IN THE SAME PLACE BUT IN THE SAME WAY?

THE EFFECTS OF THE CHANGE IN THE EDITORIAL STRUCTURE OF INDEX.HU ON COVID NEWS COVERAGE

by

Attila Balla

Submitted to

Department of Political Science,

Central European University

In partial fulfillment of the requirements for the degree of Master of Arts

Supervisor: Gábor Simonovits

Vienna, Austria

2023

In the Same Place but in the Same Way?

The Effects of the Change in the Editorial Structure of Index.hu on COVID News Coverage

Attila Balla

Abstract: In July 2020, in one of Hungary's most prominent online news sites (index.hu), the entire editorial staff resigned because of political pressure from the autocratic government. While the site has not started to publish hard propaganda with the new editorial team, strong doubts remain as to whether it remained independent. I hypothesize that after the change they are using sophisticated forms of information manipulation. Reporting less bad news (censorship), showing a more positive sentiment of the new (farming bias), or framing the government as a competent actor by attributing positive events to them while blaming external actors for the negative event (selective attribution) are possible ways to implement less visible news manipulation. With a corpus of COVID-19-related news - a salient topic both before and after the editorial change - from index.hu and four other news sites I analyzed news manipulation. The frequency of reports on indicators, news headlines, and news fragments with relevant governmental or external actors are analyzed, using historical data and human coders. The findings support only the claim for framing bias, and while a larger sample or different techniques can reveal media bias of the site, it also implies that the regime intends to maintain the credibility of the site in the public eye.

Keywords: media bias, media capture, propaganda, censorship, selective attribution, natural language processing, difference-in-differences, Hungary

Copyright notice

I hereby declare that the present MA thesis is exclusively my own work based on my research and analysis. All external data sources and information is appropriately referenced with in-text citations and in the bibliography. An earlier version of this work limited to parts of the first four chapters, including the literature review and the research design, was submitted as part of the assessment for the "Introduction to Causal Inference" course at CEU's Doctoral School of Political Science, Public Policy, and International Relations in the Winter term of the academic year 2021-2022.

Attila Balla

Vienna, Austria

14 September, 2023

Contents

List of Tables **6**
List of Figures **6**
List of Abbreviations **7**
1 Introduction **8**
1.1 Background 8
1.2 Outline 9
2 Literature review **11**
2.1 Media bias 11
2.1.1 The Definition of Media Bias 11
2.1.2 The Types of Media Bias 12
2.1.3 Challenges of Testing for Media Bias 13
2.1.4 Measurement of Media Bias 15
2.1.5 Political effects of biased media consumption 17
2.2 Media manipulation in hybrid regimes 19
2.3 The Case of Hungary 21
2.3.1 The Political System 21
2.3.2 The State of the Media 22
2.3.3 The governmental response to the Covid-19 pandemic 24
3 Assumptions **25**
4 Research Design **26**
4.1 Overview of the Research Design 26
4.2 Data and Measurement 27
4.3 Measuring Selection Bias 28
4.4 Measuring Framing Bias 29
4.5 Measuring Selective Attribution 30
4.6 Pilot Study 31
4.7 Crowdsourcing 33
4.8 Difference-in-differences Method 34
5 Analysis **38**
5.1 Analysis of Censorship 38
5.2 Analysis of Framing Bias 40

5.3 Analysis of Selective Attribution 42
5.4 Analysis of Selective Attribution by Association 45
5.5 Spillover Effect and Imperfect Compliance 49
6 Discussion 51
6.1 Sample Size 52
7 Conclusion 53
7.1 Summary of the Findings 53
7.2 External Validity ..... **Error! Bookmark not defined.**
7.3 Directions for Future Research 54
References 56
Appendices 61
Appendix A - Replication Code 61
Appendix B – Coding Questionnaire 62
Appendix C – Difference-in-Difference analyses 65

List of Tables

Table 1: The number of COVID articles for every media site 28

Table 2 Basic statistics about the number of sentences in an article on the five sites 31

Table 3: The ratings of the one- and two-sentence-long fragments in the pilot study 32

Table 4: Maximal difference between the ratings in the pilot study 33

Table 3: Summary statistics of the human coders 33

Table 4 Comparison of the frequency of coverage of good and bad news 39

Table 5: The descriptive statistics of the ratings of the news titles 41

Table 6: The descriptive statistics of the ratings of the governmental actors 43

Table 7: The descriptive statistics of the ratings of the opposition actors 44

Table 8: The descriptive statistics of the ratings of the fragments with governmental actors . 46

Table 9: The descriptive statistics of the ratings of the fragments with opposition actors 47

List of Figures

Figure 1: Number of COVID-related news per month in the five media outlets 28

Figure 3: The frequency of improving and worsening days of the pandemic 40

Figure 4: The mean of the title ratings 42

Figure 5: The mean of the actor ratings for governmental actors 43

Figure 6: The mean of the actor ratings for opposition actors 44

Figure 7: The mean of the fragment ratings for news with governmental actors 46

Figure 8: The means of the fragment ratings for news with opposition actors 47

List of Abbreviations

ATE	Average Treatment Effect
DiD	Difference in Differences
LLM	Large Language Models
NLP	Natural Language Processing

1 Introduction

1.1 Background

This thesis analyses the consequences of the media capture of a major Hungarian online news media site, index.hu that resulted in a complete change of its editorial staff in September 2020. The site's management along with the new chief editor stated that the site remained an independent media outlet after the resignation of all of the journalists and editors from the previous staff, which inspired the research question of this thesis.

This research focuses on the coverage of news related to the COVID-19 pandemic. The reason behind that is that the first coronavirus case was identified on the 3rd of March with lockdown and the second wave of the pandemic finished at the end of February, it can be said that the COVID-19 pandemic was a salient issue both before and after the editorial change at index.hu.

In July 2020, after the dismissal of the chief editor of index.hu, more than eighty employees, practically the entire editorial staff handed in their resignation which was considered an act of press historic significance (Tamas, 2020). It was one of the most visited online news portals in Hungary (NMHH, 2021) and was known as an independent and critical outlet. The news site continued its operation but from September 2020 with a completely new editorial staff. The owners and the new chief editor communicated that the previous board made a wrong decision on false assumptions and promised they would remain an independent news site. They even started a campaign with the slogan „In the Same Place, in the Same Way" (Média1, 2020).

The concerns about the site's independence were not groundless. In the previous decade, business actors connected to the ruling Orbán system bought several outlets, and alongside the

state media, they built a media empire of state propaganda (Polyák, 2019). For example, the editor-in-chief of Origo, once the biggest market competitor of Index, was fired for political reasons in 2014 and later the language of the site radically changed and they even started to use terms and language promoted by Russian state propaganda (Corruption Research Center Budapest, 2018).

Although the new Index did not become an open propaganda site and it remained one of the most visited online media sites in the country there are already enough qualitative pieces of evidence to say that the declaration about the site's orientation is not true. For example, they published an article about a sensitive land deal of a minister's wife under a pseudonym and made a friendly interview with Hungary's richest person who is connected to the prime minister and is known to financialize the outlet - as the latter was known to the public through leaked footage (Direkt36, 2022).

In this thesis, I would like to gain quantitative empirical evidence to answer the question of whether the new Index favors the Orbán regime's interests with a more pro-government editorial practice than the previous one – even if they seem to be not functioning as a hard propaganda site. I use a quasi-experimental synthetic difference-in-differences design and compare the coverage of Index with two of the biggest independent online media outlets and two propaganda sites that are proved to be tied to the government.

1.2 Outline

This thesis will follow the following structure: In the next chapter, I summarize the relevant literature that will serve as the theoretical background of the research design that follows. I will cover the concept of media bias, censorship, and academic works on the current state of the political and media system in Hungary and research on the governmental responses to the

COVID-19 pandemic. Then in the third chapter, I form the main assumptions of the study. In the fourth chapter, I show the research design of this study and the findings of the pilot study. It will be followed by the statistical data analysis in chapter five. It will begin with the analysis of censorship that compares news frequencies with historical data on the Hungarian COVID-19 pandemic. Then I follow with the statistical analyses that rely on the human-rated coding data, first the sentiment analysis of the news titles, then the sentiment and actor ratings from the news fragments. After that, chapter six will be devoted to the qualitative discussions and possible interpretation of the results. The seventh and closing conclusion chapter will summarize the findings, underline some limitations, and outline possible directions for future research. The thesis also contains supplementary documentation, which can be found in the appendix. Appendix A contains the links to the replication codes, and the raw data and. Appendix B contains a sample of the survey of the crowdsourcing data collection, while Appendix C contains the regression tables of the DiD analyses.

2 Literature review

The goal of this research study is to analyze the effect of the change on the editorial staff of one of Hungary's leading online news media, index.hu in August of 2020. The impact of the direction change is mainly measured through the selective attribution of the coverage of the COVID pandemic in Hungary. This literature review chapter starts with the concept of media bias in political science studies, the different types of media slants, and the main challenges and measurement strategies for measuring media bias. Then the next section, the impact of biased media is discussed, which can justify the societal relevance of analyzing a shift in coverage. In contrast to the United States, where most of the discussed media bias studies were conducted, Hungary is not a stable democracy, but rather some sort of hybrid regime, so it is worth taking a look at the specificities of media at autocracies and dictatorships, and then discuss the current political and media landscape of Hungary. Lastly, as this study analyzes coverage of the COVID pandemic, I gave a brief overview of the research on the media representation of the pandemic and how it affected governments.

2.1 Media bias

2.1.1 The Definition of Media Bias

In this thesis, Groeling's definition of media bias will be used, which states that media bias is "a portrayal of reality that is significantly and systematically (not randomly) distorted" (Groeling, 2013, p. 133). Some definitions focus on the mental process of the journalists which leads to definitions similar to "prejudice". As Gentzkow and Shapiro define: "All the accounts are based on the same set of underlying facts. Yet by selective omission, choice of words, and varying credibility ascribed to the primary source, each conveys a radically different impression

of what actually happened. The choice to slant information in this way is what we will mean in this paper by media bias." (Gentzkow & Shapiro, 2006, p. 281).

Groeling's definition of media bias also requires the researcher to prove that the bias is systematic, which means it is not anecdotal or episodic. Also, some definitions concentrate on partisan media bias and therefore exclude other aspects that can be distorted (e.g. class, gender, race, religion, etc.). One example is the one from Waldman and Devitt: "Bias can be defined as any systematic slant favoring one candidate or ideology over another." (Waldman & Devitt, 1998, p. 302).

2.1.2 The Types of Media Bias

Groeling identifies two categories of choices that reality can distort: selection bias and presentation bias. Selection bias in media coverage is similar to biased sampling in statistics. It means that not every unit in the target population (events and information) has an equal chance to be included in the news media. It can be either because the media is not looking for them to the same extent or because these units are available to them, but they are not publishing it with the same chance as other units. Partisan selection bias is a special form of it, that Groeling (2013, p. 134). defined as "choosing news stories that present a significantly distorted sample of reality that systematically and disproportionately favors one party over the other"

The other category of media bias is presentation bias, which skews the content of the stories based on the available events of information. The content of the news can be analyzed through several dimensions (e.g., length, visual presentation, framing, attribution). According to Groeling's definition, partisan presentation bias is "composing news stories in a manner that presents a significantly distorted view of reality, which systematically and disproportionately favors one party over the other."

There are other existing typologies of media bias. One of the most influential is the one from D'Alessio and Allen (2000). They define gatekeeping bias (like selection bias), coverage bias (visibility of topics or entities in media coverage, assuming equal coverage between the parties – a term that is not measurable outside bipartisan electoral politics), and statement bias (similar to presentation bias).

The interdisciplinary review of Hamborg et al. (2019) processes the state-of-the-art studies of media slant in social sciences and for each form of media bias, they propose different approaches from computer science that are suitable to (semi-)automate the existing analysis methods. They identified eight forms of media bias that can be identified using automated identification methods. In the stage of gathering, the three different forms can be event selection, source selection, or commission and omission. Then, in the stage of writing it can be labeling and word choice. During the stage of editing four forms can appear: story placement, size allocation, picture selection, and picture explanation. In all phases, the spin, e.g. the overall slant of the news can be the eight form of bias, which can be composed of various types of biases.

2.1.3 Challenges of Testing for Media Bias

The two major challenges of measuring media bias are the problem of the unobserved population and the problem of subjectivity. In this section, I will present these two concepts and some notable empirical research that aimed to measure them.

The problem of the unobserved population is a challenge of measuring selection bias. As researchers have no access to newsrooms, so only the final product of the newsgathering can be analyzed, which means that it is unobservable when events were not sought by the media or finished unpublished. Groeling and Kernell (1998, p. 1067) describe the problem as the following: "research-based exclusively on content analysis of reported news commits the

fallacy of drawing inferences from data that has been selected on the dependent variable. The issue of selection bias presents this research with a serious conundrum. How can it assess the representativeness of the sample when the population is comprised mostly of stories that were never reported and thereby elude observation?"

