price for a TV owner resulted 100 times larger than the bribing price for an opposition politician or a magistrate, providing a clear indication of the extent to which media independence matters to democracy as an accountability mechanism. Even countries with solid constitutional safeguards are not immune from milder forms of control (Gilens and Hertzman, 2000). In this case, strategies may involve currying favor, or, the discouragement of criticisms through the disparagement and delegitimization of journalists and the media establishment (Besley and Prat, 2006; Prat, 2015). Either way, political control of the media leads to unbalanced, unfair, or otherwise non-neutral coverage of political issues in the news.

Political influence is not the only source of bias in the news. In fact, bias can also be arise for market incentives, as when media use it to signal reputation to their audience (Gentzkow and Shapiro, 2006). Moreover, not all media environments are similarly exposed to media bias. In particular, concentrated media markets, such as television markets, may be more susceptible to political influence as the lack of competition among providers makes political control more effective. Indeed, concentrated media markets are characterized by a significantly larger degree of biased reporting than fragmented markets (Hanretty, 2014).

Assessing the consequences of political control over media represents an important yet challenging task. The main difficulty in trying to gauge the extent bias is the choice of a neutral reference for judging the bias. (Puglisi and Snyder, 2015) distinguishes between measures of explicit bias, such as open endorsements for political candidates (Ansolabehere, Snowberg and Snyder, 2006) or imbalances in the speaking time allocated in TV newscasts to politicians from different

parties (Durante and Knight, 2012; Bellucci and De Angelis, 2013), and measures of implicit bias, in which the political support is not openly expressed but operates indirectly, through a differential framing of news stories and of the airtime allocated to issues. Groseclose and Milyo (2005) consider a group of 200 prominent think tanks counting the times a particular member of the US Congress cited one of them. Then, they perform the same procedure for a number of newspapers and other media outlets, estimating an ideology score for each of them on the basis of the frequency each think tank was mentioned. This procedure thus allows creating a bridge between the ideological bias of media outlets and the bias of others actors, such as think tanks. This represents one feasible strategy enabling researchers to quantify ideological bias in the media. Gentzkow and Shapiro (2010) resort to the similarity between the language used by congressmen and newspapers as a measure of implicit bias. Other approaches involve measuring the coverage of issues seen as favorable for the endorsed parties. Puglisi and Snyder (2011) gauges media bias through the underreporting (overreporting) of scandals involving media supported (opposed) political parties. Finally, bias can also be assessed through the tone of the media in certain episodes, such as the release of economic bulletins (Lott and Hassett, 2014). In all these cases, the measurement approaches demanded some extent of arbitrary decisions on the side of the researcher (about which anchoring observations, issues or episodes to use as comparison) and therefore they leave the question open as to whether such measures would be robust to eventual modifications of these choices.

A related challenge involves assessing the consequences of slanted media content. Media effects have been mostly explored with experimental studies. Documented media effects (Druckman and Bolsen, 2011) include learning (Lenz, 2009), framing (Entman, 2007; Chong and Druckman, 2007; Esser and Strömbäck, 2014), priming (Iyengar and Kinder, 1987), and agenda setting effects (McCombs and Shaw, 1972; Russell Neuman et al., 2014; Wolfe, Jones and Baumgartner, 2013). Framing involves presenting news stories by emphasizing aspects that make them favorable to a specific political side. Agenda setting, refers to choosing the selection of news stories so to focus on political issue deemed more favorable to a political side, i.e. to *set the agenda*. Understanding the impact of these mechanisms requires an exogenous source of variation in media content, as

media users will typically self-select which content to receive. Most feasible approaches usually rely on designs with limited external validity, such as lotteries (Gerber, Karlan and Bergan, 2009), natural experiments (DellaVigna and Kaplan, 2007), and laboratory experiments (Iyengar and Kinder, 1987). Laboratory and survey experiments, in particular, have become the workhorse of contemporary media effects literature for their ability to control for omitted confounders of media effects (e.g. attending political events, or political discussions) and reversed-causality. However, caution is still required due to likely pretreatment effects (Slothuus, 2010). Moreover, estimated quantities of interest may remain artifact measures if the treatment (control) groups may remain hypothetical and never (always) receive the actual exposure in the real world (Gaines, Kuklinski and Quirk, 2007; Barabas and Jerit, 2010). For instance, the size of a non-treated group of TV watchers may remain negligible in real world. Alleviating these problematic aspects may require a more elaborated approach. For instance, (King, Schneer and White, 2017) documents the causal effect of news exposure on public opinion while leaving subjects to freely self-select into their preferred content. To this end, the authors recruited a pool of local media outlets to coordinate the randomized interventions. However, such large-scale designs may be hard to reproduce due to financial and time constraints.

In the next section, we introduce a novel research strategy that allows to shed light on various aspects relating the political influence on media. Our implementation targets the question of how Silvio Berlusconi influenced Italian public TV during his incumbency, and whether this had an impact on Italian voters. More broadly, we provide non-experimental evidence of how political actors can influence media, and how distorted information may enter in the electoral calculus.

1.3 Research Design

The previous section highlighted the difficulties in assessing the political determinants and consequences of news media slant: first, the problem of detecting political capture and media bias; second, disentangling the specific mechanisms that news outlets may adopt to deliver their mes-

sage to the public; third, the question around the actual consequences of media bias. In this section, we propose a research design focusing on detecting potential news distortions deriving from executive turnover, the actual mechanisms of news differentiation, and the downstream effects on the voters. The design demands multiple data sources, involving the text corpus of news stories, an exogenous source of variation in coverage, and a representative panel survey. We present it in three subsections, preceded by a brief introduction of the Italian context as well as of our data.

1.3.1 The Italian Natural Experiment and the Teche RAI Corpus

The Italian Context Italian television market is dominated by two prominent groups and one independent channel, all operating nationally and free-to-view. RAI is the public state-owned television network that introduced the medium in the country in 1953 and owns three main channels, each coming with a related newscast program (in Italian *telegiornale*, short "Tg"): RAI1 (Tg1), RAI2 (Tg2), and RAI3 (Tg3). Mediaset, which started its operations in 1985, is a private conglomerate owned by the former prime minister Silvio Berlusconi, leader of the center-right *Popolo della Libertà* (People's Freedom Party, PdL). Mediaset too owns three main channels, which also contain respective news programs: Rete4 (Tg4), Canale5 (Tg5), and Italia1 (Studio Aperto). All channels have similar scope and broadcast generalist content. Finally, La7 is private channel at the time controlled by Telecom Italia, a private telecommunication company.

Before 1994, throughout the Italian First Republic, TG1 was controlled by the Christian Democratic Party, TG2 by the Socialist Party, and TG3 by the Italian Communist Party. After 1994, the color of political information became more affected by the current majority in the parliament but generally leaves RAI3 in control of the left. Public channels also generally focus more on political and economic issues compared to Mediaset, due to their public service function. TG7, while owned by a private group, is directed by journalist Enrico Mentana, who was previously director of the main Mediaset newscast, TG5, until he was laid off in 2009 for his politically-independent direction.

1.3. RESEARCH DESIGN

RAI is officially controlled by the Parliament, but the executive power relies on a special parliamentary commission, whose members are, by majority, appointed by the government. Specifically, seven members of the RAI board are appointed by the parliamentary commission and two, including the President of the board, are indicated by the Ministry of Economic Affairs, which also indicates the RAI general director. This feature has historically given place to a phenomenon, called *lottizzazione*, that wants the newscasts and network directors to be replace with every new government. Specifically, newscasts director are assigned to each public network as a representative of a specific political group. Typically, the director of Rai1, the public channel with biggest audience, is nominated by the government majority party. The directors of Rai2 and Rai3, are instead nominated by either the government minority party or the opposition, depending on the political ideology of the government so to ensure that Rai3 is always left wing (Durante and Knight, 2012).

In this paper we exploit a quasi-exogenous variation in media control deriving from exceptional political circumstances: in April 2008 the center-right coalition formed by Mr Berlusconi's PdL and the right-wing populist Northern League (now renamed League) won the Italian general elections and formed the Berlusconi IV cabinet. In the Summer 2011 the Italian Bonds were severely hit by a speculative attack on the financial markets, and three months later the extreme financial and political pressure forced Mr. Berlusconi to step down. His failures to circumscribe the situation and reduce the Italian bond interest spread over the Euro zone led to his resignation on November, 12 2011. The 16th of November 2011 a technocratic government backed by a Grand Coalition of the main center-left and center-right parties was installed.

Contextually to his resignation, Berlusconi lost control of the public television thus providing us with an exogenous source of variation to the media market. When in power, Berlusconi never divested his properties in Mediaset, thus controlling either directly or indirectly almost 93% of the television market (DellaVigna et al., 2016). With this sudden loss of power, Berlusconi relinquished control over the public television, whose directors were swiftly replaced by the new, technical, government. By comparing TV news segments before and after his resignation we are able to provide evidence of his political influence on the media.

1.3. RESEARCH DESIGN

Teche RAI data We study Italian TV news outlets relying on an original dataset obtained from Teche RAI, the historical archive of the Italian public broadcasting. The data contains verbatim transcripts of all newscasts recorded between 2010 and 2014, thus covering the end Berlusconi's IV cabinet. In particular, the text corpus results from a specialized transcript system (ANTS), developed by RAI to perform speech recognition, and a validator that performs a segmentation of the signal based on the speech footprint of the speaker. The resulting accuracy is very high, with approximately 90% correct transcriptions.

The dataset is a unique verbatim source for text analysis, providing the highest level of granularity possible. All seven Italian national broadcasting networks are covered in the data. The data, for each segment of a newscast, also includes the channel, the topic, the date and starting and ending time. Topics are indicated in the form of content tags[2]. Each channel typically runs three daily newscasts of 30 minutes (morning, noon and evening service). The total amount of data collected by this database amounts to about 20,000 hours of broadcasting divided into 388,460 segments (news stories). However, in the analysis we pool the single segments into daily documents to reduce the parameter space.

Teche RAI data has an additional unique key feature: all politicians' interviews, or otherwise recorded speeches, are excluded from the corpus. This means that our scores reflect uniquely the implicit bias deriving from the content of the news, or what is being reported by journalists, and not from direct presence of politicians on TV[3].

Table 1.1 reports descriptive statistics for our data. For each network reports its newscast title, the media group, the edition time, the number of segments (a single news story), the segments per edition and their length. At times, when breaking news emerge, they may have special editions,

[2]This circumstance facilitates our task directly providing the topic of the news story, but researchers working without content tags could identify the topics through statistical topic models or correlated topic models (Blei, Ng and Jordan, 2003; Blei and Lafferty, 2007; Roberts et al., 2014).

[3]The Italian Communication Authority (AGCOM) bases its monitoring of TV political pluralism on the speaking time of politicians. Therefore, all our results are produced with data corresponding to a hypothetical case of perfect pluralism for AGCOM standards. Detailed information regarding the monitoring of political pluralism, the functions of the AGCOM authority, and the Law n. 249/1997 ("*par condicio*" law) regulating political and social pluralism on the media is available (in Italian) at the following AGCOM link: https://www.agcom.it/ldisciplina-della-par-condicio

which are not included in our data. We finally have to register occasional missing transcriptions from the corpus in the pre-treatment period. These missing values referring to the end months of 2010 and for both treatment and control groups come directly from the transcriptions data and are the results of idiosyncratic interruptions in the functioning of the ANTS system.

1.4 Data Preprocessing

Using news transcripts as data for the scaling model required a number of preprocessing steps. First, we subset the news stories (a grand total of $406,387$ stories) to include only newscasts, thus excluding other TV programs ($388,460$ stories). Second, text analysis with `WORDFISH`, as in general with "bag of words" approaches, involves converting the corpus of raw documents into a document-feature matrix of discrete occurrences of all the tokens in the documents. To this end we relied on the *tm* R package.

The creation of the data-feature matrix proceeds as follows:

1. We merge all the news stories' text by day for each TV channel. This effectively shifts the unit of analysis from the single news story to the [TV channel $\times$ day] level. An alternative aggregation scheme could have been the single TV news edition (three per day). Yet, the daily solution in our opinion is more efficient. The news of the day are generally repeated in the following editions of the day, so having a daily estimate has practically the only effect of shrinking the number of parameters to be estimated.

2. Italian language includes apostrophes, thus conventional punctuation removal tools [4] would behave stripping off the apostrophe and linking the leading character to the previous one[5].

[4] Such as the $removePunctuation()$ function from the *tm* R package.

[5] This works on English text corpora because apostrophes indicate possession and are thus always following the word, as in "*the professor's hat*". The punctuation removal would result in "*professors hat*", and once the words are stemmed the additional "s" character would be removed as well. In the Italian language apostrophes very often are associated with the elision of the article when the following word starts with a vowel. Thus, the punctuation removal of "*l'amica*" [the female friend] and "*un'amica*" [a female friend] would wrongly result in two different tokens: "*lamica*" and "*unamica*", and this would be unaffected by the following stemming process.

Table 1.1: Descriptive Statistics of TV Newscasts from Teche RAI data

Newscast	Network	Ownership	Broadcasts Time	Segments
Studio Aperto	Italia 1	Mediaset	12:30	26619
			18:30	23149
TG1	Rai 1	RAI	08:30	17463
			13:30	22383
			17:00	313
			20:00	27203
TG2	Rai 2	RAI	13:00	22840
			20:30	18649
TG3	Rai 3	RAI	12:00	14433
			14:30	12340
			19:00	18324
TG4	Rete 4	Mediaset	11:30	20016
			13:30	994
			19:00	29688
TG5	Canale 5	Mediaset	08:30	28123
			13:00	27062
			20:00	25961
TG7	la7	Telecom	07:30	15454
			12:30	27062
			13:30	25961
			20:00	19422

Broadcast times are in hh:mm (24h) format. Segments refers to a single news story broadcast in the corresponding time slot. Segments numbers are total frequency over the entire 2010-2014 period. Segments exclude direct quotes from politicians.

We thus preliminarily apply a regular expression to substitute all punctuation characters with a single space character.

3. We transform all text into lower case characters.
4. We remove stopwords in the text[6].
5. We remove words of one and two letters.
6. We remove all numbers.
7. We stem the document using the Porter's stemming algorithm.
8. We create the document-feature matrix including all unigrams in the text corpus.
9. We remove the top 5% and the bottom 5% tokens by frequency[7].
10. We filter out documents with less than 20 tokens (evaluated to be too short and thus uselessly making the algorithm's convergence difficult).

The final document-feature matrix has thus dimensions $[17,456 \times 5,430]$, with $17,456$ single tokens and $5,430$ documents (one for each TV channel per day) and matrix entries represented by the absolute frequency of tokens' occurrence. The data covers the period from the week starting on the 26 of July 2010, until the week starting on the 30 of September 2014.

1.4.1 Gauging News Content with Unsupervised Scaling of Media Content

We adopt a flexible framework that allows the study of news content without requiring arbitrary references such as the introduction of either additional political actors for comparison purpose, the

[6]We have created a custom Italian stopwords list composed as the union of: 1) all the *snowball* stopwords list that are typically included in the most common R packages (as in *tm* and *quanteda*); 2) all the words included in the *Ranks NL* stopwords list; 3) the following list of 34 additional stopwords chosen after visual inspection: `{l, poi, far, quest, qual, tant, quel, dic, so, quell, avev, piu, fa, vorrebb, gia, puo, s, sar, d, nun, ce, n, foss, x, b, va, ogni, vuol, andar, propr, fatt, vann, www, fonte}`.

[7]This is a standard procedure in quantitative text analysis that aims at cutting non-informative long tails in the distribution of words.

selection of specific issues, or the focus on the tone of the news for selected episodes. All these arbitrary choices would also deviate the focus from the actual inferential target represented by the full content of the news. We overcome arbitrary selection by applying text scaling. In particular, we use WORDFISH, an unsupervised learning technique first introduced by Slapin and Proksch (2008) to estimate party positions from the text contained in their manifestos.

Applying text scaling to television news follows the intuition that the news content results from the tension between the actual news stories of the day and the editors' take on these stories. Since each outlet is exposed to the same set of real-world events, we argue that by comparing the news' content across channel it becomes possible to identify the editorial slant on what is actually reported.

We therefore apply WORDFISH scaling model to estimate pure content-based scores on a daily basis. Following text-analysis literature, we assume that individual words are independent and identically distributed[8]. WORDFISH models word frequencies' distribution via Poisson distribution. The λ parameter is modeled as follows:

$$Y_{ijt} \sim Poisson(\lambda_{ijt}) \tag{1.1}$$

$$\lambda_{ijt} = exp(\alpha_{it} + \psi_j + \beta_j \cdot \omega_{it}) \tag{1.2}$$

where: j indexes words, i TV news programs, and t the day of the newscast. Y_{ijt} is the data feature matrix—and sole data input: each row (document) is a vector of word occurrences for a daily merge of news transcripts from channel i. The λ_{ijt} parameter is modelled as a function of four latent quantities: α_{it} is a document-level fixed-effect term controlling for document length; ψ_j is a word-specific fixed-effect term controlling for the overall frequency of the words; β_j captures the ability of word j to discriminate across document positions; finally ω_{it} is our main quantity of

[8]This may seem a radical simplification, but it was proven to work extremely well in empirical applications (Eyheramendy, Lewis and Madigan, 2003; Grimmer and Stewart, 2013).

interest measuring the latent position of news transcript it in the news' content space. In other words, `WORDFISH` partitions the total variation in word frequencies across actors and over time into a component relating words' general usage rate in the language and a component relating the differential length of considered documents, so that the residual variation is informative of the editorial control over news content[9]. All parameters are simultaneously estimated via expectation maximization (EM) algorithm[10].

The resulting ω scores, represent purely content-based positions over the latent space of televised news. Importantly, this procedure does not require arbitrary categorizations (Laver, Benoit and Garry, 2003), thus remaining neutral regarding which dimension will differentiate the actual news transcripts. When the method is applied to political manifestos (Slapin and Proksch, 2008), the latent space refers to party ideology ideology, since the content of manifestos is largely dealing with parties' policy packages and political vision. When applied to television news the interpretation is less straightforward. Three decades of political communication and journalism literature have pointed at two broad categories of news, soft news and hard news (Scott and Gobetz, 1992; Baum, 2002; Prior, 2003; Lehman-Wilzig and Seletzky, 2010). Television newscasts, usually provide a standard mix of hard and soft news, covering both the national and international political situation and some more light subjects pertaining to entertainment and sport[11]. Our expectation, therefore, is that the more salient underlying dimension that would differentiate newscasts would be related to the concentration of soft and hard news on each channel.

Our application represents, to the best of our knowledge, the first attempt to scale the

[9]In particular, the algorithm divides this residual variation in two components, β_j represents the ability of word j to discriminate between documents belonging to opposite poles of the latent continuum, and ω_{it} represents the position of each document on the latent space. For instance, if at t the word *soccer* appears more frequently in the transcripts from newscast i, the algorithm would assign more discriminatory weight β_j to that word, and based on how other actors use the word *soccer*, derive the location of the newscast i at time t (ω_{it}).

[10]The model specification does not assume that word frequency may be affected by previous word usage. This is appropriate due to the nature of the data, news, which is very volatile. Differently, one could introduce time-dependence in the ω scores. However, document-level time-dependence would artificially increase the likelihood of finding systematic trends of (non-overlapping) content scores for news programs, thus inflating the estimated size of media bias. For this reason, news transcripts' scores are left completely independent over time.

[11]Looking at the topic tag from our dataset this seems to be clearly the case.

content of television news. We validate our measure by exploring the the β_j parameters and the relative words, by reading multiple documents associated with the most extreme ω scores, and by comparing our scores and the RAI topic labels (see Section 1.5 for additional detail regarding validation). We argue that our validation supports using the labels 'soft-news', dealing with various current affairs and gossip stories, and 'hard-news', covering political and economic news, as the two endpoints of the news' latent content space.

A diff-in-diff approach to detect media capture To tackle the question of whether Berlusconi influenced public TV during his tenure, we rely on a difference-in-difference design corresponding to a two-by-two matrix of states. Mediaset, Berlusconi's private channels, represents the control group, while Rai, the public television, is the treatment group. The treatment is represented by the end of Berlusconi's government. Berlusconi maintained control of his televisions throughout the period under consideration, while his resignation interrupted his active control over public TV. Given the unanticipated end of his cabinet for reason external to the media market, we can consider as exogenous the source of variation on media content, supporting the interpretation of our estimates as causal.

Another advantage of our setting is that treatment and control groups are clearly defined and comparable. Both television groups have national reach and while the audience may be different across channel, they, as a group, have similar market shares. A source of concern could be that the editorial strategy of the control group, Mediaset, varies depending on whether Berlusconi could control or not public channels. This would imply that potential outcomes are not stable with respect to treatment assignment, thus generating a SUTVA violation. In order to address this concern, we resort to a *within-subject design* (Gerber and Green, 2012*a*) in which the timing of the intervention is assumed as random and we compare outcomes within each channel.

