NORMALIZATION AND DIFFERENTIATION IN GOOGLE NEWS: A MULTI-METHOD ANALYSIS OF THE WORLD'S LARGEST NEWS AGGREGATOR

By

QUN WANG

A dissertation submitted to the

School of Graduate Studies

Rutgers, The State University of New Jersey

In partial fulfillment of the requirements

For the degree of

Doctor of Philosophy

Graduate Program in Communication, Information and Media

Written under the direction of

Susan Keith

And approved by

_____

_____

_____

_____

New Brunswick, New Jersey

January 2020

ABSTRACT OF THE DISSERTATION

NORMALIZATION AND DIFFERENTIATION IN GOOGLE NEWS:

A MULTI-METHOD ANALYSIS OF THE WORLD'S LARGEST NEWS AGGREGATOR

by QUN WANG

Dissertation Director:

Susan Keith

This dissertation examines the history, evolution, and influence of Google News—Google's news aggregation service—from the late 1990s to 2019. There are scattered studies about Google News, but no systematic and substantive research on what is, at this writing, its nearly two decades of existence. Those two decades happened to be a period of time that witnessed the intense digital transformation of the media industry and our society. Both the object of the study and the timeframe this study examines are significant to understand the changing media and technology environment in the 21st century. Drawing on normalization and differentiation theoretical frameworks, this dissertation combines traditional research methods and computational approaches to conduct historical research, web archival analysis, legal analysis, algorithm analysis, and more. In six chapters, this dissertation traces the origin and early history of Google News; the structural, visual, and functional evolution on the Google News homepage design

since its launch in 2002; disputes about Google and its news aggregation service in different parts of the world; Google's news-related technologies and algorithms; and Google's systematic initiatives in the news area and their influences. Based on the analysis of the normalization and differentiation trends and the driving forces behind these trends, this study proposes an N-D-N theoretical model that conceptualizes Google's development in the news area and the interaction between Google and the news industry. The dissertation concludes with a discussion about the implications of the N-D-N model for policymaking on platform governance and the future of journalism. Using Google News as a case study, this dissertation provides a snapshot of the changing media landscape in the digital era. It also makes theoretical and empirical contributions to the ongoing conversation about the interrelationship between digital platforms and the traditional media industry.

Table of contents

Chapter 1. Introduction and Theoretical Frameworks

Introduction

As one of the world's leading technology companies, Google owns a wide range of products. Among the earliest and most long-lasting is Google News (Shiels, 2018), Google's news aggregation service and the central component of Google's news business. As a digital news aggregator, Google News does not produce news itself. Instead, it aggregates news from different sources, presenting news to users through a mostly automatic process based on Google's computational algorithms. Since its launch in 2002, Google News—the world's largest news aggregator—has had profound impact on the news media industry and people's everyday lives. As of January 2019, Google News covers more than 80,000 news publishers around the world (Google, 2019). Google reports that Google News contributes more than 10 billion clicks per month to publishers' websites (Schindler, 2018). By 2018, Google News had approximately 150 million unique monthly visitors in the U.S. (Helmore, 2019).

Google News is also one of the most controversial of Google's products. As it grew into a major actor in the early 21st century media ecosystem, tensions between Google News and traditional news media escalated around the world. For example, in the United Kingdom, Rupert Murdoch, the founder of News Corp., publicly accused Google and other technology companies of "stealing," for they "simply just pick up everything and run with it, steal our stories-—we say they steal our stories—they just take them" (Johnson, 2009, para. 12). In the United States, at a 2010 Federal Communications Commission event "The Future of Media and Information Needs of Communities: Serving the Public Interest in the Digital Era," news executives called news aggregators

associated with search engines "the enemy." For example, Associated Press General Counsel Srinandan Kasi argued that their practices threatened the value of original reporting by manipulating the monetization of the news content (Anderson, 2013).

The development of Google News and the tension between the "older" and "newer" media sectors reveal complex dynamics in the media and technology ecosystem in the digital era. To better understand these dynamics and the forces that drive these dynamics, this dissertation explores Google News in terms of its history, evolution, and influence, as well as Google's news business evolving around it. There are a few studies about Google News (See Das, Datar, Garg, & Rajaram, 2007; Schroeder & Kralemann, 2005; Weaver & Bimber, 2008) but no systematic and substantive examination of its seventeen years of existence since its launch in 2002. These years happen to be a period that has witnessed the most intense digital transformation of the media industry and our societies. Therefore, both the object and the timeframe that this dissertation examines are important for understanding the negotiations and relationships that shape the changing media landscape.

This study of Google News is conducted against the backdrop of the ongoing conversation about digital platforms, in terms of their roles, influence, and regulation (*e.g.* Bell & Albright, 2018; Gillespie, 2017; Newman, Fletcher, Kalogeropoulos, Levy, & Niesen, 2018). The debate in the regulatory domain is especially heated surrounding the questions of whether digital platforms should be governed and, if so, how. To answer these questions, one needs to first be clear about the answers to a series of related questions: What are these digital platforms? How do they work? Are they technology companies or media companies? The question about the nature of digital platforms is

especially puzzling: many scholars and observers point out that platforms strategically position themselves and construct the public discourse for economic, political, and legal considerations (Bogust, 2016; Gillespie, 2010; Napoli & Caplan, 2017). Digital platforms define themselves as technology companies as this identity makes them appealing to the investment community and helps them avoid legal and regulatory constraints, but they play the role of the media companies because of their power in determining what information the public receives through algorithmic gatekeeping and their close relationship with the media industry (Baram, 2017; Napoli & Caplan, 2017; Napoli, 2015).

The identity question matters because it defines the public's perception of digital platforms. It also determines the rationales underlying law and policymakers' decision-making. Studying Google News provides a timely opportunity to contribute to the ongoing conversation in terms of the role of digital platforms and their implications for journalism, democracy, and policymaking. In this dissertation, "What is Google News?" also serves as a driving question that runs throughout the study as it examines how Google News emerged, how it works and has evolved, and how it wields influences.

"What is Google News?" is a question with no simple answers. As the name implies, Google News is a hybrid of "Google"—a technology giant and a representative of "new" media actors—and "news," the product traditionally created by mass media, "old" media actors. The hybrid nature of Google News determines that it carries genes from both sides, giving the news aggregator interesting characteristics that are both similar to and different from either sector. To understand how the genes are edited in Google News' DNA in terms of its operations, functionalities, and roles, this dissertation

synthesizes the theoretical frameworks of normalization and differentiation to examine the co-existence and complex interplay of the similarities and the difference that Google News absorbs from both the media and the technology sectors.

Online News Aggregation

What is News Aggregation

Generally speaking, aggregation is the practice of "collecting information from various sources and piecing it together into a (hopefully, more or less) coherent whole" (Schweigert, 2012, para. 1). Aggregation as a news-making practice has a history as long as the history of reporting—from the reprinted letters and columns widely adopted in early American newspapers in the eighteenth and early nineteenth centuries to legacy news media's common practices in the 21st century, such as round-up newsletters and summaries of trending stories on social media, aggregation has become a "staple" of news and newswork (Coddington, 2019, location 254). In addition to traditional news organizations that use news aggregation techniques in their news-making process, in the 21st century news ecosystem, there is a group of young, small organizations who base their business primarily on news aggregation—these organizations manually produce second-hand stories based on existing news stories through practices such as selecting, editing, and remixing; and they distribute these second-hand stories via varying platforms and approaches, such as mobile apps, social media, and newsletter, for profit (Coddington, 2019).

Online news aggregation studied in this section, while sharing some of the characteristics described above, is a particular practice carried out in the digital context. Kimberley Isbell (2010), an expert in policy and international affairs at the U.S.

Copyright Office, who also worked with the Citizen Media Law Project, defined it as the practice of a digital enterprise that "takes information from multiple sources and displays it in a single place" (p. 2). In the first decade of the 2000s, news aggregators were also called "news readers", "feed reader" or "RSS aggregator" to refer to the practice of "a website or application that collates feeds into a customised newspaper/home page (Isbell, 2010, p. 22), or redistributing news content from different established news outlets on a single website (Lee and Chyi, 2015, p. 5).

According to existing studies, news aggregation in different contexts shares some common characteristics. First, news aggregation involves little, if at all, original content production. Content is usually gathered through practices such as "taking," "collating," or "collecting" existing news from various sources. Second, the final product of news aggregation is usually a collection of news, which characterizes the wholeness of the collection, such as the ideas of "a single place," "a single website," or "a coherent whole". Third, the process of news aggregation involves such practices as reading, manually or digitally; judgment making, such as "collating," which, according to Merriam-Webster dictionary, involves critical comparison, careful verification, and rational arrangement in proper order; assembling, such as "piecing it together"; and customizing, repurposing, and redistributing.

Through aggregation, news content is unbundled and rebundled. A news aggregator unbundles content as it de-aggregates prepackaged news and reshuffles individual pieces of news across multiple news sources. As a result, the previous arrangement of news is disorganized. The news aggregator also rebundles news because it selectively gathers news from different sources and presents the collected news as a

whole in a new order. Existing definitions of news aggregation, however, leave many questions unanswered in the process of debudling and rebundling, for example, the decision-making, whether its human- or technology-based, behind these processes is not explained; and the relationships between aggregator and sources are not addressed.

Who are News Aggregators

These unaddressed questions in defining news aggregation may be due to the situation that there are different types of online aggregators as well as different criteria to categorize the news aggregators. Isbell (2010) identifies four types of news aggregators based on news source: feed aggregators, such as Google News and Yahoo News, are those that collect news from news feeds across different news sources and topics; specialty aggregators, such as Taegan Goddard's Political Wire, that collect news surrounding a particular theme; user-curated aggregators, such as digg.com (sometimes called social news website that allows users to share and vote on news, http://digg.com/), that aggregate news provided by users and usually cover a wider variety of sources; and blog aggregators, such as Huffington Post and Gawker, that repurpose third-party content into content for a blog site. In 2019, however, when this dissertation is written, some of these categorizations seem outdated, e.g. Gawker no longer exists while Huffington Post's role is beyond a blog aggregator. At the same time, more players have entered the news aggregation market, including Apple News, Bing News, DuckDuckGo, and Facebook News. International players, news publishers, and news aggregation startups have also joined the news aggregation market (Facebook News, 2019).

News aggregators apply different methods to manage news. Generally speaking, there are human-based and machine-based aggregators. While some aggregators, such as

Google News, depend on automated, computational algorithms to gather, sort and rank news, others, such as Yahoo News, combines human editors and algorithms to do news aggregation (Lee & Chyi, 2015). Sasseen and colleagues (2013) divided news aggregators into two categories based on the degree of originality, i.e. those aggregators, such as Google News, The Examiner, Topix, and Bing News, that depend on automation without producing original content are defined as "pure news aggregators". News aggregators, such as Yahoo News, AOL News, and Huffington Post, that mix automated aggregation and original reporting are hybrid news aggregators. Aggregators can also be distinguished by different aggregation techniques, such as aggregators that provide single stream and those that provide single page aggregation, i.e. multiple sources are combined into a single stream or a single (web)page; aggregators that allow meme aggregation, i.e. news is sorted based on certain criteria, such as popularity, keywords, etc.; people powered aggregators, i.e. users are invited to submit and vote on news; and more (Catone, 2007).

Some news aggregators are associated with portals or search engines – such as Yahoo News, Google News, and Bing News – while others, such as Huffington Post, are not. Also, some news aggregators are platform-specific, for example, Pulse is designed for Android mobile devices. Others are community-specific, e.g. some local news aggregators serve a particular community, for example, there are over one hundred local news aggregators in 20 Dutch local communities (Bakker, 2012). More broadly speaking, social media websites, such as Facebook and Twitter, play the role of news aggregators too. Social media websites compile trending news from different sources through algorithms. They also provide a venue for users to share content from different news

organizations. By doing so, news can be socially aggregated by individual users of these platforms through following, reposting, liking and recommending, which Singer (2014) has termed “secondary gatekeeping”.

The U.S. Aggregator Landscape

Nationally, leading online news aggregators such as Yahoo News, Google News, and the Huffington Post share the U.S. news market with legacy news media. For example, a study from the Pew Research Center showed that among the top 25 online news providers in the U.S. in 2011, at least seven of them were news aggregators (Sasseen, Olmstead, & Mitchell, 2013). Studies in 2013 and 2015 showed that Yahoo News, Google News, and the Huffington Post are among the most popular news websites (Sasseen, Olmstead, & Mitchell, 2013, Lee & Chyi, 2015). On mobile platforms, Apple Inc. introduced Apple News in 2015, a news aggregator mobile app that is developed for Apple’s operating systems, but Apple News has a focus on magazines more than newspapers (Owen, 2017). Other news aggregator apps, such as Pulse, Zite, News.me, are growing too (Bakker, 2012), but controversies come along, for example, Pulse, a news aggregator app released in 2010 became the best-selling iPad RSS app in that year, but soon encountered the New York Times’ objection of using its RSS feed (Kiss, 2010).

Google News as the World’s Largest News Aggregator

Different ranking systems and individual studies produced different results. There is almost no way to *quantify* which news aggregator is the “largest”. In this study, Google News is considered the world’s largest news aggregator for the following reasons. First, Google News has been one of the most prominent news aggregators across different ranking criteria and at different times. Data from 2018 shows that Google News had

about 150 million unique monthly visitors in the U.S., almost double the numbers for CNN and the New York Times combined (Helmore, 2019). Globally, Google News reported in September 2018 that its service covered 127 countries and 65 languages worldwide (Stier, 2018). As of January 2019, Google News indexed more than 80,000 news publishers around the world (Google, 2019). Put together, Google News leads the news aggregation market in multiple areas. Second, among the "pure aggregators", i.e., those that focus on news aggregation only without producing any original content, Google News is the largest with the longest history. Google News was launched in 2002, much earlier than other large news aggregators, such as Bing News launched in 2008 and Apple News rolled out in 2015. Third, Google News is associated with Google, the world's largest search engine (alexa.com). Google has overwhelmingly dominant market share on the global search market (Clement, 2019). This study has a focus on the interaction between Google News and the news industry. Google News became the major target of the news industry largely because its dual role as a news aggregator-search engine (e.g. Anderson, 2013). As unfolded later in this study, Google New, both its practices and influences, cannot be considered separately from Google. Thus, combining quantitative and qualitative measures, this study argues that Google News can be systematically considered the world's largest news aggregator.

Google News: An Algorithm-Driven News Aggregator-Search Engine

So far, Google News has three characteristics that contribute to its uniqueness and its controversial-ness: 1). Google News functions based on computational algorithms. 2). Google News is the world's largest news aggregator. 3). It is associated with Google, the world's leading search engine, and Google News itself has search functions.

Computational algorithms are often distinguished from human approaches. In the case of news, since computational algorithms are automatic and free from human editorial interventions, they are sometimes considered bias-free. Some scholars, however, argue that new media players, such as online news aggregators, control what news the public is exposed to (Vos and Heinderyckx, 2015). Like the journalistic gatekeeping, which controls the process of transforming "countless bits of information into the limited number of messages that reach people each day" (Shoemaker & Vos, 2009, p. 1), digital media's practices involve similar decision-making and judgment making. Therefore, they share the news media's gatekeeping role through "algorithmic gatekeeping." (Napoli, 2015) In the same vein, search engines are considered the primary gatekeeper of online content as well (Hargittai, 2007; Introna & Nissenbaum, 2000). "Search engines are media companies," argued Goldman (2005, p. 189). Empirical studies found evidence to support such statement. For example, search engines in different countries are found influenced by Western perspectives in a similar way to newspapers. In fact, Yahoo News is found significantly less likely to cover developing countries than newspaper sites (Watanabe, 2013). Search engines in different political and economic systems are also found return different search results with little overlap, indicating biased operation in different contexts (Jiang, 2014).

Some scholars pointed out that algorithm, designed by human logic, already carries bias. For example, PageRank, the core algorithm of Google to sort and rank information, is criticized to bias toward popular, authoritative sites (Granka, 2010). PageRank makes judgment based on the link structures on the web. In PageRank algorithm, a site's inbound links are treated as votes for this site. These links, however,

are not of equal worth: links from popular sites have higher ranking scores while links from less popular sites have lower scores. Scholars pointed out that the popularity and authoritativeness of sites are often determined by economic power, i.e., sites having more economic power are able to use more resources for marketing and prominence, and will, these scholars were concerned, lead to a self-reinforcing process (Goldman, 2005) or the "rich-get-richer" phenomenon (Cho, Roy & Adams, 2005) encouraging power imbalances online and offline. Defenders of search engines respond that popularity and quality cannot be separated. Popular sites require effort, time, and skills they maintain (Potts, 2007). From a technical point of view, search engines are not only concerned about keywords and links, their function also depends on components such as metadata and anchor texts (Granka, 2010). Granka also argues that criticisms of search engine bias are often based on observations at aggregate level. Studies are needed to examine search engines' role in different contexts, for example, whether search engine actually benefits the long tail over the head in particular cases.

Unclear Google News Effects

The birth and growth of Google News occur against the backdrop of the decline of the traditional news industry. Newspaper circulation, advertising revenue, and investment in journalism have fallen since the turn of the 21st century (Edmond, 2015). While news aggregators associated with search engine was considered "enemy" of the traditional news industry, so far, there are only scattered studies about Google News, but no systematic research that studies its almost two decades of existence. These existing studies provided mixed, sometimes even contradictory, findings in terms of its effects on traditional news media.

Existing empirical studies revealed unclear and conditional Google News effects on traditional news media. For example, from January 2010 to February 2010, news content of The Associated Press was shortly removed from Google News due to a contractual dispute. Chiou and Tucker (2011) use this incident as a natural experiment to investigate the effect of Google News before and after the removal. The authors compared the user behavior of Google News that discontinued hosting the content of The AP during the contractual dispute and that of Yahoo News that continued to host the AP's content. The study found that when compared to Yahoo News users, visits to news websites among Google News users declined during the removal. The study thus suggested Google News benefits news organizations by bringing them more visits, because users were more likely to seek out for further and in-depth information after they saw the news on Google News. Similarly, Calzada and Gil (2016) found that news site visits declined after Google News shut down its service in Spain in December 2014 and after Axel Springer, the leading publisher of Germany's VG Media chose to opt-out of Google News service during October and November 2014. Calzada and Gil found that news site visits decreased in various ways in both cases, including direct visits, search visits, referrals, and socials.

There was also a debate about whether Google News has a competitive or non-competitive relationship with newspaper web sites. Jeon and Nasr's (2016) study found news aggregators can promote traffic to newspaper web sites, but the study revealed both "business-stealing" and "readership-expansion" effects. The authors argued that on the one hand news aggregators seem to "steal" some homepage consumption from newspapers when users use news aggregators as the starting point of their news

consumption; on the other hand, news aggregators increase readership at the article level, especially for high-quality articles. To determine whether news aggregators "steal" the profit of news organizations, George and Hogendorn's (2012) study pointed out that news aggregators may influence news organizations' profit not by direct "stealing" but by affecting users' advertisement demands. News aggregators produce the multi-homing effect, i.e. users consume news across different news sources rather than being loyalty to a fixed news brand. The multi-homing effect lowers the efficiency of the advertiser-consumer match on newspaper websites, which may reduce users' overall demands for advertising. This effect, this study argued, has strong effect on larger media outlets' advertising revenue, but this effect is also related to user characteristics, the nature of their media experience, as well as the type of advertisers, their products, and their marketing strategies. Huang and colleagues (2013) studied news aggregators and their impacts on newspaper sites in Taiwan. Their study indicates non-competitive relationship between news aggregators and online newspapers. But findings also suggest that news aggregators could have a powerful competitive replacement effect to news organizations when the given news aggregator's penetration and market share reach a very high point as in the case of Yahoo News in Taiwan (Yahoo News reaches 77.8% of the respondents in this study).

Scholars also have different observations regarding whether or to what extent Google News is biased against diversity. Schroeder and Kralemann (2005) find Google News promoted diversity, especially transnational perspectives, by including news sources worldwide. Watanabe (2013), however, pointed out that Western perspectives still dominate news aggregators, since even though Google News covers a large number

of news sources all over the world, there is a high degree of concentration in terms of the distribution of these sources on Google News. Carlson (2007) was concerned that Google News regularly recycles content from mainstream news wires. He held that when Google News links to news sources that commonly share the content of these dominant news agencies, it promotes the influence of these news wires.

Moreover, some studies found Google News has different impacts on different news genres and news outlets of different sizes. For example, the study of Calzada and Gil (2016) found larger decline for sports outlets after the Google News shutdown in Spain, but the shutdown seemed to have no significant effect on business outlets. Similarly, Athey, Mobius and Pál (2017) found users consumed less breaking news, hard news, and scarcity news (news that were not widely covered) after the shutdown, indicating that Google News may have different impacts on different news genres. While some studies found Google News benefits small news organizations more with a long tail effect because these small news organizations may otherwise have difficulties attracting visitors due to limited awareness (e.g. Athey et al., 2017). The Google News effect, argued in different studies, is often conditional. For example, certain features, e.g. location sharing (Athey and Mobius, 2012) and the use of links and geo-targeted information (George and Hogendorn, 2013) may benefit local news outlets. But Athey et al. (2017) held that to understand the effects on local newsrooms overall, one has to have a closer examination in terms of what kind of journalism these small local news outlets produce. If they produce alternative perspectives and original reporting, Google News' long tail effect could benefit these small outlets with more readership; otherwise, this very effect only lowers the incentives to produce original content.

When all these findings are put together, they complicated rather than clarified the effects of Google News. Considering that new media players often have multi-dimensional impacts on incumbents (Athey et al., 2017), media effects studies alone would not be enough to understand the complexities in the changing media and technology environment. While most of the aforementioned studies focus on a relatively short time in the course of the history of Google News, short-term effects may not explain relationships and trends in the long run. This dissertation that examines Google's news aggregation service over the past two decades, therefore, does not focus so much on media effects of Google News, but more on the dynamics, relationships, and negotiations between Google and the news industry.

Existing studies reviewed above provide valuable information about the complexity of the media ecosystem in the digital era. Such complexity is largely due to the dynamics brought in through the participation of new media players as well as the interrelationship between older and newer media sectors. Next, this chapter will provide a discussion about the normalization and differentiation frameworks to lay a theoretical foundation for this study in understanding the dynamics and relationships involved in the news ecosystem in the 21st century.

Theoretical Frameworks

This section draws on theoretical frameworks of normalization and differentiation to lay a theoretical foundation for this dissertation in investigating how Google News becomes what and where it is, especially how Google News' development has responded to and has been shaped by relationships with the news media. These frameworks allow a relational view in studying Google News' development and influence as well as the

negotiations between Google and the news industry. Normalization and differentiation, two seemingly paradoxical forces, also provide a dialectical perspective to understand the dynamics and relationships between older and newer media sectors in the changing media and technology environment.

Normalization

In the book *Politics as Usual* (2000), political scientists of the University of Cincinnati Michael Margolis and David Resnick argued against a revolutionary view of the internet, stating that internet politics are politics as usual. The authors pointed out that cyberspace is no longer a "strange realm," as it has been normalized by established rules and practices that are applied in ordinary politics and commercial activities in the real world. The normalization process, according to the authors, has transformed the internet from a "laissez-faire and libertarian" virtual community in its early history to a mass medium that is used and governed by different parties to pursue their own agendas. The authors argue that a newer social sector's development is shaped by existing norms and rules of established social sectors. As a result, certain characteristics of the latter are incorporated into the former's norms and routine practices. So as the news media industry moved into cyberspace, the political and economic forces that have shaped its structures offline remained at play.

This notion of normalization has been adopted by media scholars in understanding how traditional media have adopted new media technologies and formats. For example, City University of London professor Jane Singer (2005) examined how traditional political journalists used blogs and found that although the blog emerged as a novel online media format that challenged existing normative journalistic roles and

practices, by encouraging transparency and participation, journalists tended to normalize blogs to reflect and enhance traditional journalistic norms and identities. For instance, j-bloggers allowed only limited space for user input on blog sites, which served to secure journalists' gatekeeping role. A study of how mainstream journalists adopted Twitter found similar evidence of normalization (Lasorsa, Lewis, & Holton, 2012). Although journalists in general had adopted certain microblogging features, those who worked for legacy media were more likely to, when they use Twitter, defend their professional role and boundaries and showed less inclination to change their power relations with audience, competitors, and the general public.

On the one hand, the process of normalization tends to lead to homogeneity; on the other hand, normalization is found not a uniform process. For example, the studies by Singer (2005) and Lasora, Lewis, and Holton (2012) identified different levels of normalization between national and local media. Scholars have also found that in the news media's adoption of new communication technologies or media formats, certain affordances are more normalized than others. For example, in their use of Twitter during the 2016 U.S. presidential debate, political journalists normalized older types of interaction features, such as tweets and retweets, more than newer interaction features, such as replies and quote tweets, to conform to traditional journalistic norms (Molyneux & Mourão, 2019). Journalism studies scholars Edson Tandoc and Tim P. Vos (2016) and Coddington (2014) found that social media have been normalized into American newsrooms' journalistic routines to fit into their gatekeeping role. They also found that news professionals use social media to enable the reworking of traditional journalistic

norms and define new roles, such as the roles of news marketer and communication facilitator.

Normalization—adoption and appropriation

These studies suggest that the process of normalization involves two types of practices and decision-making: adoption and appropriation. For existing social sectors, the "adoption" process often starts with the action of acceptance, an acceptance of other social sectors' practices and ideas. Such acceptance, however, is not necessarily affirmative—often assumed without rigorous, empirical, and industry-specific evidence. More often than not, the acceptance that drives the adoption process is a decision to follow the trending practices in an industry. Scholars find that organizations tend to pursue isomorphism when adapting to innovations. They often mimic dominant practices that are widely accepted by members of the industry to lower the risk of change, maintain professional legitimacy, meet public expectations, and conform to needs of other institutions (Lowrey, 2011). For example, the common adoption of blogs by Slovene news media was motivated by a mindset that "they [the news media] *estimate* that blogging is one of the most widespread and popular forms of online communication." For those media organizations that failed to join the early adopters, "they *estimate* that they have missed the initial demand and that there is a sufficient supply of blogs" (Vobič, 2007, pp. 74-75, emphasis added). In this case, the technological power of blogging, although without substantial evidence, was assumed due to the industry-wide influence.