Different perceptions of the content of a story may arise from the different presentations of the news, although it can be also caused by the different attitudes of the news consumers towards the media. This leads to the other major challenge of media bias research which is called the subjectivity problem. Many cognitive biases can lead people to be affected by their stance when making judgments about a news source (e.g. Hastorf & Cantril, 1954). As Groelling and Kernell (1998) note, the subjectivity problem can be present when researchers aim to train human coders to judge media contents that they give an arbitrary set of rules to the raters, while it can be assumed that with alternative guidelines the distribution of good and bad news can be different. They consider the lack of weighing of the media news as another important factor as content analysis usually categorizes the news without weighing its importance, while some news is more important than others.

A famous example of the subjectivity problem is the hostile media phenomenon. It states that identical stories were perceived as having opposed biases depending on who viewed them. (Vallone et al., 1985). Further experimental studies showed evidence of the presence of this bias. Baum and Gussin (2008) asked their participants to evaluate the content of news about the 2004 US presidential election identified as originating from CNN, FOX, or a fictional TV station. They found that media outlets and their reputations send an ideological cue about the news content, which leads to biased processing. Another experiment came to the same conclusion using an experimental setup with different news labeled CNN or FOX as the source (Turner, 2007).

2.1.4 Measurement of Media Bias

The dominant research strategy to measure media bias is conducting content analysis. It can be used to identify and quantify media bias. Another method is frame analysis, which is a special version of content analysis. Third, meta-analysis offers another possible strategy, where researchers can collect and combine the existing findings of previous studies to reach further results. One example is the study of McCarthy et al. (1996), who analyzed several researches about media reports of demonstrations and their findings suggest that the best predictors are the size of the protest and the previous attention of the news media towards the topic.

Following the review article of Groeling (2013), I present what are the more innovative research in recent times that addressed the previously introduced two major problems of measuring media bias: the unobserved population problem and the subjectivity problem. For the former, the widespread way to tackle it is to use explicit or implicit assumptions about how the unobserved population would look like. These methods can be finding equivalent cases, controlling for structural factors, comparing them to other news organizations' coverage, and observing or creating the unobserved population. In the following, I will present studies for all of these approaches.

First, it can be assumed that there are some cases where political leaders can be expected to have approximately the same amount of results or notable political behavior, therefore non-biased coverage would be even-handed (Niven, 2002). One case can be when politicians switch parties, but stay in the same political context in the same voting district, therefore one can assume that the shift in their coverage must be related to a partisan bias (Niven, 2003). Another study assumed that the Republican and Democratic Party national party conventions are equally important events, and thus the unequal coverage can be considered as biased (Morris & Francia, 2010).

While these are creative approaches to measuring media slant, two remarks have to be made. First, these studies are from the US, which is a country with two dominant parties. In a plural system, where there are several relevant actors with differing sizes and popularity, it is much harder to make this sort of assumption. Secondly, given the heuristic nature of these assumptions, the equivalent nature of these cases can be questioned (e.g. the party switch can be an event that can draw interest, independent from partisan bias, or the political context of a political convention can lead to differences on their importance).

Another approach to controlling for structural factors is based on trying to account for all other structural factors that seem relevant, such as the candidates' incumbency, money, standing in opinion polls, scandals, and national conditions and market factors. The study of Schiffer (2006) used this approach to newspapers' Senate election coverage and concluded that the Democratic candidates remain in the tone of coverage, but after controlling for structural biases, this effect decreases. The obvious disadvantage of this approach is that it is difficult to identify and measure all the structural factors that are likely to be relevant.

The third way is to compare the coverage across different news outlets. The reason behind this is that while the researchers have no access to the complete population of events, they can assume that different media actors are selecting their stories from the same pool. Among others, these studies can compare the tone of various local news coverage by using a national outlet as the baseline (Barrett & Peake, 2007), analyze openly partisan outlets to give benchmarks for liberal or conservative coverage (Covert & Wasburn, 2007), or compare the content of the political newsfeed of two wire services (Baum & Groeling, 2008).

The fourth and last approach to address the problem of the unobserved population is to aim to observe the elusive population of political events. While the previous methods aim to estimate the unobserved population through assumptions, in these branches of studies, the

researchers try to narrow down the whole population into the subset of content, where they are able to observe every potential event that journalists can decide to cover. For example, Aday (2010) collected a set of real-world indicators from the US wars in Iraq and Afghanistan (e.g. missile strikes, battles, suicide bombings) and compared it with the coverage of NBC and Fox News.

Turning to the attempts that target to solve the issue of subjectivity that can be connected to content analysis. One direction is to compare the news organizations' content directly to political texts and assume that if an outlet shares more similarities with one political than with the others, then it is slanted towards it. The similarity can be measured by

2.1.5 Political effects of biased media consumption

There is much evidence in the literature that politically motivated media outlets can impact the voter's attitudes toward policy issues or voting intentions, and also the decision-making of the politicians, thus affecting the political landscape of a country. In this section, I present some research from this field of study.

Berhardt et al. (Bernhardt et al., 2008) designed an integrated model of media bias, where it is optimal for news media firms to have a partisan bias in their coverage to maximize their profits. Consumers of biased news outlets are not informed about every important fact that is unfavorable to the supported political candidate. It means that even an assumed completely rational voter who is aware of media bias can not have all the necessary information, which can lead to casting their ballot for the wrong candidate. They also conclude that even if the centrist median voters get access to unbiased news, media bias can create disproportionate "cross-over" voting.

A study covering more than 60 US Senator campaigns through three election years showed that voters evaluate endorsed candidates more favorably than candidates who fail to secure an editorial endorsement. The impact of the endorsement decision on voters' evaluations is most powerful in races with significant press attention and among citizens who are regular consumers of daily newspapers (Kahn & Kenney, 2002). This finding is also supported by a study that analyzes a single US Senate campaign in the United States in 2000, where researchers compared relative editorial slant in two local newspapers and found that it indeed affected the voters at the poll. (Druckman & Parkin, 2005).

A study from Israel showed that a businessman launching a free local media supporting Likud indicated electoral influence (Grossman et al., 2022).

In a quasi-experiment, researchers analyzed the entry of a new channel in cable markets and its impact on voting. Between October 1996 and November 2000, the conservative Fox News was introduced in the cable programming of 20% of the towns in the USA. They found a significant effect of the introduction of Fox News on the vote share in Presidential elections between 1996 and 2000. Although, they were not sure whether the “Fox News effect” is a temporary learning effect for rational voters, or is it a permanent effect for nonrational voters, who can be subject to persuasion (DellaVigna & Kaplan, 2007).

However, there is evidence that voters are not necessarily following bluntly their media’s stance. In a field experiment during the 2005 Virginia gubernatorial election, both experimental groups of people who were given Washington Times or Washington Post subscriptions showed more support for the candidate of the Democrats than those in the control group that did not receive any subscriptions (Gerber et al., 2009).

However, there are also academic works that suggest that media bias's effects are not as significant. A recent study of the news diet of American citizens showed that (Guess, 2021). In another example, survey research about the Italian state media showed that even if there are biased online news outlets, there is a huge level of overlapping in their consumption as most of the US citizens Another example can be a survey study conducted in Italy that finds that citizens can be sensitive to a shift in the political orientation of a media outlet: after the 2001 national elections when a right-wing government came into power to control the state TV, many of the left-wing viewers switched to another public channel (Durante & Knight, 2012).

2.2 Media manipulation in hybrid regimes

The study of Guriev & Treisman (2015) sets up a formal model of informational autocracies with the Orbán-regime being one of their examples. They argue that an emerging new group of modern autocrats differ from the classic dictatorships as they do not reject Western-style liberal systems, thus they pretend to be democratic and build on popularity as a source of legitimacy. They avoid open violence, more likely to conceal their involvement or camouflage its purpose. This approach has important implications for the use of media. As they are less ideological, the emphasis during the use of propaganda is not on indoctrination, but on improving evaluations of their leadership through signaling competence, favorably manipulating news, and increasing their responsiveness by allowing some space for critical views in the media sphere.

One way of manipulation is censorship. It can be done to silent criticism of the government (the "state critique" theory) or prevent the attempts to physical moves (the "collective action potential" theory). The study of King, Pan, and Roberts (2014) aimed to answer this question in China in 2013. They watched whether published social media posts were taken down later, and created posts to see whether they were prevented from publication. The results support the collective action potential theory and not the state critique theory, as the

Chinese regime lets the citizens criticize the system (even the leaders, or sensitive topics like corruption), but does not tolerate the attempts to create an offline crowd.

Another study from China (Huang, 2018) found that using hard propaganda can worsen people's opinions of the system, while simultaneously signaling the regime's power and reducing citizens' willingness to protest. It means that hard propaganda can have positive short-term effects, but it can damage the long-term prospects. While China is considered a more authoritarian country than Hungary, it is an interesting implication that even a harder system, censorship is not general, and the regime has to be careful with using hard propaganda.

A quasi-experiment from former East Germany (Gläßel & Paula, 2020) can support the previous finding. The researchers found that in the presence of alternative media sources, the censorship of the state media backfired on the regime. It can be assumed that if it was not manageable in the 20th century to prevent every form of conflicting news, in the era of the internet and social media it is much less possible option, and therefore it is a more sustainable strategy to focus on the most damaging forms of contents only. This backfiring nature of news censorship can be explained by the so-called Streisand effect: drawing attention to the forbidden information can cause political outrage toward the censor, and raise public awareness of the censored content (Jansen & Martin, 2015).

According to Rozenas and Stukal (2019), there are some cases when real-world facts and events are harder to hide from the popular because they can benchmark them through their personal experiences. Analysing Russian state media they found that the government is not using censorship, but rather selective attribution when presenting economic news. The idea is that the economy can be benchmarked by the citizens (e.g. through salaries and prices), so censoring bad news is not effective. However, the people's attribution can be shaped by connecting the positive news to the government and the negative ones to external actors (e.g.

domestic opposition, international affects). This blame-shifting technique is called selective attribution by the authors, which can be either direct (the actors are directly accused of an event), or it can be by association (they are mentioned in a positive or a negative context).

However, analyzing another competing authoritarian regime, namely Erdoğan's Turkey, a population-based survey experiment study concludes that during an economic crisis, the incumbent's effort to shift the blame to other actors was not as effective as another tactic, then changing the political agenda away from the economy to an issue area that is more favorable for the incumbent (Aytac, 2021).

To conclude this brief literature review, we can say that contemporary authoritarian systems, like the Orbán-system in Hungary can be described with various practices regarding media manipulation. We can expect that they use media not primarily for indoctrination and they are likely to avoid hard propaganda, but they manipulate the news through sophisticated methods. These can be among others censorship of specific types of news, shifting agenda setting, or shifting the blame by selectively attributing positive and negative news. The existing pieces of evidence from index.hu suggests that the government is using the site for these kinds of news manipulation.

2.3 The Case of Hungary

This subsection starts an overview of the Hungarian political and media system, with a brief summary of the contemporary history of the country and the characterization of the current situation. Finally, the Hungarian government's political reaction to the COVID-19 pandemic will be presented.

2.3.1 The Political System

After the revolutions of 1989 in Eastern Europe, the transition of Hungary from a Warsaw Pact state into a Western market democracy was considered among the more successful ones, where the new governments were led by pluralism, liberal government, with marginalizing ethnic nationalist and extremist parties (Vachudová & Snyder, 1997). However, after the Great Recession starting in 2008 and the legitimacy crisis of the socialist government opened up opportunities for radical change in the political landscape. The right-wing Fidesz party led by Viktor Orbán won by a landslide in the general election of 2010, winning a supermajority in the house of representatives that allowed the new power to submit a new constitution, and also a new media law (see more detailed in the next subchapter). This political transformation created much attention in political science as part of the democratic backsliding debate (Cianetti et al., 2018).

There has been a diversity in categorizing the Orbán-system, as it has been interpreted as a competitive authoritarianist regime (Bieber, 2018), a diffusely defective democracy (Bogaards, 2018), or an externally constrained hybrid regime (Bozóki & Hegedűs, 2018). These academic works explain the democratic backsliding in Hungary while underlining the role of power concentration, violation of the rule of law, systemic corruption, restriction on civil society, and electoral manipulation. Currently, Freedom House rates Hungary as a transitional or hybrid regime (Freedom House, 2022), which is often characterized as a novel type of authoritarian system (Levitsky & Way, 2010). Based on these academic works I treat Hungary as a hybrid regime, and that is why I mostly relied on literature on media, news manipulation, and propaganda in hybrid regimes.