1.4.2 Assessing the Mechanisms of Media Bias

Extant literature has identified two main processes driving media content. On the one hand, differences may result from quantitative variation in coverage assigned to the specific issues covered in the news. This source of news differentiation would be in line with agenda setting theory, in that the content would reflect the different saliency assigned by the editors to specific issues. On the other hand, it could be that the differentiation in the content of the news is originated by choosing different words when presenting those same issues. This explanation is consistent with framing theory, in that differential choice of words would point at TV editors as offering a specific frame of reference to their audience.

Moreover, differential issue emphasis and framing are distinct but not necessarily competing theoretical mechanisms: framing deals with *how* issues are covered, and agenda setting with *how much* they are represented. Yet, the two mechanism may not be empirically mutually exclusive, since news content could reflects a mixture of the two processes.

In line with this set up, we test for the presence of differential issue emphasis and framing mechanism with separated regression models. In the first place, we assess quantitative differences in issue coverage by inspecting cumulated airtime of news stories, aggregated by issue. Different airtime between channels, once controlling for time fixed-effects capturing the impact of real world events, would provide evidence for differential issue emphasis. In the second place, differential framing can be detected by inspecting whether TV channels correlate with the ω scores *within-issue*, where systematic differences would provide evidence for the existence of a scale of news frames[12].

[12]This approach has two caveats. First, it can potentially overestimate framing, since it could be the case that word differentiation operates through differential language complexity rather than via actual differences in the frame of reference. For instance, one TV outlet may describe the same fact referring to the same underlying interpretative frame, but just using more abstract and sophisticated terminology, or using more indirect and allegoric styles of communication. If this is the case, absent any difference in the underlying frame of reference, we could still detect systematic differences in within-issue content scores. Second, it can underestimate the agenda setting mechanism since we are only measuring one specific agenda setting mechanism, which is the time allocated to the various topics. However, additional mechanism may be in place: editors may emphasize (de-emphasize) stories by moving them earlier (later) in the newscast, even if the time allocated is the same. We are therefore assuming that the saliency effect uniquely relates airtime and not the emphasis proceeding from other sources. Finally, the reader should note that since we identified the nature of the latent news space as proceeding from soft to hard issues, our expectation is that differential issue

1.4.3 The Effects of Media Bias

In the final stage of our empirical strategy we asks whether media slant affects the voters. Identifying the effects of news content is particularly challenging: while observational studies of media effects are typically plagued by self-selection, experimental studies may estimate causal effects for hypothetical populations that could be practically negligible, and may also lack external validity. To overcome these difficulties we match ω scores with a representative panel of Italian voters: our identification strategy rests on the idea that, that while viewers can deliberately choose which channel to watch, once they cannot influence the content of that channel. This translates into regression models of political attitudes in which we control both for the individuals' self-reported favorite news program, and the estimated news scores on that specific programs in the days before the interview was taken. This way, we can separate the effect induced on individuals' attitudes by their prior media preference (endogenous self-selection), and the effect relating exogenous changes in the content scores (i.e. the viewers' TV news diet).

1.5 Results

We report the empirical findings in four subsections mirroring the structure of our research design. We start by presenting the estimated news content scores and the differentiation in soft and hard news content. Then, we explore the mechanisms of media bias by addressing agenda setting and framing effects. Next, we assess causally the extent to which Berlusconi manipulated public TV information. Finally, we document the effects of variation in the content scores on Italian voters' attitudes and preferences.

emphasis should be operating to some extent.

1.5.1 Content Scores for Italian News Programs

Once we applied text-scaling to the Teche RAI corpus to estimate daily channel-specific news' content scores we could effectively identify systematic differences between channels. Moreover, the average scores match an informed expectation regarding the typology of news program for each channel. The average values of the estimated scores are respectively: -0.113 (TG1), -0.071 (TG2), -0.739 (TG3), 0.927 (TG4), 0.460 (TG5), 1.380 (Studio Aperto), -1.208 (TG7). As presented in Table 1.2, the differences in news content are systematic and robust to the inclusion of monthly dummy variables.

The ω scores point at meaningful differences in the Italian news, since RAI channels and TG7 display a negative average of the scores, while Mediaset channels display a positive average score[13]. We emphasize that these estimates are not indicating an ideological slant of TV channels, as our corpus contains news stories and not programmatic policy statements. Nevertheless, the fact that the ranking of the channels loosely correspond to their expected political leanings lends support to our approach. In fact, Mediaset channels appear on one side, with Canale 5 showing moderate scores, and RAI programs on the other side, with RAI3 showing more extreme scores. TG7 lies close to TG3 and, together with the public channels, is characterized by harder news content.

To assess whether the content scores reflect the soft-hard dimension in the news latent space, we undertook a validation exercise involving multiple strategies: first, we inspected the word parameters (i.e. the βs) that are closer to the two endpoints of the identified continuum trying to infer the underlying content domain; second, we read multiple documents associated with the most extreme scores; finally, we compared our scores and the RAI topic labels. We provide the first validation exercise by showing the words that better discriminate across documents: tables 1.3 and 1.4 present two set of words, respectively characterized by the lowest (negative) and the

[13]The scaling model does not exclude the possibility of detecting inconsistent and chaotic patterns of nearly-random content scores. Thus, this indirectly implies that scaling models can detect systematic differences not only in the more structured and written language of party manifestos, but also in the spoken language of news reports.

Table 1.2: Differences in the estimated ω scores across newscasts

	Dependent variable: ω-scores	
	(1)	(2)
TG1 (reference)		
TG2	0.042*	0.027
	(0.017)	(0.015)
TG3	−0.626***	−0.608***
	(0.015)	(0.014)
TG4	1.041***	1.036***
	(0.016)	(0.015)
TG5	0.573***	0.585***
	(0.014)	(0.013)
Studio Aperto	1.493***	1.511***
	(0.016)	(0.014)
TG7	−1.094***	−1.075***
	(0.014)	(0.013)
Constant	−0.114***	0.061
	(0.010)	(0.137)
Monthly FE		✓
N	17,456	17,456
R^2	0.710	0.759
Residual Std. Error	0.539 (df = 17449)	0.492 (df = 17404)
F Statistic	7,112.187*** (df = 6; 17449)	1,075.335*** (df = 51; 17404)

Note: Linear regression coefficients with standard errors in parentheses. The dependent variable represents the estimated ω content scores for the whole corpus of news transcripts.

highest (positive) scores. Overall, these parameters, as well as the other strategies of the validation exercise, confirm that news programs can be laid down on a spectrum between soft and hard news content: the former is characterized by stories regarding (attempted or actual) murders, rapes, or other violent crimes; natural disasters and extreme weather conditions; celebrities, fashion and shows. These stories are exemplified by words like *sedicenne* (sixteen-years-old), *ubriaco* (drunk), linking to car accidents, or *stilista* (fashion designer), *omicida* (killer), and *Hollywood*. On the other side, hard-news stories are associated to political and socio-economic content, with words including ministers' names, banks (e.g. *San Paolo* bank), or economic terms such as *percentuali* (percentage), *valutazioni* (evaluations) and the like.

1.5.2 Detecting the Mechanisms of Media Bias

We explore the process leading to content differentiation by inspecting the news stories *within-issue*. We focus on two of the most salient issue categories at the time: a category usually relating hard news, economic issues, including all news stories relating economic policy and macroeconomic changes, banking, finance, industry and job market, and a category generally including soft news, current affairs, mainly relating gossip, celebrities, law & order (crime) stories and natural disasters. Immigration stories, strikes and protest events are also included among the soft news, although their scores are somewhat in between the two poles.

For each issue, we first re-estimate the ω scores in order to detect systematic differences in the framing of the two issues. Next, we explore the cumulative airtime for each issue to detect differential emphasis across channels.

Economic issues. We count 30,545 news stories relating the economy and estimate the scaling model on this subset of stories to identify potential heterogeneity in word usage. Estimates, reported in Table 1.5, show that *economic* content scores tend to overlap, since no systematic difference can be detected. This pattern is consistent with the view that newscasts were covering economic facts largely with a similar language.

Table 1.3: List of top 15 words associated with positive beta scores

Tokens		b	psi
anticiclon	*anticyclon*	1.21	-2.05
ragazzina	*little girl*	1.03	-1.03
ubriaco	*drunk*	1.03	-2.14
grammi	*grams*	0.95	-1.39
sedicenn	*16 y.o.*	0.93	-2.09
discoteca	*club*	0.82	-1.53
cioccolato	*chocolate*	0.74	-1.43
copertina	*cover*	0.74	-1.37
studentessa	*student (f.)*	0.72	-1.17
piover	*rain*	0.71	-1.88
stilista	*designer*	0.71	-1.84
nuvola	*cloud*	0.67	-1.91
nubifragio	*cloudburst*	0.61	-1.76
omicida	*killer*	0.57	-1.58
hollywood	”	0.57	-1.72

Table 1.4: List of top 15 words associated with negative beta scores

Tokens		b	psi
secca	*draught*	-0.31	-1.75
mora	*arrear*	-0.32	-1.44
occident	*West*	-0.32	-1.17
visco	*Visco*[a]	-0.36	-1.38
scade	*expires*	-0.38	-1.53
indipendenza	*independence*	-0.38	-1.46
valutazioni	*evaluation*	-0.39	-1.10
riunioni	*meetings*	-0.40	-1.77
salari	*wages*	-0.42	-1.71
cuperlo	*Cuperlo*[b]	-0.44	-1.33
sanpaolo	*San Paolo*[c]	-0.45	-1.44
capigruppo	*speaker*	-0.46	-1.20
crucial	*crucial*	-0.51	-1.20
percentuali	*percent*	-0.53	-1.52
rendimento	*yield*	-0.86	-1.25

[a]Vincenzo Visco, then head of the Bank of Italy.
[b]Gianni Cuperlo, then head of the Democratic Party.
[c]Banca San Paolo, a national bank involved in a scandal.

1.5. RESULTS

Next, we turn our attention towards the quantity of economic news that were broadcast. In this case, estimates reveal wide differences between channels, thus pointing to a mechanism of economic news' selection being exerted by news editors.

The average monthly hours covering economic news were respectively 2.971 (TG1), 1.737 (TG2), 3.111 (TG3), 1.175 (TG4), 4.097 (TG5), 0.474 (Studio Aperto), 4.693 (TG7). This wide differences are robust to the statistical tests performed in table 1.6 and to the inclusion of monthly dummy variables.

Current events news. We count 103,320 gossip and current affairs stories. Tables 1.7 and 1.8 report the regression analyses performed to detect systematic variation in –respectively– the ω scores estimated for this second subset of news stories and in the airtime allocated for current affairs stories.

For this second issue category, we could detect both systematic divergences in the content scores within-issue and systematic differences in the quantity of current affairs. Table 1.7 shows how TG4, TG5, and Studio Aperto are associated with larger current affairs scores, TG3 and TG7 have smaller (negative) scores, while the remaining channels are approximately located in between. These results suggest that the language used to describe the events differs across channels. Since the pool of current affairs event is shared by the news programs, we read this evidence as pointing at the presence of differential framing. Finally, Table 1.8 systematic differences in coverage, thus suggesting that framing and agenda setting tend to co-occur.

Overall, our analyses reveals interesting issue-specific differences in the extent of agenda setting and framing potential effects by television news programs. In particular, economic news stories appear to be characterized by more similar language, probably deriving from comments of the same economic bulletins. Thus, programs cannot differentiate their content changing the frame, but can still do so allocating larger or smaller coverage to economic news. Yet, in the case of current affairs, we could detect both mechanisms, thus calling for more domain-specific theorizing of media effects.

In the following subsection 1.5.3, we thus use the content scores to estimate the political

Table 1.5: Differences in economic news content scores

	Dependent Variable: ω-scores for economic news	
	(1)	(2)
TG1 (reference)		
TG2	0.018	0.010
	(0.022)	(0.021)
TG3	0.035	0.014
	(0.020)	(0.019)
TG4	−0.015	0.003
	(0.025)	(0.024)
TG5	0.003	−0.002
	(0.018)	(0.017)
Studio Aperto	−0.007	0.001
	(0.034)	(0.033)
TG7	−0.026	−0.018
	(0.018)	(0.018)
Constant	−0.001	−0.561**
	(0.013)	(0.217)
Monthly FE		✓
N	30,461	30,461
R^2	0.0004	0.061
Residual Std. Error	1.000 (df = 30454)	0.970 (df = 30409)
F Statistic	1.960 (df = 6; 30454)	38.586*** (df = 51; 30409)

Note: Linear regression coefficients with standard errors in parentheses. Significance stars: *p<0.1; **p<0.05; ***p<0.01. The dependent variable represents the estimated ω-scores for the subset of news stories dealing with economic issues.

Table 1.6: Differences in economic news airtime

	Dependent variable: monthly airtime covering economic news	
	(1)	(2)
TG1 (reference)		
TG2	−1.234***	−1.234***
	(0.334)	(0.201)
TG3	0.140	0.011
	(0.338)	(0.204)
TG4	−1.796***	−1.869***
	(0.336)	(0.203)
TG5	1.125***	1.125***
	(0.334)	(0.201)
Studio Aperto	−2.497***	−2.570***
	(0.336)	(0.203)
TG7	1.722***	1.722***
	(0.334)	(0.201)
Constant	2.971***	−0.286
	(0.236)	(0.498)
Monthly FE		✓
N	318	318
R^2	0.450	0.829
Residual Std. Error	1.601 (df = 311)	0.966 (df = 266)
F Statistic	42.370*** (df = 6; 311)	25.234*** (df = 51; 266)

Note: Linear regression coefficients with standard errors in parentheses. Significance stars: *p<0.1; **p<0.05; ***p<0.01. The dependent variable represents the monthly airtime allocated to cover economic news.

Table 1.7: Differences in current affairs' content scores

	Dependent Variable: ω-scores for current affairs	
	(1)	(2)
TG1 (reference)		
TG2	0.143***	0.185***
	(0.011)	(0.010)
TG3	−0.371***	−0.363***
	(0.012)	(0.011)
TG4	1.120***	1.089***
	(0.009)	(0.008)
TG5	0.627***	0.611***
	(0.009)	(0.008)
Studio Aperto	1.315***	1.305***
	(0.009)	(0.008)
TG7	−0.541***	−0.546***
	(0.014)	(0.013)
Constant	−0.203***	−0.832***
	(0.007)	(0.091)
Monthly FE		✓
N	101,765	101,765
R^2	0.337	0.460
Residual Std. Error	0.862 (df = 101758)	0.778 (df = 101713)
F Statistic	8,631.919*** (df = 6; 101758)	1,697.871*** (df = 51; 101713)

Note: Linear regression coefficients with standard errors in parentheses. Significance stars: *p<0.1; **p<0.05; ***p<0.01. The dependent variable represents the estimated ω-scores for the subset of news stories dealing with gossip and current events.

Table 1.8: Differences in current affairs news airtime

	Dependent variable: monthly airtime covering current affairs' news	
	(1)	(2)
TG1 (reference)		
TG2	−3.109***	−3.109***
	(0.773)	(0.455)
TG3	−3.520***	−3.520***
	(0.773)	(0.455)
TG4	1.636*	1.470**
	(0.777)	(0.458)
TG5	4.514***	4.514***
	(0.773)	(0.455)
Studio Aperto	3.176***	3.176***
	(0.773)	(0.455)
TG7	−4.874***	−4.874***
	(0.773)	(0.455)
Constant	8.370***	0.910
	(0.547)	(0.939)
Monthly FE		✓
N	321	321
R^2	0.454	0.838
Residual Std. Error	3.708 (df = 314)	2.184 (df = 269)
F Statistic	43.601*** (df = 6; 314)	27.247*** (df = 51; 269)

Note: Linear regression coefficients with standard errors in parentheses. Significance stars: *p<0.1; **p<0.05; ***p<0.01. The dependent variable represents the monthly airtime allocated to cover current affairs.

influence exerted by Mr Berlusconi on Italian public news. Next, in subsection 1.5.4, we match the content scores with a representative panel survey of Italian voters to document the attitudinal consequences of the varying news content.

1.5.3 Difference-in-Difference Estimates

Having identified a systematic measure scoring news content, we can now exploit it to detect Berlusconi's control on public TV. We proceed by aggregating the scores for Rai and Mediaset: figure 1.1 reports a plot of the daily average omega score by network over time. Black bars indicate positive scores associated to softer news stories, while grey bars indicate negative scores consisting of harder news content. Public and private television appear to have distinct preferences toward hard and soft news, with Mediaset programs being similar to tabloids in terms of their content, and Rai broadcasts similar to quality papers. Furthermore, public television coverage dramatically changed after Berlusconi resigned. During the period of Berlusconi's tenure, RAI was broadcasting news content similar to his own newscasts, while public news content tilted toward hard news after he stepped down.

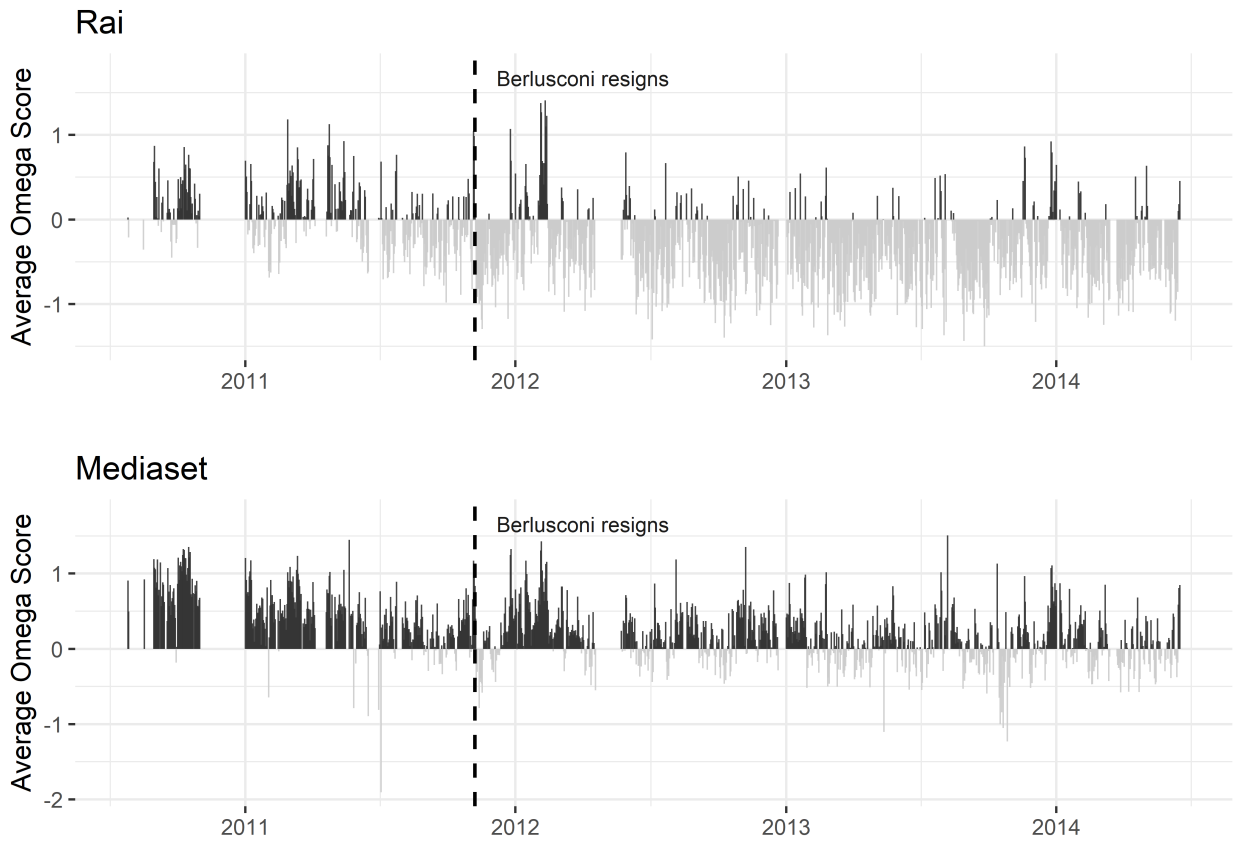
We further illustrate this point in Figure 1.2, reporting the total change in the distribution of omega scores by TV group, before and after Berlusconi stepped down. We notice that the distribution for RAI (i.e. the treatment group in our design) shifts significantly leftward (i.e. toward more hard news) in the post-treatment period, passing from an average of -0.021 to -0.448, while Mediaset scores remained virtually the same (from 0.990 to 0.800).