For emerging social sectors, their normalization also involves such adoption, because young social sectors often lack legitimacy. Scholars of administrative science found evidence that organizational legitimacy corresponds to a significant reduction in

the death rate of young organizations. An important source of such legitimacy is external relationship building and recognition, e.g., the exchange relationships with other organizations and the endorsement by powerful collective actors, that could help organizations become "a part of the power hierarchy" (Singh, Tucker, & House, 1986, p. 173; Stinchcombe, 1987). Naturally, since emerging social sectors do not have enough legitimacy, they choose to adopt existing practices and norms of established social sectors that do have strong legitimacy. Such adoption indicates existing social sectors' institutional power (Scott, 2008). When a new medium emerges, it is also a practical strategy during its early stage of initial practical implementation (Manovich, 2013). Therefore, for scholars of media and technology evolution, they tended to consider such retrieves—newer generation of medium or technology retrieves some older forms from the older generation of medium or technology—an evolutional process (Scolari, 2018), as "technology creates itself out of itself. It builds itself piece by piece from the collective of existing technologies" (Arthur, 2009, p. 176, cited in Scolari, 2018, p. 154). For example, computers, according to Manovich (2013), are the first metamedium because they grew out of "a wide range of already-existing and not-yet-invented media." (p. 105)

In the news media's digital transformation, there are many examples of adoption. From early media convergence (*e.g.* Domingo & Paterson, 2011; Paterson and Domingo, 2008) to the trend of data-driven journalism in recent years (Coddington, 2015; Hammond, 2015; Wang, 2018a), the news media sector has adopted digital technologies such as website, blog, social media, audience analytics, and big data (Anderson, 2011; Bright & Nicholls, 2014), as well as digital ideals, such as participatory culture (Jenkins, 2006; Wang, 2017) and open-source ethics (Lewis & Usher, 2013; Parasie & Dagiral,

2013). While the process of adoption is often uncritical, in boundary studies, it also represents an expansion phase in which one social sector opens its boundary and extends into new domains (Carlson, 2015). In journalism studies, little research focus, however, has been paid to the other direction of adoption, namely, how newer media sectors adopt existing practices and ideas from established media sectors.

Another types of common practice in the process of normalization is appropriation. Compared to adoption, appropriation is a process that involves more reflective decision-making. The adopter needs to make the decision in terms of what aspects of existing practices and norms should be adopted and for what purpose, as these norms and practices may afford various means and ends. For example, journalists may choose to adopt Twitter for brand building but not for audience engagement, or they may normalize certain features of Twitter, such as tweeting, but not others, such as replying, allowing them to distribute news produced by their organizations keeping them from promoting different perspectives across news organizations (Molyneux & Mourão, 2019). In another example, while bloggers proposed alternative practices or agendas, these alternatives were more likely to correlate with traditional journalistic ideas. In some cases, bloggers sought to legitimize these alternatives through established journalistic norms (Lowrey, Parrott, & Meade, 2011). Coddington (2014) found that when professional journalists normalized the use of hyperlinks in their journalistic routines, they tended to welcome certain new norms associated with the web culture, such as transparency and networked connection. Such new norms, according to the author, are those that "mesh well with laudable journalistic aims" and in practice they are also "transformed as they are filtered through" established journalistic requirements (p. 152). For existing social

sectors, although appropriation involves the introduction of new, different practices and ideas, it is still part of the normalization process, as the action of appropriation aims at convergence, rather than difference, to maintain the status quo of the existing social sector.

During the media and technology evolution, Manovich (2013) argued, after its initial stage of simulation, the new medium enters a new stage of media hybridization. The new medium and technology selectively absorb aspects of existing media through "a particularly novel combination of media types" to grow into "new species." (Scolari, 2018, p. 161) Compare to the adoption stage, the hybridization stage involves more appropriation, or, "deep remixability," in that new technology remixes "not only content from different media but also their fundamental techniques, working methods, and ways of representation and expression." (Manovich, 2013, p. 267, p. 268)

Since the appropriation process involves the decision about what should or should not be adopted, this process often comes along with conflicts, such as the incompatibilities between different media cultures seen in print-broadcast partnerships in the early 2000s (Silcock & Keith, 2006), or the conflicts between the identities of traditional and online journalists during early media convergence (Paterson & Domingo, 2008; Wang, 2018a), and the disputes between digital platforms and traditional media industry. These struggles and incompatibilities reflect the protection of boundaries and identities, which plants the seed for differentiation, a countertrend accompanying normalization.

Differentiation

Differentiation theory is a framework that originated in sociology. Sociologists such as David Émile Durkheim, Talcott Parsons, and Niklas Luhmann held that modern societies evolve into specialized social structures, with social units tending to differentiate from one another during the process of modernization (Alexander & Colomy, 1990; Hallin, 2005).

The modern theory of differentiation has been shaped by the systems theory of Niklas Luhmann, a German sociologist. Luhmann (1997) examined the shift of the mode of differentiation—from segmental to hierarchical and then to functional differentiation—in the evolution of society. In the early history of human beings, social structures were segmented in a center-peripheral model, or tribal differentiation. The industrial era featured hierarchically structured social stratification caused by power imbalances and hegemony between different social classes. Luhmann argued that society has evolved into another stage where differentiation occurs at the functional level of social systems. Social systems differentiated from one another by their exclusive, specialized functions. For example, politics specializes in governing; economy manages production, consumption, and distribution; science provides knowledge of and for society; and so on. As Alexander Görke and Armin Scholl (2006) note, "All these spheres existed prior to modern society, of course, but they were not autonomous because they did not fulfill their societal function exclusively" (p. 647). The exclusive, specialized function gives social systems a certain degree of autonomy; it also requires interdependence of these social systems to make sense of the existence of each social system and to ensure the society runs smoothly as a whole.

Differentiation—specialization and autonomy

Scholars who study the general patterns in the process of media evolution pointed out that in media industry, as new media sectors grow, existing media institutions often show strong resistance to such disruption. When the effort of resistance is no longer effective, however, these media institutions seek to change in order to survive. This process is considered differentiation in media evolution theories (Napoli, 1998). At the stage of media differentiation, incumbent media institutions depend on specialization, whether specialization in media content or target audience, to distinguish their resources from other media sectors for competition advantage (Napoli, 1998). From this perspective, specialization is an important means that drives media differentiation. Such differentiation is a strategic response to a situation when the status quo of the given media sectors was challenged. At this stage, the goal of incumbent media institutions is not anymore convergence but divergence. For the media institutions that felt threatened, this is a matter of survival, survival as a social institution with autonomy.

An institution's autonomy, according to Vos and Russell (2019), "is a matter of the institutional actors' ability to pursue their own institutional ends with manageable pressure from other social institutions." (p. 2334) To have autonomy means the social institution is able to "largely set their own course." (p. 2335) The authors held that autonomy is a fundamental condition for a social institution to exist as "the absence of autonomy would mean the absence of an institution." (p. 2334) Vos and Russell's study of institutional autonomy largely depends on how a social institution handles pressures from other social institutions. The authors proposed six factors to assess how a social institution's autonomy is structured, including the given social institution's power position relative to other social institutions, dominant form of pressure, dominant

incentive of the social institution that bears the pressure, the actors that are directly pressured, the level the pressure affects the most (e.g. at institutional level or practical level), and the resources of resistance to pressure that a given social institution has. Based on the analysis of the six factors, the study concluded that journalism's institutional autonomy is in crisis under the pressure of Silicon Valley as an institution.

Dimmick and Rothenbuhler's (1984) (American scholars of communication and journalism) niche theory argues that when new population invades an existing ecosystem, competitions occur between new and old populations for the same and finite resources. The superior competitor could either edge out the disadvantaged competitors from the ecosystem, i.e. the competitive exclusion effect, or force competitors to alter their niches in order to lower the degree of competition and to increase the likeliness of survival, i.e. the competitive displacement effect. In the media industry, each media sector defines the boundary of its resource space, or the niche, including the type, dimension, and utilization of resources, based on which the given media sector determines its "position in the multidimensional resource space of the environment" (Dimmick, Chen and Li, 2004, p. 22). In the modern history of the media industry, the media environment has been invaded by several then-new populations, including radio, television and cable. Dimmick and Rothenbuhler illustrated how the old populations adjust their strategies in terms of the breadth and overlap of their advertising niche when new population invades, for example, diversifying categories of advertising or shifting from relying more on national advertising to local advertising. This kind of competition displacement can be considered a practice of specializing resources and their utilization.

When media sociologists study media's development from a macro level by considering the media in relation to other social sectors, their studies pointed to the relationship between specialization and autonomy as two determining factors in the social process of differentiation. For example, Jeffrey Alexander (1990), an American professor of sociology, examined the process through which the media differentiated from other social institutions, such as political parties and religion, in the United States and Europe. He identified this media differentiation as the historical specification of the media. In the 17th century, politics, religion, and civil cultures reshuffled their resources and power, which produced new social groups that demanded their own voices and standards. The media sector was born to respond to such a demand. Meanwhile, the development of professionalization of the media sector, which demanded prestige and autonomy for occupations, contributed to media differentiation as well. The growth of professional norms and self-regulation led to varying sources of media differentiation, such as role differentiation, goal differentiation, and differentiation in institutional structure.

Media differentiation can also be examined through the historical developments of media institutions, as Dan Hallin, an American professor of communication, and Paolo Mancini, an Italian professor of sociology and communication, do in examining media in eighteen Western European and North American countries (Hallin & Mancini, 2004). They identify three different models of media system in terms of the relation among the media, politics, economics, and professionalism. Hallin and Mancini argue that media differentiation is not the media's break-off from other social institutions but the rise of the media logic that gradually dominated over the logics of other established social institutions, which once defined and controlled social orders and norms. The

differentiation process was related to the decline of these established social institutions in their importance, effectiveness, and power, which gave the media the opportunity to take over some of their informational and socio-political functions, which grew into a social specialization with more independency and autonomy.

Beyond North America and Europe, on which Hallin and Mancini's original work focused, studies have found different modes of media differentiation (Hallin & Mancini, 2012). In China, for example, a functional differentiation was found between state media and online commercial media, due to sector-specific media and internet regulations (Wang, 2018b). In this process, Chinese state media and online commercial media identified their unique resources and limitations, which the two segments distinguish themselves. For instance, members of Chinese state media are granted an official press pass that online commercial media do not have, which allows them to cover news and conduct interviews. State media are thus able to specialize their role in setting the agenda for other media. However, since the regulation of their online presence is stricter, related functions, such as online comments and user participation, are weak for state media. On the other hand, online commercial media—without the official press pass—focus on interactive features and entertainment. That gives them the opportunity to use public agenda to leverage the media agenda defined by state media. The media differentiation in this case allows these media segments a certain degree of autonomy, however limited, through specialized resources and positioning.

In these studies, specialization and autonomy are two major characteristics of differentiation. When a social sector differentiates itself from other social sectors, it often does this through specialization that supports this given social sector to be unique and

important enough to be independent from other social sectors and their social respective functions. Through specialization, a social sector is able to make the best use of its unique resources to maximize its advantages relative to other social sectors. As specialization grows, it also results in a stronger demand for autonomy, as the given social sector refuses to be incorporated into other social sectors so it can sustain its independence. Therefore, in the process of differentiation, specialization is the means while autonomy is the end.

The Dialectics between Normalization and Differentiation—Negotiation

The dynamics and decision-making involved in normalization and differentiation reflect the negotiations of social sectors about their boundaries. In different contexts, social bodies use different "boundary objects," including social and material objects, to define the components of boundary work, including participants, practices, and goals and interests (Carlson & Lewis, 2015). The process of normalization collapses the boundaries between different social sectors, especially at the adoption phase, and differentiation contributes to boundary building by distinguishing one social sector from another. More often than not, the two seemingly opposite forces intertwine in a dialectical way.

Both normalization and differentiation involve "the process of reproducing systems within systems, boundaries within boundaries" (Luhmann, 1997, p. 71). Normalization aims to co-opt the difference by redefining the existing boundary. The process of normalization reflects the negotiation about the boundaries between the adopting party and the adopted party. Such negotiation rises in the appropriation phase. Similarly, the differentiation process emerges within other already differentiated social boundaries. Take the relationship between news media and other social institutions, such

as religion, for example, in the 17th century in Europe, some churches had their own newspapers (Alexander, 1990). Earlier "newspapers serve[d] modern man as a substitute for morning prayers" crafting an imagined community and a shared social identity, which paved the way for the emergence of the nation as a political community (Anderson, 2006, p. 35). In contemporary societies, the differentiation between media and politics is still being negotiated. For example, news media have to depend on government as their primary source; on the other hand, governments incorporate the media's logic into their political agendas (Cook, 1998). In the digital age, new media actors, such as digital platforms that have played the role of digital consultant in American electoral politics, have also joined the negotiation (Kreiss & McGregor, 2018).

In addition, normalization and differentiation are the results of the negotiation between internal and external influences to which a given social sector is subject. French sociologist Pierre Bourdieu (1993, 1998) proposed field theory by seeing societal components as differentiated "universes" or fields. Each field, he wrote, obeys its own rules but is also subject to external influences. Therefore, each field is pillared by both the heteronomous pole that carries pressures from outside the field and the autonomous pole that carries influences from inside the field. By negotiating with heteronomous and autonomous influences, each field becomes a semi-autonomous microcosm embedded in the macrocosm (Benson, 2005). When a field normalizes new norms, practices, and objects, it has to negotiate with itself and cross-field influences (Wang, 2018a). Organization scholars also pointed out that organizational routines contribute to both organizational stability, through coordination and preservation to the past, as well as flexibility and change, through adaptability, interdependence, and flexibility (Feldman &

Rafaeli, 2002; Becker, Lazaric, Nelson, & Winter, 2005). On the one hand, routines define appropriate actions; on the other hand, they also make connections between organization members. Such connections are dynamic in nature as they require shared understanding and the transfer of information.

Hallin (2005) argued that differentiation is not the evitable direction of social structures because de-differentiation forces are also at play. In studying the differentiation of media systems in Europe and North America, Hallin and Mancini (2004) identified various forms of de-differentiation, including external de-differentiation influences, such as commercialization, globalization, Americanization, and cultural imperialism, and internal de-differentiation forces, such as secularization and the increase of individual agency, that have contributed to the homogenization of the media worldwide. But differentiation forces—such as the different political, legal, and ideological systems and the development of professionalism—have also challenged media homogenization.

Therefore, negotiation is another important characteristic in the processes of normalization and differentiation. That is because social sectors develop and evolve in relation to other social sectors and the broad social environment. Influences from inside and outside of the given social universe often determine under what circumstances the social sector chooses to normalize or differentiate. Thus, normalization and differentiation are not two separate trends but often overlap. When one trend is dominant over the other depends on the pressure of these influences and the given social sector's self-adjustment.

Five Characteristics

Normalization and differentiation are social processes that shape the interrelationship between different social sectors as our society develops. Normalization is a process, in which one social sector tends to integrate certain existing routines that other social sectors apply into its own social routines. Differentiation theory comes from the field of sociology arguing that modern societies evolve into specialized social structures, with social units tending to differentiate from one another during the process of modernization. Based on the theoretical discussion about normalization and differentiation so far, I identify five important characteristics that define these social processes. They are adoption and appropriation in the process of normalization, specialization and autonomy in the process of differentiation, and negotiation that characterizes the interrelationship between the two social processes.

1). Adoption: A characteristic of the process of normalization in which a social sector accepts and incorporates existing practices, norms, and ideas into its own social routines.

2). Appropriation: A characteristic of the process of normalization in which the adopting party selectively adopts existing social routines of the adopted party and uses them with different affordances or to serve different purposes.

Adoption is more simulative that the adopting party tends to take the existing routine as is while appropriation is more selective that certain existing practices and ideas are selectively adopted or repurposed. Both adoption and appropriation serve the goal of convergence, rather than divergence, during the process of normalization.

3). Specialization: A characteristic of the process of differentiation in which a social sector identifies and optimizes unique resources that could distinguish the given

social sector from other social sectors and help it gain competition advantages over other social sectors.

4). Autonomy: A characteristic of the process of differentiation in which a social sector demands a higher degree of self-control and independence from other social sectors or resists being incorporated into other social sectors.

When a social sector differentiates itself from other social sectors, it often does this through specialization that supports this given social sector to be unique and important enough to exist legitimately and independently from other social sectors as well as their respective social functions. As specialization grows, it results in a stronger demand for autonomy, so the given social sector can sustain its independence. Therefore, in the process of differentiation, specialization is the means while autonomy is the end.

5). Negotiation: The dialectical dynamics between normalization and differentiation, which depend on the interrelationship between social sectors and the internal and external influences that they are subject to.

The dynamics involved in normalization and differentiation determine that older and newer social sectors are intertwined in terms of their social routine and development trajectory. Studies show that normalization and differentiation are not linear processes. Whether a social sector takes the approach of normalization or differentiation is often a result of the negotiation between different social sectors and how these social sectors respond to and how they are shaped by internal and external conditions.

In the following chapters, this dissertation will use these characteristics as analytical points to examine Google News' development over the past two decades from different perspectives. The normalization and differentiation frameworks provide a

historical, relational, and sociological view to understand the nature of Google News, how it works, the relationship between Google News and traditional news media, and the forces shaping the media and technology ecosystem in the digital age.

In the following chapters, this dissertation will explore four research questions:

RQ1: How did Google News emerge and how has it evolved since its launch, historically, in terms of its front-end presentation, back-end technology, and its relationship with the news media industry?

RQ2: What influences have driven Google News' evolution? And in what ways?

RQ3: How has the evolution of Google News involved, if any, normalization and differentiation in terms of the five identified characteristics?

RQ4: What are the implications of the evolution of Google News for journalism, policymaking, and democracy?

Plan of This Dissertation

Taking a historical approach, these research questions will be explored from various perspectives in the following chapters. In Chapter 2, the dissertation traces the origin of Google News to understand the historical circumstances in which Google News was born. Taking the attacks of September 11, 2001, as a turning point, this chapter studies the status of the search landscape pre-9/11, the "news paralysis" and the online news demand on 9/11, and the early form of Google News in the immediately post-9/11 time.

Chapter 3 examines the structural, visual, and functional evolution of Google News homepage design over the past seventeen years. This chapter uses web archival analysis to identify various types of changes. This close examination of Google News

homepage design is combined with analysis of blogs, news articles, and historical scholarly work to understand what has driven the changes on Google News homepage.

Chapter 4 examines the tensions and negotiations between Google and traditional news media organizations on three continents: Europe, North America, and Asia. Using traditional interpretive legal analysis methods, this chapter examines eight international disputes over Google and its news aggregation service to identify the focal points of contention, investigate the legal frameworks pursued in these cases, and explore the implications of these disputes.

Chapter 5 goes behind the scenes to explore Google's news-related technologies and algorithms. This chapter combines a computational approach and in-depth textual analysis of 171 Google patent filings to identify news-related technological trends in Google's news aggregation service over time and key factors that have shaped Google's news-related algorithms.

Chapter 6, the final chapter of this dissertation, provides an analysis of Google's systematic ambitions in the news area, the Google News Initiative. This chapter reviews GNI's programs, technologies, and partnerships to discuss how these initiatives could influence the future of the news media industry. The dissertation concludes with a discussion of the implications of this research on Google News for policymaking on digital platform regulation. The organization of the dissertation is listed as below:

Chapter 1. Introduction and Theoretical Frameworks (p. 1-35)

- Introduction
- Online News Aggregation
- Google News as the World's Largest News Aggregator

- Theoretical Frameworks
- Plan of This Dissertation

Chapter 2. Origin and Early History: 9/11, A Turning Point (p. 36-65)

- Pre-911
- 911
- 9/11 and the Web
- Immediately Post-911
- Discussions

Chapter 3. Structural, Visual, and Functional Trends on Google News Homepage Design (p. 66-124)

- When “Google News” Was “Google News Search”
- Major Redesigns and Implications
- Gradual Homepage Changes Over Time and Implications
- Personalization
- Discussions

Chapter 4. Battles: Google News and the News Media Industry (p. 125-177)

- France
- Belgium
- Italy
- United States
- UK
- China
- Germany

Chapter 2. Origin and Early History: 9/11, A Turning Point

Google News was launched in 2002. Why this specific year? Why did Google want to roll out a news aggregation service in the first place? To understand the nature of Google News and how it works, one needs to first examine how it was born. Chapter 2 of this dissertation conducts web archival research using the Internet Archive (https://archive.org/) and draws on relevant texts, such as blogs, news articles, and scholarly literature, to trace the origin and early history of Google News.

The Internet Archive (IA) is a digital archive that records websites and their content from 1995 to the present. IA is considered the "the largest digital source for historical research pertaining to the Web and its contents over time" (Weber, 2014, p. 1031). Using IA to explore the evolution of a digital entity is an innovative approach—especially for projects that are navigational and temporal in nature—that has served a small group of scholars so far (Costa & Silva, 2010; Holzmann, Nejdl, & Anand, 2017). For example, communication scholars Matt Weber and Peter Monge (2017) used IA data to examine how news organizations adopt hyperlinks and how the hyperlink strategies shape the interrelationship between news organizations. Similarly, the News Measures Research Project led by Duke University professor Philip Napoli used IA to extract data to investigate the ecosystem of local journalism in the United States (Oliver, 2017).

Using the Internet Archive, this historical chapter uses the event of September 11, 2001, as a turning point to investigate the early history of Google and Google News in three historical periods, through the following questions:

- How did Google work before 9/11 and how Google News emerged?

- How did Google News transit from a news folder under google.com to a standalone website in the wake of 9/11?
- What was Google News like in its early history in the immediately post-9/11 era?

To answer these questions, I examined all archived web pages available in the Internet Archive's digital archive that were related to Google and Google News' early history. These web pages are associated with three URLs: google.com, google.com/news, and news.google.com.

The pages examined cover a period from December 1998, when the Internet Archive recorded Google's webpage for the first time, to March 2004 when Google's news folder and Google Directory no longer appeared on Google's homepage. (Archived webpages after 2004 are examined in the next chapter.) I downloaded and analyzed 614 webpages[1] available in the Internet Archive's database as being captured from the three relevant URLs: 430 webpages from google.com, 106 webpages from google.com/news, and 78 webpages from news.google.com. In addition to doing the web archival research, I also reviewed blogs, news articles, web articles, videos, and scholarly literature for a comprehensive picture of the historical background of Google and Google News.

Pre-911

The Early Search Landscape

Google was launched in 1998, when the search market was dominated by web portals such as Yahoo! and the early generation of search engines, such as InfoSeek. Yahoo!, launched in 1994, provided the Yahoo! directory that many searchers used then

[1] When multiple captures for one day are archived in the IA database, I recorded the most recent.

as an internet starting point. Yahoo! editors manually classified information on the Web into categories by topic, which were organized in a hierarchical structure, such as Arts and Humanities/Humanities/History/U.S. History (Sullivan, 2014). Users had to go through the hierarchy to find a relevant category and then the information for which they were looking.

As the internet grew rapidly, the two models of web information service—search engine model vs. directory model—started to show limitations that constrained the effectiveness of information search. For example, early search engines increasingly returned irrelevant information due to flaws in algorithm design and keyword matching techniques. Some of the results were criticized as commercially biased because they were auctioned to the highest bidder. Another limitation was that these search engines did not pay particular attention to news sources. Michigan State University senior information technologist Richard Wiggins (2001) shared his experience of having to spend hours to find breaking news, such as information about the 1997 fatal car accident of Diana, Princess of Wales, on the then-popular search engines. The Yahoo! directory, on the other hand, did have a "news and media" category in its hierarchy, but the Yahoo model, which adopted traditional journalistic gatekeeping techniques by depending on human editors to select and compile web pages, had trouble keeping pace with the rapid growth of the Web. As a result, the human-edited Yahoo! directory returned a growing number of links that were unfound, broken, or erroneous, exhibiting what is known as "link rot." Yahoo!'s human-maintained lists were also accused of being subjective and expensive (Sherman, 2000).

The ODP

It was against this backdrop that the Open Directory Project (ODP) and Google went live in 1998 one after the other. Unlike Yahoo!, the ODP recruited a vast army of volunteer editors to work with its editorial team to categorize web pages. Inspired by the open-source model, the ODP envisioned a "republic of the Web":

> Instead of fighting the explosive growth of the Internet, the Open Directory provides the means for the Internet to organize itself. As the Internet grows, so do the number of net-citizens. These citizens can each organize a small portion of the Web and present it back to the rest of the population, culling out the bad and useless and keeping only the best content. (About the Open Directory Project, 2009)

By the time the ODP project ended in 2017, it had nearly 92,000 editors who listed more than 3.8 million websites in over one million categories in 90 languages (Organize the Web, n.d.). In the following couple of years since its launch, the ODP directory was adopted by many search engines, among them Google.

Google and PageRank

By the time Google was founded, many early search engines had been purchased by traditional media or telecommunication companies and had become portals supported by corporate advertising (Van Couvering, 2011). In contrast to these early search engines, which were normalized by traditional communication industries, Google sought—mostly through technological means—to take a path that differentiated it from peer companies and other industries. In 1998, Google developed a new algorithmic system called PageRank as its key web search tool. PageRank was one of the algorithms Google used in "bringing order to the Web" (Page, Brin, Motwani, & Winograd, 1998). Instead of seeing

the Web as a hierarchical structure, PageRank paid attention to the interlinked nature of the Web by calculating the number, importance, and relation of the links, the linked parties, and the linking directions. Based on such calculations and other algorithms that measure the relevance, Google was able to rank search results in terms of their importance and relevance scores. The ranking algorithms were an important technological preparation for Google News to emerge after 9/11.

In 2000, Google introduced Google Web Directory on its homepage by integrating PageRank-based ranking algorithms and the data of the ODP. Two options were provided when users clicked on one of the categories of the Google Web Directory. Users could choose to view the results in Google PageRank order or in alphabetical order (Figure 2-1). Alphabetical order was the convention of the ODP, while the Google PageRank order, according to Google, ranked websites based on their importance allowing "the highest quality pages to appear first as top results for any Google directory category" (About the Google Directory, 2000).