2.3.2 The State of the Media

After 2010 the state media replaced public radio and television channels. Their programs heavily under-represented opposition politicians and intellectuals. The new media

law created a media supervisory authority loyal to the governing party. The media authority can issue financial penalties to radio and television, print or electronic media, and even to bloggers in an amount that can silence them completely. The government can influence the media through personnel policies and governmental advertising. These measures can control the policies of news agencies and state television, affect the editing culture (e.g. forgery, manipulation), and can create an atmosphere of fear and self-censorship among the media workers (Bozóki, 2011).

The online market is the most balanced among the media markets in the country. However, the online news portal with the greatest reach in Hungary, Origo.hu, was acquired by a businessman with ties to the government in February 2016. Another portal of the same actor, vs.hu was awarded over half a billion forints from foundations connected to the Hungarian National Bank. (Máriás et al., 2017)

Griffen (2020) identifies four principal elements of the media capture of the Orbán regime: (1) through the forced closure or takeover of previously independent media; (2) manipulation of the media market through state resources and regulatory power; (3). delegitimization and exclusion of independent journalists; and (4) maintaining the illusion of media freedom in the country.

It is worth mentioning that the Hungarian media were highly integrated with politics with biased journalists before the political change of 2010. Two possible reasons for this: (1) the Hungarian political culture instrumentalized and expected journalists to be loyal to political sides; (2) the media market of the country is small, so political media needed political support to survive. (Sipos, 2013).

The research of Szeidl and Szűcs (2021) gains evidence about the effect of ownership change and targeted government advertising causing pro-governmental media content, while another study (Bátorfy & Urbán, 2019) highlights the market-distorting effect of the government as clients of public advertisements.

2.3.3 The governmental response to the Covid-19 pandemic

Until now there is no comprehensive analysis of how the Hungarian governmental propaganda handled the COVID crisis. However, as a recent study points out, during the pandemic, the Hungarian government aimed to control the crisis-related information and create a political discourse that can legitimize and endorse their political ambitions (Hajnal et al., 2021).

Although the Hungarian government tended to ignore or refuse media inquiries and interview requests by independent journalists, the Covid-19 pandemic caused further centralization and restrictions on their communications. For example, journalists cannot get information about the price, quality, and certifications of the protective equipment and Covid-19 tests bought from China, even if they should be public information (Serdült, 2020). The interview-based study of Bleyer-Simon (2021) examined how journalists have seen the domestic pandemic measures, and what was the eventual effect of them on their work. The study concludes that media pluralism was further decreased in the country, mainly because the elite connected to the government interfered in the media economy.

3 Assumptions

Hypothesis 1: The editorial change in Index caused more intensive coverage when the COVID-19 pandemic in Hungary was worsening and less intensive coverage when it was subsided (censorship).

Hypothesis 2: The editorial change in Index caused a more positive coverage of the COVID-19 pandemic in Hungary (framing bias).

Hypothesis 3: The editorial change in Index caused more positive coverage of governmental actors and more negative coverage of external actors regarding COVID-related news (selective attribution)

Hypothesis 4: The editorial change in Index caused more coverage of governmental actors regarding good news, and less coverage of external actors regarding bad news about the pandemic (selective attribution by association).

All assumed effects are understood in relation to other players in the online market. It means I do not make the assumption that without the takeover, the management of index.hu would have reported on the coronavirus epidemic in the same way as it did before the takeover. I do this because the challenges and political responses of the coronavirus epidemic, or society's expectations were also radically different in the different epidemic waves. therefore, the potential impact of the editorial change is analyzed with the help of the DiD method during comparison with other major online media market actors.

The literature on propaganda and media in illiberal systems and autocracies does not give us clear implications, and our existing knowledge about the new direction of Index and its expected coverage of the COVID pandemic. There are also good reasons to believe that Index

has not started to use tools of hard propaganda, and has not changed radically its overall editorial practice. So, these assumptions are formed based on the knowledge of how sophisticated ways of news manipulation can work in hybrid regimes. Therefore, testing four these four various techniques in the literature can be considered as an exploratory approach. With that being said, alternative hypotheses can be also formed stating that the media capture did not lead to introducing any of the four forms of the slanted report (censorship, framing bias, selective attribution, or selective attribution by association).

The assumptions will be tested in a DiD analysis using historical data, and rated news titles and news fragments as input data generated by crowdsourcing data collection. The next chapter provides an overlook of the research design of the quasi-experiment, which elaborates on how the assumptions are converted into measurable variables that can be tested for the research's queries.

4 Research Design

4.1 Overview of the Research Design

The study analyzes the effects of the editorial shift in the Hungarian news outlet, index.hu, that happened in August of 2020. I am using a difference-in-differences design to analyze the Covid news coverage of the site before and after six months of the change and compare it with four other major national online news outlets. Of these four sites, two of them are considered independent sites (24.hu, 444.hu), while the other two are related to the Orbán government (magyarnemzet.hu, origo.hu). This research proxy the slant of the media outlet in four different ways. The first one is measuring the differences in the frequency of domestic COVID-19 coverage when the pandemic is worsening or improving in the country (censorship). The second one is using the sentiment of the title of the pandemic-related news (farming bias). The third

one is based on how favorable picture the news portrays of governmental and opposition political actors based on news fragments (selective attribution). The fourth one is based on the overall sentiment of the news fragments that contain the given actors (selective attribution by association).

4.2 Data and Measurement

The analyses of this thesis project are based on news from index.hu and four other Hungarian online media platforms: two of them considered independent and critical of the government (24.hu, 444.hu) and two others that are affiliated with the Orbán-system (magyarnemzet.hu, origo.hu). The decision on case selection and categorization as independent or propaganda is made by the traffic of the news sites (NMHH, 2021), their profile, and previous academic works on Hungarian online news media (Eszter et al., 2021; G. Szabó & Bene, 2016; Tóth, 2021). The original corpus of the pre-processed raw data was scraped by Simonovits and Vig (2023) and it is used with their consent. The initial dataset consists of all available news from these sites in a span of 14 months, between the beginning of January 2020 and the end of February 2021. The whole corpus of news from the five sites contains 237,582 articles. Each of the news entries in the corpus contains the URL of the article, the time when the article was published, the title, the textual content of the article, and the tags that were assigned to it by the editors.

Using these sites' classification tags, I aimed to find the news that reports about the COVID-19 pandemic, selecting the articles that cover domestic news and contain at least one of the following keywords: “COVID”, “COVID-19”, “koronavírus” [“coronavirus”], and “pandémia” [“pandemic”]. Using these keywords N = 41,676 news were filtered. The number of filtered articles for every news site can be seen in Table 1, while the distribution of COVID-related news by month is presented in Figure 1.

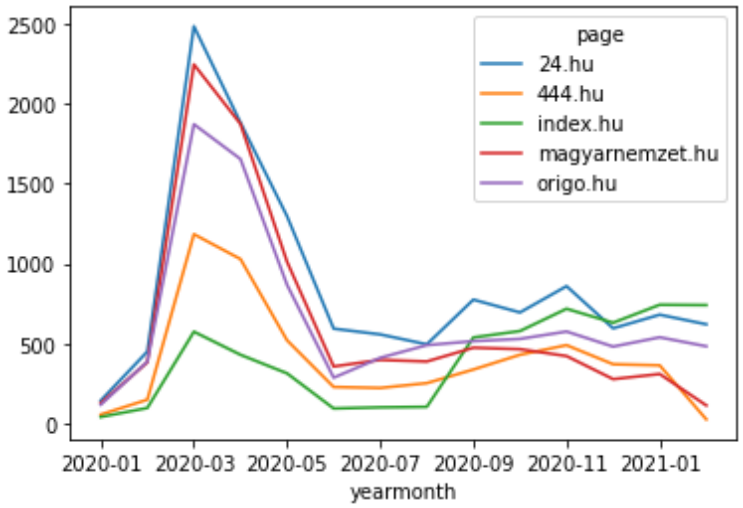

Figure 1: Number of COVID-related news per month in the five media outlets

Media site	**COVID articles**
24.hu	12 139
444.hu	5 700
index.hu	5 733
magyarnemzet.hu	8 873
origo.hu	9 231
SUM	**41 676**

Table 1: The number of COVID articles for every media site

4.3 Measuring Selection Bias

The measurement of selection bias is often hampered because censorship is rarely directly observable by the researcher (see also the problem of unobserved population discussed earlier). In this research, I measure selection bias indirectly following the methodology of Rozenas and Stukal (2019). This is based on the assumption that biased coverage is less likely to cover domestic COVID-19 news when the epidemic situation worsens, compared to when the epidemic situation improves.

To implement this strategy, I have chosen two domestic indicators that are relevant to the Hungarian coronavirus epidemic: the daily number of 1) new cases, 2) hospitalized patients.

Since vaccines were not yet available for most of the period under study, the number of newly vaccinated persons was not included in the indicators to be examined. These data were all publicly available, for this research I used the database of Our World in Data (Mathieu et al., 2020). It is worth mentioning that the first COVID-19 vaccine was given in Hungary on the 26th of December 2020 (Á. Szabó, 2020). That means that vaccines were not yet available for most of the period under study, So that is the reason why the number of newly vaccinated persons was not included in the indicators to be examined for measuring selection bias.

In order to determine whether a given news item is a report on indicators, I filtered for the occurrence of the following keywords in their various conjugated forms: 1) for new cases “új eset” ["new case"], “új fertőzött” ["new infected"], “új megbetegedé” ["new disease"]; 2) for hospitalized patients: “kórházi ápolt”, “kórházi kezelt”, [both means "hospitalized"].

4.4 Measuring Framing Bias

In order to measure the possible bias due to framing, I conducted a sentiment analysis of news headlines using the crowdsourcing method. I chose to analyze the titles because they are short and can therefore be evaluated quickly, as opposed to a longer news item or the full text of the news. In addition, as they are the most visible part of the article, as they appear on the front page, the first time the reader encounters them when reading a piece of news, it can be heuristically assumed that they contain the most important information the journalist or the editor wants to convey to the reader. Furthermore, compared to shorter fragments, they are more likely to be comprehensive by their own. There are several examples in the literature of media bias studies that relied on the method of evaluating news titles (e.g. Budak et al., 2016; Dallmann et al., 2015; Stefanov et al., 2020).

From the available 41,676 COVID-19 news, I selected 20,000 news with lengths approximately between the first and the second quartiles. The purpose of this restriction is that there are unlikely to be sufficient human resources to classify all the titles, so I have chosen titles that are possibly the easiest to interpret. A subset of these titles are rated by human raters in a way that is described below, in the section on the crowdsourcing workflow.

4.5 Measuring Selective Attribution

To measure selective attribution in the COVID-related news two pieces of information are needed: whether the reported news is good or bad about the pandemic situation in Hungary (selective attribution by association) and what is the evaluation of the actor that is related to this event (selective attribution). I decided to rate news fragments that contain mentions of relevant political actors. The selection of the actors and the news fragments are described below.

First, I identified the relevant political actors in the corpus. For that, I collected every proper noun in the text that forms a unigram or a bigram and then selected the ones that are relevant political actors using the following two categories. The first one is labeled as governmental actors. This category consists of Prime Minister Orbán Viktor and other people and institutions that can be connected to the government, e.g. the Minister of Human Resources, Miklós Kásler or the Chief Medical Officer Cecília Müller, or the major governing party, Fidesz. The second group is assigned to the opposition actors, which can mean both domestic opposition politicians (government (e.g. the High Major of Budapest, Gergely Karácsony or DK, the most supported opposition party) and external actors that are often considered as political opponents of the government (e.g. the European Union). From this method, I got a list of 86 political actors that got mentioned on average at least two times in a month in the covered news.

In the next step, I created the news fragments that mention these actors using the following selection method. For every political actor in the previously created list, I went through every news in the corpus and if the actor was mentioned in that sentence, then selected the first sentence where the actor appeared and the sentence that followed it. If it was the last sentence of the news, so no two-sentence long fragments are possible, then I dropped it. The logic behind this selection process is that it can be assumed that the most important statement regarding an actor is more likely to be made in the first mention of the actor than in latter mentions. Using this method 64,657 two-sentence long fragments were created. As mentioned in previous studies using a similar design (e.g. Rozenas and Stukal, 2019), shorter fragments can lead to a sparsity problem, while the longer fragments can cause weak intercoder reliability, and also take more time for the human raters to code them.

A subset of these fragments is rated by human raters in a way that is described below, in the section on the crowdsourcing workflow.