The visual inspection provides a first glance at our results. We formally test whether the end Berlusconi's cabinet influenced RAI news production by estimating the following diff-in-diff regression model:

$$Y_{ist} = \alpha + \lambda_{RAI} + \delta_t + \beta(\lambda_{RAI} \times Post_t) + \gamma_i + Splines_t + \epsilon_{ist} \tag{1.3}$$

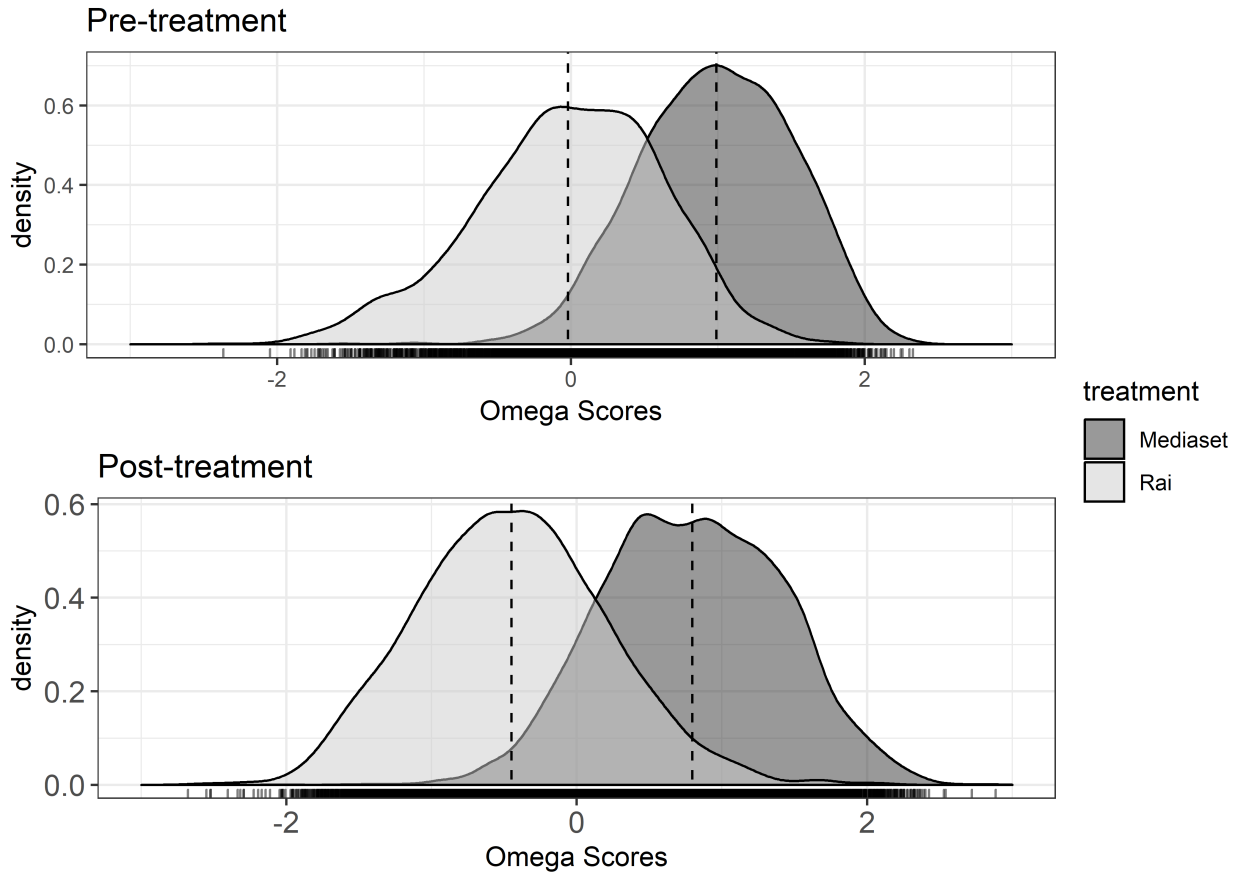
where Y_{ist} represent the omega scores for channel i in group s and at time t. λ_{RAI}

Figure 1.1: Content Scores by network, 2010-2014



represents a dummy variable for the public TV, RAI, and $Post_t$ is an indicator function that takes value 1 if the newscast occurred after November 12 2011, that is in the periods after Berlusconi lost control of the public television. We also control for non-linear time trends using cubic B-splines evaluated monthly (Wood 2006). Our parameter of interest β captures the effect on coverage of RAI not captured by Berlusconi. We include in our model also channel-specific fixed effects γ_i in order to control for heterogeneity in channel audience population and its consequential expected difference in coverage.

Figure 1.2: Distribution of Omega Scores Pre (top) and Post (bottom) by TV Group



Results Our estimates, presented in Table 1.9, clarify that the visual evidence stands the statistical test: after Mr Berlusconi lost control of the public television, the content of public programs shifted toward hard-news content news, as compared to Mediaset newscasts. In fact, the ω scores, on average, decreased by 0.19 points, corresponding to 0.37 standard deviations.

Interpretation. To gain a more substantive sense of this change we can inspect how an average Rai 30-min newscast would be before and after Berlusconi lost his power. Table 1.10 reports the average change in coverage for Public channels before and after Berlusconi lost power. The substitution between soft and hard news that was taking place in the public television under Berlusconi's

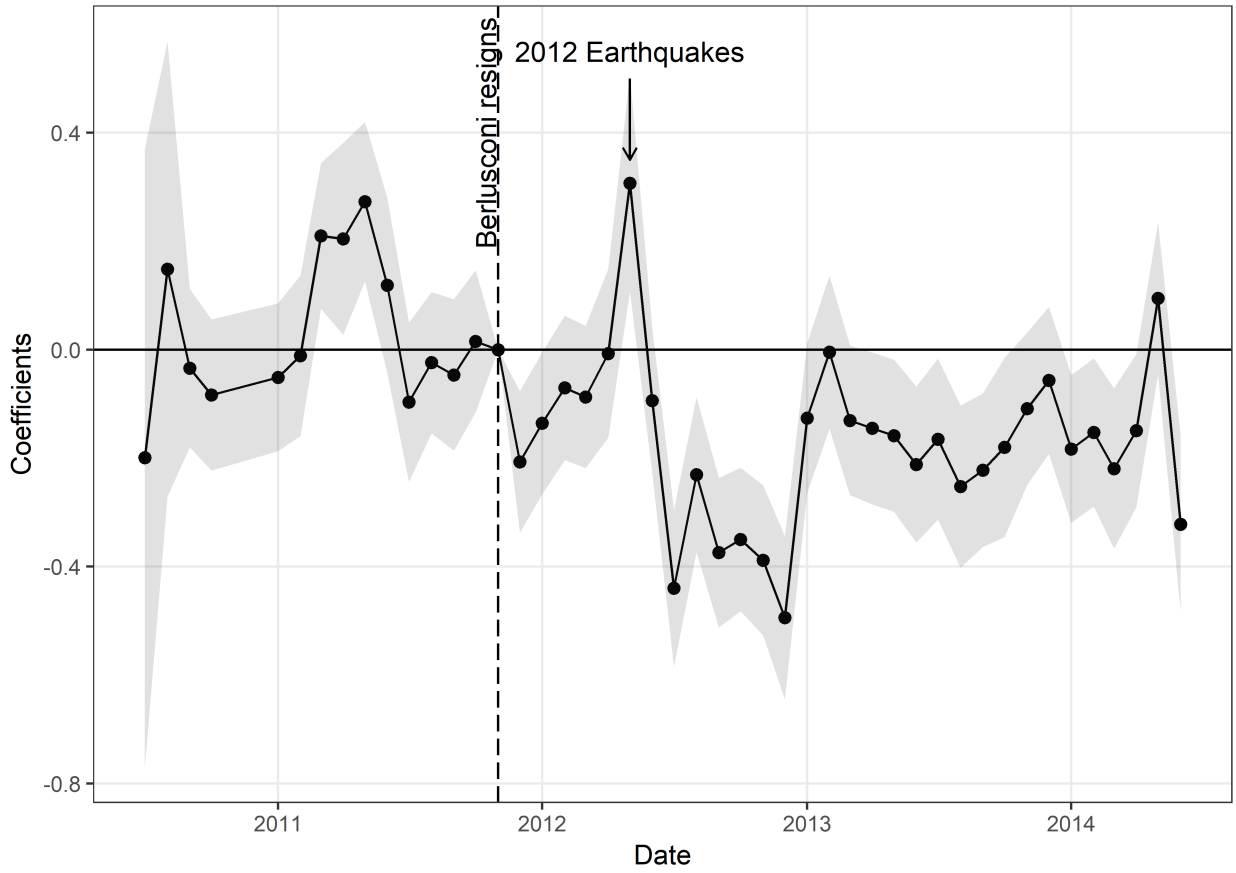
Table 1.9: Difference-in-Difference Estimates

	Dependent variable:		
	Omega Scores		
	OLS		*GLM*
	(1)	(2)	(3)
RAI × Post	−0.232***	−0.194**	−0.197**
	(0.005)	(0.026)	(0.017)
Channel FE		✓	✓
Splines			✓
Constant	0.988***	1.529***	2.422***
	(0.014)	(0.015)	(0.107)
Observations	14,265	14,265	14,265
R^2	0.486	0.646	
Adjusted R^2	0.486	0.646	
Log Likelihood			-9,866.694
Akaike Inf. Crit.			19,843.390
Residual Std. Error	0.624 (df = 14261)	0.518 (df = 14257)	
F Statistic	4,489.981*** (df = 3; 14261)	3,711.879*** (df = 7; 14257)	

Note: Standard errors are clustered at the TV channel level. Significance stars: *p<0.1; **p<0.05; ***p<0.01. Splines are calculated using monthly knots.

tenure, crowding out economic and political news in favor of entertainment and law & order-demanding coverage, emerges more clearly from this exercise: for TG1, the main public newscast,

Figure 1.3: Coefficient Plot for DID



Notes: This figure provides a graphical illustration of the parallel trend assumption. Coefficients represent results from equation (4) with 95% confidence interval for 14 periods (months) lags and leads.

the share of soft news after Berlusconi steps downs decreases from 54.7% to 44.9%.

Parallel trend The crucial assumption of this model is that during Berlusconi's tenure the variation in news content from Rai (treatment group) and Mediaset (control group) programs was similar (*parallel trend*), and would remain as such in the absence of the treatment. In order to test this assumption we regress the *omega* scores with leads and lags for months around our discontinuity.

Specifically we estimate the following model:

$$Y_{ist} = \alpha + \lambda_{RAI} + \delta_t + \sum_{t-14}^{t+14} \beta_t(\lambda_{RAI} \times Post_t) + \gamma_i + \epsilon_{ist} \tag{1.4}$$

Results from this estimations are shown in Figure 1.3: coefficients show that content scores were broadly similar before Berlusconi stepped down, while significant differences appear after Berlusconi's resignation. Intuitively, this means that RAI changed its coverage towards hard news after Berlusconi lost his power to influence public television. The effect persists for the period after, and notably attenuates as the new government is established in 2013.

1.5.4 Downstream Effects of News Content on Voters

Having established that the end Berlusconi's cabinet affected public news' content, the question remains as to whether this political influence on Italian public media affected voters' preferences and attitudes. To answer this question we match the ω content scores with information on individuals' favorite news program from a representative panel survey of Italian voters: the ITANES 2011-2013 panel, the official Italian election survey. ITANES data is representative of Italian voters and contains five waves of interviews covering a wide array of general questions about voters' political attitudes. Respondents were asked about their preferred newscast for three out the five waves. Table 1.11 reports the descriptive statistics of the ITANES dataset. Similarly to the actual Italian population, ITANES respondents tend to be heavy TV-watchers, declaring television as first source of information.

We identify the causal effect of news content variation by estimating within-individual (two-ways) fixed-effects regression models that also control for the favorite news channel: since media audiences tends to self-select toward their favorite news channel, by conditioning out of this obvious source of endogeneity we can focus on the purely content-driven within-individual variation in political preferences. The model is specified as follows:

$$Y_{it} = \alpha + \beta\omega_{it} + \gamma\mathbf{X}_{it} + \delta Channel_{it} + \lambda_i + \eta_t + \epsilon_{it} \tag{1.5}$$

1.5. RESULTS

Where our dependent variables are items tapping support for Berlusconi's party, the evaluation of the government, and the likability of Berlusconi as political leader measured for each respondent i at interview time t. Our main parameter of interest, β, captures the within-individual effect of change in news coverage; δ controls for the effect of the favorite newscast on political attitudes, thus controlling for the main source of reversed-causality. By additionally including individual fixed-effects γ_i we are, in fact, estimating how change in news coverage affected each individual, accounting for all observable and unobservable time-invariant individual characteristics that may induce self-selection toward a particular channel. We add to the specification time fixed-effects to control for preexisting trends that might impact both on coverage and public opinion.

Results. Table 1.12 reports the estimates introducing electoral support for Popolo della Liberta' (Berlusconi's party) as dependent variable: voters declare their probability of voting, from 0 to 10, for the PDL party[14].

The main independent variable is the average of the ω scores for the respondent's preferred news channel for the 30 days previous to the interview, standardized. We use this measure to isolate meaningful change in coverage while filtering out cyclical and short-term random variation. We estimate three main specifications: model (1) reports the ω content score coefficient controlling for wave and individual fixed-effects; model (2) also includes dummies controlling for the favorite newscast; model (3) finally adds two well-established controls of voting propensities: partisan identification and the ideological distance from the PDL party[15]. In all three specifications, the effect of the content scores is positive and precisely estimated: one standard-deviation increase in the content towards soft news is estimated to increase the propensity to vote for Berlus-

[14]The highest Propensity to Vote is correlated at 90% with reported vote intention. The 0-10 scale of this item allows to avoid the well-known incidental parameter problem of the Fixed Effects Logit estimator.

[15]Party identification takes the classic strength component ranging from 0, for independents and other partisans, to 3, signaling 'strong' partisan attachment. Left-right distance is computed taking the absolute value of the difference between the self-reported left-right position and the perceived left-right position of PDL. Both measures were standardized to allow for more direct comparisons. Also, the reader should note that the model is effectively introducing the within-individual variation in these variables.

coni by 0.335 (on a 0-10 scale). The effect is both statistically and substantially significant, as it corresponds to half of the estimated effect of partisan identification, and it is larger than the effect of ideological distance ($\beta = -0.221$).

Robustness: frequency of news exposure. The ITANES panel provides information about the general frequency of exposure to newscasts. Respondents not answering "Never" are then asked to indicate their favorite news program. While subjective measures of news exposure have been shown to be severely flawed (Prior, 2009), and thus we refrain from using their proportions as representative figures, we still expect our content scores to exert a stronger effect on the subgroup of voters reporting high, as opposed to low, exposure to the favorite newscast. To explore this heterogeneity of the content scores' effects we opt for a between-individual random-effects model. In fact, this allows to better assess the moderating effect of variables, such as the frequency of watching newscasts, reflecting long-term individuals' habits and that are thus unlikely to change over-time in the short term if not at random. Additionally, the test will also rule out the possibility that our results are driven by the choice of a fixed or random effects estimator (Clark and Linzer, 2015). Table 1.13 presents estimates from a random effects model: the interaction effect is positive and significant, indicating that ω scores increase support for PDL only for heavy news watchers.

Other attitudinal effects. We expand the analysis by looking at three additional key attitudinal measures: voters' support for the Berlusconi's government, Berlusconi's likability as political leader, and the voters' self-reported left-right position[16]. We again resort to two-way fixed-effects models

[16]The government evaluation item measures individuals' general support for the incumbent government, asking respondents to report how they "[E]valuate the overall performance of the government up to this moment?". Answers reported on a 0-10 scale. Since Berlusconi stepped down after the second wave, we drop observations from the third wave as answer would refer to the Monti government. Berlusconi's overall likeability is evaluated through the question: "Can you tell me, for each of the following political leaders, what is your overall judgement on a scale from 0, meaning completely negative, to 10, meaning completely positive?". Finally, the left-right position is the standard ideology scale item: "Many people use the words "left" and "right" when dealing with politics. Thinking about your political opinions, where would you place yourself on a scale from 0 to 10 where 0 means "left" and 0 means "rigth"?"

for causal identification.

Results are reported on Table 1.14. We do not find a significant effect on support for the government, although the coefficient remains positive and it is statistically significant before introducing the additional political controls. Differently, we find a positive effect of soft news coverage on voters' rating of Berlusconi as a leader ($\beta = 0.552$, larger than the effect of party ID). Finally, we find a significant effect even for a supposedly long-term construct such as ideological position: a one standard deviation increase in content scores is observed to produce a 0.137 right-wing shift in ideology[17].

Overall, these model estimates provide supportive evidence for the view that, not only the news content was politically influenced during Berlusconi incumbency, but that the external pressure to favor soft news content, hiding the government troubles and promoting support for the law & order policies of the right-wing government, had a wider and significant impact on voters' preferences and attitudes. Importantly, while experiment-backed causal estimates refer to hypothetical treatment groups, this paper offers evidence leveraging on real TV news and voters allowed to self-select into their favorite newscast.

1.6 Conclusions

A central tenet of democratic theory is that a free press is necessary to ensure participation, representation, and accountability (Przeworski, Stokes and Manin, 1999). Independent media allow for the free circulation of information that can support effective governance. However, the independence of the media can be compromised in a number of ways even when institutions are pluralistic and democratic. The Italian case shows that established democracies can also experience the detrimental influence of political power, at times in unexpected ways. In such scenarios, incentives for

[17]The reader should note however that the latter model differs from the previous due to the omission of the ideological distance control.

optimal policy action become tenuous and the likely outcome is political and economic decline.

While our setting takes advantage of a specific case of media capture, its results paradoxically speak to the effect of media commercialization. If, as it is commonly believed, soft news drive demand, new outlets scrambling for audience may be expected to increase their soft news concentration with the potential effects described in this paper. Our evidence is consistent with recent work by Durante, Pinotti and Tesei (Forthcoming) describing how the introduction of entertainment television in Italy changed fostered Mr. Berlusconi political success. The joint effect of citizens entertainment-orientation and wider content availability is likely to become the crowding out hard news consumption, and therefore, we might expect an overall decline in democratic competence.

While the evidence provided here is centered in Italy, a wealthy democracy, we believe that the mechanisms that we identified can be operating elsewhere. Anecdotal evidence from Mexico, for instance, reveals that when the government is facing particular scandals, often *Televisa*, the majour Mexican broadcaster, pivots from compromising stories to ever more fictionalized ones, often involving entertaintment figures. This common practice is called *caja china*, (the chinese box), and it such a prominent strategy that has been recently the subject of the satirical movie *The Perfect Dictatoriship*.

An important limitation of our study is the fact that we are able to assess only short term effects. While our strategy is able to overcome a lot of the difficulties that normally plague this literature, we cannot assess whether the impact of news' content on voters would be long lasting. Therefore more work is necessary to understand how permanent shifts in the media environment affect political life.

Panem et circences (literally, bread and circuses) is a Latin expression coined by Juvenal, the Roman poet of the First Century AD, to decry the generation of public approval through distraction and the mere satisfaction of immediate needs (Juvenal, *Satires*, Book XVI). Modern means of communication, like television, may be serving these ancient strategies by shaping voters awareness, competence and behavior. This paper shows that politicians can limit their accountability in previously undocumented ways, such as resorting to distracting practices like focusing on soft news'

stories, and that these *are* fostering their electoral strength. Voters, on the other hand, may not be able to avoid these tactics overall. Our results show that even when they are given freedom to choose what to watch, their agency role is not sufficient to offset the politicians' strategies, adding nuance to the theory of media power.

Table 1.10: Change in Coverage

	Pre			*Post*	
Topic	*Share*	*Share Hard/Soft*	*Topic*	*Share*	*Share Hard/Soft*
Rai 1					
Economy	8.1%		Economy	10.8%	
Foreign Affairs	8.8%	43.3%	Foreign Affairs	9.9%	55.1%
Judicial	14.0%		Judicial	12.1%	
Politics	14.4%		Politics	22.3%	
Crime	23.3%	54.7%	Crime	25.6%	44.9%
Entertainment	31.4%		Entertainment	19.3%	
Rai 2					
Economy	1.2%		Economy	11.6%	
Foreign Affairs	13.2%	49%	Foreign Affairs	8.3%	57.6%
Judicial	15.0%		Judicial	12.5%	
Politics	19.6%		Politics	25.2%	
Crime	24.4%	51%	Crime	27.8%	42.4%
Entertainment	26.6%		Entertainment	14.6%	
Rai 3					
Economy	11.1%		Economy	14.0%	
Foreign Affairs	14.3%	59.8%	Foreign Affairs	11.2%	59.2%
Judicial	13.8%		Judicial	13.9%	
Politics	20.6%		Politics	20.1%	
Crime	17.3%	40.2%	Crime	19.8%	40.8%
Entertainment	22.9%		Entertainment	21.0%	

Note: Topic and hard/soft news category are taken from the aggregation original labelling of each segment. Aggregates are computed by using half standard deviation intervals around the mean omega score in the two periods for the treatment group. Total amounts to 95, 222 segments in the *Pre* period and 263, 538 in the *Post* period.