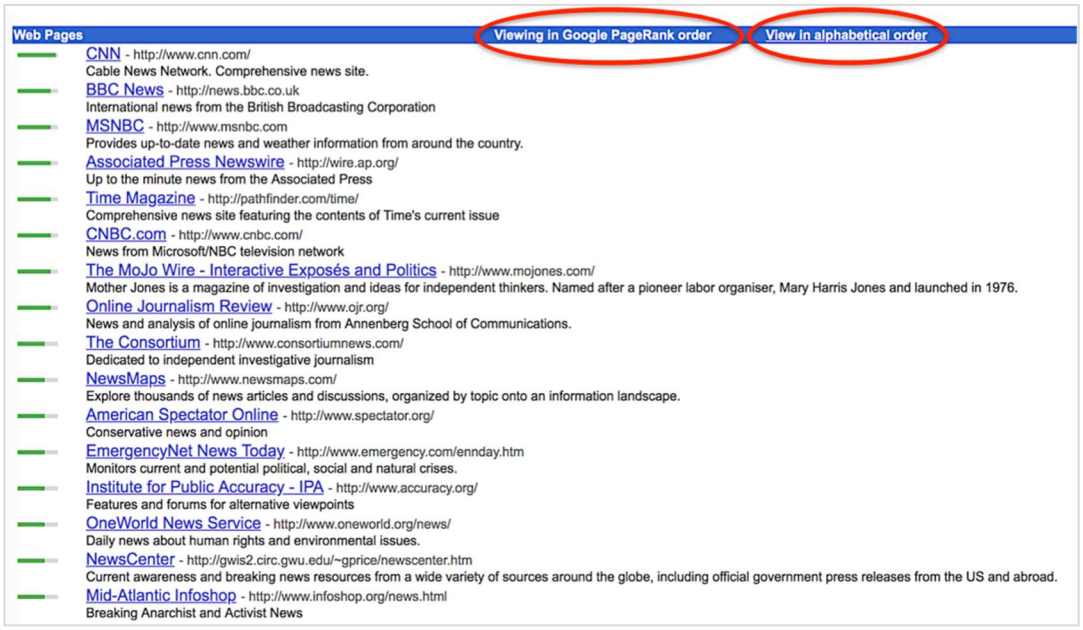

***Figure 2-1*. View in Google PageRank Order in 2000. Source: Internet Archive, emphasis mine.**

In addition to the importance ranking, Google Web Directory also integrated its advanced search technology on top of the ODP data. Like the ODP, the Google Web Directory listed "News" as one of the categories. In the "News" category, there were as many as 23 sub-categories, from different news genres such as breaking news, sports, and satire, to different media forms, such as newspapers, magazines, televisions, and radio. Google's advanced search had the function of category search, *e.g.*, "Search only in News" (Figure 2-2), which allowed users to search within a specific category, such as news. When advanced search was combined with Google's ranking algorithms, Google users were able to browse news ranked by importance and relevance scores. This meant, according to Google, that, "the most relevant and highly-regarded sites on any topic are listed first" and users will not be "buried deep within a list of other pages" (About the Google Directory, 2000). These features technically distinguished Google Web Directory from directories managed by human editors.

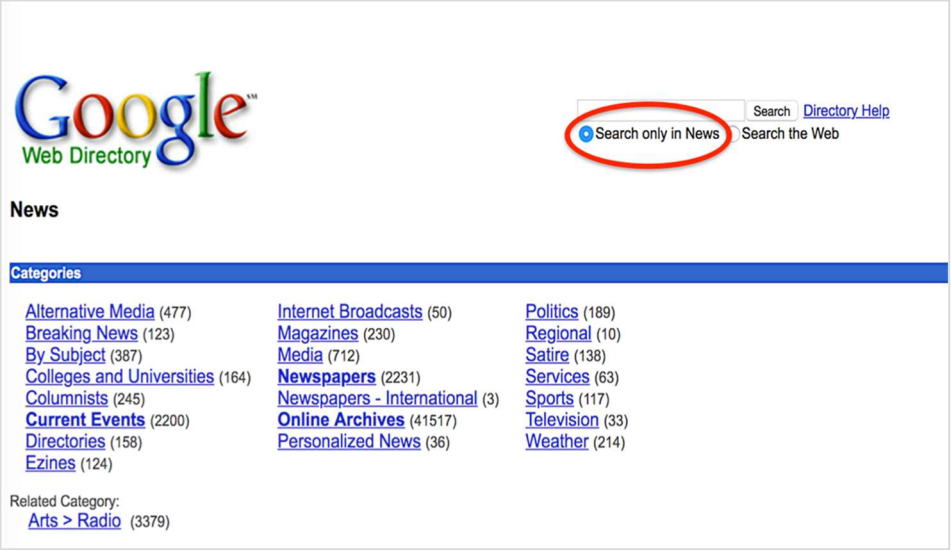

***Figure 2-2*. Google Web Directory—Category Search. Source: Internet Archive, emphasis mine.**

The search function, especially the category search that allowed users to search only in the category of news, prepared another necessary technological condition for Google News to emerge. In its early years, however, Google's interest in news was limited to treating news as just one of many categories provided by ODP. It was only later, particularly during and after the 9/11 tragedy, that Google started to realize that news had special market and social values that other information did not have.

Two Models—Human Editor and Ranking Algorithm

The directory model and the algorithm model are two distinct ways to organize the Web. The directory model depended on human editors to select, compile, and arrange the content on the Web, as an editor in journalism would, while Google's model used computational algorithms to rank news sources. When explaining under what circumstance a user should use directory instead of Google's regular web search, Google put that, "you might prefer to use the directory when you only want to see sites that have been evaluated by an editor" (About the Google Directory, 2000).

From 2000 to 2004, Google introduced Google Web Directory on its homepage as one of the key features of the search engine. In those years, although Google considered its systems to be able to provide "likely the fastest way to find information on a specific subject," it also admitted the strengths of the ODP model, which Google believed could help users to understand "how topics within a specific area are related" and "the scope of a given category, such as the number of newspapers in California." Given that Google's algorithms were not then mature enough to completely replace the directory mode, the

goal of Google's own web directory was to provide users with "both human judgment and a sophisticated ranking algorithm" (About the Google Directory, 2000). Such convergence put more emphasis on adoption and appropriation than differentiation.

As Google's search technologies became more sophisticated, this process of normalization wound down. By early 2004, although Google Web Directory still existed, it no longer appeared on Google's homepage, indicating a lowered significance of the directory model and the takeover of Google search that depended on automatic, computational algorithms to process online information. The directory mode was not completely removed from Google until 2011. In a statement published in that year, Google announced with confidence that, "Google Directory is no longer available. We believe that Web Search is the fastest way to find the information you need on the Web." (Schwartz, 2011) At that point, Google's search technologies and the logics underlying them had replaced those that supported the directory model. Google's technologies and logics became the norm on the search market. This process took time, but once these technologies and logics firmly took hold industry wide, Google became and still is the dominating player in the search industry. The story is a bit different outside the search industry, where a historical moment gave Google the opportunity to realize news as a specific type of information and started its decades-long interaction with the news media industry.

9/11

In several interviews, Krishna Bharat, the founder of Google News, has said that Google News was born in the wake of the 9/11 tragedy. The launch of Google News represented a significant shift in terms of Google's perception of news, which directly

affected Google's market direction and the introduction of its news aggregation service. A few contingencies made 9/11 a historical moment for Google News to emerge.

News Paralysis

On the morning of September 11, 2001, two hijacked airplanes were crashed into the twin towers of the World Trade Center in lower Manhattan in New York City, followed by a series of terror attacks in in Arlington County, Virginia, and Washington D.C. The attacks killed 2,993 people and injured over 8,900 others. 9/11 was the largest terrorist attack on U.S. soil (September 11th Attacks, n.d.)

After the World Trade Center's twin towers were hit by two hijacked airplanes, many Americans swarmed onto news websites to find out what exactly happened. They were quickly disappointed by what has been described as a news paralysis (Bemis, 2001). Websites affiliated with major news outlets—such as cnn.com, abcnews.com, and msnbc.com—were down due to the excessive online traffic. Later that day, some news sites decided to keep their sites in plain text as much as they could to reduce pressure on servers. But going to plain text didn't help much. It was reported that between 8 p.m. September 11 and 6 p.m. September 12, nytimes.com was available only 77% of the time, and it took an average of 100 seconds to download something from the site, much slower than usual (Taylor, 2001). When some news sites were available, only limited information was provided, which Bemis (2001) described as "posttraumatic haiku,": "The election was called off. The airports were closed. The United Nations building was evacuated" (para. 4).

Meanwhile, millions of users searched on Google for information about the attacks. Google was already many users' primary "internet resource locator" before 9/11

(Wiggins, 2001). With the combination of its search and ranking technology, users largely trusted Google to effectively identify the highly popular sites and deliver them through Google's search results. During an emergency like 9/11, Google's web search model made more sense than the directory model in finding relevant information quickly. It could save a user time and effort if the user knew the exact keyword regarding the event he/she is searching for, such as "world trade center" in the case of 9/11. Using the directory model, the same user would have had to go through Business>International Business and Trade>Services>Information News>Current Events>Economy and Business>Globalization and Free Trade to find information about world trade center.

However, after the attacks happened, Google was not able to provide users information relevant to the attacks. Given the search query "New York Twin Towers" (another name for the World Trade Center) for example, Google returned only search results that had nothing to do with 9/11 attacks (Figure 2-3). That's because Google's index was crawled a month earlier (Evolution of search, 2011). In 2001, Google's search system treated news just as it treated other kinds of information. While timeliness matters significantly for news, most information does not require the recentness for relevance.

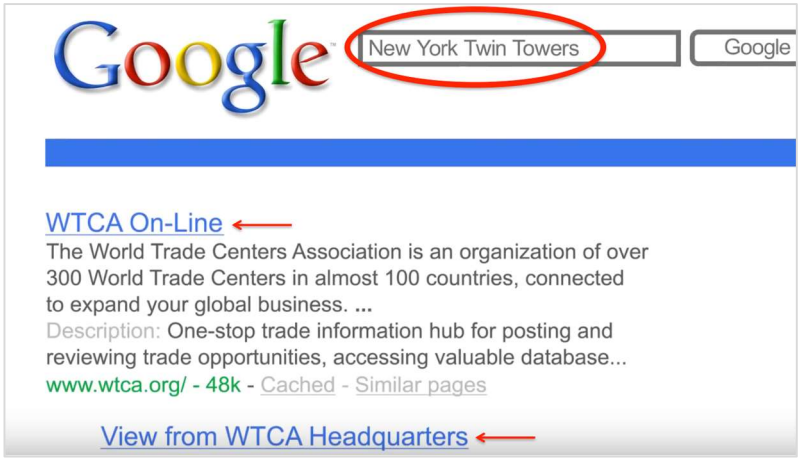

***Figure 2-3*. Google's Search Results for "New York Twin Towers" on September 11, 2001. Source: YouTube, 2011, emphasis mine.**

Online News Market

While digital news outlets were paralyzed, the demand for online news skyrocketed. MSNBC.com had a record of 6.5 million unique visitors after the presidential election in 2000, but that number doubled on September 11, 2001, when the site had estimated 12.5 million unique visitors. Nytimes.com also had a surge in its page views on that day, reaching 11.5 million, compared with its 5.7 million daily average and its previous record of 9.6 million set on the day after the 2000 presidential election (Taylor, 2001).

On the Web, news-related search also soared. Google statistics disclosed that on September 11, 2001, over 80% of the top 500 search queries on Google were related to the 9/11 attacks. Among the top 10 query terms were CNN, World Trade Center, BBC, Pentagon, MSNBC, Osama Bin Laden, Nostradamus[2], American Airlines, FBI, and Barbara Olson[3]. In this list, three of the ten top search queries were news media outlets, and they held top positions on the list as well: CNN, No. 1 on the list; the BBC, No. 3; and MSNBC, No. 5. On that morning, between 6:26 a.m. and 7:06 a.m. Pacific Daylight Time—between 40 minutes and 1 hour and 20 minutes after the attacks—the number of searches for "cnn" reached a peak, averagely 6,000 queries per minute (Figure 2-4).

---

[2] Nostradamus published a book in 1555 containing 942 quatrains that allegedly predict famous future events, including the 9/11 attacks (Cain, 2018).

[3] Barbara Kay Olson was an American lawyer and television commentator. She was a passenger on American Airlines Flight 77 when it crashed into the Pentagon in the September 11 attacks. She had delayed a trip to California on September 10, so that she could be in town with her husband on his birthday on September 11 (Ted Olson on loss and love in the decade since 9/11, 2011).

Google also found that on September 11, many more users searched news. In particular, among the top 200 search queries, news-related searches were 60 times greater than the previous day (Google, 2001). These search patterns demonstrated a strong online news demand or a huge online news market.

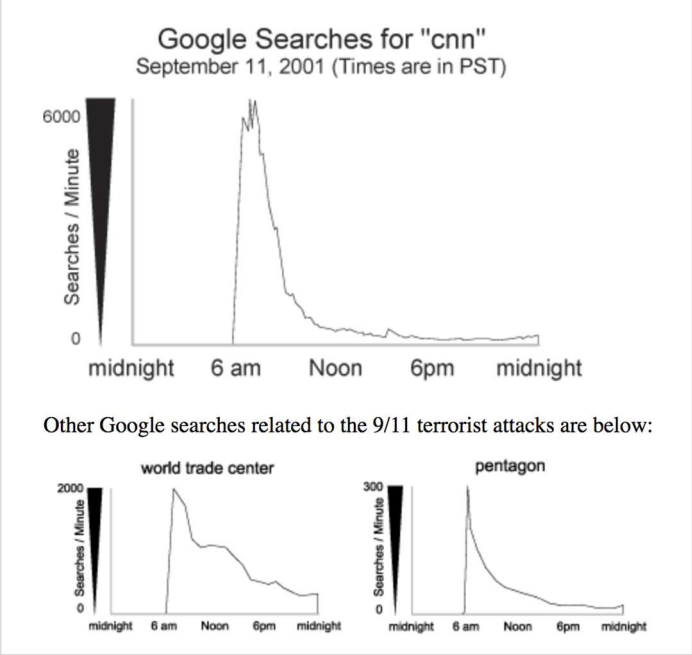

***Figure 2-4*. Google Search Statistics from 9/11/2001. Source: Google.**

What happened on September 11—the news paralysis and the surge of Google searches for news—told Google at least two things:

- In cyberspace, traditional news media had significant technical limitations.
- The demand for online news was tremendous.

Given social and economic considerations, both needed to be addressed urgently; and Google did move quickly. Google realized the first thing immediately and the second after a short while. On September 11, after the attacks, Google displayed on its homepage a short message: “If you are looking for news, you will find the most current information on TV or radio.” This sentence read as if Google was pushing users away to traditional

media, rather than keeping them on its site (Figure 2-5). Below that sentence, Google continued, “Many online news services are not available, because of extremely high demand. Below are links to news sites, including cached copies as they appeared earlier today.” (Figure 2-5) As a search engine, Google stores cached information so when a searcher requests certain information, that data can be provided faster, even if the information is not available anymore on the original source website. In the news paralysis on September 11, the cache function was a technical solution for technical problems.

***Figure 2-5*. Google Homepage on September 11, 2001. Source: YouTube, 2011, emphasis mine.**

At first, Google provided links to only two news sources—the Washington Post and CNN.com—on its home page. Compared with the thousands of news sources worldwide that Google News later indexed, providing links to only two news sources was a conservative decision, probably the result of Google defining itself as a pure search engine, rather than a portal, because it believed “search is king” (Wiggins, 2001, para. 5).

Search engines and portals provide information for users in different ways. Searching is a user-activated model, in which the search engine returns results only when

the user makes a request. Portals, on the other hand, actively present information even if the user did not request anything. In this regard, portals have a role in gatekeeping because by presenting selective information, portals' decision-makers are telling users what is so important that they should know about it. When Google presented links to two news websites on its homepage on September 11—even though there were only two of them—it was a special decision Google made *as a search engine*. This decision was probably the result of Google's leadership in realizing Google's responsibility as users' primary "internet resource locator" during the national emergency. The suggestion that "If you are looking for news, you will find the most current information on TV or radio" made it seem as though Google did not have the intention to enter the news market at that time but only provided a temporary solution to respond to the news paralysis. Things started to change later that day.

Wiggins (2001), the Michigan State senior information technologist, recorded how Google changed its homepage during 9/11. His documentation shows that later that day Google added one more news source, Yahoo! News, on its main page, bringing the total number of news sources linked to up to three (Figure 2-6). By later afternoon September 11, the short message shown on Google's homepage had been changed to this: "Many online news services are not available because of high demand. Below are links to news sites, including cached copies as they appeared earlier." The previous sentence—"If you are looking for news, you will find the most current information on TV or radio"—had been removed. News sources listed increased to six: *The Washington Post*, Yahoo! News, CNN, ABC News, Yahoo! News Photos, and *The New York Times*. Google had expanded this list of news sources since then. The removal of the message "If you are

looking for news, you will find the most current information on TV or radio" reflected Google's change of marketing strategy to one that kept users on its own site by providing them more links to news sources, rather than pushing them away to traditional media. The addition of the link "Make Google Your Homepage!" at the bottom of Google's homepage made such a marketing motivation even more evident (See Figure 2-6). It seems fair to say that these decisions were results of Google's realization of news' market value.

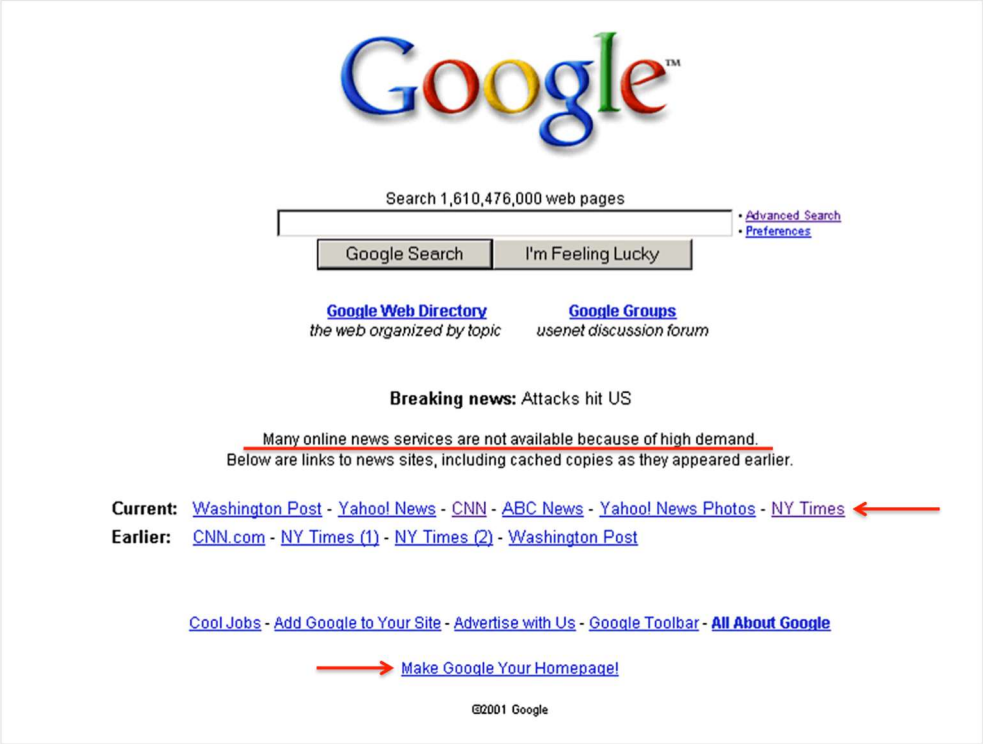

***Figure 2-6*. Google Homepage on September 11, 2001. Source: scripping.com, emphasis mine.**

Moreover, by listing more news websites on its homepage, Google took a step away from its self-position as a pure search engine. The search engine's way of providing users information about 9/11 at the beginning of the day was to wait till the searcher input a search request for, say, CNN or World Trade Center, allowing the search engine to

retrieve and return relevant search results. By listing the news sites on its homepage, Google chose to inform searchers, even those who had not yet composed a query, that these sites were important and should be visited for news. In this way, Google shared the journalistic role of informing the public. This social role of journalism was received by Google due to the public's demand for authoritative and reliable breaking news during emergency.

Google's actions on September 11, 2001, constituted adoption and appropriation, two characteristics of normalization. Note that the news media listed on Google homepage were all well-known news media, which indicated Google's acknowledgement of the authoritativeness and quality of these media outlets. Many of these news sources were also the top news outlets that users searched on 9/11. Google also adopted journalism's role in informing the public, even though this role was not fully in line with its role as search engine. Google appropriated this role to serve its own purpose: to have more users use Google as their primary internet resource locator.

The /news Folder

These efforts of adoption and appropriation escalated on the following day, when a news folder—"google.com/news"—was set up under google.com. On Google's homepage, a link with the words "News links and support information regarding attacks" led users to google.com/news (Figure 2-7) where Google page provided "links to news sites and support resources related to the terrorist attacks on the US" (Figure 2-8). Later in September, the link was renamed "News and Resources." Strictly speaking, the link was about "news *sources* and resources," as it listed a collection of news sources with their links rather than any news stories. The "News and Resources" was on Google's

homepage for about a year, until Google News (beta) was officially introduced in September 2002.

On September 12, 2001, the day after the attacks, news sources listed on google.com/news increased to 34. Over half were news sources outside the U.S., including internationally known news brands, such as BBC News, as well as other news outlets from around the world, such as *Pravda* of Russia and *Asahi Shimbun* of Japan. U.S. news sources CNET.com, FOX News, MSNBC, NPR, the *New York Post*, *Salon, Time.com*, and *USA Today* were added to the list. By September 12, news sources that were among Google's top 10 search queries on September 11—CNN, BBC, and MSNBC—were all included in Google's /news folder.

In addition to news sources, the /news folder also provided a list of "resources" (Figure 2-8), including emergency assistance, such as the American Red Cross and America's Blood Centers; transportation, such as American Airlines, United Airlines, and Amtrak; government and public databases, such as Pentagon Updates and the World Trade Center Survivor Database; user input sites, such as Report Terrorist Activity and Report You're Safe; donation sites, such as Donate via Yahoo! and Donate via Amazon; and other emergency contact info. Providing access to these public resources strengthened Google's adoption of journalism's public-service role, given that the general public usually learns about these public resources from news media.

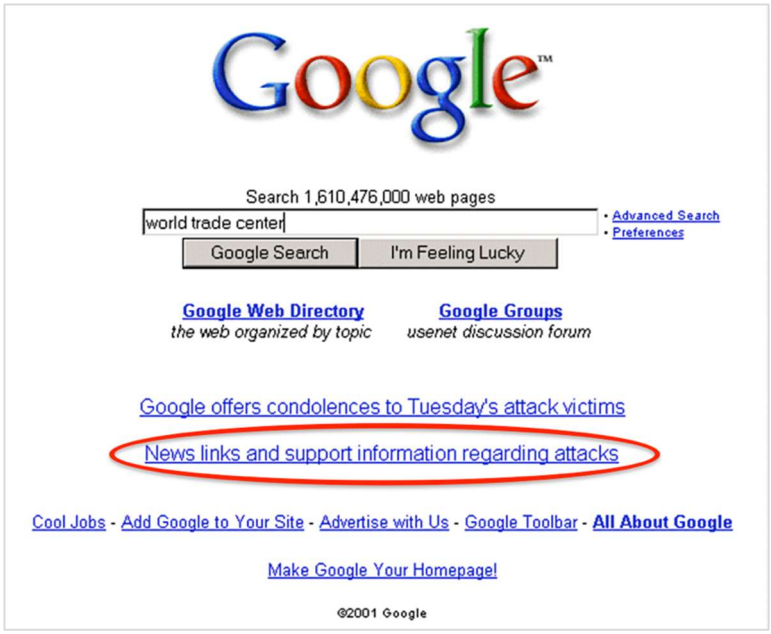

***Figure 2-7*: Google Homepage on September 12, 2001. Source: Wiggins, 2001, emphasis mine.**

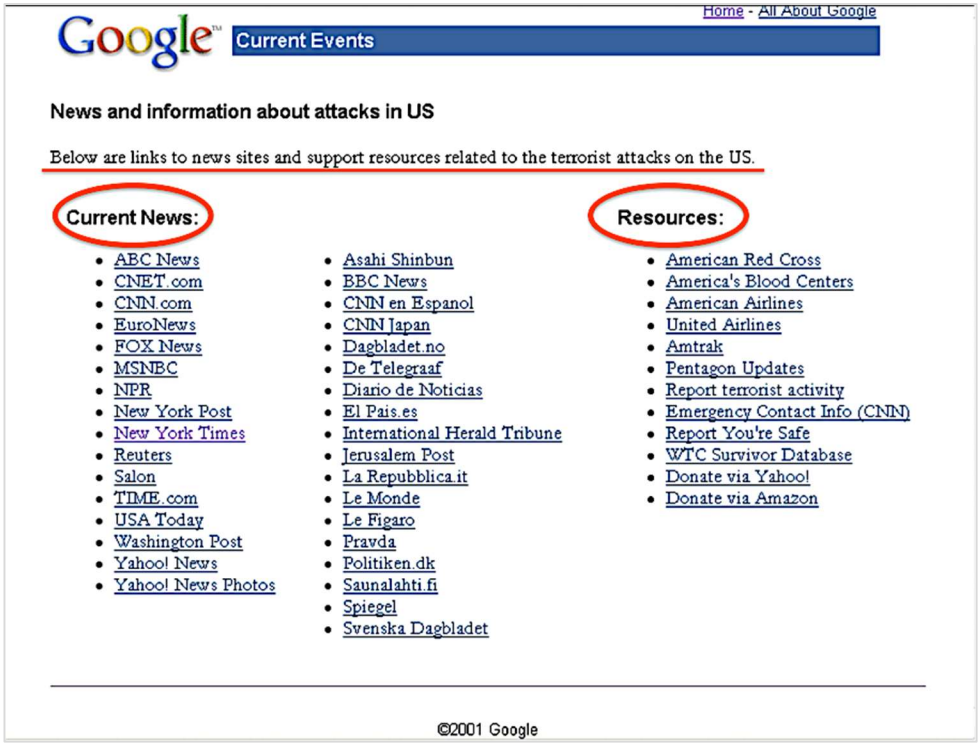

***Figure 2-8*: google.com/news on September 12, 2001. Source: Wiggins, 2001, emphasis mine.**

9/11 and the Web

Media need some watershed moments to claim their role and legitimacy, especially in their early history. 9/11 was one of such defining moments for the Web to come into its own (Halavais, 2002). The massive and urgent demand for informational, emotional, and practical support during and after 9/11 transformed many websites temporarily or permanently. On September 11, 2001, these informational, emotional, and practical needs were translated into the demand for news, as news could accommodate all these dimensions.