Sentences	24.hu	444.hu	index.hu	magyarnemzet.hu	origo.hu
Mean	14.5	13.1	16.3	17.1	16.0
Median	11	9	12	14	12
Mode	8	6	8	12	7

Table 2 Basic statistics about the number of sentences in an article on the five sites

4.6 Pilot Study

Before the crowdsourcing data collection of the main study, in January of 2023, a pilot study was conducted. The goal of this pre-test was to answer the following questions. First, there seems to be a trade-off dilemma between selecting one-sentence or two-sentence-long fragments for the main study. One sentence-long fragment can be processed faster, but there is

a chance that the lack of a broader context will make it harder to rate them as positive or negative or the mentioned political actor's portrayal as favorable or non-favorable.

During the pilot study, Hungarian adults were recruited on social media platforms to voluntarily fill out an online survey. After reading the information and consent form and agreeing to fill out the survey, each participant was asked to rate fifteen fragments in the survey. The fragments were randomly assigned to the participants from the pool of 3600 news fragments with one-sentence and two-sentence long fragments evenly divided among them. A total of 1532 fragment ratings were recorded from 102 participants (with data from unfinished data also counted). Furthermore, after running a regression within the two-sentence-long fragments: longer sentences are more understandable with alpha = 0.02, so one additional character means an additional 2% chance of being able to categorize the fragment (meaning not rating as "Neutral.").

	Overall sentiment							**Actors' favorability ratings**						
Group	**1**	**2**	**3**	**4**	**5**	**SUM**	**Ne. %**	**1**	**2**	**3**	**4**	**5**	**SUM**	**Ne. %**
1-sen	106	206	287	128	65	792	36,24%	74	91	400	140	87	792	50,51%
2-sen	144	234	171	132	59	740	23,11%	90	124	272	166	88	740	36,76%

SUM: number of all ratings in the category; Ne.%: percentage of neutral (3) ratings.

Table 3: The ratings of the one- and two-sentence-long fragments in the pilot study

Completion time is an other issue, as tow long surveys may decrease the validity of the data. Fifty percent of the participants completed the fifteen fragments between 6 and 16 mins, so it is fare to assume that even with only having tow-sentence long sentences or slightly longer sentences in the main study, the rating time is below 90 seconds. It is a gross completion time, so the net time for a single was not analyzed.

The fragments were randomly distributed from a pool of 3600 fragments without a limitation of how many times a fragment could be rated during the pilot study. 1292 fragments were rated at least once, while 240 fragments were rated at least by two participants. The inter-coder reliability is satisfactory (Kappa = 0.82).

Ratings	SUM	The maximal difference between the ratings				
		0	1	2	3	4
2	207	74	91	33	8	1
3	31	3	17	8	3	0
4	1	0	0	1	0	0
5	1	0	0	1	0	0
		32,1%	45,0%	17,9%	4,6%	0,4%

Table 4: Maximal difference between the ratings in the pilot study

4.7 Crowdsourcing

The data collection process through crowdsourcing was the following. The human raters were recruited in social media groups and paid advertisements and were asked to fill out an online survey. Adult Hungarians were eligible to participate on the basis of the selection criteria. During the May of 2023, 554 participants took the online survey. The descriptive statistics of the human coders can be seen in Table 5.

Gender	Count	Percentage	Education	Count	Percentage
woman	268	38.07%	higher education	202	57.39%
man	82	11.65%	high school	143	40.63%
other	2	0.28%	elementary school	4	1.14%
			less than elementary	3	0.85%
Age	**Count**	**Percentage**	**Residence**	**Count**	**Percentage**
65+	91	25.85%	large town	142	40.34%
45-54	84	23.86%	Budapest	124	35.23%
55-64	73	20.74%	small town	42	11.93%
25-34	52	14.77%	municipality	42	11.93%
35-44	34	9.66%	abroad	2	0.57%
18-24	18	5.11%			

Note: demographic questions were not mandatory, 202 participants did not answer.

Table 5: Summary statistics of the human coders

After consenting to the participation, in the first block, participants were given news headlines one by one. Their task was to rate the dominant mood of the headline on a Likert scale of one to five. A total of fifteen titles were graded and randomly selected. A title could be graded more than once during the sampling, but only once by a single human rater.

In the second part of the questionnaire, participants were asked to classify different news items of two sentences in length. Again, they were asked to give positive and negative ratings on a five-point Likert scale, but here they had to rate on two different criteria. First, the prevailing mood of the news item, and then the tone in which a named actor (politician, party, institution, etc.) in the news item was portrayed in relation to the event in the news item. In this part, they were shown ten randomly selected news fragments with the same randomization method as in the previous phase.

It was mandatory to answer the questions, although if the quit before completion, their answer were still recorded and used for the analysis. It is important to note that the respondents were not aware of the context of the news headlines and news items, e.g. which news portal and when. In this sense, we can speak of blind testing. A sample questionnaire can be found in the appendices. At the end of the survey participants were asked to answer some sociodemographic questions, and they were free to leave their comments as well. The participants were also incentivized to take a survey with the possibility of participating in a lottery where they could win vouchers of small value.

4.8 Difference-in-differences Method

I use the difference-in-differences method to estimate the average effect of the change in the editorial staff of Index. In this subchapter, it will be briefly presented the method (relying on how this method can be applied to address the research question, and how it makes it possible

to make a causal claim regarding the assumptions. It also describes what are the important conditions that have to be checked, namely the assumption of the parallel-trends, the stable unit treatment value assumption, and the perfect compliance assumption.

The advantage of the DiD design is that we can have a treatment group (Index) and comparison groups with different baseline means regarding COVID-related news, as we do not expect the treatment (change in the editorial staff) to be independent. The examined outcome of the analysis is the selective attribution of COVID-related events in online media articles. The treatment is the change of the editorial staff of index.hu, which makes the news from that site the experimental unit. The other four media outlets (two independent and two propaganda sites) serve as the control group, as their editorial staff did not experience major changes during the examined period. Articles from six months before and after the treatment are examined.

Let t_0 be the period before the editorial change of index (before 01/09/2020) and t_1 be the period after it and let mark the ratio of positive and negative attribution for a given actor as x_{ti} for period i (i = 0,1). If T is a dichotomous variable that marks whether there is a treatment present (so T = 1 for events from index.hu and T = 0 for the other outlets). So, the change in selective attribution for the treatment group is:

$$\Delta x(T=1) = x_{t1} (T = 1) – x_{t0} (T = 1);$$

and the difference between the two periods for the control group is:

$$\Delta x(T=0) = x_{t1} (T = 0) – x_{t0} (T = 0);$$

The ATE is then $\Delta x(T=1) - \Delta x(T=0)$, which is the difference in differences. This formula can be calculated for all the actors both direct and associative attributions. If the hypothesis of this research is right, then after the change of the editorial staff of index.hu, the

site was more likely to have a biased presentation in favor of the government regarding the handling of the COVID-19 pandemic through selective attribution.

Lastly, I consider the potential assumptions that need to be checked in order to make a valid clausal claim with the Difference-in-differences analysis. First, for the assumption of the parallel-trends (or equal-trends). While the DiD method is effective at addressing differences between treatment and comparison groups that remain constant over time, it cannot account for differences that change over time. To obtain a valid estimate of the counterfactual, we must assume that there are no time-varying differences between the treatment and comparison groups, which means that in the absence of the treatment, outcomes would show equal trends.

While it is not possible to prove the assumptions of equal trends as there is no data in the real world for that, there are several methods that can gain support for such an assumption. One possibility is conducting placebo tests, and checking whether it has no significant effect on the dependent variable.

Another important criterion is the stable unit treatment value assumption, that the potential outcomes for any unit do not depend on the treatment status of other units. This assumption is violated if there are spillover effects in the sample. It occurs when the treatment applied to the treatment group has unintended consequences on individuals or entities in the control group or other neighboring groups. In the case of this study it would mean that the overtake of Index would have a direct or indirect effect on the news creation at the other Hungarian online outlets that serve as the comparison group for this study.

Lastly, the problem of imperfect compliance is also a factor that should have be considered. If the assumption of perfect compliance is not met when we have “always takers”

or “never takers”. In this case, it would mean that some of the articles are published without having been influenced by the takeover, and the change of the editorial staff.

5 Analysis

In this chapter, I present the results of this research project. I begin by presenting the findings of the censorship analysis. Then I move forward with the description of the human coding approach. Then, I move forward with the sentiment analysis based on the unsupervised classifier. After that, the results of the supervised classification are presented, which uses the human-coded ratings as the training dataset. I conclude this statistical analysis chapter by comparing the findings of the different approaches. The raw datafiles and the replication codes for all the analyses that are presented here can be found in an online repository (see Appendix A - Replication Code).

5.1 Analysis of Censorship

It is assumed that major changes in the domestic virus situation deserve more media attention than minor changes. It is also assumed that a media portal that serves pro-government interests will be more inclined to report on the spread of the virus and the number of victims when there is an improving trend, compared to independent news portals. This has led to the assumption that index.hu, after its pro-government turn, will be more inclined to report on the decline in the incidence and death rate of the coronavirus when indicators show improvement.

To determine whether some indicators of the coronavirus epidemic are positive or negative, a 10% threshold was selected. There is no rule of thumb for how to select the threshold for categorizing a day as good or bad regarding the given indicator, although using the threshold of 5 or 15 percent led to similar results (see Appendix). With this method, I categorized 183 days as bad, 100 days as good, and 82 days are neutrals, and therefore left out for further analysis.

For the one-year period that is under study, I take the one-week moving average of the number of new cases reported and daily deaths as the starting point between 1 March 2020 and 28 February 2021. However, for the spread of the virus, it is more expressive to consider proportions, so I have assigned to each day the percentage difference from one week ago.

The results of the independent sample T-tests are summarized in Table 6. It shows that it is true for every media outlet that they more frequently covered the days when the COVID situation was worsening in the country, rather than days when it was getting better. This finding contradicts hypothesis one. Overall, it can be seen that all outlets published more articles on the pandemic, when it was worsening based on the indicators (only the case of 444. hu shows non-significant differences, also this is the site that covered the pandemic news less frequently). Also, the findings highlight, that after the overtake increased the number of news on the pandemic indicators for index.hu on both types of days.

To conclude, there is no sign of censorship in the case of index.hu after the takeover, so H1 is not supported by the empirical evidence of this research.

Media outlet	Category	Mean	SD	T-statistic	P-value
index.hu (before change)	good	1.305	0.641	-3.46	0.001
	bad	1.706	0.707		
index.hu (after change)	good	3.331	0.793	-0.321	0.749
	bad	3.380	0.823		
24.hu	good	3.151	0.997	-4.889	0.000
	bad	3.721	0.896		
444.hu	good	1.767	0.589	-1.916	0.056
	bad	1.965	0.932		
magyarnemzet.hu	good	1.979	0.754	-4.592	0.000
	bad	2.580	1.180		
origo.hu	good	4.167	1.469	-5.017	0.000
	bad	5.034	1.337		

Table 6 Comparison of the frequency of coverage of good and bad news

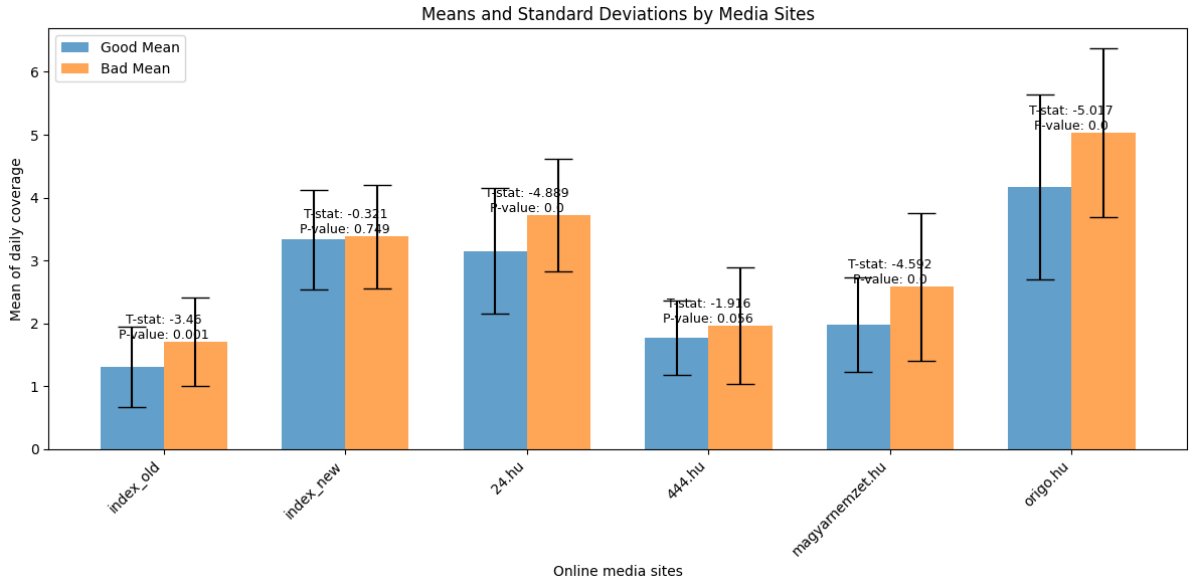

Figure 2: The frequency of improving and worsening days of the pandemic

5.2 Analysis of Framing Bias

In this section, I will present the statistical analysis of the approach that uses human coders to rate the sentiments of news titles. To determine the effect of treatment, I employ a difference-in-difference design.