Table 1.11: Descriptive Statistics of ITANES data

Statistic	N	Mean	St. Dev.	Min	Max
Age	3,343	49.006	17.291	18	98
Favourite Newscast (Tg) - Wave 1	3,222	4.239	2.393	1	8
Favourite Newscast (Tg) - Wave 2	3,215	4.388	2.442	1	8
Favourite Newscast (Tg) - Wave 3	2,764	4.376	2.440	1	8
Frequency watching favorite newscast - Wave 1	3,330	2.680	0.711	0	3
Frequency watching favorite newscast - Wave 2	3,324	2.644	0.748	0	3
Frequency watching favorite newscast - Wave 3	2,841	2.681	0.700	0	3
Propensity To Vote for PDL - Wave 1	3,111	2.997	3.632	0	10
Propensity To Vote for PDL - Wave 2	3,117	2.600	3.321	0	10
Propensity To Vote for PDL - Wave 3	2,720	2.157	2.974	0	10
Likeability of Berlusconi - Wave 1	3,182	3.363	3.265	0	10
Likeability of Berlusconi - Wave 2	3,225	2.749	2.957	0	10
Likeability of Berlusconi - Wave 3	2,752	2.452	2.677	0	10
Left-right self-positioning - Wave 1	3,011	4.854	2.976	0	10
Left-right self-positioning - Wave 2	2,999	4.702	2.927	0	10
Left-right self-positioning - Wave 3	2,479	4.511	2.707	0	10
Ideological distance with PDL - Wave 1	2,799	3.918	3.196	0	10
Ideological distance with PDL - Wave 2	2,796	4.012	3.209	0	10
Ideological distance with PDL - Wave 3	2,349	4.012	3.034	0	10
Overall government evaluation - Wave 1	3,236	3.215	2.869	0	10
Overall government evaluation - Wave 2	3,274	3.021	2.562	0	10
Overall government evaluation - Wave 3	3,272	3.344	2.506	0	10

Table 1.12: Effect of ω scores, within-individual fixed-effect regression

	Dependent variable: Propensity to Vote for PDL		
	(1)	(2)	(3)
Omega 30 days, Std	0.213***	0.379***	0.335***
	(0.047)	(0.075)	(0.076)
PartyID, PDL	-	-	0.644***
			(0.044)
LR distance with PDL, Std	-	-	−0.221***
			(0.057)
Individual FE	✓	✓	✓
Wave FE	✓	✓	✓
Tg FE		✓	✓
N	6,191	6,191	5,456
R^2	0.006	0.009	0.091

Note: Table entries are standardized coefficients from a within-individual fixed-effect regression model, standard errors in parentheses. Significance stars: *p<0.05; **p<0.01; ***p<0.001.

Table 1.13: Effect of ω scores, within-individual fixed-effect regression

	Dependent variable: Propensity to Vote for PDL		
	(1)	(2)	(3)
Omega (30 days average), Std	0.873***	0.380***	0.290***
	(0.040)	(0.076)	(0.076)
Tg frequency	0.063**	0.079***	0.048*
	(0.023)	(0.023)	(0.022)
PartyID PDL, Std			1.200***
			(0.032)
LR distance PDL, Std			−1.005***
			(0.038)
Omega × Tg frequency	0.079***	0.083***	0.049*
	(0.020)	(0.020)	(0.019)
Constant	2.860***	3.542***	3.608***
	(0.056)	(0.226)	(0.217)
Individual RE	✓	✓	✓
Tg FE		✓	✓
N	6,189	6,189	5,455
Log Likelihood	−14,823.580	−14,736.950	−12,028.120
AIC	29,659.150	29,497.900	24,084.250
BIC	29,699.530	29,578.670	24,176.710

Note: Table entries are standardized coefficients from a between-individual random-effect linear model, standard errors in parentheses. Significance stars: *p<0.05; **p<0.01; ***p<0.001. The interacting variable *TG frequency* indicates the general reported frequency of watching newscasts, rescaled to have 0 at the "Every day" category (81% of the sample).

Table 1.14: Effect of ω scores, within-individual fixed-effect regressions

	Dependent variables:		
	Government evaluation	Berlusconi rating	Left-right position
	(1)	(2)	(3)
Omega 30 days, Std	0.095†	0.552***	0.137*
	(0.076)	(0.066)	(0.063)
PartyID PDL, Std	0.306***	0.456***	0.203***
	(0.050)	(0.040)	(0.039)
LR distance with PDL, Std	−0.147**	−0.277***	-
	(0.056)	(0.049)	
Individual FE	✓	✓	✓
Year FE	✓	✓	✓
Tg FE	✓	✓	✓
N	4,805	5,505	5,939
R^2	0.025	0.089	0.013

Note: Table entries are standardized coefficients from a within-individual fixed-effect regression model, standard errors in parentheses. Significance stars: *p<0.05; **p<0.01; ***p<0.001.

† ω scores positively and significantly affect government evaluations in specifications (1) and (2), but statistical significant is no longer achieved when PartyID and LR distance are included in the model. Full table reported in online appendix.

Chapter 2

Alternative Media and Political Activism: Evidence from the Black Press and the NAACP

2.1 Introduction

Several studies have found that media presence is associated with more political participation (Gentzkow, Shapiro and Sinkinson, 2011; Drago, Nannicini and Sobbrio, 2014; Cagé and Rueda, 2016). The sheer distribution of media outlets keeps voters informed and motivated. Surprisingly, however, the partisan press doesn't affect political behavior significantly, as studies have shown that neither Republican or Democratic newspapers were able to foster turnout for their party (Gentzkow, Shapiro and Sinkinson, 2011).

I argue that this result may be limited to mainstream publications that, whichever their leaning, inevitably do extensive reporting on elections, and therefore activate voters. Little is known, instead, of how alternative media, that is publications of small-scale, counter-hegemonic groups and individuals, affect political outcomes and, in particular, how effective they are in eliciting the activism

of fringe groups (Atton, 2002).

I argue that any new political initiative needs visibility to recruit members and supporters, and the presence of sympathetic media could help diffuse the message and agenda. As the media can act as gatekeepers of political ideas, the lack of press interest may thwart the efforts of marginalized groups (White, 1950; Schudson, 2002; Shoemaker and Vos, 2009; Soroka, 2012). For these reasons, alternative media may play a crucial role in the success of political mobilizations.

I test this hypothesis by exploiting the historical case of the foundation of the National Association for the Advancement of Colored People (NAACP hereafter) in the U.S. and by looking at its relationship with the black press (Kellogg, 1967). Founded in 1909 in New York City by, among others, W.E.B. Du Bois, the NAACP sprang out of a web of smaller initiatives that were voted to the protection and enfranchisement of the African American community in the U.S. From its foundation, the NAACP advocated for the rights of the African American community and ended up becoming one of the most important actors in the Civil Rights Movement (Morris, 1984). Its activities grew exponentially in the territory, with donors emerging from both the black and the white community, and so did its agenda, adding every year ever more important and ambitious goals. If at the beginning its activities were limited to monitoring and reporting cases related to lynchings of members of the African American community, its role was reportedly crucial in the final battles of the civil rights movement and the desegregation of the black population (Morris, 1984).

In the beginning, the NAACP limited its activity to the monitoring of the violence perpetrated to the black population and circulating it in the community. This dispatching of stories and exposès was done relying, in no small part, on the dense and active networks of dedicated black press (Kellogg, 1967). While these emissions were sent to many newspapers, including prominent publications such as *the New York Times*, the mainstream media rarely reported about the activities of the association or the content of those dispatches (Kellogg, 1967). The Black Press newspapers were, instead, regularly helping the NAACP achieve visibility as it is was fitting with their militant role.

2.1. INTRODUCTION

One common goal associating the NAACP and the black press was the battle against the lynchings of the black population. Since the foundation by Frederick Douglass of the first black newspaper, the Freedom's Journal, the black press was very active in denouncing the practice and in trying to investigate the cases, often preposterous, backing up the sentence of lynching (Detweiler, 1922; Evan, 2007).

This paper evaluates how the relationship between the black press and the NAACP helped the success of the newfound organization. It tests how the sudden entry and exit of black newspapers affected the speed and efficacy of geographic diffusion of NAACP chapters across the U.S. It exploits the plausibly exogenous variation coming from the sudden entry or exit of black press newspapers in each county market (Gentzkow, Shapiro and Sinkinson, 2011), and it evaluates how changes in the number newspaper in the market affected the probability that the NAACP would establish a chapter in the following four years. To understand better the correlates of entry and exit of both a newspaper and of an NAACP chapter, I explicitly analyze the drivers of entry and exit of both separately. Consistently with historical accounts Baldasty (1992), newspaper entry is mostly driven by black population and income, while the NAACP by the level of urbanization and income.

I find that the presence of black press acted as a *deterrant* to the diffusion of NAACP chapters, differently to the distribution of mainstream media that are positively correlated with the NAACP. The effect is non-linearly related to the number of black press newspapers available with the strongest effect to the entry of the first black press newspaper in the county. The results are robust to the presence of pre-trends and several observable and unobservable characteristics of the county. The results are suggestive of a *substitution effect* between the NAACP and the black press.

I then evaluate the implications of this results by looking at how the sorting of the black press newspapers and NAACP chapters into a different location, affected the well-being of the African American community, specifically for lynching. I find that, while the entry of an NAACP chapter did not affect lynching in the following periods, the entry of a black press newspaper led to a spike in lynching for the three years following entry. These results are consistent with historical accounts

that report evidence of a *backlash* reaction by the white population to the circulation of black press publications Detweiler (1922).

The results presented in this paper provide evidence of new dynamics existing between the press and political participation, adding important implications for the effectiveness of political protests. The evidence reinforces the importance of the media in their role of gatekeepers and adds evidence on alternative media, a segment of publishing that has been completely neglected by political scientists so far.

The paper proceeds as follows. Section 2 depicts the historical case and report the documentation available. Section 3 describes the data. Section 4 analyses the drivers of entry and exit of newspapers and NAACP. Section 5 describes the empirical strategy. Section 6 and 7 report the results, and Section 8 concludes.

2.2 Lynching, the Black Press and the NAACP

The liberation from slavery and consequent enfranchisement of the black community in the U.S. was helped, in part, by a web of independent initiatives that promoted the civil rights cause. Among these, there was a group of publications, generally called the black publications, that were distributing pamphlets and news stories about the black community. The first was the *Freedom's Journal*, which was published in New York since 1827 by Frederick Douglass and can be considered the first newspaper dedicated to the African American community. In the years since a burgeoning number of other daily and weekly publications emerged in support of the African American community, and currently, it is possible to count more than 200 black press publications in the U.S. with a circulation of about 15 million copies a day (Giles, 2003). These alternative sources of information intended to document the experiences of the African American communities, not unlike what the mainstream media were doing for the white population, but their reporting, essays, and opinion pieces were explicitly political, and were often signed by prominent figures in the battle for civil rights, such as Frederick Douglass, W.E. B. DuBois, and Ida B. Wells.

2.2. LYNCHING, THE BLACK PRESS AND THE NAACP

The black press publications, except for few such as the Chicago Defender that at its peak had more than ten-thousands dealers across the United States had only local circulation and smaller readership with respect to mainstream sources (Grossman, 1989). They were also often characterized by limited run since they were, for the most part, family-owned and therefore many of them were forced to rapidly close when families did not have enough money to keep them going (Detweiler, 1922). Throughout the history of the black press, we can count about 346 newspapers, and as of today, there are about 200, with an overall circulation of 15 millions[1].

The average content of the black press papers is similar, in all respects, to that of the mainstream, white, newspapers. Black newspapers included information and news that were of particular interest for the community, often collected by specialized reporters or clipped from other newspapers. Such information included stories and practical issues, like which shops to avoid or where segregation took place (Danky and Hady, 1998). Detweiler (1922) provides a comprehensive analysis of the content of a list of forty black newspapers and reports that most of the content in each is dedicated to local news, while editorials occupy an average of 4% and about one-seventh goes to out-of-town news. Black press papers were widely read in the community, and typically their readership was extending to five times the reported circulation (Detweiler, 1922). A prominent black press publication was *The Crisis*, the official magazine of the N.A.A.C.P. that started its circulation in 1910, the year after the organization was founded. It was a 48 pages literary magazine and propaganda outlet. The organization was devoted to the fight for civil rights for the African-American community and the abolition of race discrimination. To diffuse its message, the NAACP made frequent use of the press, especially outside its magazine. To counterbalance the mainstream focus on the white population, the NAACP press office was consistently diffusing dispatches of stories involving the black community and the activities of the association to all newspapers and magazines, which, for the majority of cases, did not run them (Kellogg, 1967). Black press newspapers were, however, significantly more prone to the diffusion and publication of these stories, helping

[1]For comparison, latest data from the New York Times ascribe its circulation to 2.5 million a week (print and digital).

the promotion of the activities of the NAACP.

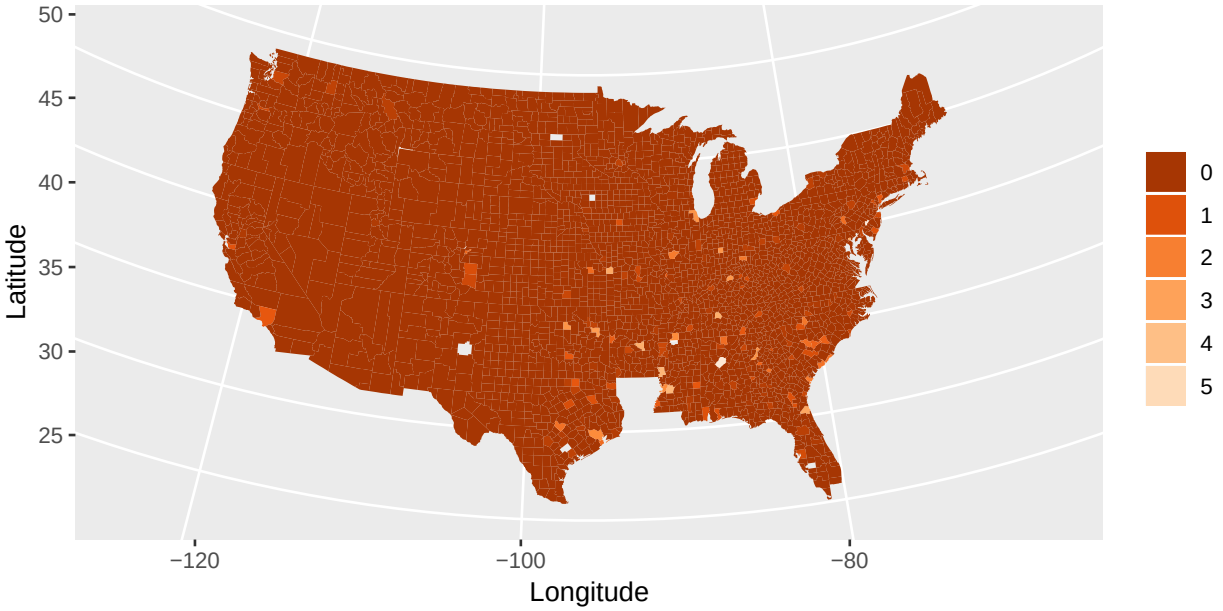
Both the black press and the NAACP were united in the prominent effort to monitor, if not limiting, the practice of lynching of the black community that was widespread in the South of the United States (Detweiler, 1922; Kellogg, 1967). The black press publications were unflinching in their condemnation of lynching, and the NAACP often investigated them directly as murders in trying to reassess the culpability of the victim and in identifying the perpetrators. This paper provides direct evidence of the relationship existing between the black press and the NAACP, and their consequences on lynching.

2.3 Data

2.3.1 Black Press Newspaper Panel

I have collected data about Black press publications from annual directories of US newspapers from 1908 to 1920, inclusive. These are the years in which the NAACP initiated its activities across the U.S. Most of the data collected come from the *African American Newspapers and Periodicals: A National Bibliography* (Danky and Hady, 1998) which represent the most comprehensive historical collection of African American newspapers in the U.S. The newspaper directories were compiled for potential advertisers interested in different markets and they were intended to be exhaustive of the possibile publication. For each year reported, I extract the name, city, publication day, number of issues, and circulation (when available) of every English-language newspapers (Some black press newspapers are in French). Figure 2.1 depicts the average number of black press newspaper circulating in each county. It is easy to see that most counties did not have a black press paper circulation and that they were mostly concentrated in the South East of the county. I merge this dataset with the data on mainstream newspapers from Gentzkow, Shapiro and Sinkinson (2011), that for the period of interest in this paper, reports data collected in N. W. Ayer and Son's (Ayer's) *American Newspaper Annual*. I then aggregate the number of observations at the county-year level

Figure 2.1: Average Number of Black Press Newspapers by County (1908 - 1918)

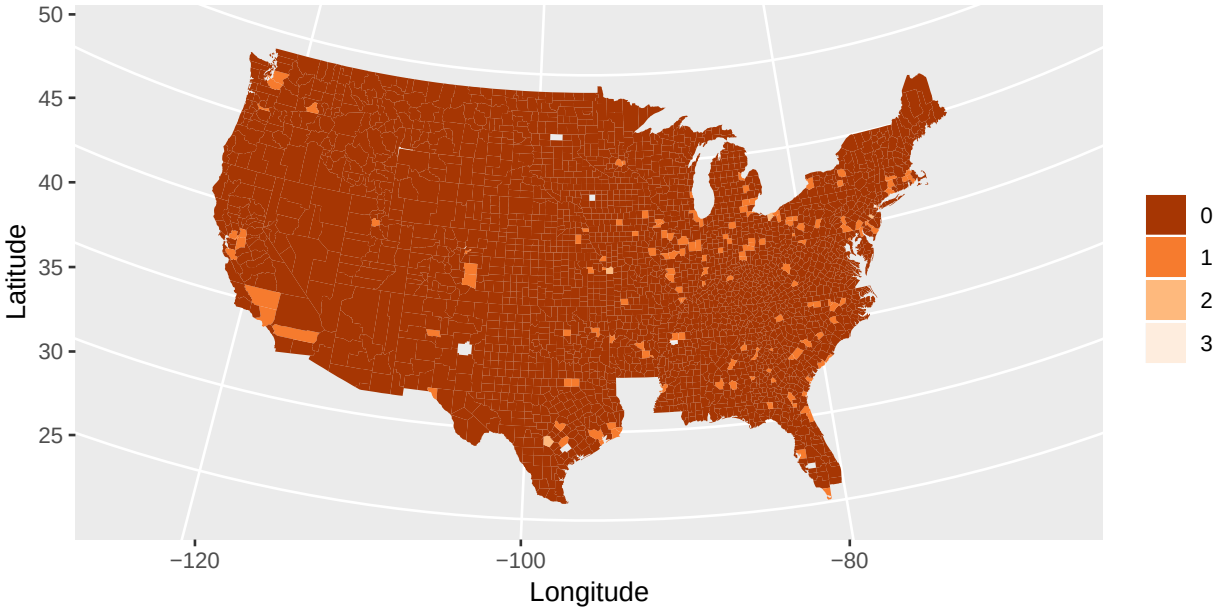


and determine the total number of newspapers circulating in each county-year pair, which represent the main dependent variable in this study.

2.3.2 NAACP Data

Data are collected from Mapping American Social Movements Project at the University of Washington. The data are disaggregated at the branch level and contain information about date of opening, city, state, years of operation after opening, membership levels from 1912 to 1977 and a detail of the activities that each branch was pursuing. Each time a new branch was opened, its existence needed to be ratified by the national council Kellogg (1967). The data are then collected based on the archives of the minutes of the meetings of the NAACP national council, which are available at the Library of Congress. Figure 2.2 reports the average number of branches present in each county during the period under study. The diffusion of the NAACP extends across the United States. In

Figure 2.2: Average Number NAACP Branches by County (1908 - 1918)



the period under study, I can account for more than three hundreds branches of the organziation, while, currently the NAACP has more than one thousand.

2.3.3 Lynching Data

The lynching data come from the NAACP historical lynching records and the database HAL, the historical American lynching database run by scholars, which used original NAACP lynching records as its basis and then verified the records from 1882 to 1930 when most lynchings occurred. All data used here have been additionally verified with secondary sources.

2.3.4 Demographic Data

I match each county-year observation to data on demographic caracteristics from US Census and County Data Books (ICPSR 2896 by Michael R. Haines 2006) supplemented with data from

the National Historical Geographic Information Systems (NHGIS.org). I compute the share of population that is black, the share of population that is foreing-born, the share of population that is males and 21 or older, the share of population living in cities with 25,000+ residents, the share of the population liniivng in towns with 2,500+ residents and the population employed in manufacturing as a share of males 21 and older. I interpolate nominator and denominator across census year by using a natural cubic spline Herriot and Reinsch (2002) and divide the two to obtain an estimate of the relevant share. I also use data from the census on the definition of Primary Metropolitan Statistical Areas (FIPS) as of 1990. Specific county-level data on educational attainment do not exist so I use census variables on literacy rate among the black and white population as a proxy for education.