News Was the "King"

In the wake of the 9/11 attacks, many actors on the Web that were not news-oriented in nature—from blogs, humor sites, advertisers, and local sites—provided information that had characteristics of news, or what informatics scholar Alexander Halavais called "do-it-yourself journalism," to meet the public's demand. For example, Slashdot.com, a site featuring discussion forums that primarily focused on technology and science-related "news for nerds," was one of the earliest sites online to posted about the 9/11 attacks. On 9:12AM, only 23 minutes after the first attack happened, Rob Malda, also known as "CmdrTaco," the founder of the site, shared a post on Slashdot (Figure 2-9). Be briefly reported what happened to the World Trade Towers and noted, "Normally I wouldn't consider posting this on Slashdot, but I'm taking an exception this time because I can't get news through any of the conventional websites, and I assume I'm not alone." (Slashdot.org, 2001) On Fark.com, a weblog that allowed users to submit and comment on comedy stories, editor Drew Curtis also made a statement addressing the transformation of the role of the site during 911, "One thing we've had trouble with in the

past few days is making a smooth transition from a comedy news site to a *real* news site." The blog site posted 157 entries regarding 9/11 attacks during the days following September 11. Curtis claimed, "We really hadn't ever thought we would need to, and probably wouldn't have except for the fact that on 9/11 every major news site went down and someone had to pick up the slack" (Halavais, 2002, Part 3, p. 3-4, emphasis added)

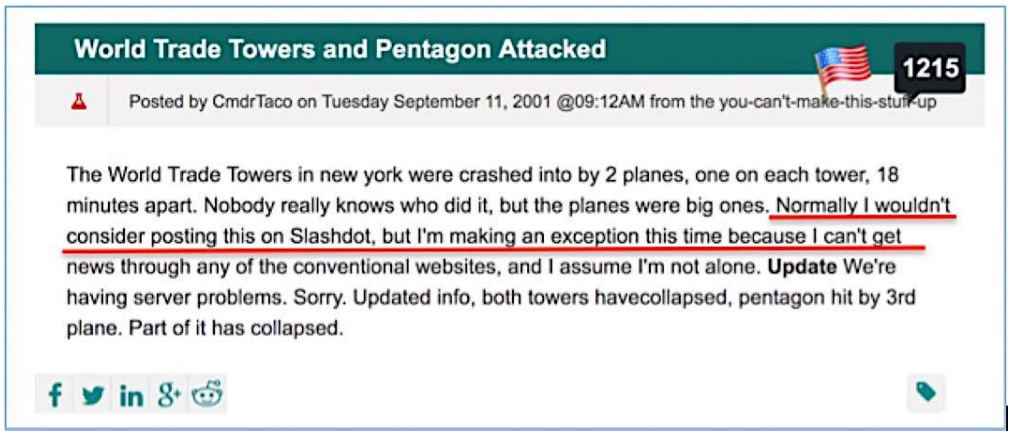

***Figure 2-9*: Slashdot.com on September 11, 2001. Source: Slashdot.org, emphasis mine.**

In a national crisis, it might be the sense of duty that encouraged these non-news sites to take up the journalistic role in public service. The public was eager to learn more about the attacks through information of different sources. Both Slashdot and Fark had significant surge in users, tripling their normal peak-hour traffic. Slashdot estimated that its editors received at least five times as many emails as FoxNews.com did between September 11 and September 12 (Miller, 2001). "This week," many software, open-source, and information sites realized, "news has been the undisputed traffic king." (para. 16) An online news market that was both socially and commercially important became more and more visible in the aftermath of 9/11.

Online News Aggregation

Slashdot and Fark were among the actors on the Web who provided 9/11-related information, including first-hand accounts, images, analysis, and background, during and after the attacks. A common practice on their sites was online aggregation. That is, they pooled, put together, organized, and presented information online that they saw as important to help meet users' informational, emotional, and practice needs in a national crisis. Aggregation was a common practice of news media as well. For example, BBC America compiled eyewitness accounts and information about 911 from thousands of audience emails (Halavais, 2002). Many local media aggregated, too, noted Crawford (2002), "the San Francisco Chronicle, did a magnificent job on September 12 and beyond—as did most other metropolitan newspapers—packaging news, background, and commentary in ways that only a major print newspaper can do effectively" (para. 5). In an emergency like 9/11, the practice of news aggregation required huge responsibility in terms of the accuracy and reliability of the information. On September 13, the site posted a rumored story about CNN using a ten-year old footage to fake the images of Palestinians dancing in the street after the 911 attacks. The story was proven untrue as Reuters and CNN both confirmed that the footage was taken on September 11, 2001. In apologizing for the post of an article that carried untrue claims and received public criticism, the editor of Fark.com stated, "that was first time that I realized we'd passed from being a fun silly website to being a real source of news for people. This whole journalistic integrity thing really hadn't applied before." (Halavais, 2002, Part 3, p. 4)

News aggregation existed before 9/11, but it became a common practice of many websites in light of 9/11 to meet the public's need for information. While the wide adoption of news aggregation in the digital environment prepared a context for Google

before it rolled out its own news aggregation service in 2002, the journalistic responsibilities associated with the practice of news aggregation required Google's news aggregation service to have a good method for aggregating news professionally and ethically, in other words, to normalize journalistic norms and values.

Immediately Post-911

The attacks of September 11, 2001 had profound impact on internet users' online behavior, especially the way they used the internet for news. About a year after 9/11, the Pew Research Center issued a report, based on findings of a daily tracking survey among 1,527 internet users, that found that 66% of American internet users had used the internet to get news (Pew Research Center, 2002). Among the respondents, 32% reported that they had gotten news online more frequently after September 11, 2001. This percentage in the use of the internet to "get news" was much higher than all other categories, including "send or read email" (13%), "get health information" (9%), "use government agency websites" (17%), "get mental health information" (12%), and "make donations online" (26%). Among the internet users who reported they had increased their level of using the internet to get news, 43% of them cited 9/11 as a major reason for the change. One year after 9/11, internet users' demand for online news was increasing. It looked like a good time for Google to enter the news market.

Google News Founder and Motivation

Google News founder Krishna Bharat has told interviewers that two things motivated him to create the aggregator: the inefficiency of news media and the idea of news as experience. The inefficiency of news media, according to Bharat, is reflected in two areas. First, technically, the issues that caused the "news paralysis" on September 11,

such as server capacity, loading speed, and search effectiveness, concerned Bharat. These issues also let Bharat think that Google's technological advantages could solve the problems that failed news media on 9/11. Second, journalistically, Bharat pointed out news media failed to provide new developments about 9/11 in a timely and in-depth way. "After Sept. 11," said Bharat, "when all the newspapers were recording who, what, when, where—there was a big question of why. Why did this happen? What's going to happen in the future?" (Kramer, 2003, para. 22) To solve this problem, Bharat called for multi-perspectival news, allowing the public to learn about a topic from different sources and points of view. "Bringing those views together seemed like a good social function. Helping people understand multiple points of view, and hence becoming wiser for it — whether they agree with it or not — just understanding there is another point of view is enlightening" (Glaser, 2010, para. 7).

Bharat also talked about the idea of news as experience. According to Bharat, in the digital context, journalists' job is to create news as an overall experience rather than a simple product. It's like consuming at a Starbucks store, Bharat said: "You think you're there for coffee, but you're really there for the full experience," including "the music, the collegiality, the reading space, the aroma, [and] the brand. You pay for coffee, sure; but 'the whole experience is what counts'" (Garber, 2011, para. 7). For Bharat, a Ph.D. in computer science and a member of Google's then 10-person research lab, the solution to solving the efficiency and experience problems was technology. "Fundamentally," Bharat said in one interview, "I wanted to build a tool that would automate this" (Kramer, 2003, para. 22). That ambition started shortly after 9/11.

Early Form of Google News

In December 2001, only three months after 9/11 attacks, Bharat had already demonstrated internally his dynamic news project, which became the prototype of Google News. In the pilot project, 20 news sources were crawled once an hour. Then the most recent stories on selected topics were returned, with the entire process automated. The success of an internal presentation allowed Bharat and his team to continue to work on the project, which they expanded to cover more news sources. Figure 2-10 below shows the internal test version of Bharat's dynamic news project in December 2001, which was the embryonic form of Google News (Beta) introduced publicly the following year (Shiels, 2018).

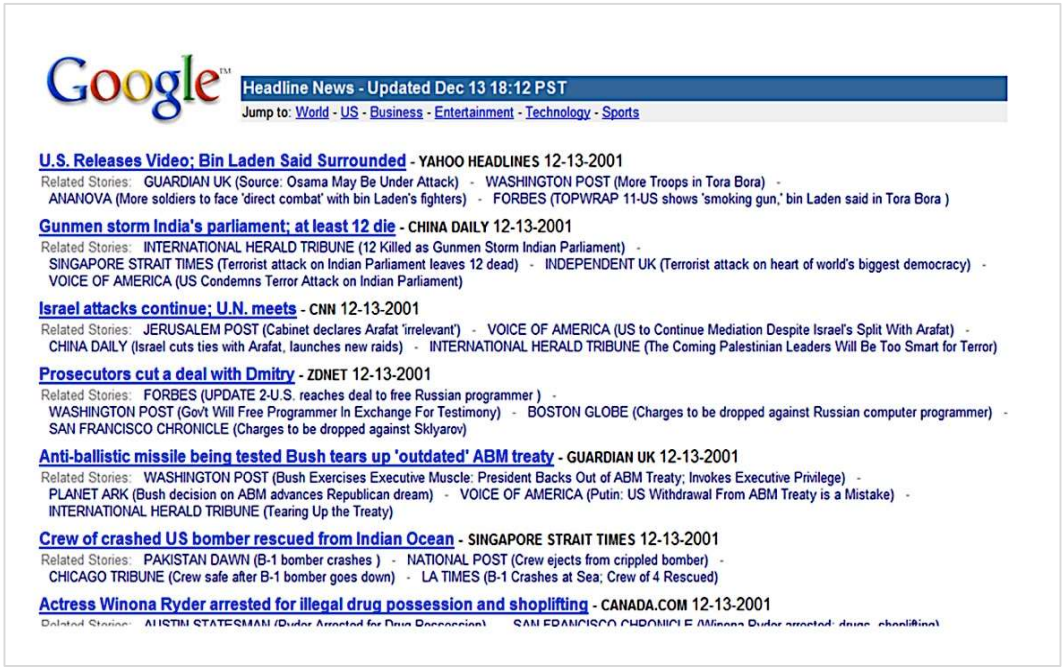

***Figure 2-10*. An Internal Test Version of Bharat's Dynamic News Project in December 2001. Source: Shiels, 2018.**

Bharat's dynamic news project was very different from Google's /news folder (google.com/news) introduced by Google on September 12, 2001. The latter listed only a collection of news sources with their links. When a user clicked on the link, the user was

directed to the news source's website, rather than any specific news story. Therefore, Google's /news folder was more of a *news source* aggregation rather than a *news* aggregation service. Bharat's model was a genuine news aggregation service, as it aggregated news *articles* from different news sources, which also required more powerful algorithms. News stories were presented by headlines rather than news sources. When a user clicked on a news headline, the user was directed to that specific news story on its original news site. This kind of deep linking fragmented news into individual articles that debundled the news packages that news media outlets tried to sell to audience. On the one hand, Google's news aggregation service provided a new way to experience news—it was automatic, story-based, and run by Google's ranking algorithms. On the other hand, this approach tends to lose the logics that bundle up news packages. These logics may involve thoughtful editorial decision-making when news media play their gatekeeper role as well as news media's economic consideration to increase revenue by bundling together content that audience may or may not be interested.

In late January 2002, four months after 9/11, Google gave the news aggregation page a public trial by putting a text-based version on Google's public site (Figure 2-11). This was the first public appearance of the early form of Google News, but it was not yet an independent service. A link on google.com/news that read "Click here for headline news" directed users to the news aggregation page called "Headline News." "Headline News" adopted the section-based structure that news media used to organize news on their websites. News stories from 155 news sources were categorized into Headline News, World News, US News, Business News, Entertainment News, Science and Technology News, and Sports News. Most of these news categories were kept the same

on Google News throughout its history.

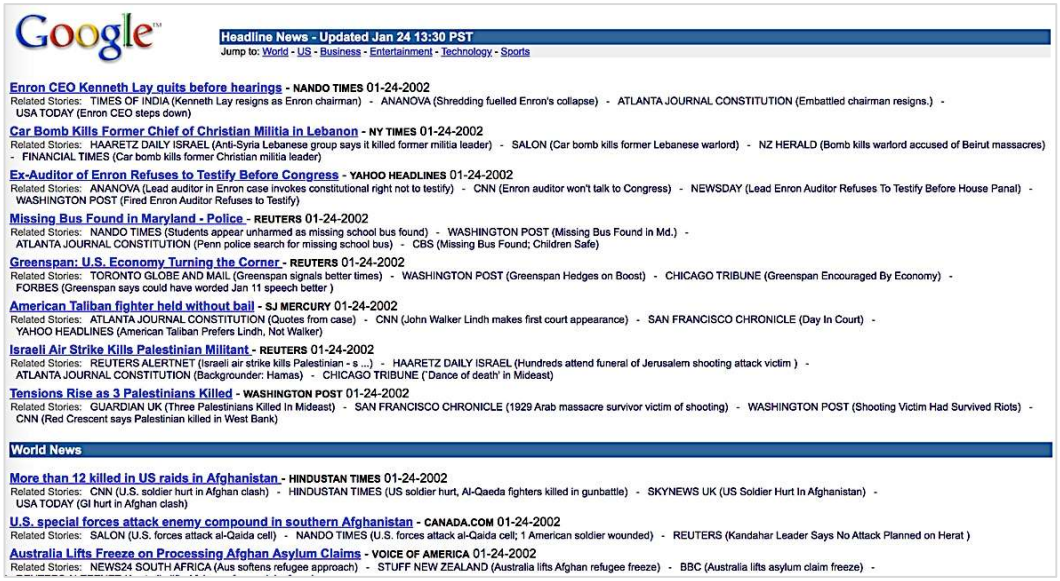

***Figure 2-11*. The Public Trial of Headline News in January 2002. Source: Internet Archive.**

After the public trial, Google's news aggregation service quickly attracted a large group of interested users, about 70,000 a day (Hammonds, 2003). Within the following few months, the news team coped with users' feedback, upgraded hardware, designed a user-friendly interface, and fine-tuned algorithms. As early as March 2002, Google News Search (Beta) was available under the URL news.google.com. In September 2002, when Americans were commemorating the first anniversary of 9/11, Google News was formally introduced on Google homepage with the slogan "Get your news the Google way—Try Google News." That was when Google News became a standalone service. Within just one year, from September 11, 2001 to September 2002, Google made adjustments about its news aggregation service for multiple times until it formalized it in 2002 (Table 2-1). These frequent adjustments made in a short period of time revealed Google's changing perception of news as a type of information and as a business. They

also demonstrated Google's capability in making fast judgment about the information market and turning such judgment into actions agilely and effectively. Since then, Google News has had many changes on its website, as shown in the next chapter.

Time	Version	URL	Content	Layout
9/11/2001	Homepage info and links	google.com	Info and links to small number of news sources	Short info and links displayed on homepage
9/12/2001	/news folder	google.com/news	Lists of news sources and resources	Lists
Dec. 2001	Dynamic news project	n/a internal testing	7 news sections, including headline, publication date, related stories. From 20 news sources	Text-only List-style
Jan. 2002	Headline News	google.com/news public trial	7 news sections. From 155 news sources	Text-only List-style
Mar. 2002	Google News Search (Beta)	news.google.com	7 news sections. Added search function (more details in Ch. 3)	Text-only List-style
Sep. 2002	Google News (Beta)	news.google.com	7 news sections. Added news snippet, news image, and more, from 4,000 news sources (more details in Ch. 3)	Text+image Segment-based

***Table 2-1*. Different Versions of Google's News Aggregation Service in its Early History. Source: author.**

Discussions

This chapter traced the origin and early history of Google News through three historical periods: pre-9/11, 9/11, and immediately post-9/11. In the pre-911 era, Google

adopted the directory model from early search engines and the ODP that had human editors organize the Web. In 2000, when Google introduced its own Web Directory, it appropriated ODP by using PageRank algorithms to rank the ODP data. Google used the ODP data as raw data to promote a different, algorithm-based model to manage information on the Web. In the meanwhile, Google gave users the opportunity to view the information by both models, considering that algorithms were not able to completely replace the directory model managed by human editors. In 2004, when Google's search function became more powerful, Google Web Directory was no longer a main feature on Google homepage, indicating the decline of the directory model, which was finally retired in 2011. During this process, Google used its specialized searching technology to differentiate itself from peer companies and became the dominant player in the search market.

By 2001, when the 9/11 attacks occurred, Google was still in the appropriation stage, having introduced its own Web Directory, run by the PageRank algorithm. The algorithm model, however, was not yet mature enough to replace the human editor model. On September 11, 2001, having witnessed the news paralysis and the strong demand for online news, Google started to take up a part of journalism's role in informing the public, although in a very limited way. This was a deliberate decision that changed Google's position as a pure search engine. While Google previously had not prioritized news, when Google set up a new /news folder on September 12, 2001, the move indicated that Google had started to treat news as a special type of information. By listing news sources and public resources for users to access in the aftermath of 9/11, Google's adoption of the journalistic role was strengthened.

The national emergency gave Google an opportunity to realize a new information market associated with news—a special type of information with both economic and social values. Google was not a news media outlet, but it moved fast to integrate news into its Google business by appropriating news with its technologies. Three months after 9/11, when the early form of Google News was rolled out inside Google, and later when it was introduced to the public, Google News transitioned from an adoption stage to an appropriation stage, in which it integrated news and the specialized technology underlying Google's web search model. When Google News was promoted on Google's homepage for the first time, Google announced, "Get your news the Google way." The "Google way" was a very different idea from how news media handled news—news media have humans report, write, and edit news and display news online while Google uses these human-handled news stories as an information pool from which a small body of news stories are selected by automatic, computational algorithms.

The "Google way," as the founder of Google News envisioned, was a new way to "experience" news. For example, it was automatic, it presented news across different media sources worldwide, and it also helped users to pinpoint the specific news story that they wanted to see without having to going through the whole news bundle that news media tried to sell to the audience. Having adopted journalism's role in informing the public and appropriated news with computational technologies, Google gave birth to Google News. Google News' early history was characterized with normalization in relation to the news media industry.

Since then, Google News has made many changes, big and small, on its website. The next chapter will provide a close examination of how Google News' homepage

changed between 2002 and 2019, which provides an account of Google News' evolutionary trajectory and the driving forces behind its evolution.

Chapter 3. Structural, Visual, and Functional Trends on Google News Homepage Design

Google presents its news aggregation service via Google News website. This website is a product of Google; it's also a platform that connects Google, news media, and users. The analysis of this website provides an important window through which the author can explore the evolution of Google News over time in terms of how Google News perceives and presents news, how Google News defines its role in the news ecosystem, and how Google deals with its relationship with the news industry. This chapter focuses on the website's interface design, especially the design of its homepage, over the past seventeen years by examining the structural, visual, and functional changes and trends.

Studies found that the design of website interface affects the quality and effectiveness of the website, such as the website's usability and functionality, including the visualization, interactivity, and navigation tasks the given website attempts (Klett, 2002). The design can have an effect on users' experience on the website as well, such as users' first impressions (Tuch, Presslaber, Stöcklin, Opwis, & Bargas-Avila, 2012), learning activities (Klett, 2002), and online purchase behavior (Blanco, Sarasa, & Sanclemente, 2010). In addition to these functional considerations, scholars also found interface design carries aesthetical and emotional implications (Lee, Kim, & Choe, 2001). In the case of websites that focus on news, Shaw's (2010) study provided evidence that the design of news websites influenced users' perception of news credibility. The author argued that news media applied agenda setting and framing tactics through the use of visual characters, not just textual content, in their online design to influence audience's

reception. Zettl (2013) stressed that the visual factors on the website had impact on users' perception of message or content. While a large body of current literature on homepage design focuses on effect analysis through the examination of visual elements, Cooke (2005) examined the structural elements, in addition to graphic design, on news websites. The author identified styles of news website design in the 1990s and early 2000s, such as the three-panel layout, the multiple points of entry model, and the information module structure. Cooke pointed out that these design trends incorporated the ideas and techniques that traditional news media used in designing the visual presentation for newspapers and broadcasting programs.

In studying website design, scholars such as Hoffmann and Krauss (2004) argued that interpretive research benefits visual studies in investigating the meanings, dynamics, and relationships behind isolated visual elements. Echoing Hoffmann and Krauss, the goal of this chapter is not to measure the effect of homepage design, but to provide a historical analysis of the changes and trends in terms of the structural, visual, and functional evolution on Google News homepage.

This chapter provides a close examination of how Google News has presented itself to users, via its homepage, over the past seventeen years. The chapter is based, in part, on manual examination and comparison of about 6,000 Google News home pages available in the Internet Archive's digital archive. In those pages, which were from the years 2002 to 2019, there were about 180 days[4] when changes, big or small, were observed on the Google News homepage. All observed changes were recorded, examined, and classified in a manner explained in detail in this chapter.

[4] Since there may be multiple changes on a given day, this study counts the number of days when changes are observed.

In addition to the web archival research, this chapter also relies on historical materials, including blogs, news articles, web articles, and scholarly literature, to contextualize changes in how Google News presented itself. Some of these materials were also located via pages archived in the Internet Archive. On the archived webpages, if there were clickable links to material that could aid in understanding the historical background of Google and Google News—*e.g*. Google's announcements, instructions, demonstrations, etc.—those links were followed. If linked information appeared important but was no longer clickable on the archived webpage, the author searched online for it separately, in an attempt to collect as many historical materials as possible.

When "Google News" Was "Google News Search"

Google News is located at the URL news.google.com. While Google News is publicly known for being launched in September 2002, as early as March 2002 the Beta version of news.google.com appeared online. The webpage had a simplified header, a single-page, long-scrolling main body section, and a very brief footer. Except for the Google logo on the very top of the page—rendered in the classic Google color combination of six letters in blue, red, yellow, blue, green, and red—the page featured mostly blue and black text, with hyperlinks that turned red when the cursor hovered over them. Other informational text appeared reversed out, with white text on a red background (*e.g*., the name of news sections on red dividers) or red texts on white background, as in the slogan "Search and browse the latest headlines" (Figure 3-1).

The website's logo —"Google News Search, rather than "Google News" (See Figure 3-1, black arrow), appeared in the header section of the webpage, with the search box right below. Next to the search box, there were links to general tips about using

Google as a search engine. The slogan “Search and browse the latest headlines” appeared under the search box. The slogan seemed to indicate that searching, rather than browsing, was the main function of the service (Figure 3-1). The rest of the webpage was dominated by the main body section containing headlines, news sources, and publish dates, which was followed by a very simplified footer section, with only two links directing users to the homepage of Google’s general search and all other information about Google (Figure 3-2).

***Figure 3-1*. Google News Search (BETA) Homepage in March 2002. Source: Internet Archive, emphasis mine.**

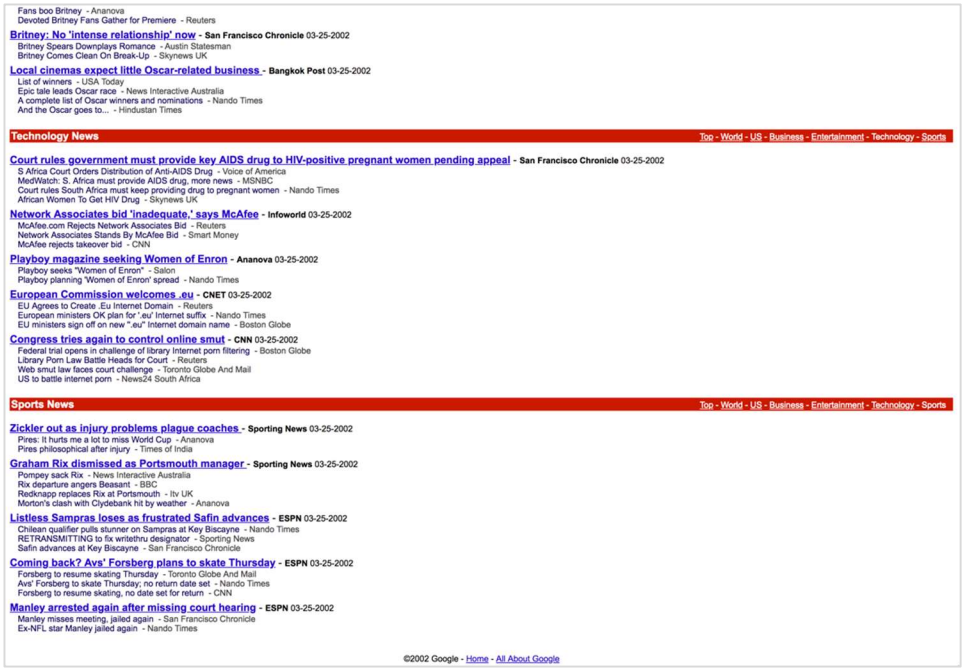

***Figure 3-2*: Google News Search (BETA) Homepage in March 2002—continued. Source: Internet Archive.**

The search function in Google News Search was new. In the early testing version that Bharat presented at Google in December 2001 (See Figure 2-9 in Chapter 2), users were only able to browse news articles arranged by algorithm that were listed on the webpage. In the March 2002 version, with the addition of the search function, Google normalized news with its search technology by using news content as candidate information, as any other types of information, for searchers. The introduction of this service, "a novel approach to news," described it as a news aggregator that "presents information culled from many of the world's news sources" but also as a search engine that "automatically put related stories together in the same search result." (About news search, 2002, para. 1). In addition, it stressed the service's capability of the "continuous updates throughout the day" that allowed users "keep up to date with what's happening now and learn about the stories that led to the most recent developments." (About news

search, 2002, para. 1). This change, made several months after the 9/11 attacks, seemed to respond to the 9/11 context when Google failed to provide relevant information about 9/11 attacks because its index was browsed a month earlier. (See Chapter 2.)