A crucial assumption for the DiD estimates is that without any treatment both the treatment group and the control group would produce the same trend over time. This criterium is referred to in the literature as the assumption of parallel trend (see e.g. Gertler et al., 2016). The assumption of the parallel trend can be examined by testing the significance of placebo treatment in the time before the actual treatment happened. I chose the 1st of May 2020 and 1st of July 2020 as the time of placebo treatments as the real treatment happened on the 1st of September 2020. The coefficient is close to zero for both cases with non-significant p-values, which provides evidence for the assumption of a parallel trend.

The DiD analysis showed that there is a significant treatment effect (β = 0.227, p = 0.026), which means the treatment caused a significant increase in the rating of the titles. After the treatment, an increase of 0.227 points can be predicted compared to the non-rated groups. Although looking at the statistics measure of the goodness-of-fit of the regression model, it can be seen that the R-squared value is very low (0.008). It indicates that the editorial shift (the quasi-independent variable) only explains a very small proportion of the variation in the sentiment of the headlines (the independent variable).

This finding proves a piece of evidence for H2, although it must be mentioned that the average title ratings after the takeover only reached the independent outlets' mean. It means that while it can be said that the treatment led to a more favorable presentation of the news related to the pandemic, on the other hand, the picture shows us that index.hu is still more comparable to the independent outlets that are generally critical with the government, rather than the propaganda sites, that are showing an even more positive picture. Further

	index.hu		propaganda		independent	
	mean	SD	mean	SD	mean	SD
before treatment	2.382	1.106	2.632	1.263	2.805	1.235
	(N = 248)		(N = 1852)		(N = 1500)	
after treatment	2.443	1.220	2.430	1.257	2.684	1.210
	(N = 565)		(N = 853)		(N = 790)	

Table 7: The descriptive statistics of the ratings of the news titles

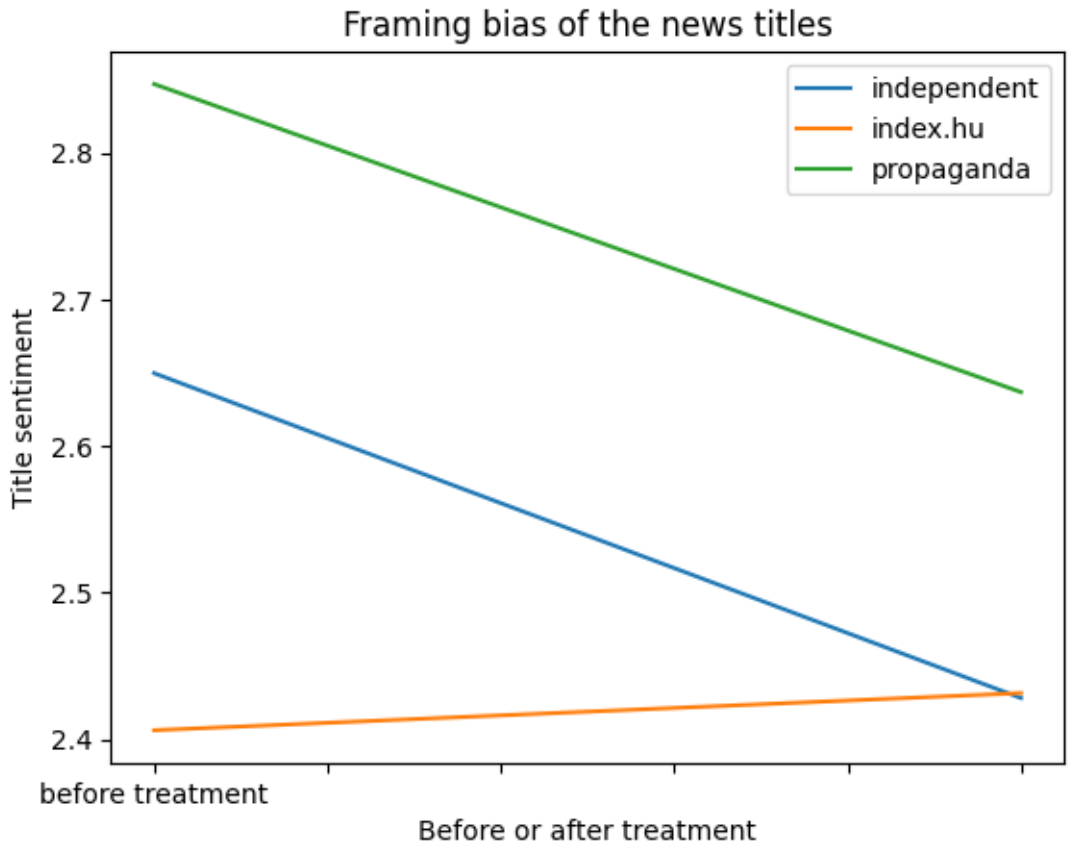

Figure 3: The mean of the title ratings

5.3 Analysis of Selective Attribution

In this part, I test the third assumption by processing the actor ratings from the crowdsourcing data collection. This assumption is based on the concept of selective attribution, which is a form of a sophisticated form of news manipulation. It means that even if there is a somewhat balanced coverage of news, there is a tendency in some sort of outlets to attribute the positive events to one group of actors, and negative for others. This means that I assume that propaganda outlets are more likely to present the Hungarian government as responsible for the positive events of the pandemic and connect the bad events more likely to external actors that oppose the government – all compared to independent media outlets.

Starting with the governmental actors, it seems that before the treatment the fragments from index.hu portrayed a picture of the governmental actors that is between the propaganda outlets and the other independent sites, and after the takeover, there was a decrease in the ratings of the governmental actors, while the difference increased between the pro-governmental and the critical websites. The Difference-in-Differences analysis shows that there is no significant effect of the treatment ($\beta = -0.069$, $p = 0.851$), which implies that based on the actor ratings of the fragments, we cannot make a claim that has an effect on how Index portrayed governmental actors in COVID-related news.

	index.hu		propaganda		independent	
	mean	SD	mean	SD	mean	SD
before treatment	2.531	2.432	2.231	2.479	3.220	2.495
	(N = 16)		(N = 118)		(N = 59)	
after treatment	3.780	2.416	1.883	2.371	2.476	2.467
	(N = 25)		(N = 47)		(N = 41)	

Table 8: The descriptive statistics of the ratings of the governmental actors

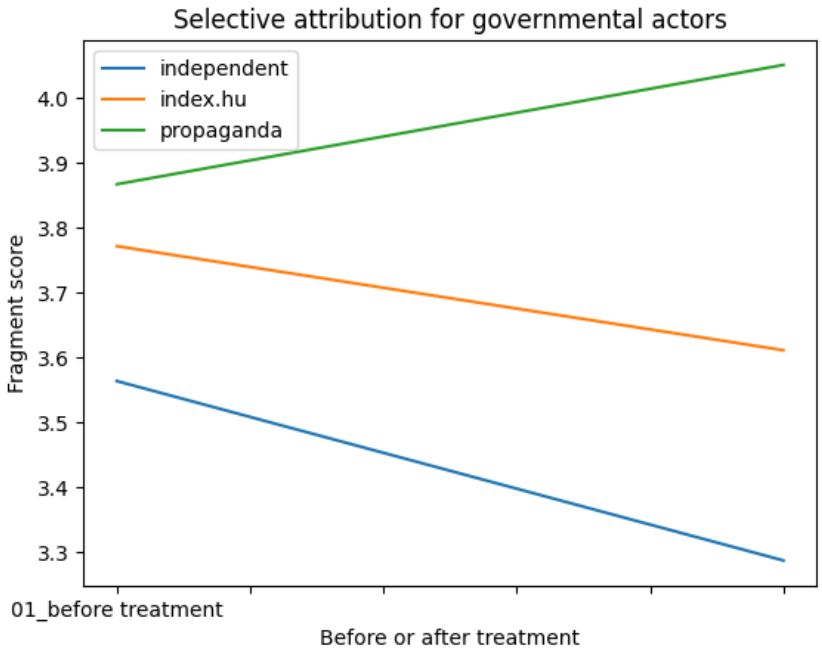

Figure 4: The mean of the actor ratings for governmental actors

For the opposition actors, it is more obvious that the expected outcome is not present, as can be seen in Table 9 and in Figure 5. The trend for both the independent and the pro-governmental outlets that in the second half of the analyzed period, they portrayed the opposition actors in a more negative way, and while I expected the new editorial staff of index.hu to be more critical of the actors of the external actors, the opposite can be seen, as they average rating score increased (from 2.531 to 3.780). Interestingly, the DiD analysis indicates weak evidence against the null hypothesis ($\beta = 0.1.650$, $p = 0.055$), while the placebo tests suggest that the assumption of parallel trends can be supported. However, the sample size is rather small for the treatment group (16 ratings before and 25 ratings after the treatment).

	index.hu		propaganda		independent	
	mean	SD	mean	SD	mean	SD
before treatment	2.531 (N = 16)	2.432	2.231 (N = 118)	2.479	3.220 (N = 59)	2.495
after treatment	3.780 (N = 25)	2.416	1.883 (N = 47)	2.371	2.476 (N = 41)	2.467

Table 9: The descriptive statistics of the ratings of the opposition actors

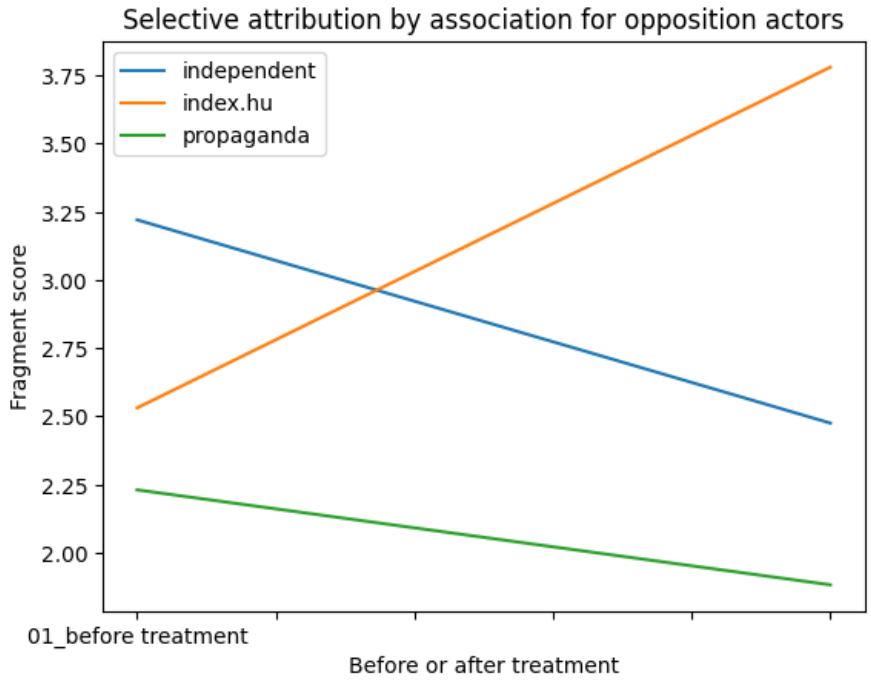

Figure 5: The mean of the actor ratings for opposition actors

The conclusion of the analysis of selective attribution is that the assumption cannot be accepted based on the processed data. It cannot be proved that there is an effect of more positive coverage of governmental actors, and in the case of the opposition actors, there is weak evidence of an opposite effect to that expected.

5.4 Analysis of Selective Attribution by Association

The next assumption I am testing is selective attribution by association using the fragment's overall sentiment ratings given by the human coders. This bias is an even more sophisticated way of news manipulation, which would mean that a given actor tends to verbalize negative or positive news at a given media outlet. If the hypothesis is correct, after the editorial change of the index, it was more typical to mention opposition actors in bad news, while government actors were more likely to be mentioned in connection with good news.