2.3.5 Market Definition

I define the news market to be a county. I do this because the county is the smallest unit at which I can disaggregate Census data over such a long period of time. In fact, some counties contain multiples news markets (cities) and newspapers circulate across nearby counties (some, such as the *Chicago Defender* have almost national reach). However, Gentzkow, Shapiro and Sinkinson (2011) show that today the median newspaper sells more than 80 percent of its copies in the county in which its headquareted, and the median county in which at least one newspaper is headquarted gets more than 80 percent of its copies from in-county newspapers. Given improvement in transportation led the circulation to expand, for the historical focus of this paper, the county is probably a good approximation of the relevant market.

2.4 Determinants of Entry and Exit

2.4.1 Black Press Newspapers

To justify the identification strategy used in this paper, I clarify the forces driving entry and exit of newspapers. Historical accounts of this period indentify two main drivers (Baldasty, 1992). The first is population, an specifically literate population. Since newspaper have large entry cost, and relatively low price tag, the potential number of consumer is extremely relevant for their business. The second is income. Advertisers typically care about the purchasing power of their costumers and therefore richer areas have higher advertisment revenue. I test these conjectures with my data and evaluate the demographic caracteristics associated with newspaper presence and entry. Table 2.1, column (2) reports the cross-section correlates of newspaper presence. This table reports coefficients from an OLS regression in which all the yearly observation are pooled together. I dint that both income and population are indeed positively associated with newspaper presence, and, more specifically, that a urban, literate population is conducive to a more florid newspaper market. I replicate this analysis by using the panel structure of my data, and looking at whether changes in demographic characteristics affect the drivers of entry and exit. Results reported in Table 2.2 reveal that only changes in the literate black population affect the entry of black newspapers, while changes in income only slightly affect entry.

2.4.2 NAACP Chapters

I replicate the analysis above for the entry and exit of NAACP chapters. Historical accounts reveal that the decision to install a branch of the NAACP in a particular community was relatively top-down, and coming from the initiative of particular donors willing to contribute to the activities of the association (Kellogg, 1967). My analysis reveal that the NAACP was typically entering in urban, homogenous, and relatively richer areas, perhaps where the availability of potential donors was higher. Surprisingly, instead, its presence was not associated to movement in the concentration

Table 2.1: Correlates of NAACP and Black Press

	NAACP	Black Press
	(1)	(2)
Urb 25K	0.055***	0.069***
	(0.003)	(0.010)
Urb	0.005	0.129***
	(0.003)	(0.011)
Foreign	−0.036***	−0.185***
	(0.008)	(0.024)
Blacks	−0.003	0.578***
	(0.007)	(0.021)
Black Ill	0.031	−2.666***
	(0.041)	(0.133)
Manuf Output pca	0.0001***	0.0001***
	(0.00000)	(0.00000)
N	35,412	35,412
R^2	0.056	0.071
Adjusted R^2	0.055	0.071
Residual Std. Error (df = 35405)	0.107	0.344
F Statistic (df = 6; 35405)	347.769***	452.839***

*p < .1; **p < .05; ***p < .01

of black population. All the demographic controls used here are included in the main analysis.

Table 2.2: Drivers of Entry/Exit

	NAACP	Black Press
	(1)	(2)
Numb All	0.057***	0.016***
	(0.001)	(0.003)
Urb 25K	0.093**	−0.090
	(0.047)	(0.109)
Urb	−0.005	−0.023
	(0.067)	(0.156)
Foreign	−0.942***	0.634
	(0.180)	(0.417)
Blacks	−0.243	2.057***
	(0.252)	(0.584)
Black Ill	0.078	−0.995*
	(0.225)	(0.521)
Manuf Output pca	0.00002***	−0.00002*
	(0.00001)	(0.00001)
N	32,688	32,688
R^2	0.078	0.002
Adjusted R^2	0.077	0.001
F Statistic (df = 6; 32681)	456.540***	9.064***

*p < .1; **p < .05; ***p < .01

2.5 Empirical Framework

2.5.1 Identification

Let c index counties, s index states, and $t \in \{1, \ldots, T\}$ index years. My outcome of interest is the presence (and number) of NAACP branches within each county c at time t. My key independent variable of interest is n_{cst}, which is defined as the number of black press newspapers in county c, state s, at time t. Formally the specification can be written as follows:

$$\text{NAACP}_{ct} = \beta_1 n_{ct}^{black} + \beta_2 n_{ct}^{all} + \delta \mathbf{X}_{ct} + \rho_c + \gamma_{st} + \epsilon_{ct} \tag{2.1}$$

where β are county fixed effects, γ_{st} are state-year fixed effects, $\mathbf{X}_{cst}$ is a vector of observable characteristics, δ is a vector of parameters, and ϵ_{ct} is a county-year shock. β is the causal effect of n_{ct} on NAACP_{ct}. Unless differently specified, the vector $\mathbf{X}_{ct}$ the share of population that is Black, the share of population that is black and illiterate, the share of population that is foreing-born, the share of population that is males and 21 or older, the share of population living in cities with 25,000+ residents, the share of the population liniivng in towns with 2,500+ residents and the population employed in manufacturing as a share of males 21 and older.

I estimate the model above in first differences because of high prevalence of persistent shocks. We let the Δ operator be the first-difference operator and the result come from the estimation of:

$$\Delta\text{NAACP}_{ct} = \beta_1 \Delta n_{ct}^{black} + \beta_2 \Delta n_{ct}^{all} + \Delta\gamma_{st} + \delta \Delta \mathbf{X}_{ct} + \Delta\epsilon_{ct} \tag{2.2}$$

where county fixed effect drop out because of the differencing.

This specification highlights that sudden, yearly, changes in the number of total newspapers, and number of black newspaper, affect the probability that, in the following year the NAACP installs a branch of the organization. The estimates will be driven by the way in which Black politi-

cal organization is affected in counties that experience changes in newspaper media market relative to counties in the same state and year that do not. I cluster standard error at the level of the county to allow for correlation over time within a county.

Similarly to Gentzkow, Shapiro and Sinkinson (2011), I assume the ϵ_{ct} is a county-year idiosyncratic shock to the outcome of interest, in this case the foundation of a NAACP branch, that is unrelated to newspaper profits and, therefore, to the newspaper decision to entry or exit the market. I test this assumption by using pre-trends, as it is standard practice for panel data. If the relationship between Δn_{ct}^{black} and ΔNAACP_{ct} comes only from the causal effect, then Δn_{ct}^{black} cannot be correlated with past values of ΔNAACP_{ct}.

2.6 Effect of Black Press Newspapers on NAACP Chapters

2.6.1 Main Analysis

The main analysis uses as independent variable the count of all mainstream and black press newspapers circulating in the county. I evaluate the general, average, effect of entry, and the effect conditional on the pre-existing number of newspapers in the county. This second analysis disentangle the effect coming from the extensive margin, the entry of the first black newspaper, from the intensive margin, the accumulation of more black press newspapers.

I present the analysis both using the two-way fixed effect model and the first-difference model introduced above. This allows to give a more comprehensive picture by exploiting the difference in market structure, and by identifying the on-impact effect of entry.

There are a number of potential mechanisms linking the presence of the black press with the one of the NAACP. If we consider political mobilization, such as the one of the NAACP, within the context of political participation more generally, we should expect the presence and entrance of a newspaper to have a positive effect on participation (Gentzkow, Shapiro and Sinkinson, 2011). However, if we are considering the NAACP as a fringe group organization, the effect may be

less clear. If the activities of the alternative newspapers and the NAACP belong to the same subpopulation, the effect on the media presence could be negative as there could be substitution among the two. Yet, in the same context, the effect could be positive if, instead, the are synergies and complementarities between the activities of the press and those of the NAACP. For instance, the black press could help the popularity of the NAACP by reporting on their activities and enhancing their visibility. I bring this hypothesis to the test in Table 2.3. The results show an overall negative effect of the number of black press newspapers on the entry of a branch of the NAACP. This effect is in net contrast the the total number of mainstream newsapapers, that instead affect positively entry. The results of the mainstream media are consistent with previous evidence that shows that media have a positive effect on political participation (Gentzkow, Shapiro and Sinkinson, 2011). Table 2.3, column (2), reports the heterogenous effects (the intensive margin) of the market structure on the entrance of a NAACP branch. The relationship seems to be non-linear, as the effect decreases and then increases again. These results reveal that the effect act both on the intensive and the extensive margin.

2.6.2 Robustness

I test the validity of the identification assumption by using a pre-trends analysis as in (3), where the specification is exactly the same as in the main analysis, but adds leads and lags for the main dependent variable.

$$\Delta \text{NAACP}_{ct} = \sum_{t-4}^{t+4} \beta_1 \Delta n_{ct}^{black} + \beta_2 \Delta n_{ct}^{all} + \Delta \gamma_{st} + \delta \Delta \mathbf{X}_{ct} + \Delta \epsilon_{ct} \tag{2.3}$$

This test evaluates whether the estimated effect is the result of spurious correlation between the dependent and the independent variable as could be coming from a common trend between the two. If the assumption of absence of common trend is valid, the coefficients should not be statistically significant before the entry or exit of newspaper. The effect could, however, persist over time, and this case could be evaluated by adding dummies for the following periods.

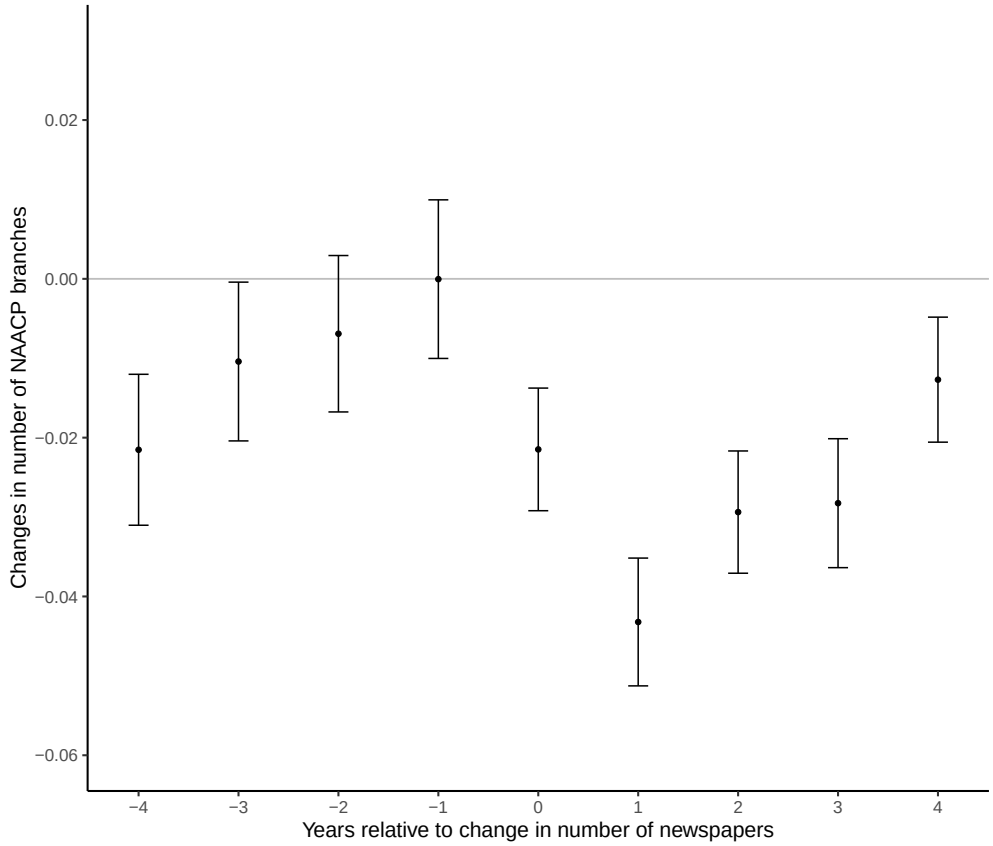
The number of period before and after treatment that is included in the specification is directly dependent on the lenght of the panel dataset. For this reason, I add to the specification four lagged and led periods.

The results are reported in Figure 2.3. They show an on-impact effect of entry that persists for the four periods after, while it is not significant in the periods before (with the expection of four periods before). The implication of this result is that, at least in this time frame, the entrance of a black press newspaper in a county leads to a persistent decrease in probability of entry of the NAACP. In a way, the persistence of the effect is not surprising, as it is easy to see that the mechanisms that may link NAACP and black press are not likely to change in the four years following the entry. Still, it reveals that the nature of the effect is not that of an immediate reaction and that is, possibly, linked to a structural relationship between the two.

2.6.3 Discussion

The evidence provided shows a negative relationship existing between the circulation of black press publications and the presence of an NAACP chapter. The results are evidence of sorting of the NAACP away from areas covered by the black press. Given the sympathetic nature of the activities of the two, the reason for sorting could not be animosity between the two. Historical accounts, in fact, present a relationship of mutual assistance between the black press and the NAACP, mostly taking the face of the black press publishing stories coming from press dispaches of the NAACP Kellogg (1967). The presence of *substitutability* could, instead, be strategic. Given the fact that, at the beginning, the NAACP was mostly a monitoring and advocacy group, by extending its activities in areas that were not serviced by the black press could increase the reach of the black protest movement. Secondly, substitution could be the result of the identity of the actors funding either the black press or the NAACP. Both organizations were dependent on the voluntary contributions of wealthy citizens, who may have decided to help the black community by picking either a newspaper or a NAACP chapter. Therefore, the results may not be linked to the nature of the two activites

Figure 2.3: NAACP Branches



but to the distribution of preferences of the actors.

2.7 Effect of Entry and Exit on Lynching

To further evaluate the political impact of the main results, this section compares the effectiveness of the two different forms of political organization on the lynchings of the black population, arguably one of the most pressing issues for the African American community at the time and an important source of motivation to political action both in the case of the NAACP and the black press.

These result will help evaluate the consequences of the sorting effect provided above and thus elevat-

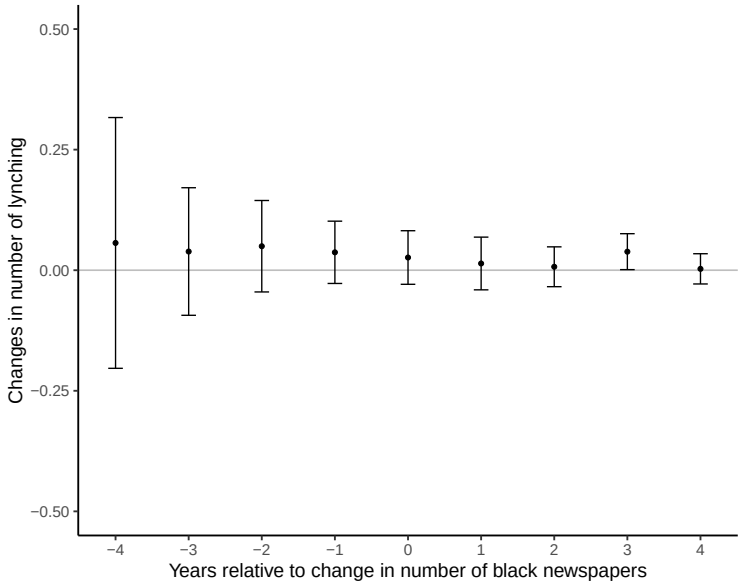
ing the significance of these results. In fact, if the activities were not equally effective at improving the living conditions of the African American community, then the impact of sorting could be quite consequential, possibly leading to conclusions about the desirability of either form of political action.

The results presented in this section resorts to a similar emprical strategy in Section 6. More specifically, I will expolit the sudden entry and exit of either NAACP chapters, mainstream, or black press newspaper to evaluate the effect on the number of lynchings in the coming periods. Analytically, this would lead to:

$$\Delta \text{Lynchings}_{ct} = \sum_{t-4}^{t+4} \beta_1 \Delta n_{ct}^{black} + \sum_{t-4}^{t+4} \beta_2 \Delta n_{ct}^{all} + \sum_{t-4}^{t+4} \beta_3 \Delta n_{ct}^{NAACP} + \Delta \gamma_{st} + \delta \Delta \mathbf{X}_{ct} + \Delta \epsilon_{ct} \quad (2.4)$$

Where the dependent variable, Lynchings is the number of lynchings in county c at time t.

Figure 2.4: Effect of Branches Entry



2.7. EFFECT OF ENTRY AND EXIT ON LYNCHING

Figure 2.5: Effect of All Mainstream Entry

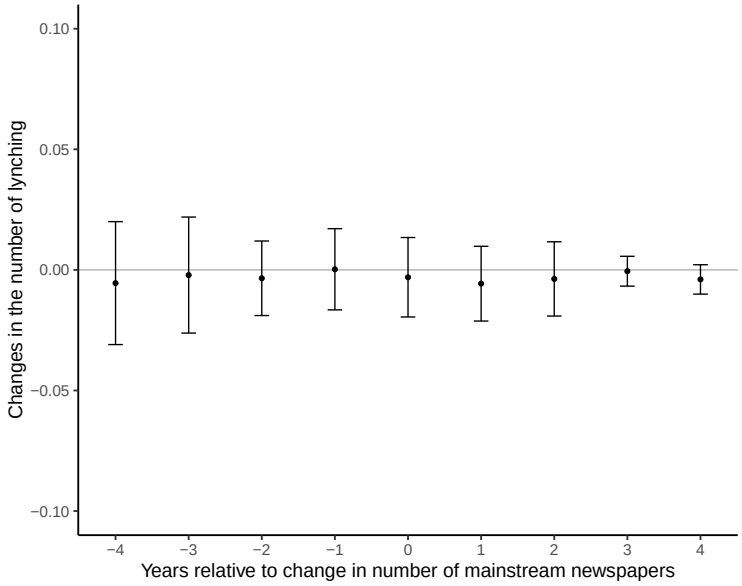
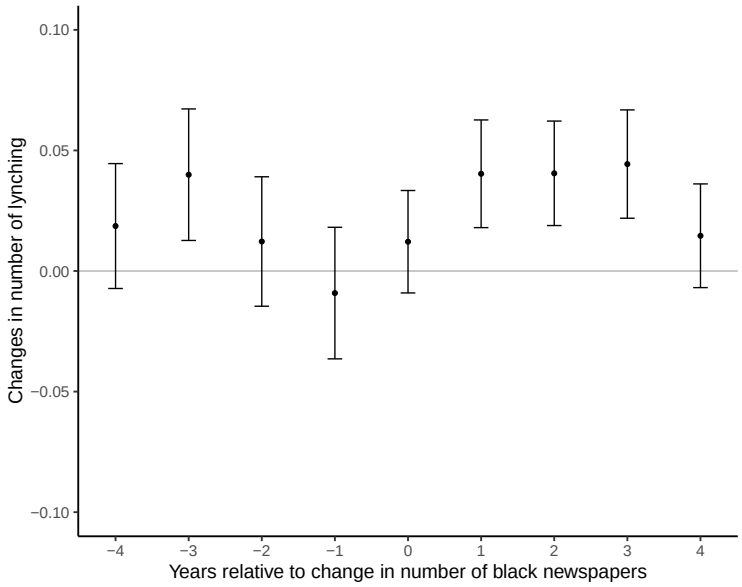


Figure 2.6: Effect of Black Press Entry



The results from the estimation are presented in Figure 2.4, 2.5, 2.6 and show that the entry of a black press newspaper had a significant and positive effect on the number of lyinchings

in the three years following the entry while neither the NAACP or the mainstream media seemed to affect them. These results could be taken as evidence of a *backlash effect* of the black press. These results are consistent with anecdotal evidence that relates the presence of the black press to lynching. As seen in Section 2, the black press was adamant in the denouciation of violence perpetrated to the black population and in many cases led to protest and attacks from the white population. In certain cases, the white population issued petition for the dismantling of the local black press paper Detweiler (1922), at times advocating with racial slur such as in:

> White people of this city have issued an order that no "colored newspapers" must be circulated in this town, but that every "darkey," the petition reads, must read the *Falcon*, a local white paper edited by a Confederate veteran. The whites stated this step was being done in order to keep the "nigger from getting besides himself, and to keep him in his place."

The petition quoted above appeared in Sommerville, Tennessee in 1919. In other instances, the protests became violent attacks toward journalists working for the black press, including cases of lynching like the following:

> Athens, Ga., Feb. 17. - John Lee Eberhardt, your staff correspondent here, was taken from the local jail and burned at the stake. Eberhardt had been arrested early yesterday as a suspect in the murder of a white woman by name of Lee. It is believed by many here that advantage was taken of the situation to "get" Eberhardt for circulating the *Philadelphia American*. He had been warned on several occasions. More than five thousand took part in the lynching.

This quote reported in Detweiler (1922) resorts to 1921 but, other historical accounts mark the practice of lynching of journalists of the black press as not rare. These anectodes could help explain the effects reported above and why the entry of a black press newspaper may lead to the paradoxical increase in lynching. However, it is unclear why the NAACP, that reportedly was

acting on the same role did not have a similar impact. Future work should evaluate the mechanisms by which these two different forms of protest have different effects on lynching.