In March 2002, the main body of the Google News Search webpage was a one-column, list of articles, which was split by red rules, containing white type, into seven topic-based blocks: top news, world, US, business, entertainment, technology, and sports (Figure 3-1, 3-2), sections that were introduced in the testing version as well. Each topic-based section covered five news events or issues, except the Headline News section that had eight news stories. Each of those news stories was represented by a headline, displayed in blue that served as a hyperlink to an article about that event. At that time, only the headline, the news source's name, and the publication date were made available on news.google.com, with the headline clickable linking to the original news story. The display did not include news snippets, photographs, or videos—features that later sparked legal disputes, as discussed in Chapter 4.

In spring 2002, the design of the site news.google.com sent the public an ambiguous message about this service: Was it a news search engine, or as a news aggregator? If news.google,com were being positioned as a news search engine, a simple search window like the one on the simplified Google home page would have been enough. Displaying headlines of news stories would not have been necessary. If news.google,com were being presented as a news aggregator, the search function would not have been necessary. The version of news.google.com available after March 2002 combined the two roles. Although the header section of the webpage defined it as a

search engine specialized in news, the main body presented the service as a news aggregator.

Such ambiguous self-position was reflected in the logo design as well. In the 2002 version of the logo, the word “Google” was big and tall while the word “News” was much smaller and placed below the word “Google” as if news were subordinate to Google as a brand (Figure 3-1). The design of the logo did not change much until May 2009, when the word “Google” and the word “news” were in similar size, which made Google News look more like a standalone news aggregator. The Google News logo changed for several time after that, for example, the word “news” gets bigger in 2009 and then gets smaller in 2011 before growing again in 2017 (Figure 3-3).

***Figure 3-3*. Google News Logo Change. From top to bottom: 2002, 2009, 2011, and 2017. Compiled by author.**

This ambiguity may reflect Google News’ adjustments in self-positioning in its early days. This ambiguity in self-positioning provided Google an opportunity, although

ultimately an unsuccessful one, for defending itself when it was involved into copyrights lawsuits in Europe. For example, in a copyright lawsuit in Belgium, Google defined Google News as just a search engine, because search engines enjoyed certain safe harbors under the European Union's copyright law while news aggregators didn't. (See Chapter 4.)

In March 2002, the news aggregation function in Google News Search was very basic. There were no extra features, just the list of news headlines and news sources. Users could simply scroll down until they found a news headline that interested them. Clicking on the news headline, users would be directed to the original news website. It was very straightforward and, somehow, boring, too, as Google focused on a limited aspect of online news, that is, online news as texts, or online news consumption as a *reading* experience only—consuming online news by reading the headline, followed by text. If, as Bharat envisioned, Google aimed at making news consumption a different "experience," this kind of scroll down-click-read design was not very appealing. Compared to Google's list-like presentation of news in early 2002, Cooke's (2005) study showed that in those years some news media applied different design trends on their websites, for example, in 2001, ABCNews.com used information modules that "eliminated cumbersome scrolling by departmentalizing information in a smaller space." (p. 40) As shown in Figure 3-4, the homepage of ABCNews.com in 2001 involved rich visual elements and a module-based, rather than list-based, structure, which looked very different from the homepage design of the early versions of Google's news aggregation service, including Google News Search in March 2002 (See Figure 3-1 in this chapter) and the internal testing version in December 2001 (See Figure 2-9 in Chapter 2). Some of

the design ideas that news media used for their websites in the early 2000s were adopted by Google in September 2002 when Google News Beta was launched (See Figure 3-5).

***Figure 3-4*. Homepage of ABCNews.com in 2001. Source: Internet Archive.**

Major Redesigns and Implications

As Google News evolved and technologies advanced, more features were added. This section examines changes to the homepage of news.google.com from its launch in 2002 to 2019, as documented in the Internet Archive. Some changes were minor, such as the change of the color of text. Some changes were gradual, such as the addition/removal of certain features. Other changes were significant, resulting in an overhaul of the homepage. Next, major redesigns in 2002, 2010, and 2017 will be investigated respectively, followed by the examination of gradual changes throughout the past seventeen years. These changes carry important implications for understanding how Google has defined the role of its news aggregator, how it has perceived news, and how it

responded to the traditional news media industry's actions and reactions in responding to the development of Google News. By examining these topics, this chapter outlines the interaction of normalization and differentiation that drove the decision-making behind Google News' website changes.

Google News in Fall 2002

Since March 2002, Google News has experienced three significant homepage overhauls. The first happened in September 2002. When entering news.google.com, the homepage looked very different from six months ago. By late that month, the website was being introduced to the public as "Google News," rather than "Google News Search." (See Figure 3-5.) The search box was still available, but the search function was not highlighted in the "About" page, which was different from the situation in March. In the introduction to Google News on the "About" page, Google described its news aggregator as "highly unusual" because "it offers a news service compiled solely by computer algorithms without human intervention" (About news search, 2002, para. 2).

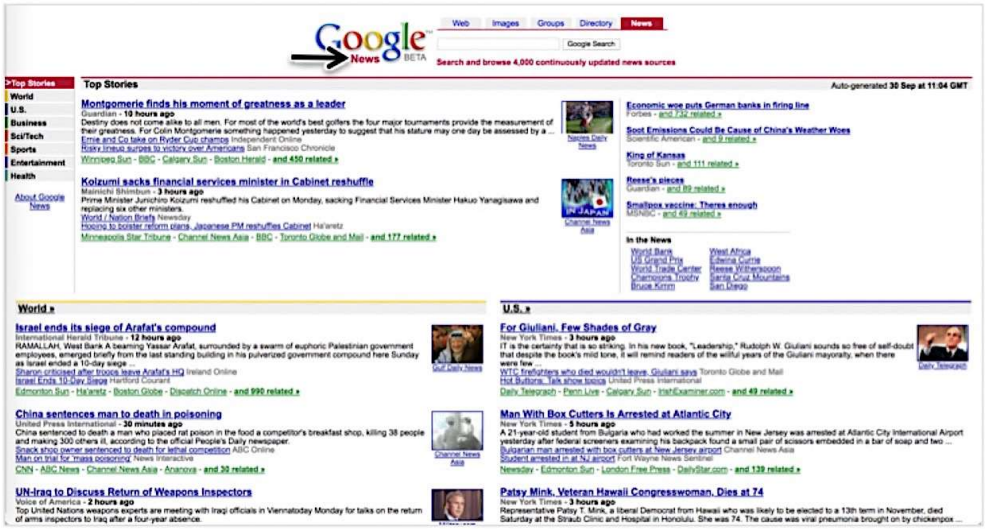

***Figure 3-5*. Google News (BETA) Homepage in September 2002. Source: Internet Archive, emphasis mine.**

The redesigned page downplayed the aggregator's search function and highlighted the news aggregation functions of Google News, including their scale, *e.g.,* "culled from approximately 4,000 news sources worldwide," In addition, it touted the frequency of updates, ("updated continuously throughout the day") and the ranking function ("present the most relevant news first"). Google highly stressed the access to multiple perspectives through an automatic process free from human interventions: "While the sources of the news vary in perspective and editorial approach, their selection for inclusion is done without regard to political viewpoint or ideology." (About news search, 2002, para. 1, 2)

All these features looked very different from the practices of traditional news media, but Google's explanation of the decision-making about what stories were published on the Google News homepage provide evidence that the workings of Google News involved both differentiation and normalization relative to traditional news media. On the one hand, how Google selected news was "very much in the tradition of Google's web search, which relies heavily on the collective judgment of web publishers to determine which sites offer the most valuable and relevant information." On the other hand, Google News "relies in a similar fashion on the editorial judgment of online news organizations to determine which stories are most deserving of inclusion and prominence on the Google News page." (About news search, 2002, FAQ #7) It was also during this time that Google defined its news aggregator as "a valuable source of information on the important issues of the day," (para. 2) a role similar to journalism's gatekeeping role. How Google did both normalization and differentiation algorithmically will be elaborated

in Chapter 5. Here, let's focus on Google News homepage design to find some clues about the two trends.

In Google News (BETA) in September 2002, the homepage design abandoned the previous one-column, list-like style, in favor of a two-column design featured "modules" as in the case of the homepage design of ABCNews.com in 2001 (See Figure 3-4 and Figure 3-5). In the header, News was listed with other four Google services: Web, Images, Groups, and Directory. These were also the five services promoted on Google's main page. This change indicated an upward organizational status of Google News, as Google introduces new products from time to time but only promotes only some of them on its main page.

In the main body of the Google News homepage, the Top Stories section was a relatively independent area, which contained two main stories (in a bigger, bold font) as well as links to related stories (in a smaller, regular font) and news sources. On the left, a navigation bar with links to eight news sections was added. The rest of the main body area was divided into two columns to reflect the news sections listed in the navigation bar: World, U.S., Business, Sci/Tech, Sports, Entertainment, and Health news, the last of which had not appeared in the early 2002 version. The previous Technology section had been renamed as Sci/Tech and was moved up before the Entertainment section, indicating the growing importance of the area.

These news sections have remained mostly unchanged throughout the years, becoming the standard template of Google News. This selection and arrangement of news sections were a common practice of traditional news media. Google News adopted this practice to organize its website, but Google and news editors do this differently. News

editors sort news stories into different sections depending on their professional judgment. Google used grouping technology to automatically group related news stories into news sections. Google News' grouping technology also differentiated from news editors' grouping practice in the way that news editors seldom publish different accounts of the same event from different news organizations. Google's grouping technology broke the organizational boundary as its robot crawled thousands of news sources worldwide across the web.

Unlike the earlier version, which listed only news headlines, sources, and times, the September 2002 Google News (BETA) version included a short paragraph of news snippet and news photos. These elements were the controversial ones that involved Google in copyright disputes later. (See Chapter 4.) A column on the right highlighted five repeated articles selected from the eight news sections.

Below that was a new "In the News" feature, which included ten terms, known as "named entities" in the field of natural language processing. These terms—names of people (*e.g.*, President Bush), places (*e.g.*, Gaza City), and topics (*e.g.*, homeland security)—were the trending search terms about news on that day. The use of named entities was another application of Google News' grouping technology, which organized news stories surrounding specific—and usually trending—people, places, locations, things, and events across different news sources. Using grouping technology and named entities, Google un-bundled the news arrangements. Studies show that how news was selected, arranged, and packaged is often the results of news media's gatekeeping and agenda setting functions (White, 1964; Shoemaker and Vos, 2009). By de-bundling the news, Google challenged the news media's traditional roles in these areas. Google also

re-bundled news the "Google way" through its own technical rules, for example, using named entities to cluster news stories. Both de-bundling and re-bundling involved the adoption of certain news elements; they also generated new norms. For example, named entities could help users build a habit of consuming news based on trending topics rather than editorial decisions; news production could change its patterns accordingly as well, for example, by using SEO rules.

The interface design adopted in the September 2002 version continued to be used as a template—with gradual adjustments over years—until the Google News homepage was redesigned in 2010.

Google News Redesign in 2010

The second time the Google News website was given a major redesign was in 2010. After this redesign, which Google News considered the biggest since the site's launch in 2002 (Beckmann, 2010), the homepage looked very different[5]. One major change was that in addition to the navigation bar on the left side and the news stream in the middle column, an additional column was introduced on the right side of Google News homepage (Figure 3-6, blue arrows).

[5] Some changes in the 2010 version were not completely new by the time this redesign was carried out but had evolved over time. For example, the new "News for you" section embedded under the "Top Stories" section, which allowed users to personalize news, was introduced as the "new heart" of the homepage in this version (Stolt, 2010). As a matter of fact, this feature had a long history. This type of gradual change will be discussed separately later in this chapter.

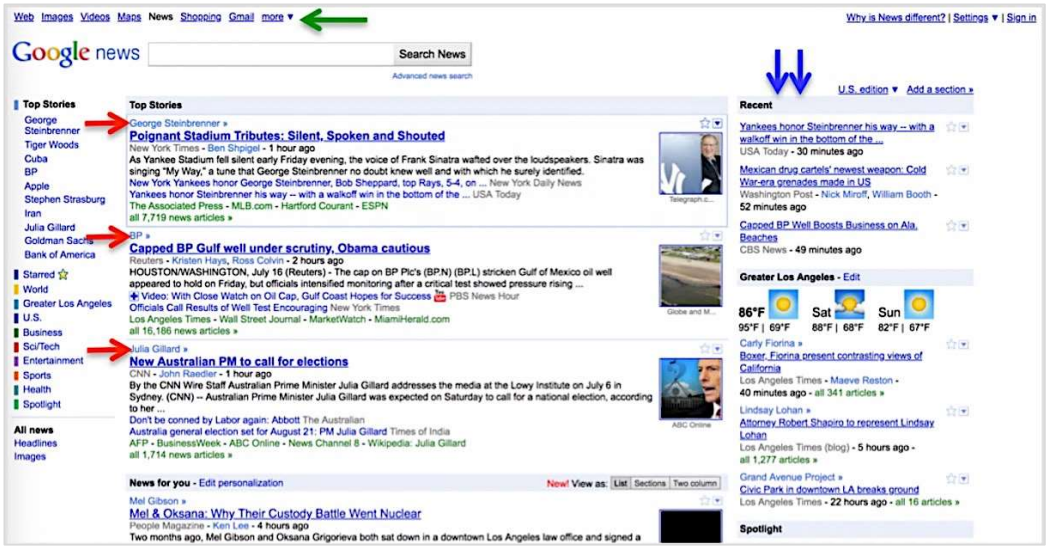

***Figure 3-6*. Google News Homepage in 2010. Source: Internet Archive, emphasis mine.**

When it was first added to Google News' redesigned homepage in July 2010, the right-hand column was a busy area with as many as six features, including:

- "Recent": A segment that displayed a small group of recent news articles shortly after they were published by their original news sources, usually added to Google News in less than an hour after their online publication.
- "2010 FIFA World Cup South Africa": A temporary special news section available during the 2010 World Cup in South Africa, with links and information about game schedules, standings, and teams.
- Local coverage: A segment including weather information and a series of local news articles tailored to the location of the user.
- "Spotlight": A segment that introduced as many as twenty long-form, in-depth feature stories of "lasting value," as opposed to quick updates or breaking news.

- "Google Fast Flip": A place to introduce Google products for new ways to consume news. For example, Google Fast Flip was a Google tool that allowed users to quickly flip through pages online like flipping through a magazine.
- "Most popular": A segment with ten stories "most popular" among readers, although Google didn't specify how "most popular" stories were defined and selected. "Most popular" stories were different from "Top Stories," as the latter were important stories defined by news editors at publications. The segment also allowed users to click on the name entities for other related stories.

The addition of this column expanded Google's news aggregation features, which normalized more journalistic practices. For example, World Cup news was not grouped into Sports News in the standard news sections but was presented as a special section in the right-hand column. Establishing special sections during major news event was a normal practice of news media, which was adopted by Google News. In a similar sense, recent updates, local coverage, and in-depth features were also separate from standard news sections to highlight different news genres. This column also had room to promote new ideas about news, *e.g.* Google products that aimed at making news a different experience and using popularity as a criterion to evaluate news. Specific features in this right-hand column have changed from time to time, but the creation of this column as a separate space on Google News homepage opened an area for flexibility. This space has been an arena of negotiation to balance interests of different stakeholders, including Google, news media, and users. For example, while this column featured "Most popular" stories among users, it has also accommodated "Editors' Picks" to allow editors of news publications to choose and highlight news stories that they define as quality journalism.

While Google promoted its own products here, it has also introduced features such as "Fact check" to respond to external criticisms.

In the 2010 version of Google News, each news story in the middle column was given a topic, such as "George Steinbrenner," "BP," "Iran," etc. (See Figure 3-6, red arrows.) The topics were also the named entities that appeared in the "In the News" feature. In this version, named entities were assigned to each news story that could help users and news publishers realize and learn the pattern in terms of the match between a specific news story and a specific named entity. As will be discussed in Chapter 5, named entities are building blocks of Google's algorithm systems to tag and cluster data. The more users use named entities to find news they were interested in, the better they could follow Google's algorithmic rules. Over time, users could also get used to use these named entities as keywords when they searched for information on Google. These data would, in turn, benefit Google as it refined its algorithms. This is an example of how Google interlocks its algorithmic design with its product design—product features often reflect and serve algorithmic designs.

This kind of interlock effect is not completely different from the traditional media industry's vertical integration strategy, in which media companies promote their own media products across business channels that they own in order to maximize profit. The media industry often does this by owning both the production and the distribution links of the supply chain, for example, the Walt Disney Company, an American media and entertainment conglomerate, owns companies that produce and market its own film and television products. In the case of Google and other tech companies, the integration is done mostly through technological means, not as much through ownership (Van

Couvering, 2011). As the example above shows, since product design and algorithmic design are often interlocked, one who uses the product also has to follow the algorithmic logics. As a result, users and content producers are captured into the infrastructural control of digital platforms.

Once users learned how to use named entities to find news that they were interested, data about their use patterns could be fed back into Google's algorithmic systems to analyze users' preferences and interests. These data would serve as training data to improve Google's development in machine learning. Over the years, Google News has trained its users as well as its algorithms for news experience driven by artificial intelligence, particularly news personalization. Such training gave rise to the AI-powered Google News that was introduced in 2018. Before that, however, Google News had another website redesign in 2017, which also introduced changes that prepared for the AI-powered Google News.

Google News Redesign in 2017

A new version of Google News, with a "clean and uncluttered" (Paka, 2017) user interface, was introduced in June 2017. In this new version, news stories were held in various "news cards"—the homepage was segmented into card-shape areas. Except news photos in the very front of each news card, most of the elements on the homepage were not colorful with white or light grey backgrounds and black text (Figure 3-7). Perhaps the most important, and unannounced, change that contributed to the "uncluttered" look was that the news snippet for each story was no longer available. In each news card, only "key elements" of online news, such as headlines, publisher names, and article labels, were shown. (See Figure 3-7, red arrows.)

This change appeared to be the result of several copyright disputes Google had been involved in, especially in Europe, since launching its news aggregator. One of the concerns of the European news publishers who sued Google was that the combination of the news headline, news photo, and news snippet that Google News made publicly available on its website provided "more than enough" for users to understand the news content and therefore significantly reduced the likelihood of users clicking on the news link and browsing the news publishers' own websites. (See Chapter 4). It was against this backdrop that Google News removed all news snippets in its 2017 redesign, after fifteen years of using them. Clearly, some of the changes on the Google News homepage reflected the tension and negotiation between Google and the traditional media industry.

Another change in the 2017 version of Google News homepage was the addition of article labels throughout the news stream. (See Figure 3-7, blue arrows.) Article labels are "predefined, generally understood terms" tagged on news stories to describe the nature of the news content (Google, n. d.). Publishers apply appropriate labels, by, for example, embedding appropriate tags in HTML markup, when they opt in to Google News. Labels shown on the Google News homepage "help users select what they want to read," according to Google (Paka, 2017). In the 2017 version of the home page, labels that were not visible to users were also applied in Google's algorithm systems to classify content. In this study, article labels were first observed as early as in May 2011. But before 2017, these labels were used only in the "Top Stories" section. In 2017, these labels were applied more widely throughout the whole news stream across different sections. According to Google, "People have told us these labels identify important facets of a story and provide more context" (Paka, 2017).

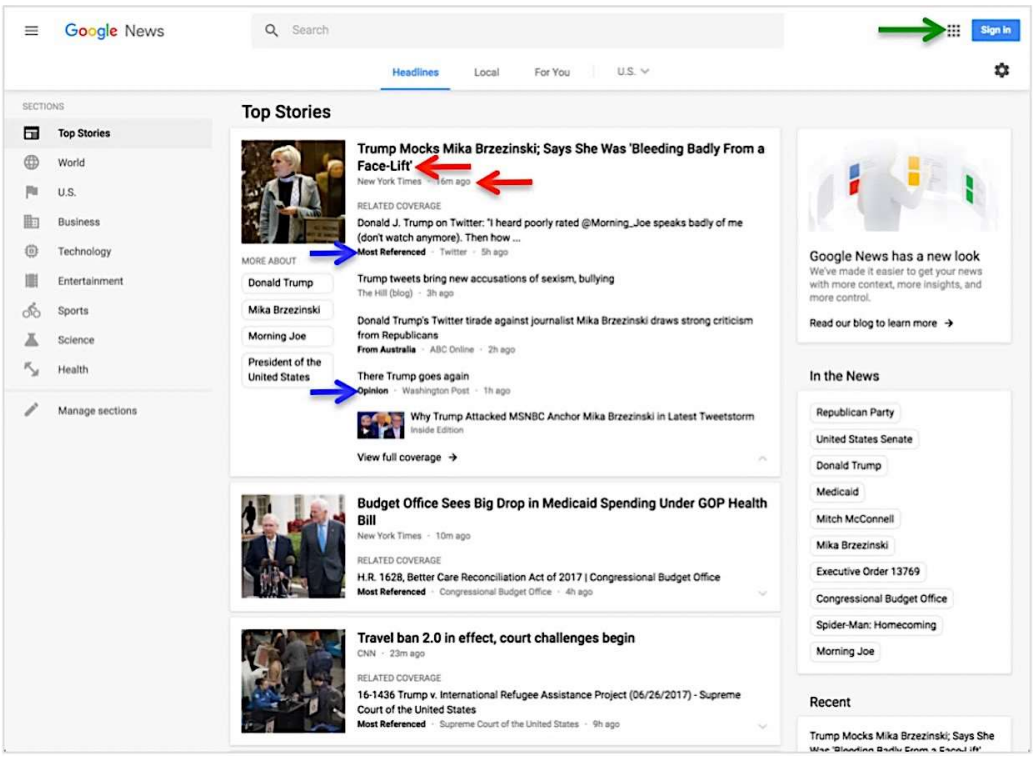

***Figure 3-7*. Google News Homepage in 2017. Source: Internet Archive, emphasis mine.**

During the web archival research, a wide range of labels used on the Google News homepage over the years was observed, including "Most Referenced, "Fact Check," "Trending," "Opinion," "Satires," "Press Release," "Featured," "Live Updating," "Highly Cited," "Local Source," "In Depth," "Trending On Google+." "Wikipedia," "From Germany (or other locations)," "International," etc. Similar labels have been used by news media to tell readers about the nature and characteristics of news content (Hågvar, 2012). Google adopted this function, as the product manager of Google News said, to let users "see additional context on stories immediately even as you are scanning" (Paka, 2017, para. 7). Furthermore, labeling information was fed into Google's algorithmic systems to decide how news would be clustered, placed, and ranked. This information was very useful to train Google's machine learning systems and enhanced

Google's search function because the more contextual information was given, the more likely users were able to pinpoint the news articles they were interested in. This "deep linking" to specific content also became, as Chapter 4 shows, a concern of news publishers, as it reduced the likelihood of users browsing a news website as a whole.

Google News in 2018

The three major redesigns—in 2002, 2010, and 2017—led to the new Google News presented in 2018. The layout of the 2018 Google News homepage was similar in appearance to the 2017 page, with news cards on a while background. The 2018 version, still in use as of this writing in 2019, has a simplified header section. On the left was the hamburger menu (in the form of three short horizontal lines), which could expand into a left-side navigation bar. Right next to it was the Google News logo. In the middle was the search box, in which a gray-text message reminded users that they can search for topics, locations, and sources. On the right side was a nine-dot menu for Google applications—such as Google Drive, Google Maps, and YouTube—and the "sign in" button. The rest of the page was the main body; there is no more the footer section. Some of the links that once were located in the footer section have been moved into the left-side navigation area, such as "Language & region", "Settings", "Send feedback", and "Help" (Figure 3-8).