For the rating of the governmental actors, it can be seen in Table 10 and in Figure 1Figure 6, that after the change in the editorial staff (the treatment), Index was more likely to report on governmental actors in a positive context. However, both before and after the takeover, Index was more likely to associate the governmental actor with positive news than the propaganda sites. Although the Difference-in-difference analysis showed that there is no significant effect of the treatment ($\beta = 0.301$, $p = 0.421$), that means that based on the fragments' ratings, we cannot make a claim that the change in the editorial staff caused Index to associate governmental actors with more positive news regarding the pandemics then with the previous editorial staff.

	index.hu		propaganda		independent	
	mean	SD	mean	SD	mean	SD
before treatment	2.173 (N = 283)	2.492	2.280 (N = 59)	2.196	1.797 (N = 425)	2.353
after treatment	2.444	2.363	2.192	2.125	1.604	2.325

(N = 259) (N = 162) (N = 240)

Table 10: The descriptive statistics of the ratings of the fragments with governmental actors

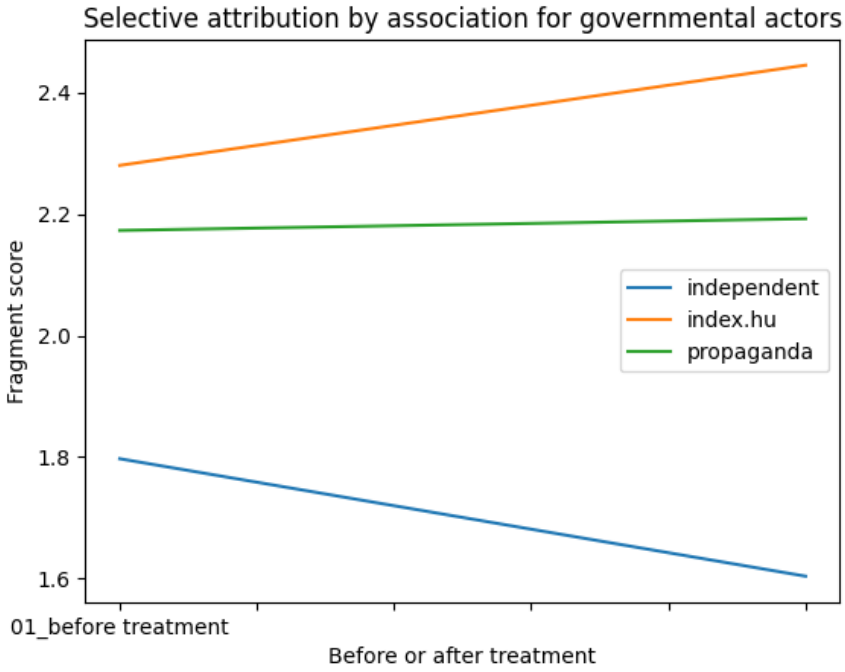

Figure 6: The mean of the fragment ratings for news with governmental actors

Moving forward, I look at the selective attribution by association regarding the opposition actors. The results are presented in Table 11 and **Error! Reference source not found.**. The assumption would imply that after the overtake, index.hu was more likely to associate the external actors that can be considered as opposed to the government with negative news regarding the pandemic. However, the empirical data contradicts this, as we can see that Index associated the opposition actors in a more positive context after the overtake, while before that it was below the level of not just the independent outlets, but also the propaganda sites. Furthermore, it is worth mentioning that the size of the sample is relatively small (before treatment: N = 16; after treatment: N = 25 in the experimental group). After that, it is not unexpected that the result of the Difference-in-Differences analysis indicates no significant

treatment effect in this case (). As in the previous sections, the DiD analyses can be seen in detail in the Appendices.

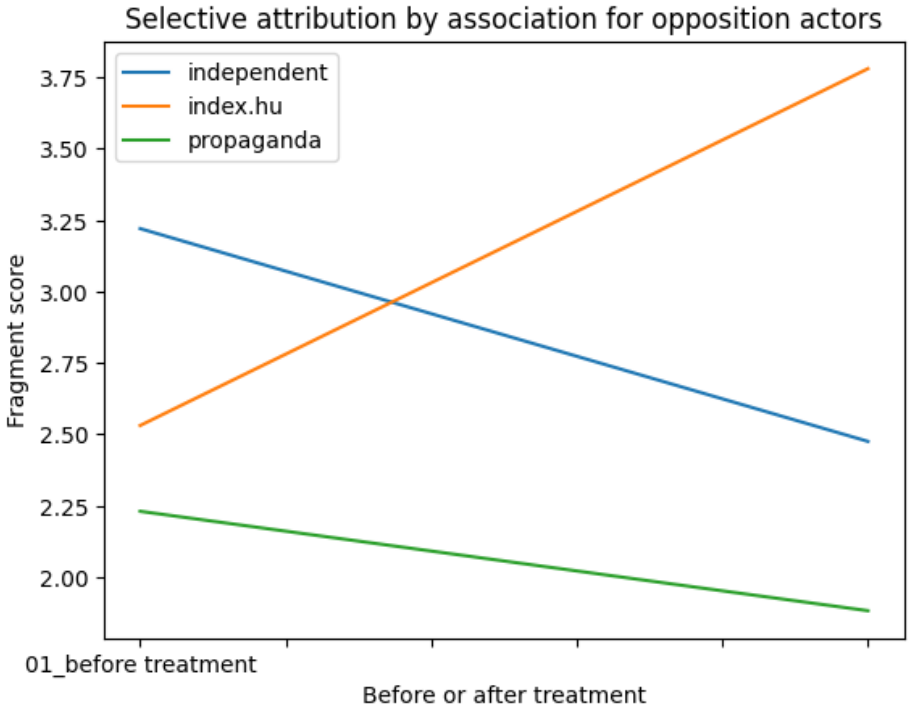

Figure 7: The means of the fragment ratings for news with opposition actors

	index.hu		propaganda		independent	
	mean	SD	mean	SD	mean	SD
before treatment	0.969	1.830	1.139	1.747	1.847	2.371
	(N = 16)		(N = 118)		(N = 59)	
after treatment	1.800	2.363	1.028	1.967	1.183	2.067
	(N = 25)		(N = 47)		(N = 41)	

Table 11: The descriptive statistics of the ratings of the fragments with opposition actors

To conclude the findings on the selective attribution by association, it can be said that both for the governmental and the opposition actors I fail to reject the null-hypotheses. I am not able to show any significant effect of the editorial change on this matter, so H4 is not supported by the findings of this study.

5.5 Spillover Effect and Imperfect Compliance

In the final part of the analysis chapter, I will examine whether there is evidence of any Spillover Effect and Imperfect Compliance in the quasi-experimental design. In order to make a causal claim, it is necessary to rule out the possibility that the effect was not caused by one of these.

I will first look at the possible presence of a spillover effect. This would mean that there is an indirect effect on a subject not directly treated by the experiment, such that the coverage of covid news significantly changed in the other major Hungarian outlets. It is not possible to control all relevant factors. First, I check the potential social interaction. A trivial spillover effect could be that journalists changed jobs. Those who were formerly employees of the index later started working for another portal So the possible difference could be that part of the effect is due to some of the index journalists leaving.

However, only one of the index's TV journalists signed up with a media outlet under study (444.hu) and published eight articles on the coronavirus between 28 October 2020 (the first article published by the new outlet) and 28 February 2021 (the end of the period under study). Compared to the 5700 articles published in the period under review, this is a negligible number to have a significant effect. I proxied the potential effect in two additional ways. I examined whether there was a change in the number of views of the pages or the volume of the articles published. I found no significant shift in either of these variables when comparing the time periods before and after the treatment. Based on these results, I assume that there was no significant spillover effect.

Finally, I examine the possibility of imperfect compliance. This would mean that part of the treatment group behaves as a "never taker". I would assume that there were articles in the index that did not touch on the changeover. However, since all members of the outgoing

editorial team left on 1 September, and the site's archive does not contain any more recent articles from the old authors, nor were they under any obligation to publish material from earlier, it could be argued that material produced after 1 September 2020 was already under the control of the new management.

6 Discussion

The lessons of the analysis are summarized in this discussion chapter. I argue that, although no evidence was found for most of the hypotheses, the observational experiment conducted in this thesis provided useful insights and conclusions about media capture in Hungary and the possibilities of quantitative analysis of media bias on a text-as-data basis.

There are several factors that can influence the media bias of a news outlet that cannot be measured with the dataset used for this study. The dataset only contained textual data, while most of the articles contained a cover photo and several other images within the text. However, pictures of a news article can highly influence the way people observe the covered topic (Rosenberg et al., 1986). In some cases, journalists can ardently manipulate images in their reports (Dearden, 2015; Estrin, 2014) Besides images, further factors that are overlooked by this thesis can be the placing of the news on the front page, the amount of time that it spent on it, or whether they shared the news on their social media site or not.

Another possible solution to measure media bias of the Covid coverage is not to focus on the positive or negative sentiment of events or specific actors, as it can be misleading or uninformative. In the case of Covid coverage, the same news can be positive or negative at different stages of the course of the pandemic. Instead of measuring the positive or negative tone of coverage, another possible solution can be to measure the tone or the framing of coverage and how close it is to different political actors' agenda and claim that if a news coverage tends to follow a political side's language then it is slanted to their direction. In that sense, it can be assumed that after the editorial change, index.hu would be more likely to use the same key phrases in their news, than what the government uses. The study of Gentkzow and

Shapiro (2010), which was already mentioned in the literature review, can be a good example of this approach.

The findings also highlighted that one crucial change caused by the takeover is the high increase in the number of articles. This can presumably be explained by the fact that the site is much less likely to publish its own analysis, but rather to publish press releases or to take over news from the wire service, and social media posts, i.e. to be much more involved in the news race. This also points to the fact that the categorization system used was not sensitive to the type of news the newspaper publishes or the journalistic interpretation (if any) it is given.

Overal, the finding of the study is in alignment with the political communication literature on modern autocratic systems and hybrid regimes, underlining the power's motivation not to just use the media for propaganda, but also for controlling the political marketing landscape as well.

6.1 Sample Size

An obvious weakness of the study that needs to be further highlighted is the low number of items in the control group. During the selection of the fragments and the titles I aimed to have a stratified random sample of the titles and fragments from the five different outlets, but unfortunately, it led to a very small experimental group, that only consisted of approximately 20% of the total cases. This has proven to be a mistake, that makes it harder to identify any effects of framing bias or the different forms of selective attribution.

7 Conclusion

This final chapter concludes this thesis study. First of all, I summarize the main findings and takeaway points of the research. Secondly, I drew attention to some of the potential limitations of the study regarding the generalization of the findings. Finally, I propose some possible directions for future research.

7.1 Summary of the Findings

This thesis analyzed the editorial change in the Hungarian news outlet Index, that happened in 2020. The motivation for this research is to understand more how this media capture works in informational autocracies, like Hungary. In an exploratory manner, the assumptions covered different types of forms of news manipulation, including censorship, framing bias, and two different forms of selective attribution. However, only framing bias gained support during the empirical analysis, while there is no clear evidence of censorship or any variation of selective attribution. The analysis also revealed a stable difference between the two sites that remained independent during the time of investigation and the other two that can be considered as propaganda. This finding implies that the applied measurements can be used to analyze media bias on news.

7.2 External Validity

Thinking about external validity, it can be said that analyzing index.hu is a case study, which means that the findings are not necessarily generalized to other cases of state capture in Hungary, or in a broader sense in hybrid regimes around the world. However, the case of Index fits into the trend that modern autocrats are likely to use more sophisticated ways of controlling media than being openly oppressive with the free media and using their publicity to spread hard

propaganda. In that regard, the lack of support for Assumptions 1, 3, and 4. can suggest not just the problem with sampling or with operationalizing the concepts of censorship, selective attribution, and selective attribution by association, but it can also imply that the case of index.hu is a novel form of government-controlled media.

7.3 Directions for Future Research

One possible way is to re-address the original research question, either with a bigger sampling or with a change in the methods. In this section, I would like to propose the possibilities of moving forward a more

This study relied on manually coded data. In a recent study, researchers ran an analysis using a sample of more than two thousand Twitter posts, and they found that the popular Large Language Models (LLM) application, ChatGPT outperforms crowd-workers across several annotation tasks, including assessing relevance, determining stance, identifying topics, and detecting frames. The zero-shot accuracy of the model surpassed crowd-workers in four out of five tasks, and the intercoder agreement outperformed both crowd-workers and trained annotators in all tasks. They also found that their model is twenty times more cost-effective than employing Amazon MTurk. These findings highlight the significant potential of LLMs in enhancing the efficiency of text classification (Gilardi et al., 2023).