2.8 Conclusions

The literature on political protest has often considered that media visibility is a primary requirement for the success of protest movements (Boyle and Schmierbach, 2009). This paper shows that the relationship between the media and newborn political movements may be more articulated than that and, especially, take aspects of substitutability instead of complementarity. The different reach on the territory of the protest movement that results from the previous allocation of substitute activites, may lead to different success outcomes for the protest itself. This could happen, for instance, when other, more prominent and established media outlets are antagonist to the protest movement and hence the presence of sympathetic media may radicalize the public perceptipon of the oppressed group. In the specific case presented here, the prexisting distribution of the black press, led to a different allocation of the NAACP branches and a backlash effect on the lynchings of the black population.

The results from this paper are consistent with a model of media as 'gatekeepers', that is media as actors that may determine the success or failure of political movement conditional on their visibility and reveals how their strategy may determine the success or failure of the protest activities. Additionally, it shows that some form of political protest, such as media activity, may be counter productive for the goal of the protest as they increase the political tension between different social group without empowering the oppressed one. This paper provides these results with some limitations to their generalizability. The first limitatation is that I study the relationship between the media and political activism in a context in which the media market is limited to the printing press. In the more competitive market now effects may be smaller but reach could be higher. It is therefore hard to predict how these results would extend to different media markets. The second limitation is that I study the two forms of protest when they were pursuing the same activities. As

the NAACP evolved into a different organization with more resources, the presence of substitution between the two may not be relevant anymore.

Nevertheless, this work provides new insights on the relationship between the media and protest movements, with important consequences for the democratic process.

Table 2.3: Panel Analysis

	Total Effect	Heterogenous	Sample Entry
	(1)	(2)	(3)
# All Press	0.024***	0.024***	0.022***
	(0.001)	(0.001)	(0.001)
# Black Press	−0.053***		
	(0.002)		
≥ 1 newspaper		−0.075***	−0.074***
		(0.005)	(0.005)
≥ 2 newspaper		−0.027***	−0.008
		(0.010)	(0.008)
≥ 3 newspaper		−0.109***	−0.046***
		(0.012)	(0.011)
% Urban 25K	0.086***	0.085***	0.079***
	(0.012)	(0.012)	(0.010)
% Urban	−0.065***	−0.065***	−0.057***
	(0.017)	(0.017)	(0.015)
% Foreign	−0.461***	−0.461***	−0.359***
	(0.086)	(0.086)	(0.074)
% Blacks	−0.029	−0.001	−0.016
	(0.083)	(0.083)	(0.072)
% Black Illiterate	−0.144**	−0.138**	−0.110**
	(0.059)	(0.059)	(0.051)
Manufacturing 1000$/capita	0.032***	0.031***	0.028***
	(0.004)	(0.004)	(0.003)
N	35,412	35,412	35,388
R^2	0.303	0.303	0.282
Residual Std. Error	0.096 (df = 32677)	0.096 (df = 32675)	0.082 (df = 32651)

*p < .1; **p < .05; ***p < .01

Table 2.4: First Difference Analysis (Robust)

	No Controls	Controls	Sample Entry
	(1)	(2)	(3)
# All Press	0.056***	0.058***	0.049***
	(0.008)	(0.008)	(0.008)
# Black Press	−0.036***	−0.035***	−0.023***
	(0.008)	(0.008)	(0.008)
% Urban 25K		0.090	0.065
		(0.060)	(0.060)
% Urban		−0.006	−0.004
		(0.078)	(0.078)
% Foreign		−0.920***	−0.645***
		(0.205)	(0.205)
% Blacks		−0.170	−0.210
		(0.301)	(0.301)
% Black Illiterate		0.042	−0.002
		(0.235)	(0.235)
Manufacturing 1000$/capita		0.024***	0.017***
		(0.005)	(0.005)
State FE	*No*	*Yes*	*Yes*

*p < .1; **p < .05; ***p < .01

White-Robust SE, clustered at county level.

Chapter 3

How Does the Media Environment Affect Readership? Evidence from an App Patient-Preferred Trial in Italy

3.1 Introduction

The introduction of online outlets has radically reshaped our media landscape. The deluge of news sources that has characterized this past decade has affected the entire market for news profoundly. It has created an unprecedented level of media competition (driven by infinitely smaller entry cost of online outlets) and diverted interest away from traditional news sources (20% of Americans say they get their news from social media often (Shearer and Gottfried, 2017)).

Researchers concerned with recent media trends have typically focused on social media. Social media report news in a very different form with respect to traditional outlets: the information posted is typically not vetted, and often individual newsfeeds are customized to the taste of the user (Prior, 2005).[1] As a result we have mounting evidence showing that social media foster the creation of

[1]Even more alarming, a recent survey of 2000 adult Americans reports that 86% does not fact-check what they see on social media. This negligence has predictably led many observers to worry about the possibility of misinformation

echo chambers and filter bubbles (Flaxman, Goel and Rao, 2016; Eady et al., 2018).

Outside the realm of social media, there is very little evidence on users' online news consumption. However, news portals are among the most significant sources of information. One the most recent studies by Guess (2016), reveals that the majority of Americans have an online media diet that relies on one of a handful of popular portals such as Google, MSN news, CNN and little else. A glance at their pages shows how crowded they are with news and advertisement. They often present up to fifty stories contemporaneously and update them constantly. Analogously to the case of social media, these portals have an impressive media power and the particular way these portals select and arrange the news on the page could be very consequential.

Together these two bits of evidence suggest that, even though the internet allows users to search for whatever news they want, most users refer to pages that provide a news feed. But, how much can media diets be manipulated online?

This paper tackles this question directly and experimentally in the context of Italy, by developing a customized news-app, and by implementing a randomized field experiment on content. In particular, to focus on overall political knowledge, this paper shies away from partisan issues and concentrates on the competition between soft and hard news. This paper tests by manipulating both the quantity and the positioning of the two kinds of news, how the combination of the two affects readership, preferences, and knowledge.

The use of soft and hard news is apt to this context for two reasons. Soft news headlines are usually more enticing than hard news and, therefore, putting the two in direct comparison allows for an assessment of how much attention users spend on hard news. Secondly, the exact definition of hard and soft news implies that only one category contains politically-relevant information and thus allow for a clear idea on how readership impacts knowledge.

Methodologically, the advantages of an approach based on mobile app are two-fold. First, custom-made apps allow for total control from the experimenter back-end on actual content viewed by the users and, therefore, for precise targeting of news as well. Second, they permit continuous tracking

and echo chambers, arguably a weaker concern on traditional outlets (Bastos, 2015; Guess, Brendan and Reifler, 2018).

of what subjects receive and read, how much time they spend on each page and the frequency they look at the news. These two features result in an exact and credible measure of news consumption and, hence, of treatment effects.

One major limitation of this study is, however, that while it collects qualitative evidence about users usual news consumption, it nevertheless monitors only the consumption of news on the app, and, therefore, it cannot discern if consumption is residual, meaning news already consumed have been disregarded, or non-residual. Generally, this concern would drive against our result and it should bias downward the treatment effect, but it is still an issue for the generalizability of the results.

Nevertheless, this study tries to improve generalizability, by resorting to a patient-preferred trial (Gaines and Kuklinski, 2011). This new method consists of splitting the subject pool among two different treatment groups. Each subject in the first group will be allowed to pick the treatment that prefers, while each subject in the second group will be assigned a treatment based on the randomization (the assignment to the full treatment pool will be randomized to avoid self-selection into treatment group). This procedure permits to parse out the different treatment effect among self-selectors and non-self-selectors and provide a more realistic estimate of the implied readership effect on the general population when users can select where to read their news.

The results show that media diets can be manipulated to a considerable extent and, consistently, that individual preferences offset external manipulation only minimally. Specifically, a high hard news diet can improve readership of hard news. No effect is found on individual preferences. These results provide new evidence of the effects of the media environment on news diets with policy implications for media regulation and the importance of public service media.

The rest of the paper proceeds as follows. Section 2 reports the experimental context. Section 3 reports the theoretical framework and the hypothesis. Section 4 describes the research design. Section 5 and 6 collect the analysis and the results. Section 7 concludes.

3.2 The Italian Context

Italy is a wealthy democracy and a founding member of the European Union. Within the European context, it ranks as 25 out of 28 regarding internet connectivity, right behind Poland and before Bulgaria. This backward position is mostly due to fixed broadband coverage and take-up, while its mobile internet use is comparable to the European average.

The Italian population is adopting the new technology at a slower pace compared to other OECD countries. Only 69% of the population is regularly (daily) online, and by 2016 only 63% of the population was reached by the Internet. This low rate of connectivity persists across age groups, and it is higher for the older population, only 27% of the Italians belonging to the age group 64-75 use the internet[2](*Rapporto Sulla Conoscenza 2018* (2018)). Interestingly for this study, the Italian population uses significantly more the Internet on mobile than on the laptop, about 35% compared to the 20% of the EU average.

According to the 2018 European Union's Report on Digital Economy and Society, Italy is the European country with the lowest percentage of the population that looks at the news online. That is, 56% compared to a European average of 72%. This outcome might be related to its historically strong television sector. While 82% of Italians regularly receive information from television, only 39% declare to follow the printing press.

Readership differs among newspapers. Among the online news, La Repubblica has the dominant share. The websites with the broadest reach are those of the leading newspaper and main TV broadcasters like Rai and Corriere Della Sera. Other news portals like Yahoo News, Notizie.Libero and MSN News have a smaller, but still relevant, reach while the other digital-born outlets generate a minimal audience (Newman (2017)). Only 12% of the Italian population pays for the news. This aspect is particularly useful for this study since it will generate a better response to the retainment strategy.

[2]This is partly due to the lower level of education, among the highly educated in this category the connectivity rate is actually higher than the overall population, up to 75%.

3.3 Theory and Previous Research

Previous research has modeled news consumption following an opportunity, motivation and ability model (OMA hereafter) (Luskin (1990); Delli Carpini and Keeter (1996); Prior (2007)). Opportunity refers to the availability of political information, and it is strongly determined by characteristics of the media environment, such as the total number of news outlets. Motivation and ability are, instead, individual characteristics. Motivation encompasses interest and curiosity, while ability refers directly to the capacity of an individual to absorb information from the environment, and it is dependent on education but also on the way the environment delivers the news.

The expansion of media outlets on the internet, including within social media, has radically reshaped the media environment and the pool of readers. Thanks to the deep engagement with online media, users that previously were not reading the news became more likely to see the headlines. Secondly, the media themselves have reacted to the increase in competition by reviewing their editorial policies in order to retain their readers.

Considering this evidence within the framework of the OMA model, then, it is possible to argue that the reshaping of the media environment may have affected the way in which voters experience the news in complex and sophisticated ways. For instance, while it is true that the increase in the number of outlets has the potential to make voters more informed (i.e., by increasing opportunity), the rise in competition for attention has led many outlets to increase their entertainment coverage (i.e., decreasing opportunity). All in all, the net effect on *opportunity* is ambiguous.

Additionally, in a swirling media environment, such as that of social media, it may be particularly taxing to focus on the news, and thus many users may prefer to scroll through headlines or other more enticing posts. The implication is that the internet media environment may affect *motivation* and *ability* too (evidence in education seems to be consistent with this claim (Raymond, Shapiro and Arnell, 1992; Rose, 2010; Shawn Green and Bavelier, 2003; Ophir, Nass and Wagner, 2009)).

3.3. THEORY AND PREVIOUS RESEARCH

This paper manipulates experimentally some features distinctive of the internet environment to test their consequences on the OMA framework and, consequently, on political knowledge and engagement. In particular, it affects the opportunity to learn about politics by manipulating the concentration of hard and soft news within the medium. Depending on each user initial preferences for soft and hard news, a media diet consistent or contrasting with the initial preferences may have different effects. The experiment allows for an estimate of the effect on the opportunity with a comparison of the media diet offered in the app and the resulting readership of the user. Secondly, the experiment will allow for an estimate on motivation with the comparison between the initial preference for soft and hard news and the percentage of soft and hard news read. In other words, this experiment will allow for the evaluation of two measures of hard news readership that correlate with the opportunity: the percentage of hard news read with respect to the total number of news seen and the percentage of hard news read with respect to the total number of hard news seen. The first measure refers to behavior, while the second measure refers to taste. Finally, to test how changes in the media environment may affect users ability to learn, I plan to manipulate the amount of distraction by changing the positioning of soft news (assuming soft-news are more enticing than hard news to the average reader).

Further, this experiment will perform a test on knowledge. By comparing how the different concentration of hard news leads to a change in political knowledge, the experiment will provide evidence of the link existing between online information and political knowledge.

This discussion leads to the following hypothesis:

Hypothesis 1: Increases in hard news coverage concentration increases hard news consumption.

Hypothesis 2: Increases in hard news coverage concentration increases preferences for hard news.

These two hypotheses will be directly tackled in the following section.

3.4 Research Design

The experiment presents two central contributions. First, it will test the effect that new technologies have on the human ability to learn from the news, and more precisely, it will provide new evidence on the effect on political knowledge. Secondly, it will enable a very granular analysis of the relationships that users have with technology and an assessment of which voters are more susceptible to be affected by it.

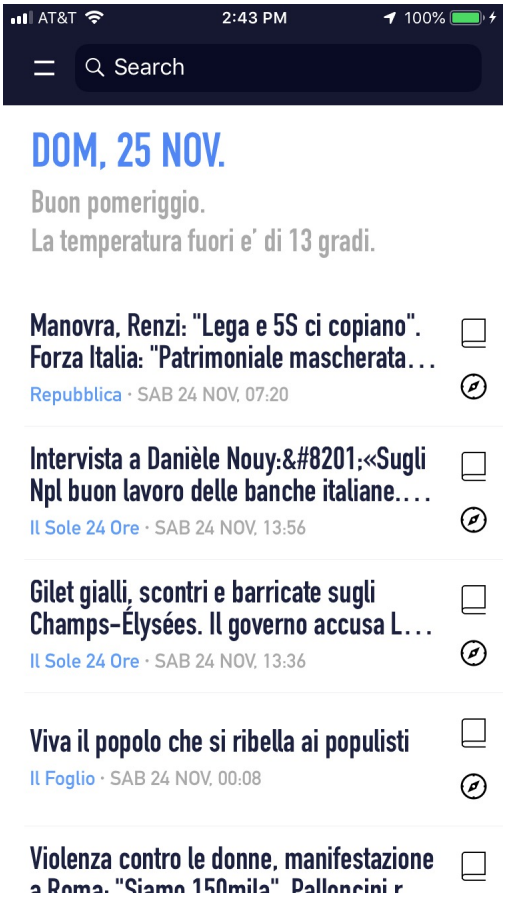
3.4.1 Dossier App Design

The UI design of Dossier involves a series of news stories with graphics and reporting of the newspaper, quite similar to other news apps. The app gathers real-time news collected from the RSS feeds of all national daily newspapers. The news feed updates each time the user enters the app and when the user scrolls up to refresh. The app creates the newsfeed based on the designated treatment. For instance, a user that is assigned to the hard news treatment receives a newsfeed that contains an infinite number of blocks of five news: four of the most recent hard news stories and one of the most recent soft news ones.
See Figure 3.1 for a rendering of the iOS version.

Online Baseline Survey

In order to overcome the issues of reliability of survey responses on behavioral measures, I will implement a strategy that combines behavioral data and surveys. Usage and voters preferences for news will be detected with the former, while information recall and political preferences with

Figure 3.1: App Design



(a) 1a: App Homepage

(b) 1b: Example of an article

the latter. I estimate baseline levels of the dependent variables by using an online baseline survey (questions will be the same as those of the endline survey, see below). Besides demographics, this analysis crucially needs to assess pre-treatment engagement with the news to identify high and low internet and app users. Recent statistics show that app users are increasing over time, mostly on the Google Play platform and mostly on social media. News apps, instead, do not score on the top 10 lists across the globe. App usage is decreasing with age. Estimates from this year show[3] that users spend on average 2.3 hours a day using apps on their phone, ranging from 3.2 hours for the

[3]http://www.businessofapps.com/data/app-statistics/

age bracket 18-24 to 1.6 hours for the over 65. In order to account for this variability, the sample selection block randomizes treatment assignment based on how different categories are expected to use their mobile (see below). Further, to gauge the pre-existing media diet and see how it differs, ideologically, from the one proposed by the app, I record the political ideology of their regular news sources by asking their more frequent sources of information.

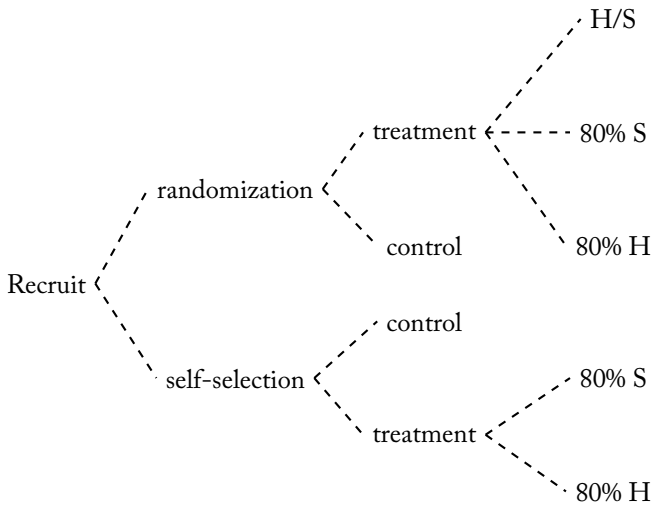
Patient Preferred Trials

This experiment is set up as a patient-preferred trial (Gaines and Kuklinski, 2011; Knox et al., 2018). All subjects are asked to state their preferred treatments, i.e., their preferred news diet, by picking the three preferred news subjects from a list of potential six including three hard news category (Economy, Italy, and International news) and three soft news categories (Entertainment and Sport and Culture). This way, whatever their selection, they would fall into one category only among hard and soft news lovers. Then, subjects are randomized into one of two conditions, either will be forced to take the randomly assigned treatment irrespective of their preferences, or they will be given the treatment consistent with their preferred diet (this aspect is crucial as by observing the actual behavior I will be able to detect discrepancies between the actual and the revealed preferences). The assignment scheme is illustrated in Figure 3.2.

3.4.2 Treatments

Soft/Hard news concentration This treatment will provide users with different concentration of high vs. soft news. Softer topics are believed to be important catalysts for news consumption, though it is effectiveness may depend on the particular type of user (if more or less attracted to them). The experiment considers two treatment conditions. The first has 20% of news pooled from the soft news basin, and the remaining 80% on the hard news one, while the second has the specular allocation (80% hard and 20% soft). The analysis will clarify how the mix of news affects

Figure 3.2: Treatment Scheme



users engagement for different types.

Soft news distraction The second treatment is operating on the positioning of the news, instead of on the concentration. Users in this treatment group receive an equal amount of hard and soft news but with each one next to each other. This mode of delivery should increase the level of distraction of the platform.

Control Users in the control condition receive an equal amount (50%) of news within each category, delivered in blocks of five per type.

Pre-Treatment Media Diet

This experiment crucially depends on the ability to ascertain two pre-treatment conditions on subjects' media diets that might impact the effectiveness of the treatment. First, the frequency of reading the news online may affect the intensity of the treatment. For avid readers of news, the increased availability of information given by the app will provide less input compared to less avid readers of news. Secondly, the impact that the app should have would be to reduce focus and at-

Table 3.1: Generation Strata

Generation	Age group
Gen Z, iGen, or Centennials	22 or younger
Millennials or Gen Y	between 23 to 41
Generation X	42 to 53
Baby Boomers	54 to 72
Traditionalists or Silent Generation	Older than 72

tention while reading the news, and this is crucially conditional on the pre-existing habits of the users as well. As mentioned above, studies have shown that devices impact our skills as our brain adapts to them. Therefore is important not only to assess how much they were reading before but also on what form, which device they were using.

In order to overcome this difficulty, the baseline survey includes measures of pre-existing media preferences, including detailed questions about each respondent's media diet and direct monitoring of consumption.

Sampling

The eligible population for this study is the Italian online population. This population is significantly younger and more educated than the overall Italian population (weights will be provided to recover the population analogs of the treatment effects).

In each of the treatment branches, treatment assignment will be block-randomized based on generation, gender and educational level for efficiency. It is reasonable to expect that different age and educational groups would have a different engagement with their phone and thus with the app (younger and more educated individuals use more the internet), and therefore stratifying the randomization will likely reduce sample variability. Further, the randomization procedure block randomizes among men and women to allow for a separate analysis (Gerber and Green (2012*b*)).

Manipulation Checks

In order to test the two manipulations, the app includes two checks in the form of a brief in-app questionnaire reporting a recently seen headline and asking the user if he or she remembers seeing it.