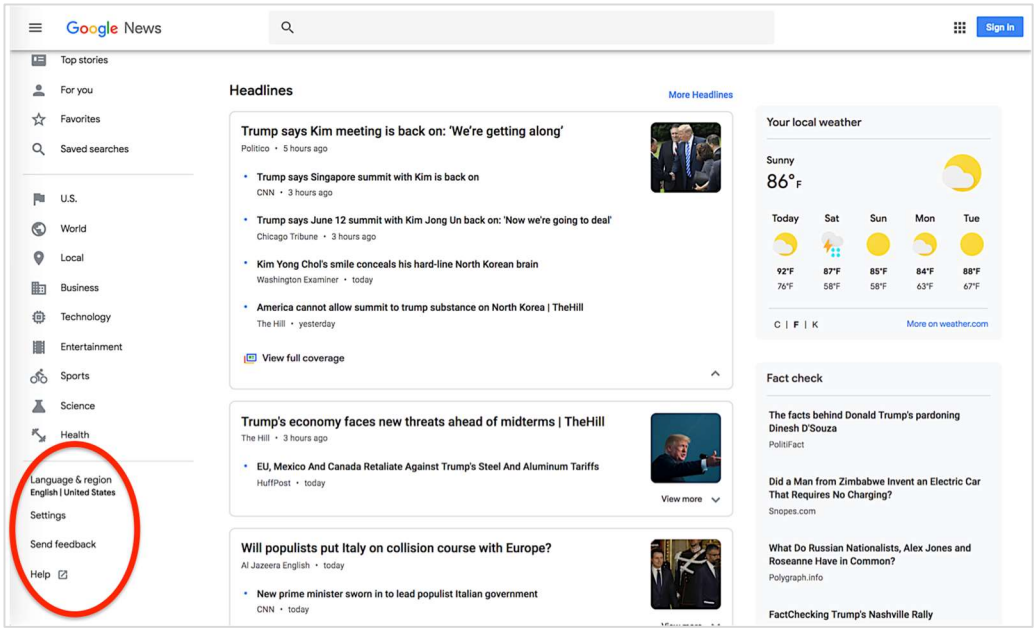

***Figure 3-8*. Google News Homepage in 2018. Source: Internet Archive, emphasis mine.**

The main body of the webpage still consisted of three columns. On the left was the navigation bar, which was divided into four parts. The top displayed "Top stories," which led to the standard news page; the "For you" section, "Favorites," and "Saved searches" all required sign-in to get personalized news. The next part had nine standard news sections, including U.S., World, Local, Business, Technology, Entertainment, Sports, Science, and Health, most of which had been standard sections since Google News was launched. The order of the sections reflected their importance. In 2018, the U.S. section was moved ahead of World and Local gained more prominence by joining the list and holding the third position. The lower part provided service-related information, including language and region, settings, links to the Android app and the iOS app, a link for users to send feedback, and a link to the help center. At the bottom of

the navigation bar was the information about the privacy policy, terms, and the about Google page.

In the middle column were news cards grouped-based on the eight standard news sections. The Local section, however, was not listed in the middle column news stream. To get local news, users had to click on the link in the left-side navigation column, which made the local section an independent section. Each news card in the middle column contained a headline, news source, and time and had one news photo or video (from YouTube or news sources) at the right side of the news card. Each news section had four news cards, and the first card usually had five news stories. Users could click on the "View full coverage" link for "a complete picture of how that story is reported from a variety of sources." (Upstill, 2018) The right-hand column had only four sections. The first displayed local weather information based on users' IP address and cookies, without requiring that they sign in. The second section, "Fact check," usually provided five fact check stories. Following "Fact check" was "Spotlight," which introduced news stories a couple of hours to several days old. Each Spotlight news story also had one news photo at the right end of the news card. At last was the "In the news" section with about ten named entities.

For each news card, including the ones in the middle column and the ones in the right-hand column, there was an option for readers to "Save for later," a share icon that let users copy a link to the article or post it to Facebook or Twitter, and a "More" menu where users could choose to view full coverage or go to the original news source directly. By signing in, users could gain more options for each story, including hiding all articles from a particular news source or having more/fewer stories like a certain article. Users

who clicked anywhere on the news card, were directed to the original news source, increasing the likelihood that users would visit the original news source's website.

In terms of the look of the homepage, the 2018 version of Google News was not a major "redesign" because it looked similar to the previous version. So, when the new Google News was introduced in 2018, it was called the "reimagined Google News" (Upstill, 2018), as it used real-time artificial intelligence and machine learning technologies to organize news. The AI-powered features highlighted in the 2018 version as well as other features, however, did not come into being overnight. The following section will discuss changes on the Google News homepage over time.

Gradual Homepage Changes Over Time and Implications

In addition to the major homepage overhauls, this study identifies changes that are not considered significant redesigns but gradual adjustments over time. These gradual changes are categorized into four types in relation to Google News' homepage structure. Google News homepage is normally structured with up to four components (Figure 3-9): 1). the header section on the very top; 2). the search and function section; 3). the main body of the homepage; and 4). the footer section. In some years, the header and the search/function section have been combined in different ways. The main body has different forms, two or three columns. The footer is removed in the redesigned 2018 version of Google News.

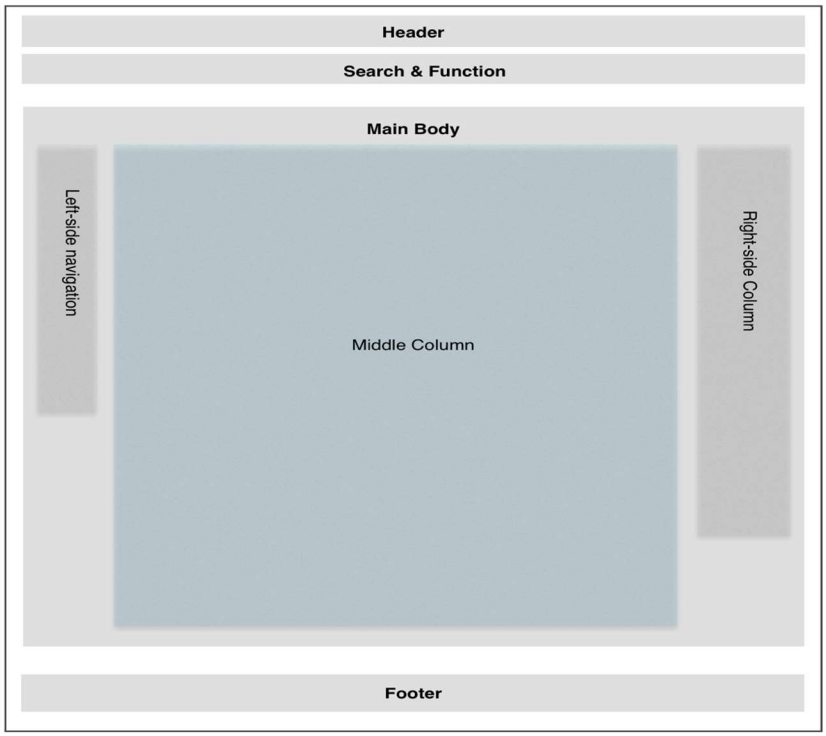

***Figure 3-9*. Google News Homepage Structure. Source: author.**

The research, which examined pages archived in the Internet Archive, identified four types of gradual home page changes: changes related to Google products, functional changes, news-related feature changes, and service-related changes. The identification of the change type relates to the location of the given change on the homepage. The following section will first offer a brief discussion of product, functional, and service changes, followed by a more detailed examination of feature changes, the biggest category with the widest implications for news and news media.

Product-related Changes

Changes to the homepage that related to Google products were made to the header section, which is used as a spot on Google News homepage to make visible and cross-promote Google's own products. When the Google News Beta version was introduced in 2002, there were links to four Google products other than news: Web, Images, Groups,

and Directory. Since then, links to as many as 25 Google products have been shown on the Google News homepage at different time, including Search, Maps, Local, Video, Products, YouTube, Play, Gmail, Drive, Calendar, Translate, Photos, Shopping, Wallet, Finance, Docs, Books, Blogger, Contacts, Hangouts, Keep, Google+, Jamboard, Earth, and Collections. Since 2017, products have not been directly listed on the header section of Google News homepage as in the 2010 version (See Figure 3-6, green arrow) but, rather, integrated under the nine-dot icon on the upper right corner (see Figure 3-7, green arrow).

Google has a wide range of products. It selectively promotes a short list of products on its main page. Over the past seventeen years, Google News has never fallen out the top 6 on this list, which indicates Google News has been a stable and strong product among internal competitors inside Google.

Functional Changes

Functional changes are located in the search and function section of the Google News homepage, which is usually below the header section. This area provides functionalities other than browsing news. For example, the search box is located in this section for users to search news, topics, locations, and sources. There were also specialized searching functions showing in this section in the early years of Google News, such as “Advanced search,” “News archive search,” and “Blog search.” Over the past seventeen years, this area has also provided functions about different kinds of settings, such as search settings, preference, layout viewing options, and personalization settings. The boundary of the header and the search and function sections is often blurred. In 2018, on the redesigned Google News homepage, these two sections were combined

into a single, simplified header section, which had a main menu (the hamburger icon) on the left, the Google News logo, the search box, the Google apps menu (the nine-dot icon), and the sign-in button (Figure 3-10, green arrows).

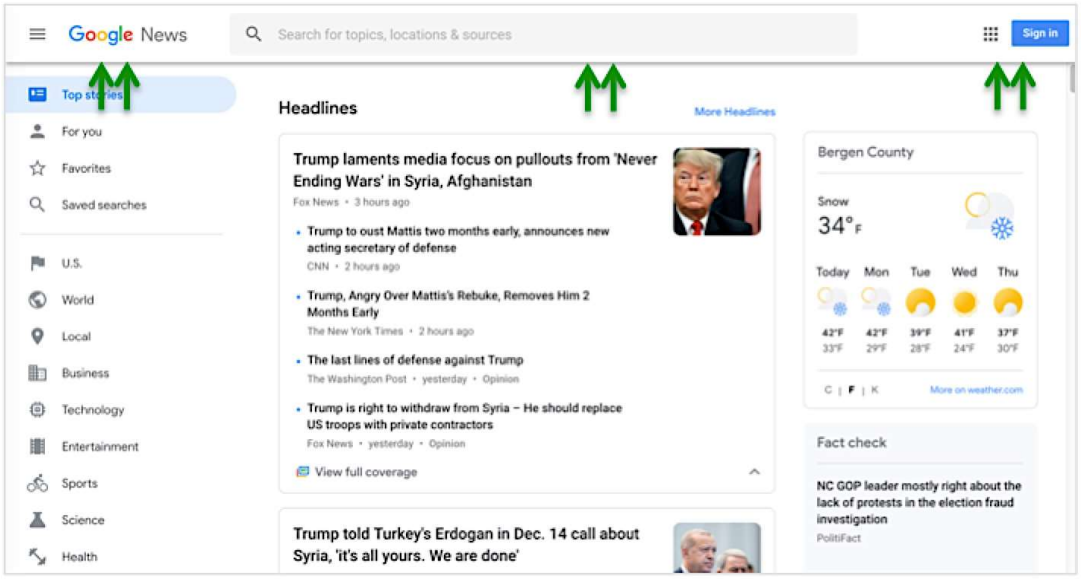

***Figure 3-10*. Google News Homepage in 2018. Source: Google News website, emphasis mine.**

Service-related Changes

The footer section is where service-related information— such as the "about" pages, privacy policy, and terms of use. Since 2006, Google News has also introduced such services as "Help," "For publishers," "Blog," "Advertising programs," "Business solutions,", "Send feedback," and more (Figure 3-11, red arrows). The footer section has played a role of customer service and public relations, with information that serves news publishers and producers, regular users, advertisers, the media, and the general public.

***Figure 3-11*. The Footer Section on Google News Homepage 2009. Source: Internet Archive, emphasis mine.**

Another important component of the footer area was the international editions through which Google has been expanding its global reach. Each international edition is specifically tailored for audience in the given region. Users could choose different editions to see news aggregated for a specific region or in a specific language. In 2003, Google News listed five international editions in the footer section: Australia, Canada, India, New Zealand, and the U.K. On its homepage, this list has been expanding since then. It covered 10 international editions in 2003, 21 in 2004, 23 in 2005, 38 in 2006, 39 in 2007, 56 in 2008, and 63 in 2009. As the list grew, individual editions were no longer listed one by one after 2009. Instead, a link indicating "Other News Editions" took users to the full list. That list expanded to 72 international editions in the redesigned Google News in 2010 and 81 in the 2017 version of Google News. In 2018, the footer section was removed from Google News homepage. A link that allows users to select language and region to access Google News' international editions, as well as a few other links that used to locate in the footer section, were moved to the left-side navigation section, where

81 international editions are listed. As of September 2018, Google News reported that its service covered 127 countries and 65 languages (Stier, 2018).

News-related Feature Changes

News-related feature changes observed in the main body of Google News homepage were the largest category among the types of identified homepage changes. These changes were related to various aspects of news, including news content (e.g., news sections, special sections, and different news genres) and news sources; news presentation, such as multimedia (e.g., image, video, mobile); news distribution, such as social features (e.g., sharing, most popular); and new forms of news experience, such as personalization and localization. Major features are examined in more detail below.

News sections

The arrangement of the news sections has been the most stable element on the main body of Google News homepage over the years. When Google News Beta was launched in 2002, there were eight news sections: Top Stories, World, U.S., Business, Sci/Tech, Sports, Entertainment, and Health. Such an arrangement reflected how news media organize and classify news. Moreover, as shown below in Figure 3-12, even the color codes adopted by Google News to mark different news sections were similar to those used by news media, such as USA Today (Garcia, 2012). These news sections have become the standard sections that Google News uses to structure its website. Over the years, the standard news sections have undergone only slight changes. For example, Entertainment was moved ahead of Sports in 2009. In 2011 Sci/Tech was separated into Technology and Science, with Technology behind Business and Science, after Sports; and U.S. was moved up ahead of World in 2018.

***Figure 3-12*. Left: Google News Color Codes in 2002. Right: USA Today Color Codes. Source: Internet Archive and USA Today, emphasis mine.**

Special sections

Over its first seventeen years, the Google News homepage presented at least 26 special sections. These special sections focused on three types of news topics:

- Major events, which have accounted for 12 of the 26 special sections. These have focused on political events (such as elections, presidential events, political protests, etc.) and other national or international events that draw wide public and media attention (such as the Academy Awards, Olympic Games, etc.).
- Crisis response, the subject of nine special sections, such as those about China earthquake in 2008, Golf of Mexico oil spill in 2010, and Pakistan floods in 2010, etc., often with resources, news coverage, and some interactive tools that were Google products, such as Google maps.
- Advocacy of journalism, which accounted for five special sections. For example, from 2011 to 2016, Google News had a few special sections

concerning journalists' rights and freedom of speech by linking to the Committee to Protect Journalists, an independent, non-profit organization that advocates press freedom and the rights of journalists to report news without the fear of reprisal.

The adoption of the special section is a typical example of normalization. News media often use special sections to introduce a collection of news stories about certain topics that editors believe especially newsworthy. A special section could contain news coverage, commentaries, background information, follow-ups, and more to help readers understand the topic comprehensively. Google News, as an automatic news aggregator, could have included news associated with these special topics into standard sections arranged by algorithms, but it chose, instead, to present them separately in the form of special sections.

This decision may be driven by two factors. First, Google may have been trying to avoid a type of bias in its news aggregator. Wide media attention on specific topics could have overwhelmed other news stories in Google's algorithmic systems. By taking these stories out of its regular algorithmic results and presenting them in special sections, Google was able to overcome algorithmic flaws with a journalistic method.

Second, Google may have adopted the idea of special sections to call attention to these topics even if they were not in a user's areas of interest. Google's technology documents reveal that, for example, Google would recommend Olympic Games-related news to a user, given that this is a trending topic that interests the wider community, even if the user might not be personally interested in sports according to his/her past interest or search history (Dolan & Liu, 2012). Furthermore, 35 percent of the Google News special

sections were about crisis response, which reflected a legacy of Google's role during 9/11. These special sections could also invoke public attention on affected people and areas. In this case, Google shared journalism's gatekeeping role in informing the public what its members should know, although what Google defined as something important was algorithmically driven, rather than subject to criteria of newsworthiness as defined by journalists. Nevertheless, its role in algorithmic gatekeeping went beyond the role of a pure news aggregator.

Another category of Google News special sections focused on journalists' rights. These special sections served several purposes. First, they could raise the public's awareness of related issues, as news stories about this topic might not be able to get prioritized by Google's algorithm. Second, these special sections served a media relations function. For example, in 2011, Krishna Bharat spoke at the Journalists Memorial's annual rededication ceremony at the Newseum, a museum in Washington, D.C., with a theme on news, press freedom, and the First Amendment (Carlson, 2011). To follow up, Google News offered several special sections on its homepage concerning journalists' rights, such as "Find out where journalists are getting killed around the world: CPJ's 2011 Impunity Index," "Number of jailed journalists skyrockets worldwide," "The number of journalists imprisoned worldwide reached a record high in 2012," and "Latest CPJ report finds a record 259 journalists jailed worldwide in 2016, with Turkey the leading jailer." By opening these special sections, Google News advocated for journalism. This advocacy seemed to go against Google News' position as "a news service compiled solely by computer algorithms without human intervention." (About news search, 2002, para. 2) These practices reflected the changing relationship between Google and the news media.

In its early history, Google tended to play a role of reformer of the news media industry in deep crisis by giving news media advice on technology, business model, and news experience as a whole (See Garber, 2011.) In more recent years, however, Google's attitude changed by glorifying journalism as "the best of human intelligence." In 2018, when Google reintroduced its news aggregation service, it said, "Of course Google News wouldn't exist without the great journalism being created every day" (Upstill, 2018, para. 13). That is true if one thinks about the news media as the content supplier for Google News. But the power relation between Google and the news media that has unfolded over the past two decades is more complex than just that between a content supplier and a content wholesaler.

News sources

After 9/11, Krishna Bharat, the founder of Google News, was concerned that it was hard for people to access different perspectives regarding a topic unless they wanted to search manually, while it was critical for people to see multiple perspectives especially during a national emergency like the 9/11 attacks (Raj, 2019). This concern motivated the creation of Google News, a news aggregator that aimed to provide users news with different perspectives across news sources worldwide. When Google News was launched in 2002, it searched and browsed 4,000 news sources. That number grew to 4,500 in 2003 and soared to 25,000 in 2009. Google News reported it crawled more than 50,000 news sources worldwide in 2012 and more than 75,000 news sources in 2016 (Bharat, 2012; Morehead, 2016). As of January 2019, Google News covered more than 80,000 news publishers around the world (Google, 2019). As Google News expanded its reach, studies found that news content presented on Google News was dominated by Western

perspectives and concentrated on large, popular, and mostly legacy media outlets (Watanabe, 2013; Nechushtai & Lewis, 2019). This seemed to be a trend going against Google News' initial intention.

It is a challenging task to gain a historical view of how Google News managed news sources in the past two decades, considering just how user interface design and homepage features have changed over years. In an attempt to understand that, a pilot study was conducted of three constructed weeks across 2002-2018. This section reports on the results of that study.

The study focused on primary stories, those articles on a Google News homepage with highlighted features (*e.g.* bold or larger text) that do not require extra action (*e.g.* clicking to expand/collapse) to be visible. Unlike secondary stories, which usually are presented in regular (unbolded) or smaller text and sometimes require extra actions on the part of the user to be seen, primary stories remain stable in terms of their characteristics throughout the three major homepage redesigns. Considering the consistent characteristics of primary stories, this pilot study focuses only on primary stories and their news sources to get a taste of the long-term trend in terms of how Google News selects news sources on its homepage over the years.

The pilot study examined the primary stories to determine whether they were from mainstream or non-mainstream sources and U.S. or non-U.S. outlets. To define mainstream[6] news sources, this section refers to several lists that rank top news websites, such as Alexa's top sites by category, Top 100 USA News Websites, and the 100 Most

[6] "Mainstream" could be a controversial concept, which sometimes refers to "conventional newspapers, television and other news sources that most people know about and regard as reliable." (Collins, n.d.) "Mainstream" in this section is defined for analytical purpose.

Important Online Publishers. News sources that are not listed on these top lists are defined as non-mainstream for this analysis. To tell whether a news source was U.S. or non-U.S., the study used web searches or went to the given source's original website to understand its nature. This study then compares the types of primary news stories in relation to the total amount of the primary stories. The results are presented in Figure 3-13:

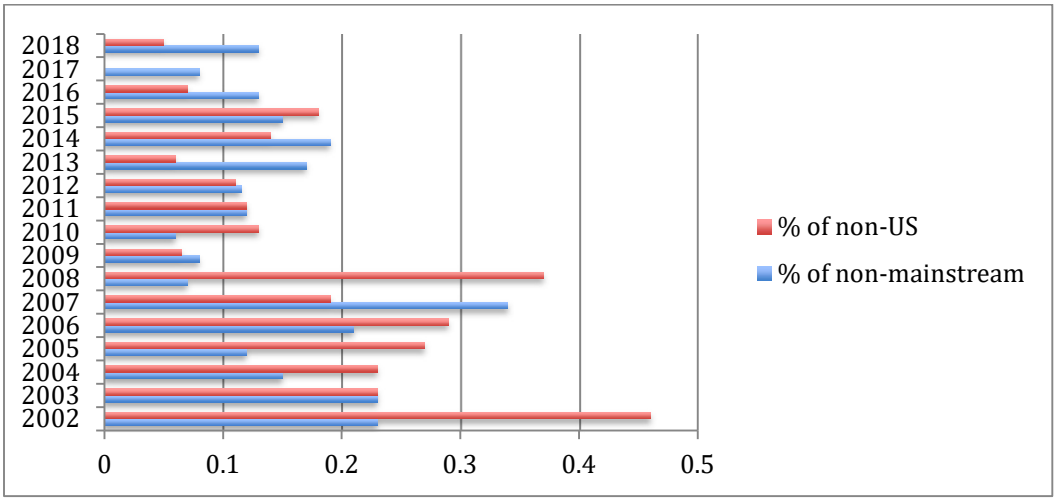

***Figure 3-13*. Proportion of Non-mainstream and Non-U.S. Sources. Source: author.**

On average, there were 32 primary stories on each Google News homepage available in the Internet Archive from 2002-2018. A total of only 14% of the stories were from non-mainstream news sources and 17% from non-U.S. news sources. The number of non-U.S. sources would be even lower if Reuters, a British news agency acquired in 2008 by the U.S. company Thomson Corporation, were considered a U.S. source. Overall, 2009 seems to have been a watershed in that the proportion of non-U.S. sources included in homepage primary news stories dropped more precipitously after that year. While there is no clear evidence to explain this drop off, this change correlated with the

global financial collapse of 2008-2009 that resulted in widespread job losses in many industries, including the news industry (Goldman, 2009).

About 11% of the non-mainstream source-produced news appeared in the Top Stories, World, and U.S. sections, which have been the top three standard news sections on Google News throughout the years. This observation indicates that in its hardest news sections, news from well-known news organizations may have a better chance, than smaller/less-well-known news outlets, to be presented to users on Google News homepage. About 20% of the non-U.S. source-produced news stories were included in these news sections. It is worth pointing out that many non-U.S. sources that appear on the Google News homepage are themselves mainstream media. For example, news from Reuters, the Guardian, and the BBC accounts for over 61% of the news stories selected from non-U.S. news sources. In this regard, having diversity of geographically located sources doesn't necessarily mean diversity of types of news organizations on Google News homepage.

This brief section provides a general picture of the long-term trend in terms of the news sources and their stories selected for Google News homepage. Future studies could use a larger sample to get more precise statistics. While this brief analysis may not be able to provide any causal evidence, the overall decline of non-mainstream, non-U.S. news sources, especially non-U.S. news sources, generally correlates with the timing of the increase of international disputes between Google and news media in different countries that will be analyzed in Chapter 4.

The rest of this section will examine other feature changes that cover social, multimedia, mobile, and news features that dealt with different news genres and forms.

The analysis is also based on web archival research through the Internet Archive's data. For each feature, changes were traced throughout the examined years, and data from that work was combined with historical research to understand the historical background of these changes and how the feature works.

Social

As will be shown in Chapter 5, Google News has not given social features[7] top priority. The "Most Popular" section was introduced on Google News homepage as early as 2006, but was changed into "Most Shared" in 2010 and disappeared in 2016. The nature of this section is unclear since Google News only briefly mentioned that the "Most Popular" section presented news stories popular with readers but did not specify the source of user data, whether it was from a social media site such as Twitter or from its own data (Parfeni, 2010). In 2009, each news story was given an "Email This Story" button, which seemed to be an early step to integrate social functions on Google News. Since then, Facebook and Twitter have been the standard social sharing features available when social functions were available on Google News. Google's own social tools, on the other hand—such as Google Buzz, "+1", etc.—came and went over the years, lacking a consistent and center social product of Google's own.

In the redesigned 2010 version, a set of social features were introduced on the Google News homepage that allowed users to email and share news via Google Reader (a web feed reader operated by Google), Facebook, Twitter, and Google Buzz (a short-lived sharing tool introduced in 2010 that allowed Google users to network with Gmail contacts). As Google Buzz was shut down in 2011, Google's "+1" button joined the

[7] "Social" here refers to the idea of information sharing and interactivity as in the concept of social media or online social network.

social features on Google News homepage. The "+1" button was designed for users to recommend content across their social network and the Web. In 2012, the "+1" feature was gone, replaced by Google+, Google's own social networking system. Google Reader was retired in 2013, and Google+ was shut down in 2018 due to a bug that could potentially risk users' privacy and the low usage and engagement (Smith, 2018). As of 2018, the Google News homepage provided limited social features that allow news to be shared via Facebook and Twitter, and/or by copying the link of the news story when clicking on the share icon.

Multimedia

In addition to text, Google News offers a few multimedia features on its homepage, such as images, video, and other forms of rich media. News photos have always been an element included on the homepage throughout its history, since Google News launched its Beta version in September 2002, but the arrangement of the images—*e.g.* the position, number, and size—have changed over the years. In 2002, news photos were located at the right end of the news clusters. These images had different sizes. Images were not necessarily attached to each news story. The inconsistence and randomness of image display seemed to indicate that images may have not been algorithmically standardized at that time.

By 2011, photos were moved to the front of each news story, and they were aligned and the same size. When Google News was redesigned in 2017, each story card was led by a larger news photo. In 2018, images were moved back to the right end of each story card. In the right-hand column, a place that used to be image-free, news photos were added to each story in the "Spotlight" section.

Before 2017, Google News offered users the options to select between text-only version and text+image versions of the home page. The text+image format became the standard, however, in 2017, and the text-only version was eliminated. As will be shown in Chapter 4, news organizations were very concerned about the combination of the three news elements (i.e. news headline, snippet, and news photo) made publicly available on Google News, in terms of copyright issues and the effect of such combination in reducing users' interest in browsing the full coverage on the news organizations' own websites. While Google News gave up news snippet after 2017, images were kept, and they were even highlighted more, indicating Google's perception of the increasing importance of the visual element in online news.