However, a recent study has also highlighted that LLMs may not only be an alternative to human rating methods but may also influence the outcome of crowdsourcing models (Veselovsky et al., 2023). Crowd workers have financial incentives to use these models in order to improve their performance, and thus their income. To verify this assumption, Veselovsky et al. conducted a case study on the prevalence of LLM usage in crowdsourcing data collection. They gave an abstract summarization task on Amazon Mechanical Turk, and their analysis

came to the conclusion that an estimated 33–46% of the crowd workers used LLMs during the completion of their assigned task. Although generalization to other tasks is not clear, this problem can be considered as LLM friendly. Nevertheless, it can be assumed that in the future it will be even more important to consider not only whether human-sourced data is considered more reliable than automatically generated data, but also to check whether the input is indeed created by human judgment.

Another opportunity can be to mix the human-rated approach with machine learning and implement a supervised learning algorithm. It would mean that some part of the data would be rated by human coders, and later that would be used as train data to set up a classifier algorithm. If it produces satisfactory performance, then it can be used to categorize the unrated items of the data as well. Overall, the conclusions of this research along with the recent development in the field suggest that further research should rely more on LLMs and machine learning algorithms.

References

Aday, S. (2010). Chasing the Bad News: An Analysis of 2005 Iraq and Afghanistan War Coverage on NBC and Fox News Channel. *Journal of Communication*, *60*(1), 144–164. https://doi.org/10.1111/J.1460-2466.2009.01472.X

Barrett, A. W., & Peake, J. S. (2007). When the President Comes to Town Examining Local Newspaper Coverage of Domestic Presidential Travel. *American Politics Research*, *35*(1). https://doi.org/10.1177/1532673X06292816

Bátorfy, A., & Urbán, Á. (2019). State advertising as an instrument of transformation of the media market in Hungary. *Https://Doi.Org/10.1080/21599165.2019.1662398*, *36*(1), 44–65. https://doi.org/10.1080/21599165.2019.1662398

Baum, M. A., & Groeling, T. (2008). New Media and the Polarization of American Political Discourse. *Political Communication*, *25*(4), 345–365. https://doi.org/10.1080/10584600802426965

Baum, M. A., & Gussin, P. (2008). In the Eye of the Beholder: How Information Shortcuts Shape Individual Perceptions of Bias in the Media. *Quarterly Journal of Political Science*, *3*(1), 1–31. https://doi.org/10.1561/100.00007010

Bernhardt, D., Krasa, S., & Polborn, M. (2008). Political polarization and the electoral effects of media bias. *Journal of Public Economics*, *92*(5–6), 1092–1104. https://doi.org/10.1016/J.JPUBECO.2008.01.006

Bieber, F. (2018). Patterns of competitive authoritarianism in the Western Balkans. *East European Politics*, *34*(3), 337–354. https://doi.org/10.1080/21599165.2018.1490272

Bleyer-Simon, K. (2021). Government repression disguised as anti-disinformation action: Digital journalists’ perception of COVID-19 policies in Hungary. *Journal of Digital Media and Policy*, *12*(1), 159–176. https://doi.org/10.1386/JDMP_00053_1

Bogaards, M. (2018). De-democratization in Hungary: diffusely defective democracy. *Democratization*, *25*(8), 1481–1499. https://doi.org/10.1080/13510347.2018.1485015

Bozóki, A. (2011). Occupy the State: The Orbán Regime in Hungary. *Debatte: Journal of Contemporary Central and Eastern Europe*, *19*(3), 649–663. https://doi.org/10.1080/0965156X.2012.703415

Bozóki, A., & Hegedűs, D. (2018). An externally constrained hybrid regime: Hungary in the European Union. *Democratization*, *25*(7), 1173–1189. https://doi.org/10.1080/13510347.2018.1455664

Budak, C., Goel, S., & Rao, J. M. (2016). Fair and Balanced? Quantifying Media Bias through Crowdsourced Content Analysis. *Public Opinion Quarterly*, *80*(S1), 250–271. https://doi.org/10.1093/poq/nfw007

Cianetti, L., Dawson, J., & Hanley, S. (2018). Rethinking "democratic backsliding" in Central and Eastern Europe – looking beyond Hungary and Poland. *East European Politics*, *34*(3), 243–256. https://doi.org/10.1080/21599165.2018.1491401

Corruption Research Center Budapest. (2018, March 19). *The impact of Russia's state-run propaganda apparatus on online media in Hungary – 2010–2017*. http://www.crcb.eu/wp-content/uploads/2018/05/crcb_2017_mrsrpphnm_English_180319_.pdf

Covert, T. J. A., & Wasburn, P. C. (2007). Measuring Media Bias: A Content Analysis of Time and Newsweek Coverage of Domestic Social Issues, 1975–2000*. *Social Science Quarterly, 88*(3), 690–706. https://doi.org/10.1111/J.1540-6237.2007.00478.X

D'Alessio, D., & Allen, M. (2000). Media bias in presidential elections: a meta-analysis. *Journal of Communication*, *50*(4), 133–156. https://doi.org/10.1111/J.1460-2466.2000.TB02866.X

Dallmann, A., Lemmerich, F., Zoller, D., & Hotho, A. (2015). Media Bias in German Online Newspapers. *Proceedings of the 26th ACM Conference on Hypertext & Social Media - HT '15*, 133–137. https://doi.org/10.1145/2700171.2791057

Dearden, L. (2015, September 16). *The fake refugee images that are being used to distort public opinion on asylum seekers*. The Independent. https://www.independent.co.uk/news/world/europe/the-fake-refugee-images-that-are-being-used-to-distort-public-opinion-on-asylum-seekers-10503703.html

DellaVigna, S., & Kaplan, E. (2007). The Fox News Effect: Media Bias and Voting. *The Quarterly Journal of Economics*, *122*(3), 1187–1234. https://doi.org/10.1162/QJEC.122.3.1187

Direkt36. (2022). *Inside Viktor Orbán's war for Index, Hungary's most influential news website*. https://www.direkt36.hu/en/az-index-ostromanak-szinfalak-mogotti-tortenete/

Druckman, J. N., & Parkin, M. (2005). The impact of media bias: How editorial slant affects voters. *Journal of Politics*, *67*(4), 1030–1049. https://doi.org/10.1111/J.1468-2508.2005.00349.X/ASSET/IMAGES/LARGE/FIG1P.JPEG

Durante, R., & Knight, B. (2012). Partisan Control, Media Bias, and Viewer Responses: Evidence from Berlusconi's Italy. *Journal of the European Economic Association*, *10*(3), 451–481. https://doi.org/10.1111/J.1542-4774.2011.01060.X

Estrin, J. (2014, May 8). *The Real Story About the Wrong Photos in #BringBackOurGirls*. The New York Times. https://archive.nytimes.com/lens.blogs.nytimes.com/2014/05/08/the-real-story-about-the-wrong-photos-in-bringbackourgirls/

Eszter, K., Zoltán, K., & Renáta, N. (2021). A korrupció hazai online médiareprezentációjának vizsgálata természetes nyelvfeldolgozással. *Médiakutató*, *22*(2), 69–88.

Freedom House. (2022, April 20). *Nations in Transit 2022 - Hungary*. https://www.ecoi.net/en/document/2088055.html

Gentzkow, M., & Shapiro, J. M. (2010). What Drives Media Slant? Evidence From U.S. Daily Newspapers. *Econometrica*, *78*(1), 35–71. https://doi.org/10.3982/ECTA7195

Gerber, A. S., Karlan, D., & Bergan, D. (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. https://doi.org/10.1257/APP.1.2.35

Gertler, P. J., Martinez, S., Premand, P., Rawlings, L. B., & Vermeersch, C. M. J. (2016). *Impact evaluation in practice*. World Bank Publications.

Gilardi, F., Alizadeh, M., & Kubli, M. (2023). Chatgpt outperforms crowd-workers for text-annotation tasks. *ArXiv Preprint ArXiv:2303.15056*.

Griffen, S. (2020). Hungary: a lesson in media control. *British Journalism Review*, *31*(1), 57–62. https://doi.org/10.1177/0956474820910071

Groeling, T. (2013). Media Bias by the Numbers: Challenges and Opportunities in the Empirical Study of Partisan News. *Http://Dx.Doi.Org/10.1146/Annurev-Polisci-040811-115123*, *16*, 129–151. https://doi.org/10.1146/ANNUREV-POLISCI-040811-115123

Groeling, T., & Kernell, S. (1998). Is Network News Coverage of the President Biased? *The Journal of Politics*, *60*(4), 1063–1087. https://doi.org/10.2307/2647731

Grossman, G., Margalit, Y., & Mitts, T. (2022). How the Ultra-Rich Use Media Ownership as a Political Investment. *The Journal of Politics*. https://doi.org/10.1086/719415

Guess, A. M. (2021). (Almost) Everything in Moderation: New Evidence on Americans' Online Media Diets. *American Journal of Political Science*, *65*(4), 1007–1022. https://doi.org/10.1111/AJPS.12589

Hajnal, G., Jeziorska, I., & Kovács, É. M. (2021). Understanding drivers of illiberal entrenchment at critical junctures: institutional responses to COVID-19 in Hungary and Poland. *International Review of Administrative Sciences*, *87*(3), 612–630. https://doi.org/10.1177/0020852320981138

Hamborg, F., Donnay, K., & Gipp, B. (2019). Automated identification of media bias in news articles: an interdisciplinary literature review. *International Journal on Digital Libraries*, *20*(4), 391–415. https://doi.org/10.1007/s00799-018-0261-y

Hastorf, A. H., & Cantril, H. (1954). They saw a game; a case study. *Journal of Abnormal and Social Psychology*, *49*(1), 129–134. https://doi.org/10.1037/H0057880

Jansen, S. C., & Martin, B. (2015). The Streisand effect and censorship backfire. *Faculty of Law, Humanities and the Arts - Papers (Archive)*. https://ro.uow.edu.au/lhapapers/1884

Kahn, K. F., & Kenney, P. J. (2002). The Slant of the News: How Editorial Endorsements Influence Campaign Coverage and Citizens' Views of Candidates. *American Political Science Review*, *96*(2), 381–394. https://doi.org/10.1017/S0003055402000230

Levitsky, S., & Way, L. A. (2010). *Competitive Authoritarianism*. Cambridge University Press. https://doi.org/10.1017/CBO9780511781353

Máriás, L., Nagy, K., Polyák, G., & Urbán, Á. (2017). Soft censorship in Hungary. When propaganda rules public discourse. *Mérték Média Monitor*.

Mathieu, E., Ritchie, H., Rodés-Guirao, L., Appel, C., Giattino, C., Hasell, J., Macdonald, B., Dattani, S., Beltekian, D., Ortiz-Ospina, E., & Roser, M. (2020). Coronavirus Pandemic (COVID-19). *Our World in Data*.

McCarthy, J. D., McPhail, C., & Smith, J. (1996). Images of Protest: Dimensions of Selection Bias in Media Coverage of Washington Demonstrations, 1982 and 1991. *American Sociological Review*, *61*(3), 478. https://doi.org/10.2307/2096360

Média1. (2020, November 18). *„Ugyanott, ugyanúgy" – Kampányt indított az Index*. https://media1.hu/2020/11/18/ugyanott-ugyanugy-kampanyt-inditott-az-index/

Morris, J. S., & Francia, P. L. (2010). Cable News, Public Opinion, and the 2004 Party Conventions. *Political Research Quarterly*, *63*(4), 834–849. https://doi.org/10.1177/1065912909338463

Niven, D. (2002). *Tilt?: The search for media bias*. Praeger. https://books.google.com/books?hl=hu&lr=&id=cm4l0pryUbUC&oi=fnd&pg=PR7&dq=Niven+D.+2002.+Tilt%3F+The+Search+for+Media+Bias.+Westport,+CT:+Praeger&ots=j0P9heZXZk&sig=bGuys39TJZiOz8W7G7vz6yWf1IA

Niven, D. (2003). Objective evidence on media bias: Newspaper coverage of congressional party switchers. *Journalism and Mass Communication Quarterly*, *80*(2), 311–326. https://doi.org/10.1177/107769900308000206

NMHH. (2021). *Internetes közönségmérési adatok (2020. IV. negyedév) • National Media and Infocommunications Authority*. https://english.nmhh.hu/article/218350/Internetes_kozonsegmeresi_adatok_2020_IV_negyedev

Polyák, G. (2019). Media in Hungary: Three Pillars of an Illiberal Democracy. *Public Service Broadcasting and Media Systems in Troubled European Democracies*, 279–303. https://doi.org/10.1007/978-3-030-02710-0_13

Rosenberg, S. W., Bohan, L., McCafferty, P., & Harris, K. (1986). The Image and the Vote: The Effect of Candidate Presentation on Voter Preference. *American Journal of Political Science*, *30*(1), 108. https://doi.org/10.2307/2111296

Rozenas, A., & Stukal, D. (2019). How autocrats manipulate economic news: Evidence from Russia's state-controlled television. *Journal of Politics*, *81*(3), 982–996. https://doi.org/10.1086/703208/ASSET/IMAGES/LARGE/FG7.JPEG

Schiffer, A. J. (2006). Assessing partisan bias in political news: The case(s) of local senate election coverage. *Political Communication*, *23*(1), 23–39. https://doi.org/10.1080/10584600500476981

Serdült, V. (2020). Inside story: Hungary's media silence: What's it like working as a journalist under the new rules introduced by Hungary's Viktor Orbán? How hard is it to report? *Index on Censorship*, *49*(2), 64–66. https://doi.org/10.1177/0306422020935806

Simonovits, G., & Vig, Á. (2023). News sharing as a measure of media alignment. *Journal of Quantitative Description: Digital Media*, *3*. https://doi.org/10.51685/jqd.2023.010

Sipos, B. (2013). Bias, partisanship, journalistic norms and ethical problems in the contemporary Hungarian political media. *Central European Journal of Communication*, *6*(1), 89–104.