3.4.3 Fieldwork

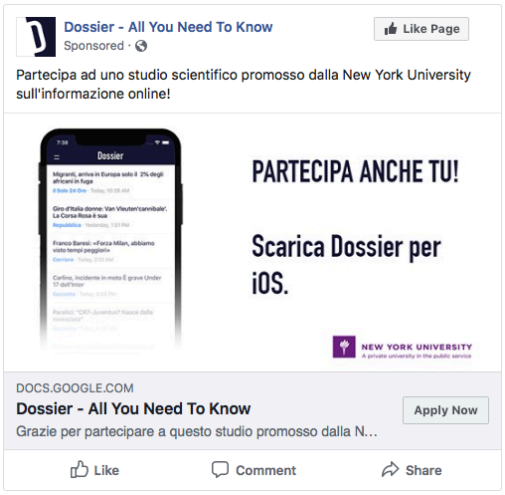
App release and participant recruitment

Users will be recruited on the pretext of beta-testing of a new app for news and a research study promoted by New York University. This feature is also introduced to reduce their concern about being monitored and justify the changes they will see in the content delivered through the app.
Participants were recruited through Facebook Inc. Figure 3.3 depicts the two ads that circulated on Facebook and Instagram portals. People downloading the app will be prompted with a screen in which they will be able to create their account and go through the consent process.
To incentivize participants to continue to participate in the study, they received weekly emails that summarize 'how they are doing,' collecting average time spent on the app reading the news, news article read and from which sources they mostly read. Also, and maybe more importantly, the app will give users free access to otherwise subscription-only content.

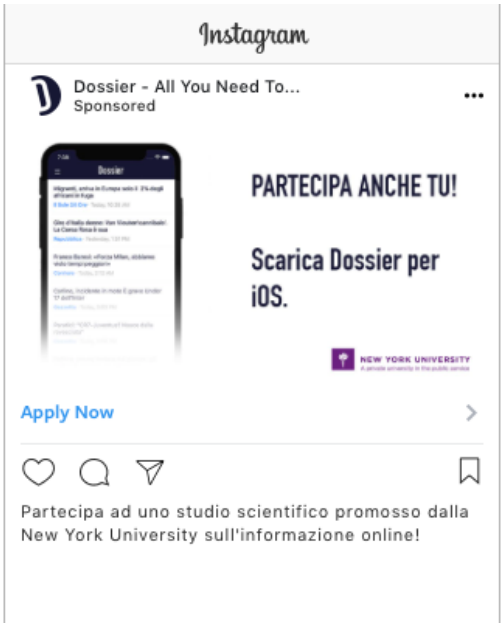
Data Collection

The entire data collection took place from mid-September to mid-November 2018. The treatment period was of 30 days, to ensure proper treatment efficacy.
The data collection was entirely digitalized and on the app directly, including the baseline and end-line survey. During the treatment period, users received an in-app survey asking for a recall of particular news in that week to test manipulation checks.
Besides the information collected during registration (see recruitment section), some of the vari-

Figure 3.3: Advertisment Campaign



(a) 1a: Facebook Ad



(b) 1b: Instagram Ad

ables of interest will be collected by the app.

- Number of minutes spent on the app per week;
- Time devoted after notification;
- Number of minutes spent reading news;
- What kind of news they read: hard or soft and how behavior changes;

Also, some variables will be measured by the baseline/endline survey as described above.

Baseline/Endline survey

In addition to demographics (age, gender, education), the survey asks the following questions:

- Party Preference

- Candidate preference at the election
- Most important issues (from a predetermined menu)
- Interest in politics
- Basic knowledge of political figures

Demographics will follow the standard practice. The other responses will be:

- news recall: prompt on the app whether the news was seen, or read.
- Issue position: pro or against certain measures
- Issue knowledge: standard things
- political behavior: vote choice, party, and L-R positioning.

3.5 Analysis

Thanks to randomization, the analysis of the data is straightforward, and the independence assumption is trivially true.

The analysis compares each treatment, T1 to T3, with respect to the control condition of having 50% soft news and soft news separated from hard news. The specification will be as following:

$$Y_{ij} = \tau D_{ij} + \mathbf{H_i X_i} + \epsilon_{\mathbf{ij}} \tag{3.1}$$

Where D_i is the treatment assignment variable for individual i in block j and $\mathbf{X_i}$ is a vector of pre-treatment characteristics to increase efficiency (Lin, 2013). Standard errors are clustered at the block j level. The outcome variables are as signaled above, including attrition, readership concerning hours, the probability of reading hard news and outcome variables from the survey. These estimates will recover the CATE on the self-selected and the randomized into treatment.

3.5.1 Main Outcomes

The main outcome of this study is the measure of news consumptions, especially how the hard news diet changes due to different media environments. In order to isolate the effect due to how the environment affects preferences and the shift in probability of encountering hard (soft) news, the study reports the difference from the observed readership and the effective share of news seen by users, any difference being interpretable as the reaction due to preference to the shift in the environment. This distinction allows for inferences on how individuals preferences can offsets changes in the environment, perhaps by compensating undesirable outcomes.

Measures To test the impact of the change in online behavior induced by the change in newsfeed, I create two different measures, one that represents *readership* and one that represents *preferences*. In the first one, I evaluate how the change in newsfeed affected how many hard news news users read, per session, by calculating the share of hard news read with respect to the total number of news, hard and soft, visualized on the screen. Analytically, this equates to calculating the following:

$$\text{Readership} = \frac{\#\text{Hard News Read}}{\#\text{Hard News} + \#\text{Soft News Seen}} \tag{3.2}$$

This measure would pick up any 'mechanical' treatment effect induced by the higher probability of visualizing a hard news due to the bias in the algorithm.

The second measure, instead, would look into changes in preferences by looking how the bias in the algorithm shifted the number of hard news read with respect to the total number of hard news seen. If exposure to hard news induces a change in taste in hard news, we should expect the proportion of news read with respect to supply to increase. Analytically, this implies a definition of preferences as:

$$\text{Preferences} = \frac{\#\text{Hard News Read}}{\#\text{Hard News Seen}} \tag{3.3}$$

3.6 Results

3.6.1 Sample Balance

Table 3.2 reports the descriptive statistics of the effective sample and the balance tests. The control and treatments do not show any statistical difference in the observable variables. The sample is much younger, on average than the Italian population whose average age is 44.3 years (2015, Istat). This feature is due to the sampling strategy that could reach only the online population that represent 59% of the total and whose average age is 36 years old. Users have an average education of high-school which is in line with the Italian population, and they have a significant interest in the news given the high frequency they report in news consumption. Finally, users have a relatively high level of political knowledge. The score from 0 to 1 reports the percentage of correct answers to the five issues to survey in the app. The decline in the score is mostly due to the last question that asks about the Italian per capita income (Islam, Draghi, Macron, Cinque Stelle, Per-Capita Income).

3.6.2 Results on Use

This section reports the statistics of the time use an average article consumption divided by treatment and generation. Table 3.3 reports the summary statistics by treatment category. The results show interesting distinctions for how the media environment affect readership. A higher soft newsfeed leads to less time spent on the app (on average, 2.76 minutes vs. 2.82 for hard news) but more app launches making overall the feed more attractive. Further, a feed of hard news leads to more article reads thus suggesting that online behavior changes according to the feed, a higher soft news environment more conducing to browsing than reading.

Table 3.4 reports the same statistics by generation. There is a significant drop in the probability of readership of news for new generations. The probability increases steadily with age. Older genera-

Table 3.2: Sample Descriptive Statistics

	Control			Treatment			Balance test
Statistic	N	Mean	St. Dev.	N	Mean	St. Dev.	(p-values)
Age	18	27.167	9.364	43	31.186	13.364	0.188
Gender	19	0.474	0.513	61	0.541	0.502	0.620
Education	19	3.684	0.885	60	3.750	1.002	0.786
Frequency TV	19	2.579	2.912	57	3.702	3.035	0.160
Frequency Online	19	1.368	0.761	59	1.492	0.679	0.534
Frequency Print	19	3.000	3.109	59	2.593	2.853	0.617
Use	14	13.000	18.890	69	10.753	19.125	0.690
Political Knowledge	15	0.707	0.225	53	0.691	0.190	0.803

Balance tests reported are two-sides t-tests on the alternative of the difference in means not equal to zero. *Education* in a 5 points scale, 1 = Elementary, 2 = Middle School, 3=High School, 4= College, 5 = More than college. *Frequencies* reported as in the questionnaire. *Political knowledge score* reported as the percentage of correct answers of the 5 questions questionnaire.

tions, especially the so-called "baby boomers" have a higher consumption of news both concerning time use (consistently with more readership) and app launches. Other statistics do not differ remarkably.

3.6.3 Political Knowledge and Issues

Table 3.5 reports descriptive statistics of political knowledge and most important issues divided by generation. Older generations, such as Gen X and baby boomers have higher political knowledge than younger generations (at least as measured by our score). While generally, users replied cor-

Table 3.3: App Use and News Consumption by Treatment

Treatment	Seconds	App launch	Articles	Min	Max
Hard	168.973	2.13	1.136	0.00	5.00
Soft	165.49	3.52	0.919	0.00	5.20
Even	201.481	4.06	0.916	0.00	4.00
Control	184.89	3.22	0.692	0.00	4.16

Articles equals the number of articles per session. *Min* and *Max* refer to articles.

Table 3.4: App Use and News Consumption by Generation

Generation	Articles	Seconds	App launch	Min	Max
Gen Z	0.722	203.05	2.5	0.00	936.76
Millenials	0.00	0.00	0.00	0.00	0.00
Gen X	1.019	218.30	2.22	0.00	903.97
Baby Boomers	2.583	271.9	2.66	0.00	911.7
Traditionalists	0.00	0.00	0.00	0.00	0.00

Articles equals the number of articles per session. *Min* and *Max* refer to articles. Generation ages as reported above.

rectly on political figures and symbols, the lowest amount of correct answers can be concerning economic issues such as Italian per capita income.

Most remarkably most Italians report that the most critical issue facing Italy right now is immigration. This finding could be an artifact of our study since immigration was on top of the list of issues we provided or the result of the media environment that reports about immigration extensively on

Table 3.5: Political Knowledge and Salient Issues

Generation	Score	Issue 1	Issue 2	Issue 3
Gen Z	0.537	Immigration	European Union	Unemployment
Millenials	0.00	0.00	0.00	0.00
Gen X	0.733	Immigration	Taxes	Corruption
Baby Boomers	0.733	Immigration	Unemployment	Inequality
Traditionalists	0.00	0.00	0.00	0.00

Aswers reported as the within-generation highest, second and third highest reported concerns from the questionnaire.

a daily basis.

3.6.4 Results on Readership and Preferences

The first set of results disaggregates demand for news according to what users reported at the beginning of the study. Overall, 87.04% of the respondents preferred a media diet based on more soft news, while the remaining 12.96% chose a more hard news diet. Unfortunately, due to missing data on the age of the participants, it is impossible to assess how preferences differ across generations. From the evidence collected, while younger generations are more prone to consumption of soft news, the study cannot conclude this definitely as they also represent a larger fraction of the sample. Further, the sample collected contains no participant that preferred hard news.
The main result of this paper will relate to the evaluation of how different treatment conditions that mimic two different kinds of media environments affect readership and preferences.

The coefficients on Figure 3.4 show the average treatment effect for the two groups of selectors and non-selectors. The results show that the coefficients are similar between selectors and

Figure 3.4: Effects on Readership of Hard News

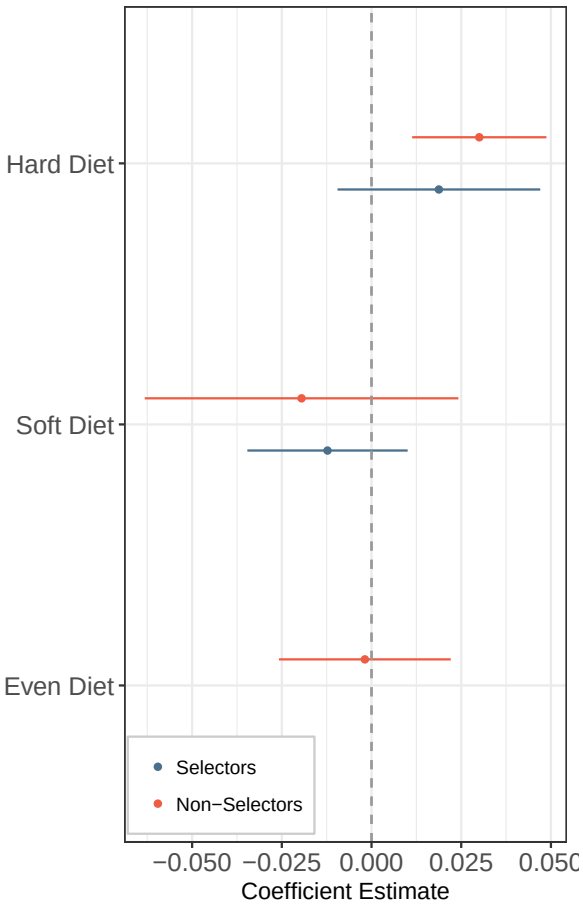
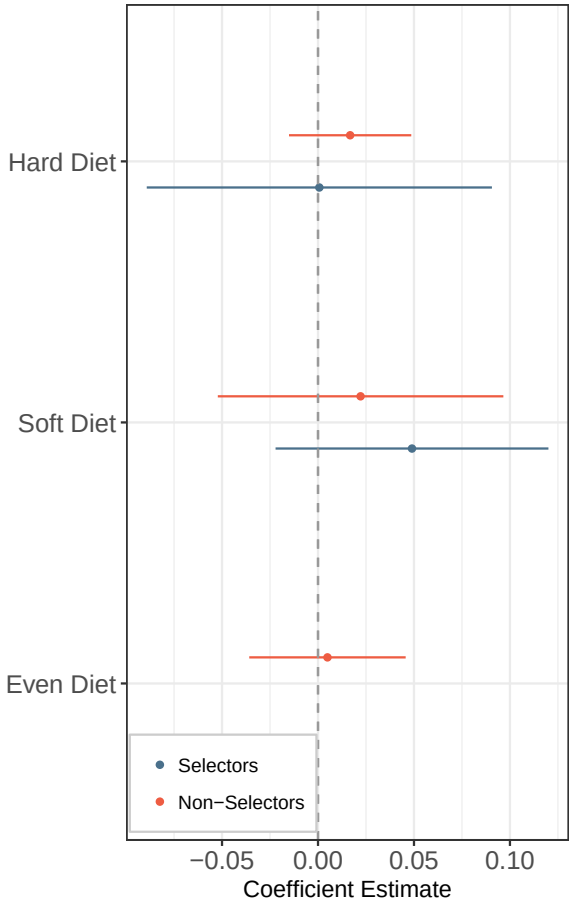


Figure 3.5: Effects on Preferences for Hard News



non-selectors in both the cases of a hard and a soft news diet. No effect is found concerning the even diet. Interestingly, the result show that a media environmnet with more hard news provision leads to an increase in hard news readeship, and therefore, consistent with results that promote the provision of public service media, it is possible to increase redearship of political news, a category included in hard news, by making them more available to voters. A more general implication of these findings is that readership is highly dependent on context and supply. More specifically, individual preferences for news do not offset radical switch in the environment. Notice, in fact, that the biggest increase is found in the sub-sample of non-selectors, i.e., the pool of individual forced to a hard news diet.

Additionally, this study documents that it is more concentration than positioning that affects hard news readership. Specifically, while I find that there is a slight decay in readership of hard news if soft news are positioned next to it, I cannot reject the null hypothesis of no effect.

The second set of results on Figure 3.5, evaluates how shifts in the media environment are able to train individual preferences. Perhaps due to the limited duration of this pilot experiment, the results show no effect on preferences of the shift in media environment.

These results limit the scope of this research, but do invalidate the wider importance of the analysis. As the fruition of news comes more and more often as the result of an algorithmic allocation, the mechanical effects implied in Figure 3.4 are a cause for concern (and hope) for how the internet, and social media in particular, guide exposure to information.

To give more sense of the coefficients, the table below reports the changes in individual averages of the readership of hard and soft news conditional on treatment condition and subject preferences. Table 3.6 reports the average individual share of consumption of hard news and how this relates to the media environment. The column referred as Δ reports the different behavior for the opportunity given by different environment due to different treatment conditions. Notice that, even though as reported above a media news supply of soft news reduces consumption of hard news (from 0.718 to 0.371), demand for hard news even among soft news lovers is high as their readership is higher than the supply (0.371 instead of 0.2).

Table 3.6: Changes in Averages

Treatment	Choice	Av. Readership	Δ
Even	Hard	0.857	0.357
Hard	Hard	0.928	0.128
Control	Hard	0.813	0.313
Soft	Hard	0.250	0.005
Even	Soft	0.552	0.052
Hard	Soft	0.923	0.123
Control	Soft	0.718	0.218
Soft	Soft	0.371	0.171

Average readership refers to the percentage of hard news (1) read by the users in each category. The Δ refers to the difference with respect to the control group.

3.7 Conclusions

The media environment has shifted radically with the introduction of the internet. While most studies have focused on how social media have led to a personalization of news exposure (the DailyMe hypothesized by Cass Sunstein), few have underlined how the increase in entertainment and competition has altered the general provision of news. This study looks at the consequences of shifts in media environment comparable to the one we experience and how they could affect mobile news readership and political knowledge. The study reports evidence that self-selection could compensate only minimally to large shifts in content as most people tend to adhere to the feed they

receive, even when they can personalize it.[4] In particular, receiving preferred content does not alter the effective diet significantly, while shifts in the overall feed have strong impacts. Soft news and hard news lovers alike read hard news in equal quantities when provided a mostly hard news feed. Similarly, albeit in the opposite direction, results show that soft news and hard news lovers read either with similar probability when provided mostly soft news.

These results attenuate concerns toward excessive personalization of news feeds while raising new concerns for the overall state of the media environment. These results suggest that, as new outlets enter the market and our attention diverts toward soft news-rich social media, hard news consumption will decline thus compromising democratic competence and deliberation.

[4]To clarify, this statement refers to the fact that even when users can suggest what they prefer, an algorithm will ultimately decide what will be on the newsfeed.