In addition, after Google acquired YouTube in 2006, video coverage was introduced in 2007 and has been focused on YouTube videos (Zhang, 2007). In 2006, mobile news was also introduced on Google News, which meant that news-related videos from YouTube could be played on mobile devices without having to worry about players and browsers. In the meantime, the vast number of news sources included in Google News were invited to join the YouTube Partner Program, which allowed YouTube news partners to participate in advertising revenue sharing and offered them featured premium placement on the YouTube news page, where Google and YouTube "feature news videos from partners related to the top news stories on Google News" (Ma, 2009, para. 4). Google News Videos included in Google News were ranked and clustered same way as articles. In 2017, Google News announced that it used improved, but undisclosed, algorithms to select top and related videos (Paka, 2017). In addition to these traditional

media forms, Google News has also advocated newer forms of rich content—mostly through its own products—such as maps, Earth images, analytics, and data visualization.

In 2018, Google News became more mobile-optimized to support multimedia. Unlike Google News' previous effort, in which multimedia elements were promoted separately—such as the separate links to the text version, image version, YouTube news section, and mobile news—a new visual format called newscasts was applied to bring together a collection of news materials by integrating the multimedia components (Upstill, 2018). This allowed Google News to further de-bundle and re-bundle news across not only different news sources but also across different medium forms.

Mobile

In 2006, Google News introduced a link on its homepage to its "New! Mobile News." To access the mobile version of Google News, users needed to visit google.com on their mobile phone's web browser and click the link to Google News (Figure 3-14) At that time, only sources that designed mobile-friendly content were included in the mobile version of Google News (Google, 2006). Compared to other features, mobile seemed not to have been given adequate promotion on Google News homepage. Mobile features were not even promoted during the years, 2010 and 2017, when Google News had major redesigns. It was only in 2012 and 2014, respectively, that Google News provided advisement on its homepage about Google News for tablet and the Google News app.

In fact, after 2006 since Google News for mobile devices was introduced, Google had been working on improving mobile-optimized version of Google News, including enhancing usability on mobile devices (e.g., making tapping, scrolling, and swiping easier), making more news features mobile-friendly (e.g. Local, Editors' Picks, etc.), and

developing mobile versions for multiple systems and platforms (e.g. Android, iOS, etc.). These efforts aimed to improve the mobile web version of Google News accessed through web browsers on mobile devices. It was until August 2014 that Google News released the app "Google News and Weather" to the Google Play Store for Android and iOS device users to download for free (Saoji, 2014). In late 2013, there was another app "Google Play Newsstand" that users could use it to read news, blogs, and magazines on their mobile device. In 2018, a new mobile app "Google News" was rolled out, which combined and replaced the previous "Google News and Weather" and the "Google Play Newsstand". The new app, with four tabs "For You", "Headlines", "Favorites", and "Newsstand", was reported to use artificial intelligence and machine learning technologies to support personalized news experience (Cipriani, 2018). On 2018 version of Google News, links to the Android app and iOS app are provided on the left-side navigation bar on its homepage.

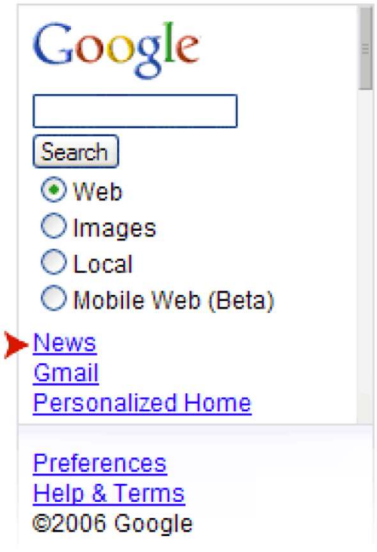

***Figure 3-14*. Access Google News on Mobile Device in 2006. Source: Google, emphasis mine.**

News features

News-related features in the main body of the homepage also have changed over time. Some of the features are longstanding and still available on Google News homepage today, such as “In the news” and “Spotlight”; others are newer and reflect important implications, such as “Fact check.” This section offers a brief examination of these features.

- **In the news:** This segment of the Google News home page, a feature since the beta version in 2002, contains terms—known as named entities in natural language processing terminology—that function as hyperlinks. When users click on a term, they are directed to a page containing a collection of news stories about the term, which might be the names of a person(s), place(s), location(s), thing(s), or event(s) in news. These terms play an important role in helping Google News design its algorithms and organize news. They are the building blocks of Google’s key technologies, including the search technology, grouping technology, and natural language processing technology—the most basic and underlying elements in Google’s algorithmic systems. Named entities usually appear in the “In the news” section. During mid-2010 to mid-2017, these terms were moved into the left-side navigation area but were moved back to the right-hand column in 2017. In 2018, “In the news” remained one of the four segments in the right-hand column, along with local weather, fact check, and spotlight. The introduction of named entities also gave news audiences a new way to consume news. If named entities are viewed as search terms users might employ, they indicate an effort to

train users to consume news in a way that reinforces and optimizes Google's algorithms.

- **Spotlight:** This feature, introduced to the Google News homepage in 2009, was temporarily removed at a few points between 2015 and 2017, but was brought back and appeared in the 2018 version as one of the four sections on the right-hand columns. It aggregates long-form and in-depth featured articles, such as "investigative journalism, opinion pieces, special-interest articles, and other stories of enduring appeal" (Google, 2009). According to Google, news featured on Spotlight is automatically selected by computer algorithms, using undisclosed algorithms. The establishment of Spotlight looked like the result of a negotiation between Google News and the traditional news media industry. This feature was considered an effort of Google News that "steps back from the ever-quickening news cycle" to value "in-depth pieces of lasting value," often seen as an important asset of quality journalism (Seward, 2009, para. 1). Compared to the homepage news overall, some observers found Spotlight news was overwhelmingly dominated by mainstream news sources (Slocum, 2010), which is not surprising as they are often the only news outlets that have the resources to do long-form and in-depth reporting.
- **News in time**: Over the years, Google News has had several features based on the chronology of news stories. For example, as early as in 2002, Google News gave users the option to sort news by date, which allowed users to view the most recent news first, by day, when clicking on a news topic. In 2009, Google News introduced "Timeline" and "Living stories," which allowed users to follow a news

topic's new developments as they unfolded. Similar features continued to show up on the Google News homepage, including "Recent" in 2010 and "Realtime coverage" in 2012. By 2018, however, timelines no longer appeared on the homepage, though it could be found on the "full coverage" page for selected news stories. Technically, time-related information is important in Google's algorithms; temporal information is parsed from news content and machine language, such as timestamps. From a journalistic view, these features served two purposes. First, they allowed users to closely follow the most recent updates in news. This approach addressed the nature of news as something time-sensitive and new, focusing on news genres such as breaking news. These were the characteristics of news that Google ignored before 9/11, when Google treated news no differently from other types of information and only crawled news sites irregularly. Second, these features were also useful for understanding the evolution of news stories as they unfolded by treating news as historical events. The second approach relates to the timelessness of news by valuing the contexts of news events, focusing on news genres such as developing news. These decisions reflect the debate in the journalistic field about fast vs. slow journalism and cold vs. hot journalism. In the digital era, different forms of journalism require differing digital approaches in terms of the presentation, distribution, and engagement. One example: "Living stories," a project for which Google News collaborated with *The New York Times* and the *Washington Post*, the three parties working together to experiment with "new ways to interact with news and the quality of reporting" and "a different format for presenting news coverage online" (Google, 2009, para. 2, 3). By

covering both the timeliness and the timelessness, these experiments could balance various interests in the journalistic field.

- **Editors' picks:** This feature, introduced in 2011, allowed Google's partner news organizations to promote their articles on Google News home page. In 2002 when Google News was launched, the home page bore this statement: "This page was generated entirely by computer algorithms without human editors. No humans were harmed or even used in the creation of this page." Shortly afterward, a new statement appeared on the homepage, where it remained until recently: "The selection and placement of stories on this page were determined automatically by a computer program." (About news search, 2002) With features such as "Editors' picks," the editor-free claim hardly stood. According to Google, this feature aimed to highlight news stories that were original, high quality, and/or innovative (Mehta, 2011). Observers pointed out, however, that publishers could submit any stories that they thought were "simply worth reading"; the point of "Editors' picks" is, "they [news editors] can choose" (Garber, 2010). "Editors' picks," as well as a few other features—such as the "outstand" tag and the "original-source" tag that news organizations can use to feature their own quality and original content (Smydra, 2011; Weigle, 2010), gave back a degree of control to news organizations and blurred the boundary of differentiation and normalization. These features were no longer available on the Google News homepage in 2018. According to Google they may be replaced by "other ways publishers can optimize visibility" (Google SearchLiaison, 2018). The addition and removal of

these features may also be a result of the negotiation and power shift between Google and news organizations.

- **Fact check:** This feature had a short history on Google News. Introduced as a news label in 2016, ahead of that year's U.S. Election Day, this feature was seen as a response to criticism that social media and other digital platforms were being targeted by fake news (Toor, 2016). To "divine fact from fiction, wisdom from spin," Google News asked news publishers to use ClaimReview[8] techniques and "commonly accepted criteria" to label fact checking content by identifying the claim being checked, the source of the claim, identification of the fact checker, and the fact-checking conclusion (true or false) (Gingras, 2016). In 2017, Fact check became an independent, "dedicated" section in the right-hand column on the homepage (Paka, 2017), and it remains one of the right-hand column sections on Google News in 2018. According to Google, "only publishers that are algorithmically determined to be an authoritative source of information will qualify for inclusion" in the fact check feature (Kosslyn & Yu, 2017). The author conducted an examination of the news sources appearing in the "Fact check" section on the Google News homepage, which revealed that they are mostly the fact checkers included in a list of global fact-checking sites identified by the Duke University Reporters' Lab (See reporterslab.org**)**. According to this list, there are

[8] ClaimReview, introduced by schema.org, is a "fact-checking review of claims made (or reported) in some creative work (referenced via itemReviewed)" that can enable a "summarized version" of fact check content (e.g. Claim: The world is flat. Claimed by: Flat World Society) to be displayed on Google. Schema.org is "a collaborative, community activity with a mission to create, maintain, and promote schemas for structured data on the Internet, on web pages, in email messages, and beyond." Google is one of its founders (ClaimReview, n.d.; Fact Check, n.d.).

50 active fact-checking sites in the United States, 15 of them affiliated with PolitiFact, created in 2007 by the *St. Petersburg* (Florida) *Times* (now the *Tampa Bay Times*) and acquired in 2018 by the non-profit Poynter Institute. The majority of the rest of the fact checkers are affiliated with news organizations, with others are run by non-governmental groups. A small group of fact checkers—including PolitiFact, Snopes.com, Washington Post, FactCheck.org, Gossip Cop, CBS News, and Polygran.info—appeared frequently in Google News' Fact check section.

- **Local**: In February 2008, a Local News section was added to the Google News homepage. Users could keep track of current events in their area by inputting their city, state, or zip code. While studies have found that Google News benefits small, local news organizations that would otherwise only have limited reach (e.g. Athey, Mobius, & Pál, 2017; George & Hogendorn, 2013), critics have accused major digital platforms of threatening the economic sustainability of local media (News Media Alliance, 2019). As early as in 2005, Google had launched a product called Local, which allowed users to search location-based information that matched a user's interest, such as "Pizza near LAX." News also could be searched on Google Local, but compared to the early versions, the Local News section was particularly designed to promote both local content and local news media. In the following year, this feature was introduced in Google News sites for the UK, India, Canada, Germany, and France. In 2014, the year that a study found that two-thirds of Americans who live in small towns depend on their local paper for news and information, Google developed a deal with the Local Media

Consortium, an organization representing more than 1,600 local media outlets in the U.S., to build its local network (Gingras, 2014). In 2010, local weather was added to Google News homepage along with local coverage. Pew Research Center's study found that local weather is the most important daily information need for local residents (Pew, 2019). In 2018, Local became one of the standard news sections on Google News homepage located in the left-side navigation bar, while local weather was the first section in the right-hand column. Joining Google, recently Facebook has also invested in building partnerships with local media globally, aiming to invest $300 million over three years (Li, 2019). Several factors have contributed to digital platforms' increasing attention to local information. First, organizations such as the News Media Alliance have been putting pressure on major digital platforms to support local media through efforts such as advocating the proposed "Journalism Competition & Preservation Act" (Chavern, 2018). The bill was introduced by Representatives David N. Cicilline (D-RI), the Chairman of the House Judiciary Antitrust Subcommittee, and Doug Collins (R-GA), the Ranking Member of the Committee on the Judiciary, on April 3, 2019, that aims to provide a temporary safe harbor for publishers of online content to collectively negotiate with dominant online platforms regarding the terms on which their content may be distributed. Google's investment in local news, especially in recent years, is a response to such pressure. At the same time, online local news markets have economic potential, given that 89% Americans get local news online (Pew, 2019). Local media have the direct and close connection to local communities, an important channel for tech companies to

access local markets and related resources, including geographic data—one of the most important data types in algorithm systems. Given these influences, local could be an area that digital platforms keep investing in the next few years. The development of the local section on Google News' homepage is another example of Google's non-algorithmic intervention since many local news stories may not have a chance to stand out in Google's algorithmic systems and to be listed on Google News homepage if they were not listed in "Local" separately.

Personalization

Personalization is a long-term, strategic goal of Google that has also been reflected in changes to the Google News homepage. Google News personalization has evolved around two models: what I call "direct personalization" and "contextual personalization" (Wang, 2018c). In the case of Google News, direct personalization requires users' conscious choice and physical action to edit certain elements of the news page so that users can make their own version of the news page based on their choices. This model is very straightforward in that users make personalized choices and are aware of these choices, the reason they made them, and the results of such choices. Direct personalization involves user labor, *e.g.,* selecting, deleting, rearranging, and saving. At the same time, it allows certain degree of user agency in the sense that users do not have to passively accept the pre-designed version of the website.

Contextual personalization is more backend that Google uses contextual user information that it collects to make personalized recommendations that are believed to match or relevant to users' interests, preferences, and historical choices. Contextual personalization makes sense of the action that users take at the moment by connecting it

to wider contexts, *e.g.*, what the user did in the past and in different social settings, what the user's social networks look like, and what the larger communities are to which the user is connected. Based on such information, Google uses its own logics and methods to contextualize who the user was, is, and will be, individually and collectively. The goal of contextualized personalization is to make recommendations based on the prediction of users' behavior and their preference patterns. Since the collection, processing, and use of contextual user information are based on Google's decision, users enjoy less agency and provide less labor in this model. Depending on the role of the participants in the process of personalization and the degree of agency users enjoy, some scholars distinguish these two models of personalization as user-initiated customization vs. system-initiated personalization or active vs. passive personalization (*e.g.* Coner, 2003; Sundar & Marathe, 2010).

Early direct personalization

Google News users were introduced to a link to "Customize this page" for the first time in March 2005. This customization functionality gave users a few options. They could add a standard section, from the eight news sections that appeared in the standard view of Google News homepage; choose an international edition that they prefer to use; add a custom section by using keywords to search a topic of interest; and decide how many news stories (one to nine) to see on the customized news page. Google News also allowed users to un-customize the news page by resetting it to default or closing the function (Figure 3-15). Before that, Google had always offered users some opportunities to make their own choices on the news page, for example, as early as 2002, users were able to choose their preferred view, *e.g.* the text version or graphical version, of Google

News, which can be seen as an early form of customization. In August 2005, Google News allowed users to choose news sections that interested them by subscribing to Google News feeds. These options in those years were mostly focused on direct personalization and did not require any "personally identifying" information. Users' personalized choices were stored in cookies on their own computers (Google, 2004). Sign-in also was not required on the Google News homepage until November 2005. That's also when "Customize this page" was changed into "Personalize this page" with similar offerings (Figure 3-16).

***Figure 3-15*. "Customize this page" in March 2005. Source: Internet Archive, emphasis mine.**

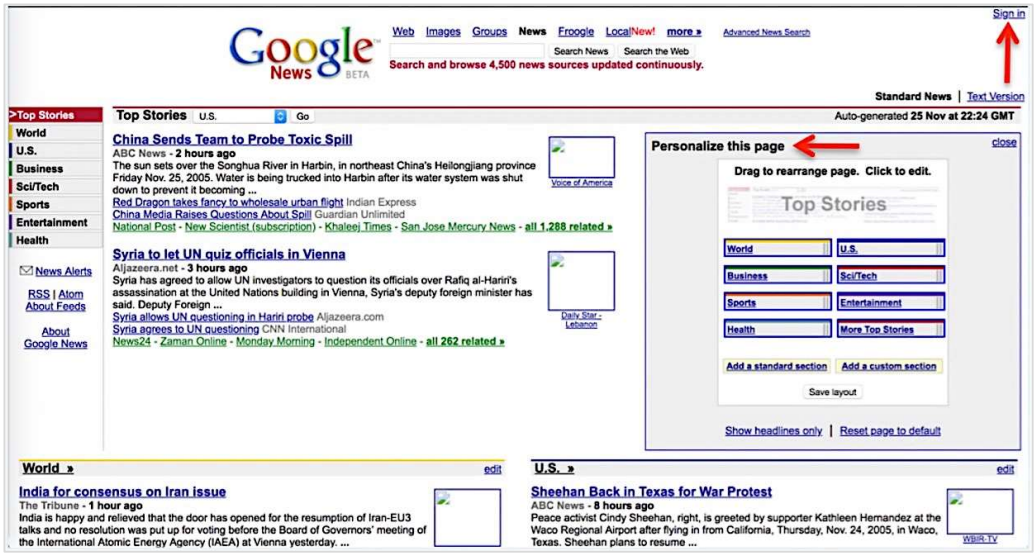

***Figure 3-16*. "Personalize this page" in November 2005. Source: Internet Archive, emphasis mine.**

Direct personalization + contextual personalization

By the time Google News exited the Beta version in January 2006, a new feature, Get Recommended Stories, had been rolled out on its homepage while Personalize This Page remained available. Recommended Stories were stories Google News recommended to users based on their past news selections. To get Recommended Stories, users had to sign in and enable personalized search. Google used algorithms to analyze the user's news selections and compare it to those of similar Google users. The recommendation function focused on contextual personalization. By signing in the Google account, Google collected various types of user information, including personal information (e.g. name, email address, account password, payment information), log information (e.g. web request, IP address, browser information, cookies), and user communications (e.g. emails). In addition, Google also collected personal information that users provided to affiliated sites, other sites owned by Google, and information traced by links. Google also

combined user information associated with Google account with information from other Google services and third parties (Google, 2005). In the Get Recommended Stories section, Google News could also recommend news stories “that have been read by many other users who've also read similar stories as you in the past” (Lenssen, 2006). Since then, while direct personalization and contextual personalization have co-existed on Google News, the latter become more and more powerful, and user behavior and data involved in direct personalization could also contribute to contextual personalization.

From 2006 to 2010, Google News promoted several personalization-related features on its homepage, such as choosing news source preferences, editing the news page, adding a section, and inputting location-related information. By July 2010, when Google News had a major redesign, personalization-related features were distributed at a number of points on the homepage, including “Sign in” and at least three sets of settings (news setting, search setting, and Google account setting) in the webpage header, the international edition, and Add a Section at the top right corner, three versions that users could choose to view the news page (All News, Headlines, and “Images”) in the left-side navigation bar, a star button, and a drop-down menu to let users choose more or fewer stories from the given news source for each story, and the new “News for you” section in the middle column as the “new heart” of the redesigned Google News (Stolt, 2010). In the “News for you” section, users could choose from three view options (list, sections, two-column), search and add news topics to follow, choose how often they would like to read news from each section (never, sometimes, always), and rearrange the order of the news sections. Users were also given the option of closing or resetting the whole section

(Figure 3-17). Engagement with any of these features would involve users in the process of personalization.

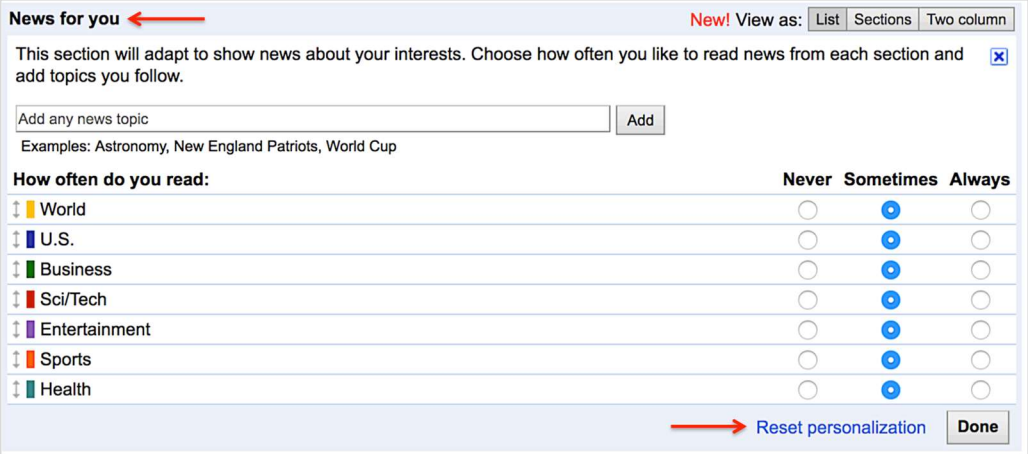

***Figure 3-17*. "News for you" in 2010. Source: Internet Archive, emphasis mine.**

In the following few years, Google News kept promoting personalization on its homepage, including introducing a video showing how to personalize the news page, embedding a slider to personalize Google News, and promoting personalization features on mobile versions. On the Google News homepage redesigned in 2017, users could manage news sections, follow topics in the "For you" section after signing in, and choose local information based on location information. In 2018, there are three sections in the left-side navigation bar, which are "For you," "Favorites," and "Saved searches." All these features required sign-in but allowed users to have recommended stories based on their interest and add/manage favorite news sections, sources, and topics. By signing in, users could personalize each story by choosing whether they would like more/less stories like this or deleting all stories from this particular news source. Over the year, options for direct personalization did not change much, mostly about a few straightforward choices of editing news sections, topics, sources, and frequency. But direct personalization has become more granulated, from personalization at news section level to personalization at

article level. Compared to direct personalization, the contextual personalization has developed more remarkably.

In 2006, Google interpreted the development of its contextual personalization to its users this way, “Already, Google News allows you to personalize your homepage, but this is a step further as it allows you to be even lazier (you don’t need to configure your preferences).” (Lenssen, 2006, para. 7) This interpretation apparently only addressed the distinction between direct personalization and contextual personalization in terms of the user labor involved in the process of personalization, but didn’t say anything about privacy and user agency. As the concerns about the privacy issue grow, Google has provided users different ways to control their privacy setting. At the same time, contextual user information that Google has collected expanded to a wider range.

An examination of all the versions of Google’s privacy policies over the past twenty years suggests that data that Google has collected, stored, analyzed, and shared come from various data sources. In recent years, in addition to users’ personal information associated with their Google accounts, user data that Google collected and manipulated include information obtained by third party applications and gadgets, location data, and advertising cookies (2009), user communications such as content of SMS messages, and deleted information (2010), user information shared by Google partners (2014), YouTube information, comments, and posts (2014), user information used for Google Analytics (2015), photos, videos, docs, spreadsheets, related information from publicly accessible sources, information stored even without signing in Google account, and more (2017). By 2018, user information that Google collected covered things that users create and provide and information collected as users use Google

services. Google uses these data to "build better services," including providing personalized services, content, and ads (Google, 2018). Today, when Google News becomes more AI-driven, contextual user information only plays an increasingly important role in training and optimizing Google's AI technology.

Meanwhile, Google News has been attempting to maintain a balance between personalization and unpersonalization. In early years, Google News always gave users the option to either close or reset personalization if users preferred unpersonalized news. In 2018, when Google News promoted its new website and mobile apps, the "Full coverage" feature was given special attention because it is a unpersonalized view of news that provides same news for everyone. At a time that there are mixed findings about news personalization in forming or popping filter bubbles (e.g. Haim, Graefe, and Brosius, 2018; Duke, 2018) and other uncertain effects of personalization on people's social life and societies, Google's effort in balancing personalization and unpersonalization is a response to the concerns of various stakeholders, including Google users, news media, policymakers, and the general public. In the meantime, in spite of the debate on personalization effects, personalization is an area that Google has been investing, as is shown in Chapter 5.

Discussions

Based on web archival research, this chapter details changes observed on the Google News homepage between 2002 and 2019 and their implications. This examination reveals that since news.google.com was launched, the website has been pretty consistent in terms of its homepage design, with only three major redesigns in 2002, 2010, and 2017. The Google News homepage was organized around standard news

sections throughout its history, a practice that Google adopted from news media. In 2002, the change from Google News Search to Google News moved toward normalization, downplaying the role of the search function and highlighting the service's role as a news aggregator. By defining itself as "a valuable source of information on the important issues of the day," Google News shared journalism's gatekeeping role. Such role was approached differently from news media. For example, Google News de-bundled news packages that news media put together. It then re-bundled news, using its own grouping and clustering technologies. These processes mixed normalization and differentiation. On the one hand, Google adopted and appropriated various journalistic practices and ideas, such as news sections, special sections, and labels that identify news genres and types; on the other hand, Google News distinguished itself from traditional media by stressing the news aggregator as an algorithm-based service without human intervention that provided multiple perspectives across news sources worldwide. Google's news service differentiated from traditional news media through specialized techniques, such as named entities, that introduced new norms and practices in terms of news distribution and consumption, such as SEO-based news, and reading news based on named entities. Overall, at its early stage, Google News focused more on normalizing news media's practices and ideas by integrating these practices and ideas into its operation and technology. While it also proposed ideas and norms that were very different from those of news media, these efforts mostly aimed at the development of its specialized technologies, rather than demanding an independent status from news media.

As Google News evolved, it went beyond being a pure news aggregator that relied solely on automatic computational algorithms. As this chapter demonstrates, many

changes on its homepage were not driven by algorithms but by Google's strategic decisions, which served as non-algorithmic interventions during the evolution of Google News. These decisions reflected the changing interrelationship between Google and news media. The addition of the separate right-hand column on Google News homepage, for example, provided a space for news content, genres, and types that were chosen with human intervention, e.g., "Editors' picks" were recommended by news editors and "Spotlight" news was promoted for long-form, in-depth news pieces. Google also used this space to promote Google's own tools and products as well as "Most popular" news with users. This column has become an arena of negotiation to balance interests of different stakeholders, including news media, Google, and users. In this column, Google also opened a number of special news sections for topics that Google believed newsworthy. Google uses special sections, an idea adopted from news media, to balance algorithmic bias, play a gatekeeping role, and maintain media relations.