Stefanov, P., Darwish, K., Atanasov, A., & Nakov, P. (2020). Predicting the Topical Stance and Political Leaning of Media using Tweets. *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, 527–537. https://doi.org/10.18653/V1/2020.ACL-MAIN.50

Szabó, Á. (2020, December 26). *Beadták az első COVID-19 elleni vakcinát a Semmelweis Egyetemen*. Semmelweis Hírek.

Szabó, G., & Bene, M. (2016). Széttöredezett vagy összekapcsolódó? A magyar médianyilvánosság hálózatszerkezete három eset tükrében. *Politikatudományi Szemle*, *25*(3.), 33–58.

Szeidl, A., & Szucs, F. (2021). Media Capture Through Favor Exchange. *Econometrica*, *89*(1), 281–310. https://doi.org/10.3982/ECTA15641

Tamas, G. M. (2020, July 25). Index! *Mérce*. https://merce.hu/2020/07/25/tgm-index/

Tóth, J. (2021). Negative and Engaged: Sentiments towards the 2016 Migrant Quota Referendum in Hungarian Online Media. *East European Politics and Societies*, *35*(02), 493–518.

Turner, J. (2007). The messenger overwhelming the message: Ideological cues and perceptions of bias in television news. *Political Behavior*, *29*(4), 441–464. https://doi.org/10.1007/S11109-007-9031-Z/FIGURES/1

Vachudová, M. A., & Snyder, T. (1997). Are transitions transitory? Two types of political change in Eastern Europe since 1989. *East European Politics and Societies*, *11*(1), 1–34. https://doi.org/10.1177/0888325497011001001/ASSET/0888325497011001001.FP.PNG_V03

Vallone, R. P., Ross, L., & Lepper, M. R. (1985). The hostile media phenomenon: Biased perception and perceptions of media bias in coverage of the Beirut massacre. *Journal of Personality and Social Psychology*, *49*(3), 577–585. https://doi.org/10.1037/0022-3514.49.3.577

Veselovsky, V., Horta Ribeiro, M., & West EPFL, R. (2023). Artificial Artificial Artificial Intelligence: Crowd Workers Widely Use Large Language Models for Text Production Tasks. *ArXiv Preprint*. https://github.com/epfl-dlab/GPTurk

Waldman, P., & Devitt, J. (1998). Newspaper Photographs and the 1996 Presidential Election: The Question of Bias. *Journalism & Mass Communication Quarterly*, *75*(2), 302–311. https://doi.org/10.1177/107769909807500206

Appendices

Appendix A - Replication Code

Please find the raw data and the Python code that was used for data processing in the following repository:

https://github.com/balla-a/ceu-ma-thesis

Appendix B – Coding Questionnaire

Original information and consent form in Hungarian

Kedves Kitöltő!

Felkérjük Önt, hogy vegyen részt egy kérdőíves kutatásban, amelyet a CEU Politikatudományi Tanszéke végez. A tanulmányt Balla Attila, a politikatudomány MA program hallgatója készíti szakdolgozati projektjének keretében Simonovits Gábor vezetésével.

A tanulmány célja: A kutatás azt vizsgálja, hogy a magyar online média hogyan tudósított a COVID-19 világjárványról, és hogyan ábrázolta a különböző politikai szereplőket.

Eljárás: Politikai szereplőkkel foglalkozó címeket (összesen 15 darabot) és rövid hírrészleteket (összesen 10 darabot) kell majd elolvasnia. A feladata az lesz, hogy ítélje meg őket megadott szempontok szerint. A vizsgálat körülbelül 15 percet vesz igénybe. Ha úgy dönt, hogy részt vesz a vizsgálatban, részletesebb és konkrétabb utasításokat fog kapni a későbbiekben.

Potenciális kockázatok: A kísérletben való részvétellel járó esetleges kockázatok megegyeznek a számítógép normál hétköznapi körülmények közötti használatának kockázataival.

Potenciális előnyök: Ha a felmérés végén megadja e-mail címét vagy telefonszámát, részt vesz egy sorsoláson, amelyen különböző értékű nyereményeket nyerhet (1 db 50 ezer Ft, 3 db 10 ezer Ft és 10 db 5 ezer Ft értékű Líra könyvutalvány kerül kiosztásra, tehát összesen 14 nyeremén 130 ezer Ft összértékben). A sorsolásra az adatgyűjtés befejezése után kerül sor (legkésőbb 2023. június 2-án), ami után a megadott kapcsolattartási adatok törlésre kerülnek.

Az Ön jogai résztvevőként: A vizsgálatban való részvétel önkéntes. Ön dönthet úgy, hogy nem kíván részt venni, és a vizsgálat során bármikor indoklás nélkül visszaléphet. A hozzájárulás visszavonása a felmérés befejezése után nem lehetséges, mivel az adatok anonimizálásra kerülnek. A hozzájárulás visszavonása a sorsolástól az adatgyűjtés után sorra kerülő sorsolás előtt lehetséges.

Adattárolás és adatvédelem: Személyes adataihoz nem férünk hozzá, csak az elérhetőségi címéhez, ha úgy dönt, hogy részt vesz a sorsoláson. Csak a feladat szempontjából releváns adatokat rögzítjük (például, hogy Ön hogyan értékelte a mondatokat, vagy mikor). Amint a nyertesek átvették a nyereményeket, minden elérhetőségi adatot véglegesen törölünk.

Kérdések: Fontos számunkra, hogy Ön minden szükséges információt megkapjon ahhoz, hogy eldönthesse, részt vesz-e a vizsgálatban vagy sem. Kérjük, bátran tegyen fel további kérdéseket a vizsgálattal kapcsolatban. Balla Attilához (Balla_Attila@student.ceu.ed) vagy Simonovits Gáborhoz (SimonovitsG@ceu.edu) fordulhat.

A kérdőív az alábbi linken érhető el, amire kattintva kijelenti, hogy elolvasta és egyetért a feltételekkel, valamint hozzájárul a kutatásban való részvételhez:

Amennyiben nem szeretne részt venni, úgy nincs egyéb teendője, mint bezárni ezt az oldalt. Semmilyen adata nem került rögzítésre.

Example questions in Hungarian:

1. cím a 15-ből

Kérem, olvassa el a következő hír címét:

Elmaradnak a tőzsdei társaságok éves rendes közgyűlései

Milyen az olvasott címnek a meghatározó hangulata?

- Pozitív.
- Inkább pozitív.
- Semleges.
- Inkább negatív.
- Negatív.

---

6. hírrészlet a 10-ből

Kérem, olvassa el a következő hírrészletet:

A háziorvosok munkájának megbecsülése mellett ezzel is az a cél, hogy felgyorsítsuk az oltást, és húsvétra a regisztráltak megkapják az első oltásukat, hiszen a járványból a védőoltás jelenti a kivezető utat - írja az Emberi Erőforrások Minisztériuma közleményében. Az oltási munkacsoport most is jelentős feladatokat bíz a háziorvosokra.

Milyen az olvasott hírrészletnek a meghatározó hangulata?

- Pozitív.
- Inkább pozitív.
- Semleges.
- Inkább negatív.
- Negatív.

Az Emberi Erőforrások Minisztériuma az előbbi hírrészletben:

- Pozitív színben van feltűntetve.
- Inkább pozitív színben van feltűntetve.
- Semleges színben van feltűntetve.
- Inkább negatív színben van feltűntetve.
- Negatív színben van feltűntetve.

Translation of the information and consent form in English

Dear Participant!

You are invited to participate in a survey conducted by the CEU Department of Political Science. The study is being conducted by Attila Balla, a student of the MA in Political Science program, under the supervision of Gábor Simonovits.

The aim of the study: The study investigates how the Hungarian online media covered the COVID-19 pandemic and how the different political actors were portrayed.

Method: You will be asked to read headlines (15 in total) and short news clips (10 in total) dealing with political actors. Your task will be to judge them according to the given criteria. The test will take about 15 minutes. If you decide to take the test, you will receive more detailed and specific instructions later on.

Potential risks: the potential risks of taking part in the trial are the same as those of using a computer under normal everyday conditions.

Potential benefits: If you provide your email address or phone number at the end of the survey, you will be entered into a draw to win prizes of different values (1 book voucher worth HUF 50,000, 3 book vouchers worth HUF 10,000, and 10 book vouchers worth HUF 5,000, for a total of 14 prizes worth HUF 130,000). The draw will take place after the data collection is completed (no later than 2 June 2023), after which the contact details provided will be deleted.

Example questions' English translation:

Title 1 of 15

Please read the title of the following news:

Annual general meetings of listed companies cancelled

What is the prevailing mood of the title you are reading?

- Positive.
- Rather positive.
- Neutral.
- Rather negative.
- Negative.

---

6 of 10 news items

Please read the following news item:

While appreciating the work of family doctors, this is also aimed at speeding up vaccination and ensuring that registered people receive their first vaccine by Easter, as vaccination is the way out of the epidemic, the Ministry of Human Resources said in a statement. The vaccination task force is now entrusting GPs with important tasks.

What is the dominant mood of the news item you read?

- Positive.
- Rather positive.
- Neutral.
- More negative.
- Negative.

The Ministry of Human Resources in the news item above:

- It is presented in a positive light.
- It's presented in a rather positive light.
- It is presented in a neutral light.
- It's presented in a rather negative light.
- It is presented in a negative light.

Appendix C – Difference-in-Difference analyses

The DiD analysis of framing bias

	coef	std-err	t	P>|t|	[0.025	0.975]
Intercept	2.728	0.021	127.383	0.000	2.686	2.770
treatment	-0.346	0.082	-4.234	0.000	-0.505	-0.186
post_period	-0.166	0.037	-4.439	0.000	-0.239	-0.093
treatment_post	0.227	0.102	2.231	0.026	0.027	0.426
R-squared:	0.008					
No. Obs.	5808					
Log-Likelihood:	-9487.2					

The DiD analysis of selective attribution for the governmental actors

	coef	std-err	t	P>|t|	[0.025	0.975]
Intercept	3.745	0.085	44.274	0.000	3.579	3.911
treatment	0.026	0.305	0.085	0.933	-0.573	0.624
post_period	-0.091	0.132	-0.692	0.489	-0.349	0.167
treatment_post	-0.069	0.367	-0.188	0.851	-0.788	0.650
R-squared:	0.001					
No. Obs.	1428					
Log-Likelihood:	-3182.9					

The DiD analysis of selective attribution for the opposition actors

	coef	std-err	t	P>|t|	[0.025	0.975]
Intercept	2.561	0.186	13.737	0.000	2.194	2.928
treatment	-0.030	0.647	-0.046	0.964	-1.304	1.245
post_period	-0.402	0.323	-1.242	0.215	-1.038	0.235
treatment_post	1.650	0.857	1.925	0.055	-0.037	3.338
R-squared:	0.027					
No. Obs.	306					
Log-Likelihood:	-710.11					

The DiD analysis of selective attribution by association for the governmental actors

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	2.023	0.086	23.469	0.000	1.854	2.192
treatment	0.257	0.311	0.827	0.408	-0.353	0.867
post_period	-0.136	0.134	-1.013	0.311	-0.399	0.127
treatment_post	0.301	0.374	0.805	0.421	-0.432	1.033
R-squared:	0.006					
No. Obs.	1428					
Log Likelihood:	-3209.4					

The DiD analysis of selective attribution by association for the opposition actors

	coef	std-err	t	P>|t|	[0.025	0.975]
Intercept	1.375	0.152	9.042	0.000	1.076	1.675
treatment	-0.407	0.528	-0.770	0.442	-1.446	0.633
post_period	-0.275	0.264	-1.041	0.299	-0.794	0.245
treatment_post	1.106	0.700	1.581	0.115	-0.270	2.483
R-squared:	0.010					
No. Obs.	306					
Log Likelihood:	-647.860					