Appendix A

Questionnaires

Questionnaire (In English)

Q1	In Italy there are many political parties that would like to get your vote. If there were elections tomorrow, which of these party would be more likely to get your vote?
	• +Europa • 10 Volte Meglio • Autodeterminazione • Blocco Nazionale Per Le Liberta' • Casapound Italia • Civica Popolare Lorenzin • Forza Italia • Fratelli D'italia Con Giorgia Meloni • Grande Nord • Il Popolo Della Famiglia • Italia Agli Italiani • Italia Europa Insieme • Italia Nel Cuore • Lega • Lista Del Popolo Per La Costituzione • Movimento 5 Stelle • Noi Con L'italia - Udc • Partito Comunista • Partito Democratico • Partito Repubblicano Italiano - Ala • Partito Valore Umano • Patto Per L'autonomia • Per Una Sinistra Rivoluzionaria • Potere Al Popolo! • Rinascimento Mir • Siamo • Svp - Patt • Liberi E Uguali
Q2	Do you usually watch TV news? If so, how frequently?
	• No, never • Less than once a week • Once a week • Twice a week • Three times a week • Four times a week • Five times a week • Six times a week • Every day • I don't know • No answer
Q3	Which TV news do you normally watch?
	• Tg1 (RAI1) • Tg2 (RAI2) • Tg3 (RAI3) • Tg4 (Rete4) • Tg5 (Canale5) • Studio Aperto (Italia1) • La 7 News (La7) • SKY TG 24 - Satellite • Local TV news • Other TV news • No Answer

Questionario (In English)

Q4	Do you look up political information online?
	• Si, more than once a day • Si, rarely (less than once a week) • No, never • I don't know • No answer
Q5	Do you usually read the print newspaper? If so, with which frequency?
	• No, never • Less than once a week • Once a week • Twice a week • Three times a week • Four times a week • Five times a week • Six times a week • Every day • I don't know • No answer
Q6	Which newspaper do you read? (Max two answers)
	• Avanti! • Avvenire • Corriere della Sera • Europa • Il Foglio • Il Gazzettino • Il Giornale • Il Giornale di Sicilia • Il Giorno • Il Lavoro • Il Manifesto • Il Mattino • Il Messaggero • Il Piccolo • Il Popolo • Il Resto del Carlino • Il Riformista • Il Secolo XIX • Il Secolo d'Italia • Il Secolo d'Italia • Il Sole-24 Ore • Il Tempo • Italia Oggi • La Gazzetta del Mezzogiorno • La Nazione • La Padania • La Repubblica • La Stampa • Liberazione • Libero • L'Unione Sarda • L'Unita' • Free Press (Metro, City, Leggo) • Non Answer
Q7	Second Newspaper
	same list

Questionario (In English)

Q8	In your opinion, which are the three most important issues for Italians right now?
	• Immigrazione • L'Unione Europea • Tasse • Disoccupazione • Inquinamento • Corruzione politica • Arretratezza del mezzogiorno • Le Pensioni • La riforma della sanita' • La riforma della giustizia • Il reddito di cittadinanza • Inflazione • Evasione fiscale • Scuola
Q9	How interested you are about Italian national and local politics?
	• A lot • Some • Little • No interest
Q10	How much influence do you think politics has on your life and the one of your family?
	• A lot • Some • Little • No influence
Q11	Political Knowledge
	• Islam symbol • Mario Draghi from the picture • Italian GDP per capita is 27 700 Euro • Italian poverty rate is at 6,5% • Mattarella from the picture • Macron from the picture • Movimento 5 stelle Flag

Questionario (in Italian)

Q1	In Italia esistono molti partiti ognuno dei quali vorrebbe avere il suo voto. Se ci fossero delle elezioni domani, quale partito avrebbe la piu' alta probabilita' di avere il suo voto? (Se viene richiesto quando?, rispondere che si tratta di una possibilità in generale)
	• +Europa • 10 Volte Meglio • Autodeterminazione • Blocco Nazionale Per Le Liberta' • Casapound Italia • Civica Popolare Lorenzin • Forza Italia • Fratelli D'italia Con Giorgia Meloni • Grande Nord • Il Popolo Della Famiglia • Italia Agli Italiani • Italia Europa Insieme • Italia Nel Cuore • Lega • Lista Del Popolo Per La Costituzione • Movimento 5 Stelle • Noi Con L'italia - Udc • Partito Comunista • Partito Democratico • Partito Repubblicano Italiano - Ala • Partito Valore Umano • Patto Per L'autonomia • Per Una Sinistra Rivoluzionaria • Potere Al Popolo! • Rinascimento Mir • Siamo • Svp - Patt • Liberi E Uguali
Q2	In genere lei guarda il telegiornale? Se si, con che frequenza?
	• No, mai • Meno di 1 volta alla settimana • 1 giorno alla settimana • 2 giorni alla settimana • 3 giorni alla settimana • 4 giorni alla settimana • 5 giorni alla settimana • 6 giorni alla settimana • Tutti i giorni • Non sa • Non risponde
Q3	Qual è il telegiornale che lei abitualmente vede di più?
	• Tg1 (RAI1) • Tg2 (RAI2) • Tg3 (RAI3) • Tg4 (Rete4) • Tg5 (Canale5) • Studio Aperto (Italia1) • La 7 News (La7) • SKY TG 24 - Satellite • Telegiornale locale • Altro telegiornale • Non risponde

Questionario (In Italian)

Q4	Lei cerca informazioni e notizie politiche navigando in internet?
	• Si, frequentemente (almeno una volta al giorno) • Si, raramente (qualche volta a settimana) • No, mai • Non so • Non risponde
Q5	In generale lei legge un giornale quotidiano (esclusi i giornali sportivi)? Se si, con quale frequenza? (INTERVISTATORE: Include anche la lettura su Internet)
	• No, mai • Meno di 1 volta alla settimana • 1 giorno alla settimana • 2 giorni alla settimana • 3 giorni alla settimana • 4 giorni alla settimana • 5 giorni alla settimana • 6 giorni alla settimana • Tutti i giorni • Non sa • Non risponde
Q6	Quale giornale legge? (Max due risposte) PRIMO GIORNALE
	• Avanti! • Avvenire • Corriere della Sera • Europa • Il Foglio • Il Gazzettino • Il Giornale • Il Giornale di Sicilia • Il Giorno • Il Lavoro • Il Manifesto • Il Mattino • Il Messaggero • Il Piccolo • Il Popolo • Il Resto del Carlino • Il Riformista • Il Secolo XIX • Il Secolo d'Italia • Il Secolo d'Italia • Il Sole-24 Ore • Il Tempo • Italia Oggi • La Gazzetta del Mezzogiorno • La Nazione • La Padania • La Repubblica • La Stampa • Liberazione • Libero • L'Unione Sarda • L'Unita' • Free Press (Metro, City, Leggo) • Non risponde
Q7	SECONDO GIORNALE
	stesso elenco

Questionario (in Italian)

Q8	Secondo lei quali sono le tre questioni piu' importanti per gli italiani in questo momento?
	• Immigrazione • L'Unione Europea • Tasse • Disoccupazione • Inquinamento • Corruzione politica • Arretratezza del mezzogiorno • Le Pensioni • La riforma della sanita' • La riforma della giustizia • Il reddito di cittadinanza • Inflazione • Evasione fiscale • Scuola
Q9	Quanto e' interessato alla political italiana e locale?
	• Molto • Abbastanza • Poco • Per niente
Q10	Quanto pensa influisca la politica sulla propria vita e quella della propria famiglia?
	• Molto • Abbastanza • Poco • Per niente
Q11	Conoscenza della politica
	• Bandiera dell'islam • Mario Draghi • Il PIL procapite italiano e' 27 700 Euro • Il tasso di poverta' italiano e' il 6,5 % • Mattarella dalla foto • Macron dalla foto • Simbolo Movimento 5 stelle

Bibliography

Acemoglu, Daron, Georgy Egorov and Konstantin Sonin. 2013. "A Political Theory of Populism." *The Quarterly Journal of Economics* 128(2):771–805.

Ansolabehere, Stephen, Erik C. Snowberg and James M. Snyder. 2006. "Television and the Incumbency Advantage in U.S. Elections." *Legislative Studies Quarterly* 31:469–490.

Atton, Chris. 2002. *Alternative Media*. London: Sage Publications.

Baldasty, Gerald J. 1992. The Commercialization of News in the Nineteenth Century. Madison, WI: University of Wisconsin Press.

Barabas, Jason and Jennifer Jerit. 2010. "Are Survey Experiments Externally Valid?" *American Political Science Review* 104(2):226–242.

Bastos, Marco Toledo. 2015. "Shares, Pins, and Tweets: News readership from daily papers to social media." *Journalism Studies* 16(3):305–325.

Baum, Matthew A. 2002. "Sex, Lies, and War: How Soft News Brings Foreign Policy to the Inattentive Public." *American Political Science Review* 96(01):91–109.

Baum, Matthew A. and Angela S. Jamison. 2006. "The Oprah Effect: How Soft News Helps Inattentive Citizens Vote Consistently." *The Journal of Politics* 68(4):946–959.

BIBLIOGRAPHY

Bellucci, Paolo and Andrea De Angelis. 2013. "Government Approval in Italy: Political Cycle, Economic Expectations and TV Coverage." *Electoral Studies* 32(3):452–459.

Besley, Timothy and Andrea Prat. 2006. "Handcuffs for the Grabbing Hand? Media Capture and Government Accountability." *American Economic Review* 96(3):720–736.

Blei, David M., Andrew Y. Ng and Michael I. Jordan. 2003. "Latent Dirichlet Allocation." *Journal of Machine Learning Research* 3:993–1022.

Blei, David M. and John D. Lafferty. 2007. "A Correlated Topic Model of Science." *The Annals of Applied Statistics* 1(1):17–35.

Boyle, Michael P. and Mike Schmierbach. 2009. "Media Use and Protest: The Role of Mainstream and Alternative Media Use in Predicting Traditional and Protest Participation." *Communication Quarterly* 57(1):1–17.

Cagé, Julia and Valeria Rueda. 2016. "The long-term effects of the printing press in sub-Saharan Africa." *American Economic Journal: Applied Economics* .

Chong, Dennis and James N. Druckman. 2007. "A Theory of Framing and Opinion Formation in Competitive Elite Environments." *Journal of Communication* 57(1):99–118.

Clark, Tom S. and Drew A. Linzer. 2015. "Should I Use Fixed or Random Effects?" *Political Science Research and Methods* 3(2):399–408.

Dahl, Robert A. 2000. *On Democracy*. New Haven: Yale University Press.

Danky, James Philip and Maureen E. Hady. 1998. *African-American Newspapers and Periodicals: a National Bibliography*. Harvard University Press.

DellaVigna, Stefano and Ethan Kaplan. 2007. "The Fox News Effect: Media Bias and Voting." *The Quarterly Journal of Economics* 122(3):1187–1234.

BIBLIOGRAPHY

DellaVigna, Stefano, Ruben Durante, Brian Knight and Eliana La Ferrara. 2016. "Market-Based Lobbying: Evidence from Advertising Spending in Italy." *American Economic Journal: Applied Economics* 8(1):224–256.

Delli Carpini, Michael X. and Scott Keeter. 1996. *What Americans Know about Politics and Why It Matters*. Yale University Press.

Deloitte. 2018. "Media and Entertainment Industry Outlook.".

Detweiler, Frederick German. 1922. *The Negro press in the United States*. Chicago: University of Chicago Press.

Drago, Francesco, Tommaso Nannicini and Francesco Sobbrio. 2014. "Meet the press: How voters and politicians respond to newspaper entry and exit." *American Economic Journal: Applied Economics* .

Druckman, James N. and Toby Bolsen. 2011. "Framing, Motivated Reasoning, and Opinions About Emergent Technologies." *Journal of Communication* 61(4):659–688.

Durante, Ruben and Brian Knight. 2012. "Partisan Control, Media Bias, and Viewer Responses: Evidence From Berlusconi's Italy." *Journal of the European Economic Association* 10(3):451–481.

Durante, Ruben, Paolo Pinotti and Andrea Tesei. Forthcoming. "The Political Legacy of Entertainment TV." *American Economic Review* .

Eady, Gregory, Jan Zilinsky, Jonathan Nagler and Joshua Tucker. 2018. "Measuring How Many People are in Media Bubbles on Twitter." *Working Paper* .

Enikolopov, Ruben and Maria Petrova. 2015. Media Capture: Empirical Evidence. In *Handbook of Media Economics*. Boston, MA: Elsevier pp. 687–700.

Entman, Robert M. 2007. "Framing Bias: Media in the Distribution of Power." *Journal of Communication* 57(1):163–173.

BIBLIOGRAPHY

Esser, F. and J. Strömbäck. 2014. *Mediatization of Politics: Understanding the Transformation of Western Democracies*. New York: Palgrave Macmillan.

Evan, Ivan. 2007. Cultures of violence : Lynching and Racial Killing in South Africa and the American South. Manchester: Manchester University Press.

Eyheramendy, Susana, David D. Lewis and David Madigan. 2003. On the Naive Bayes Model for Text Categorization. In *Proceedings of the Ninth International Workshop on Artificial Intelligence and Statistics*.

Flaxman, Seth, Sharad Goel and Justin M. Rao. 2016. "Filter bubbles, echo chambers, and online news consumption." *Public Opinion Quarterly* 80(Specialissue1):298–320.

Gaines, Brian J. and James H. Kuklinski. 2011. "Experimental estimation of heterogeneous treatment effects related to self-selection." *American Journal of Political Science* 55(3):724–736.

Gaines, Brian J., James H. Kuklinski and Paul J. Quirk. 2007. "The Logic of the Survey Experiment Reexamined." *Political Analysis* 15(1):1–20.

Gentzkow, Matthew and Jesse M. Shapiro. 2006. "Media Bias and Reputation." *Journal of Political Economy* 114(2):280–316.

Gentzkow, Matthew and Jesse M. Shapiro. 2010. "What Drives Media Slant? Evidence From U.S. Daily Newspapers." *Econometrica* 78(1):35–71.

Gentzkow, Matthew, Jesse M. Shapiro and Michael Sinkinson. 2011. "The effect of newspaper entry and exit on electoral politics." *American Economic Review* .

Gerber, Alan S., Dean Karlan and Daniel Bergan. 2009. "Does the Media Matter? A Field Experiment Measuring the Effect of Newspapers on Voting Behavior and Political Opinions." *American Economic Journal: Applied Economics* 1(2):35–52.

BIBLIOGRAPHY

Gerber, Alan S. and Donald P. Green. 2012*a*. *Field Experiments: Design, Analysis, and Interpretation*. 1st ed ed. New York: W. W. Norton.

Gerber, A.S. and D.P. Green. 2012*b*. *Field Experiments: Design, Analysis, and Interpretation*. W. W. Norton.

Gilens, Martin and Craig Hertzman. 2000. "Corporate Ownership and News Bias: Newspaper Coverage of the 1996 Telecommunications Act." *The Journal of Politics* 62(2):369–386.

Giles, Bob. 2003. "Journalism and Black America: Then and Now." *Nieman Reports* 57(3).

Grimmer, Justin and Brandon M. Stewart. 2013. "Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts." *Political Analysis* 21(3):267–297.

Groseclose, T. and J. Milyo. 2005. "A Measure of Media Bias." *The Quarterly Journal of Economics* 120(4):1191–1237.

Grossman, James R. 1989. *Land of Hope: Chicago, Black Southerners, and the Great Migration*. Chicago, Illinois: Chicago University Press.

Guess, Andrew M. 2016. "Media Choice and Moderation: Evidence from Online Tracking Data." *Working Paper* .

Guess, Andy, Nyhan Brendan and Jason Reifler. 2018. "Selective Exposure to Misinformation: Evidence from the consumption of fake news during the 2016 U.S. presidential campaign.".

Hallin, D. C. and P. Mancini. 2004. *Comparing Media Systems: Three Models of Media and Politics*. Cambridge: Cambridge University Press.

Hanretty, Chris. 2014. "Media outlets and their moguls: Why concentrated individual or family ownership is bad for editorial independence." *European Journal of Communication* 29(3):335–350.

BIBLIOGRAPHY

Herriot, John G. and Christian H. Reinsch. 2002. "Algorithm 472: procedures for natural spline interpolation [E1]." *Communications of the ACM* 16(12):763–768.

Horkheimer, Max and Theordor W. Adorno. 1947. *Dialectic of Enlightment*. Continuum.

House, Freedom. 2017. "Press Freedom's Dark Horizon - Freedom House press Report.".

Hölmstrom, Bengt. 1979. "Moral Hazard and Observability." *The Bell Journal of Economics* 10(1):74–91.

Iyengar, Shanto. and Donald R. Kinder. 1987. *News That Matters: Television and American Opinion*. Chicago: University of Chicago Press.

Kellogg, Charles F. 1967. *NAACP*. Baltimore: John Hopkins University Press.

King, Gary, Benjamin Schneer and Ariel White. 2017. "How the news media activate public expression and influence national agendas." *Science* 358(6364):776–780.

Knox, Dean, Teppei Yamamoto, Matthew Baum and Adam Berinsky. 2018. "Design, Identification, and Sensitivity Analysis for Patient Preference Trials.".

Larreguy, Horacio, John Marshall and Jim Jr. Snyder. 2018. "Publicizing Malfeasance: When The Media Structure Facilitates Electoral Accountability In Mexico." *Mimeo* .

Laver, Michael, Kenneth Benoit and John Garry. 2003. "Extracting Policy Positions from Political Texts Using Words as Data." *American Political Science Review* 97(2):311–331.

Lehman-Wilzig, Sam N. and Michal Seletzky. 2010. "Hard news, soft news, 'general' news: The necessity and utility of an intermediate classification , Hard news, soft news, 'general' news: The necessity and utility of an intermediate classification." *Journalism* 11(1):37–56.

Lenz, Gabriel S. 2009. "Learning and Opinion Change, Not Priming: Reconsidering the Priming Hypothesis." *American Journal of Political Science* 53(4):821–837.

BIBLIOGRAPHY

Lin, Winston. 2013. "Agnostic notes on regression adjustments to experimental data: Reexamining Freedman's critique." *Annals of Applied Statistics* 7(1):295–318.

Lott, John R. and Kevin A. Hassett. 2014. "Is Newspaper Coverage of Economic Events Politically Biased?" *Public Choice* 160(1-2):65–108.

Luskin, Robert C. 1990. "Explaining Political Sophistication." *Political Behavior* 12(4):331–361.

McCombs, Maxwell E. and Donald L. Shaw. 1972. "The Agenda-Setting Function of Mass Media." *The Public Opinion Quarterly* 36(2):176–187.

McMillan, John and Pablo Zoido. 2004. "How to Subvert Democracy: Montesinos in Peru." *Journal of Economic Perspectives* 18(4):69–92.

Morris, Aldon D. 1984. *The Origins of the Civil Rights Movement*. Simon and Schuster.

Newman, Nic. 2017. Reuters Institute Digital News Report 2017. Technical report Reuters Institute for the Study of Journalism.

Ophir, E., C. Nass and A. D. Wagner. 2009. "Cognitive control in media multitaskers." *Proceedings of the National Academy of Sciences* .

Prat, Andrea. 2006. The More Closely We Are Watched, the Better We Behave? In *Transparency: The Key to Better Governance?* Oxford: Oxford University Press pp. 91–103.

Prat, Andrea. 2015. Media Capture and Media Power. In *Handbook of Media Economics*. Boston, MA: Elsevier pp. 669–686.

Prior, Markus. 2003. "Any Good News in Soft News? The Impact of Soft News Preference on Political Knowledge." *Political Communication* 20(2):149–171.

Prior, Markus. 2005. "News vs. Entertainment : Choice and Widens Turnout Gaps in Increasing Media Political Knowledge." *American Journal of Political Science* 49(3):577–592.

BIBLIOGRAPHY

Prior, Markus. 2007. *Post-Broadcast Democracy: How Media Choice Increases Inequality in Political Involvement and Polarizes Elections*. Cambridge Studies in Public Opinion and Political Psychology Cambridge University Press.

Prior, Markus. 2009. "Improving Media Effects Research through Better Measurement of News Exposure." *The Journal of Politics* 71(3):893–908.

Przeworski, Adam, Susan C. Stokes and Bernard Manin, eds. 1999. *Democracy, Accountability, and Representation*. Cambridge Studies in the Theory of Democracy Cambridge University Press.

Puglisi, Riccardo and James M. Snyder. 2011. "Newspaper Coverage of Political Scandals." *The Journal of Politics* 73(3):931–950.

Puglisi, Riccardo and Jim Snyder. 2015. Empirical Studies of Media Bias. In *Handbook of Media Economics*, ed. S.P. Anderson, J. Waldfogel and D. Stromberg. Elsevier Science chapter 15, pp. 647–666.

Rapporto Sulla Conoscenza 2018. 2018. Technical report Istat.

Raymond, Jane E, Kimron L Shapiro and Karen M Arnell. 1992. "Temporary Suppression of Visual Processing in an RSVP Task: An Attentional Blink?" *Journal of Experimental Psychology: Human Perception and Performance* 18(3):849–860.

Roberts, Margaret E., Brandon M. Stewart, Dustin Tingley, Christopher Lucas, Jetson Leder-Luis, Shana Kushner Gadarian, Bethany Albertson and David G. Rand. 2014. "Structural Topic Models for Open-Ended Survey Responses." *American Journal of Political Science* 58(4):1064–1082.

Rose, Ellen. 2010. "Continuous Partial Attention: Reconsidering the Role of Online Learning in the Age of Interruption." *Educational Technology* 50(4):41–46.

BIBLIOGRAPHY

Russell Neuman, W., Lauren Guggenheim, S. Mo Jang and Soo Young Bae. 2014. "The Dynamics of Public Attention: Agenda-Setting Theory Meets Big Data." *Journal of Communication* 64(2):193–214.

Schudson, Michael. 2002. "The News Media as Political Institutions." *Annual Review of Political Science* 5(1):249–269.

Scott, David K. and Robert H. Gobetz. 1992. "Hard News/Soft News Content of the National Broadcast Networks, 1972–1987 , Hard News/Soft News Content of the National Broadcast Networks, 1972–1987." *Journalism Quarterly* 69(2):406–412.

Shawn Green, C and Daphne Bavelier. 2003. "Action video game modifies visual selective attention." *Nature* .

Shearer, Elisa and Jeffrey Gottfried. 2017. News Use Across Social Media Platforms. Technical report Pew Research Center.

Shoemaker, Pamela J. and Timothy Vos. 2009. *Gatekeeping Theory*. New York: Rutledge.

Slapin, Jonathan B. and Sven-Oliver Proksch. 2008. "A Scaling Model for Estimating Time-Series Party Positions from Texts." *American Journal of Political Science* 52(3):705–722.

Slothuus, Rune. 2010. "When Can Political Parties Lead Public Opinion? Evidence from a Natural Experiment." *Political Communication* 27(2):158–177.

Soroka, Stuart N. 2012. "The gatekeeping function: Distributions of information in media and the real world." *Journal of Politics* 74(2):514–528.

Stanig, Piero. 2015. "Regulation of Speech and Media Coverage of Corruption: An Empirical Analysis of the Mexican Press." *American Journal of Political Science* 59(1):175–193.

White, David Manning. 1950. "The "Gate Keeper": A Case Study In the Selection of News." *Journalism Quarterly* 27(4):119–126.

BIBLIOGRAPHY

Wolfe, Michelle, Bryan D. Jones and Frank R. Baumgartner. 2013. "A Failure to Communicate: Agenda Setting in Media and Policy Studies." *Political Communication* 30(2):175–192.