Since the mid-2010s, many changes on Google News homepage have been responsive changes reflecting Google's response to news media's actions and reactions. For example, the removal of news snippet in the 2017 redesign was a result of the copyright disputes between Google and news organizations; the introduction of "Fact check" responded to the news media's criticism of the role digital platforms played in the crisis of fake news in the wake of the 2016 U.S. presidential election; in response to news media's concern about their online traffic, on 2018 version of Google News, Google also made it easier to direct users to news media's own websites when users browse on Google News. In addition to these responsive changes, Google also took proactive actions to specialize its own news service. The AI-powered Google News introduced in

2018, including the back-end machine learning technologies and the personalization features on its homepage, differentiates the “Google way” from the news media’s way of conducting the business of news. To better understand these changes and the driving forces behind them, one has to have a closer look at the tensions between Google and the traditional news media. The next chapter provides such an opportunity to examine disputes involved Google News in different parts of the world.

Chapter 4. Battles: Google News and the News Industry

Google's adoption of news into its search business started its interaction with the news media industry. As Google News grew, the ideas and norms that were in line with its specialized technologies but very different from established journalistic ideas and norms started to trigger tensions between Google and traditional news media. Since 2003 as Google News has expanded its international reach by establishing more and more international editions, disputes about the aggregator's activities have broken out around the world. These disputes reveal concerns and negotiations between Google and other stakeholders in the media landscape in the digital era.

This chapter examines and compares disputes about Google News in eight countries: France, Belgium, Italy, the United States, the United Kingdom, China, Germany, and Spain. These countries have distinct media, political, economic, and legal cultures. A commonality among these countries is that news aggregator is a new addressee in their regulatory systems. News aggregators practice business about news through digital technologies. Although Google has business globally, the practice of news aggregation is relatively new, without a consistent regulatory tradition across different countries. Therefore, when disputes occurred, involved parties chose different responses, which give this chapter an opportunity to analyze these cases with a global and comparative view.

Many of the examined disputes dealt with the legal doctrine of copyright, which protects creators' interests in original creative works. The contemporary copyright law evolves from the Berne Convention for the Protection of Literary and Artistic Works. Adopted in 1886, the Berne Convention was an international agreement on copyright and

related rights to protect creative works and their creators. All eight countries in which disputes are examined are members of the Berne Convention (the U.S. didn't join the Berne Convention until 1989, more than 100 years after the other countries.) While the Berne Convention addresses copyright issues in the pre-internet era, as the world entered the digital age, the World International Property Organization (WIPO) issued international copyright standards to respond to digital developments. Among other countries, these eight countries signed the WIPO Copyright Treaty in 1996, which adopted principles of Berne Convention but also put forward agendas that addressed digital technologies (WIPO, 1996). The WIPO required its member countries to update their laws to comply with the WIPO treaties (Blythe, 2006).

In the U.S., the Digital Millennium Copyright Act (DMCA) was introduced in 1997 to implement the WIPO treaties, updating a federal copyright law last fully revised in 1976. The DCMA granted safe harbors in which internet service providers can avoid liability for violations of copyright law by their users. In Europe, the European Union Copyright Directive (EUCD) went effective in 2001, taking "network operators" and "intermediaries" into the copyright system in the information society (Directive 2001/29/EC). The EUCD offered a high level of protection to rights holders and relatively limited exceptions and limitations for online service providers. While the EUCD aimed at harmonizing copyright rules in the EU countries, different European countries have different national laws. China is also a member of the WIPO. However, the internet overall is regulated first and foremost as the extension of China's national sovereignty. Under these circumstances, disputes about Google manifested distinct characteristics in the examined cases.

The eight cases were chosen for this analysis because they were highly publicized cases widely reported by news media that had important outcomes. Using traditional interpretive legal analysis, this chapter examines and compares the eight cases in terms of involved parties, their arguments and counter arguments, legal frameworks pursued, consequences of the disputes, and more. Three bodies of documents were examined to support the legal analysis in this chapter:

1) Laws and measures based on which disputes were judged or contended and legislations passed in the wake of certain dispute, such as the Digital Millennium Act in the United States, Belgium's Copyright and Related Rights Act, Article 82 of EC Treaty of European Commission, the Ancillary Copyright Law in Germany, the Canon AEDE in Spain, and the Cybersecurity Law in China.
2) Other available case documents, including court decisions, complaints and responses, status reports, expert reports, and memos.
3) Relevant public texts, such as blogs, legal reviews, white papers, news articles, and scholarly literature for background information.

Next, the eight cases will be analyzed in chronological order, determined by the year when the given case started. The analysis compares the following areas:

- The time span of the dispute
- The parties that brought the accusation against Google News
- Approaches the involved parties used, e.g. litigation, lobbying, diplomatic, or commercial approaches
- Legal frameworks pursued by involved parties

- Focal points of contention, i.e. concerns and controversies on which the dispute focused
- Consequences, such as passage of legislation, settlement, and shutdown

France

Background

In March 2005, Agence France Presse (AFP), the world's third-largest news agency headquartered in Paris, France, filed a lawsuit in the United States against Google, claiming copyright infringement by its Google News service. The complaint claimed that Google News violated copyright law by using AFP's news content, including headlines, leading paragraphs, and images, without AFP's permission. The litigation process lasted for two years. On April 6, 2007, the case was dismissed by the United States District Court for the District of Columbia after the parties signed a licensing agreement that allowed Google News to use AFP's news content in "innovative, new ways." Details about whether and, if so, how Google News would compensate AFP for such use were not disclosed (Auchard, 2007). In the aftermath of *Agence France Presse v. Google Inc*., Google signed a €60 million agreement with French media in 2013, creating a Digital Publishing Innovation Fund that media outlets could apply to for money to help them develop their internet presences in ways that would increase online revenues. The agreement left Google free to host French publishers' news materials without having to contend with threat of legislation that would have created a "Google tax" on the company for publishing French media content without permission (Pfanner, 2013; Welch, 2013).

Focal Points of Contention

One of the foci of the lawsuit by AFP against Google News was the copyrightability of news. While AFP attempted to claim, in a U.S. court, copyright over its headlines, leads, and images, Google argued that news was not protected by U.S. copyright law. The debate, therefore, was fundamentally concentrated on the nature of journalistic content and journalistic practice. In the history of copyright, there had been inconsistent definitions of the nature of news. Article 2(8) of the Berne Convention, for example, said that the copyright protection should not apply to "news of the day or miscellaneous facts having the character of mere items of press information" (Berne Convention, 1979). In the U.S., fact-based reporting and objectivity are in the core values and norms of American journalism. However, objectivity as a journalistic value has never developed in Europe as fully as in the United States (Schudson, 1981). French journalism traditionally has a strong political-literary character, which involves considerable amount of creative work, such as writing that mixes descriptive, normative, and commentary statements (Benson, 2013). As a wire service—one that reports in French, English, Arabic, Portuguese, Spanish and German (Key statistics, 2018)—AFP was creating news meant to be read in many cultural contexts.

Nature of journalistic work

In the lawsuit, Google held that news content is not copyrighted subject matter under the U.S. Copyright Act because news stories report "facts" and facts are not copyrightable. AFP argued that journalistic work has both expressive value and factual value (Keith, 2007). The minimal threshold of copyrightability, argued AFP, is "originality," and news meets the two conditions of originality: independence and creativity. AFP claimed that news is "independently created by the author (as opposed to

copied from other works)" and it possesses "some minimal level of creativity," for example, in the headline writing (*Agence France Presse v. Google Inc.*, 2005). According to AFP, the copyright is in the original expression in news rather than the facts themselves. When Google uses AFP's news content on its news aggregation service, not only did it copy the facts but also it copied the expression of the facts, hence the infringement of copyright.

Since news headlines, snippets, and images are the three key elements of a news story that were made publicly available on Google News, there was heated debate about copyrightability in these areas. For example, in the case of news headline, Google argued that AFP headlines were too short to be protected. Google held that without a substantial length, the headlines could not hold protectable expressions and that AFP headlines were only titles that functioned as labels of news stories. In the Expert Report Google submitted to the court, Google pointed out that 37 C.F.R. § 202.1(a) stipulated that titles are words and short phrases that are not subject to copyright (1:05CV00546-GK, 2007). But AFP defended the copyrightability of headlines by claiming that headlines differ from titles because the former involves originality and creativity while the latter would be only an identifier of the news story. In the expert report that AFP prepared for the court, AFP illustrated that while "Plane crash into downtown office buildings" (a mock headline) might serve as an identifier of a news story, the actual headlines news professionals created from that information—"War at home" (an actual headline in The Dallas Morning News on September 11, 2001) and "Terror beyond belief" (an actual headline of The Star Ledger of Newark, New Jersey, on September 11, 2001)—sent readers rich information that was far beyond just facts (1:05CV00546-GK, 2006).

AFP defended journalistic practice as a professional process that produces meaningful informational packages and compilations through decisions such as how headline, text, and image should be put together and how news elements should be visualized and presented. AFP saw this compilation as a "whole constitutes an original work of authorship" that involved professional and logical process of decision-making, and that could have profound and direct impact on how users perceive reality (1:05CV00546-GK, 2005). To support its viewpoint, AFP demonstrated in its expert report that the design of headlines and visual elements played significant roles in helping readers quickly understand the meaning of the news content or producing bias in users' news experience. Based on this argument, AFP asserted that news also has commercial value. AFP presented a series of eye-tracking studies to demonstrated that after going through the combination of a headline, blurb, and image made publicly available on Google News, users were less likely than they would have been without reading the aggregator to go to the original news website and read the full news articles. Thus, the commercial value of the news was harmed (1:05CV00546-GK, 2006).

AFP's argument reflected journalism's traditional gatekeeping role. In this role, the selection, de-selection, and arrangement of news components are rational editorial decisions that determine what information the public should get and in what way the public should be informed. Google's perception of news as pure facts that do not merit copyright protection was closely related to its technological view of news. As will be shown in the next chapter, in Google's automatic algorithmic systems, news has to be "datafized." For example, Google needs to tag news by breaking it down into smaller units based on tagging entities, such as the named entities discussed in Chapter 3, and

then rank news by transforming it into pieces of quantifiable data that can be calculated and algorithmically processed. This process of datafication did not capture the inherent logics of news making. What Google's algorithms could process was only the algorithmically-identifiable information in the news, for example, information associated with the names of people, places, and events, or, simply the "facts." The journalistic logics lost in Google's algorithmic systems are either considered insignificant by Google or unsolvable by algorithm. In Google's technological view of news, journalistic practice was a series of discrete activities based on the division of labor in the newsroom, which, according to Google, had nothing to do with inherent, logical journalistic decisions. For example, editors assign tasks, reporters cover beats, and copy editors write headlines (1:05CV00546-GK, 2005), resulting in a news-making process that is no different from an algorithmic process, in which news elements only had practical functionalities, for example, news headlines, argued Google, are ordinary statements of the news events and news images are merely "another link to a story."

Consequence

In 2007, the two-year litigation between AFP and Google News ended with a settlement, in which the parties agreed on "innovative, new ways that will dramatically improve the way users experience newswire content on the Internet" (Sullivan, 2007, "Postscript"). Although no further details were made available, such "new ways" to use AFP content aimed to "highlight original journalism," "giving credit to the newswire journalists who worked hard to break the news" and to "ensure that AFP's original journalism and breaking news are easily discoverable on Google services and in particular on Google News" ("Postscript"). Law scholar Eric Goldman saw this

settlement as a signal of change in Google's "historical position" that "it can grab whatever it wants without permission" (quoted in Perez, 2007, para. 5). Observers also pointed out that considering technological advancement far outpaced legal reform, it did not seem a good idea for both parties to rely on judicial systems to solve the dispute (Perez, 2007).

In the aftermath of *Agence France Presse v. Google Inc*., the negotiation between French publishers and Google News lasted for years regarding whether Google should pay licensing fees for hosting French publishers' news content (Wojazer & Vinocur, 2013). Meanwhile, the settlement between AFP and Google News also avoided a "Google Tax" (Lasar, 2011), a tax proposed in 2010 by the Zelnik Commission, a panel headed by former Virgin executive Patrick Zelnick and charged in 2009 by French Cultural minister Frédéric Mitterrand with finding ways to ensure that artists and authors were reimbursed for the use of their creations online. Its report called for a series of measures aiming to finance French cultural industries in the digital environment and promote legal downloading, including the tax (Mortaigne, 2010). While AFP's lawsuit focused specifically on Google News, the so-called Google Tax targeted Google the search engine, and the key concern also shifted from a journalistic issue to an economic issue, particularly Google's advertising revenue. It was reported that to avoid new legislation creating a Google Tax and the licensing fee for using French publishers' content, Google signed an agreement with President Hollande of France in early 2013 for two initiatives. First, Google agreed to invest 60 million Euros ($82 million) to create a Digital Publishing Innovation Fund aiming at supporting French media's digital development. Second, Google promised to help French publishers to increase their online

advertising revenue by using Google's advertising technology (Schmidt, 2013). In exchange, Google News was able to host French publishers' news materials without paying a licensing fee. The agreement was reported as a "commercial agreement" (Google and France reach landmark agreement, 2013) with a focus on the parties' commercial interests. While copyright law addresses commercial rights, the implication of *Agence France Presse v. Google Inc.* was not only about economic interests. The debate between AFP and Google about the copyrightablity of news concerned the nature of journalistic work at professional level, a meaningful reflection that was absent in later cases.

Belgium

Background

Shortly after the launch of Google News in Belgium in 2006, Copiepresse, an association representing the publishers of the French- and German-language newspapers in Belgium, filed a descriptive distraint petition to the Brussels Court of First Instance, claiming that Google News breached its copyrights by counterfeiting (*Google Inc. v. Copiepresse*, 2011). The court granted the petition and appointed a judicial expert, Luc Golvers, to appraise Google News (Smith, 2007). In August 2006, Copiepresse filed an official lawsuit against Google. Based on Section 87 of the Belgian Copyright and Related Rights Act of June 30, 1994 and the Belgian Database Protection Act of August 31, 1998 (Michielsen, 2006), the court ruled in September 2006 that Google had violated the publications' copyright. The ruling required Google to withdraw the press content from both Google.be and Google News Belgium and post the ruling on the home pages of these sites or pay the fine (Whetstone, 2006).

Focal Points of Contention

In *Google Inc. v. Copiepresse*, the first thing the Belgian court had to decide is which law was applicable in this case. Google insisted that American law should prevail because the ownership, the location of the hardware and software, and the act called into question all occurred in the United States. The Belgian court ruled, however, that Belgian law should be applied because the affected news producers were Belgians; because Google used a Belgian domain name, i.e. google.be and news.google.be, which intentionally targeted the Belgian public; because Google worked with Belgian advertisers; and because the dissemination of the news content available on Google News Belgium occurred in Belgium. This ruling is evidence that, although digital platforms have business and influence globally, national boundaries stand erect. These boundaries contribute to the collision between local interests and digital platforms' global ambitions, which provided a context for several of the cases examined in this chapter.

Technological issues

Google Inc. v. Copiepresse had a heavy focus on technological issues, which distinguished the case from *Agence France Presse v. Google Inc*. The key debate covered three technical issues.

First, caching was a major concern of Belgian publishers. Caching is a technique that Google used to archive information and make it available to searchers even if the information was no longer accessible on the original website. In *Google Inc. v. Copiepresse,* Copiepress protested that caching allowed Google to illegally reproduce Belgian publishers' content and breach their copyright. Google defended the feature from a technical point of view by stating that caching was conducted by intelligent robots that

came across the web content automatically. Since the web and the information on the web are publicly accessible, the crawling should not, Google argued, be considered infringement. In addition, Google maintained that Google did not reproduce the news content and communicate it to the public given that it only transmitted HTML files, rather than actual news articles. It was the users, Google argued, who copied and downloaded the news articles, when their internet browsers read the HTML files, as Google provided users only with "installations aim at or facilitating communication" (*Google Inc. v. Copiepresse*, 2011). Here, both parties made arguments from technological perspective, which gave the case more digital characteristics.

The EUCD exempted caching from copyright liability by considering the technology "an integral and essential part of technological process" "widely recognized and used by industry" that "enable[s] transmission systems to function efficiently" (Directive 2001/29/EC, Article 33). The Brussels court ruled, however, that Google's technical arguments were not founded, since the infringement relates to whether, rather than how, the copyrighted content was copied and communicated to the public by Google. The court noted that the public would not have had the access to the content, in the case when the given content was no longer exist on the original website, if Google had not made it available through caching. According to the Belgian copyright law, only the author is entitled to reproduce and communicate his/her work to the public.

In *Google Inc. v. Copiepresse,* Google also attempted to defend its use of Belgian publishers' content through the debate of the technical nature of Google News. Google insisted that Google News did not reproduce or communicate protected work to the public because it was merely a search engine that functioned on automated algorithms to

generate search results. The search engine argued that Google should be exempted under Article 5.1 of the EUCD, which stipulated that "temporary acts of reproduction… whose sole purpose is to enable… a transmission in a network between third parties by an intermediary" should be exempted from the reproduction right (Directive 2001/29/EC, Article 33). This argument brings to mind the early ambiguity of Google News' self-positioning as it combined search and news aggregation services. (See Chapter 2.) That ambiguity might have been due to the service's early self-adjustment, but it is clear that in these disputes, Google realized that the technical nature of Google News matters in a legal context. Luc Golvers, the judicial expert appointed to appraise Google News, concluded in his report that Google News functioned as an information portal and not as a search engine (Court of First Instance of Brussels, 2006), because it does not merely "host" but also selects, classifies, and ranks information based on its own algorithms. Golvers concluded that Google News is not a "passive intermediary" and it cannot invoke the exception provided for under the Belgian copyright law (*Google Inc. v. Copiepresse*, 2011).

Another issue that was also technical in nature was "implied consent," an argument Google made to defend its use of Belgian publishers' news materials. Google argued that technical standards—such as metatags and robots.txt files, which are robot exclusion protocols (REP) that allow websites to automatically communicate with web crawlers that they should not take information from a site—were "nearly universally accepted and are honored by all reputable search engines." (Whetstone, 2006, para. 5) Because Belgian publishers did not use these technical standards, Google argued, they had given Google the "implied consent" to exploit their content or they chose to opt-in

Google's news service. The court ruled that this theory of implicit permission was not compatible with the principle of copyright protection, which requires explicit permission. The technical means cannot be used, the court said, to justify the assumption that "anything that is not forbidden is permitted," even in a digital context (*Google Inc. v. Copiepresse*, 2011). To support this decision, the court cited a long quote from Carine Bernault, a professor at the University of Nantes in France, condemning the idea of "subjecting copyright to technology" to create "a situation of dependency in respect of systems":

> These technological measures should therefore remain at the service of the social (judicial, if you wish) rules, of the rules of society has chosen. It would not be acceptable that software programs would become some sort of normalization tool of the so called information society, a technical 'law' that would be imposed surreptitiously!" (*Google Inc. v. Copiepresse*, 2011, #50)

The court's statement provided a picture of the conflict between judicial rules vs. technical law. In defending Belgian news publishers, the former expressed strong disapproval of the dependent relationship with the latter.

In 2008 Copiepresse and its member publishers proposed an alternative technical standard, known as ACAP (Automated Content Access Protocol), which was an extension of REP but would give publishers full control over their work by telling robots and web crawlers what search engines can or cannot use (Groklaw, 2008). Although not successful, ACAP represented Belgian publishers' demand for self-control. This demand for self-control was shared by Google. Google News resisted being integrated into the

copyright system, a legal framework advocated by news media industry. In doing so, Google defended the legitimacy of its news aggregation by depending on its specialized technology, whether technical processes or technical standards, to seek the status of autonomy. This struggle showed Google's strong demand for differentiation that declared that technology business is not news business.

Business model

Economic concerns were another focus in the dispute between Copiepresse and Google News. Copiepresse demanded over $77 million compensation for losses it claimed were caused by Google News' copyright infringement (Ricknäs, 2008). In 2006, the Court of First Instance of Brussels held that the use of Google News circumvented the advertising of Belgian publishers and allowed the evasion of paid news, which risked publishers not being able to "avail of sufficient resources to pay their journalists properly" (Court of First Instance of Brussels, 2006). In addition, by providing the essential elements of the news produced by Copiepress, the court found, users were "perfectly informed" about the essentials of the news without having to visit the original article and the news website, which could cause financial loss for Belgian publishers and therefore violate the integrity of their work. Copiepresse also criticized Google for using news aggregated for free as "decoy" to attract traffic to its own sites and gain advertising revenue (Copiepress, 2011). The court of appeal of Brussels supported this view by stating:

> Admittedly, the "Google News" service is free of charge and, in Belgium, it does not contain any advertising; this fact does however not imply that the economic balance of the interests at stake would tilt in favor of

> Google, because it must be taken into account that this free service can only be provided thanks to the significant revenue Google generates as a result of attractiveness of all its services and the horizontal sliding of revenue which this interactivity facilitates. (*Google Inc. v. Copiepresse*, 2011, #58)

While no precise figure for how much monetary value Google News contributed to Google, in 2008, Google Vice President Marissa Mayer said at the Brainstorm Tech conference in Half Moon Bay, California, that Google News might not make money on its own but it "funnels readers over to the main Google search engine, where they do searches that *do* produce ads," which drove $100 million worth of benefit to the "Google ecosystem" (Fortt, 2008, para. 2, 3). Compared to *Agence France Presse v. Google Inc*., Belgian publishers and Belgium's courts made the news media's economic interest a case that was more urgent and justifiable because their charge was not only targeted the techniques Google used in its news aggregation service but also the underlying business model.

Consequence

In September 2006, the Court of First Instance of Brussels ruled that Google violated the publications' copyright. Google appealed in 2007. In 2011, the Court of Appeal of Brussels issued a final judgment by mainly upholding the original decision to confirm Google's copyright violation. In 2012, the litigation between Copiepresse and Google was settled in an agreement in which Google agreed to partner with Belgian publishers on "a broad range of business initiatives." e.g. Google will help publishers to increase their revenue and audience engagement by optimizing their use of Google

products while publishers will advertise Google services in their media (Essers, 2012; Geerts, 2012). It was reported that although Google said it did not pay to settle the dispute, the company actually gave $6.5 million to Belgian publishers and likely agreed to buy advertising from publishers' papers was well (Byford, 2012).

In *Google Inc. v. Copiepresse*, Google accused Copiepresse of using its "dominant position" to restrict Google's access to Belgian market. Google charged the intention of Copiepresse as "first of all financial and protectionist" (*Google Inc. v. Copiepresse*, 2011). It is interesting how the power balance was shifted in later cases, in which Google became the one who was accused of abusing the dominant position while news publishers considered themselves the victim of unfair competition. While the Belgian court stated, "it is impossible to say whether the parties *are* competitors" (*Google Inc. v. Copiepresse*, 2011, 12/62, emphasis added), the tension between the "old" and "new" media actors put the two parties into a competitive relationship, which was defined by a competition of control, including the control of "what the public may or may not have access" (Copiepresse, 2011) and the control of whose standard should prevail.

Italy

Background

In August 2009, the Italian Federation of Newspaper Publishers (Federazione Italiana Editori Giornali or FIEG) filed an official compliant with the Competent Authority of Competition and Market (Autorità Garante della Concorrenza e del Mercato or AGCM), Italy's antitrust watchdog, requesting an investigation of Google News Italia in terms of the lack of transparency in its operation (Associated Press, 2009). The investigation began on August 26, 2009, based on Article 82 of the European Community

Treaty (FIEG/Google, 2010). On August 27, the financial police of AGCM inspected Google's Italy headquarter office in Milan (Associated Press, 2009). About one week later, AGCM announced that it was extending the investigation from Google News to its parent company, Google Inc. in Italy (Sayer, 2009). On March 3, 2010, the investigation was further widened to Google Ireland Limited (AGCM, 2010). AGCM ended the probe in January 2011, finding the commitments proposed by Google were sufficient for removing the anti-competitive concerns raised in FIEG's complaint (Pfanner, 2011). Google and FIEG reached an agreement in June 2016 aiming at a "strategic collaboration" through initiatives such as the investment of over 12 million Euros from Google, the creation of a digital lab, and training in innovative approaches (ANSA, 2016).

Focal Points of Contention

Unlike *Agence France Presse v. Google Inc.* and *Google Inc. v. Copiepresse*, which primarily depended on the legal framework of copyright, the dispute between FIEG and Google pursued the antitrust doctrine. The investigation of ACGM developed under the guidance of Article 82 of EC Treaty, the rules of European Commission on the abuse of a dominant market position. Article 82 targets "abusive exclusionary conduct by dominant undertakings," which aims to "exclude actual competitors from expanding or would-be competitors from entering a market" with the risk of "potentially depriving customers of more choice, more innovative goods or services and/or lower prices." (Antitrust, 2008, para. 1) Article 82 identifies common forms of "abusive exclusionary conduct," including "exclusive dealing, rebates, tying and bundling, predatory practices, refusal to supply and margin squeeze" (para. 4). One of the criteria for domination is the

market share. Given that Google accounts for 90 percent of the online search market in Italy (Cohen, Zampano, & Kiviniemi, 2009), far beyond the 40 percent threshold set by Article 82 (MEMO/08/761, 2008), the case was probed under the "dominant market position" framework to decide whether Google News Italia and its parent company violated Article 82 of the EC Treaty. In particular, Italian publishers' concerns about Google's market domination concentrated on two separate, yet closely related, issues: the links between Google News and Google Web Search and the lack of transparency and verification of advertising revenue sharing. The former was related to Google News, while the latter targeted Google's AdSense service.

Link between Google services and AdSense

In FIEG's complaint, the Italian publishers were concerned about Google News' power to "significantly affect the level of visibility" of publishers' online content and their advertising income (FIEG/Google, 2010). According to FIEG, while publishers have the option to opt out of Google News, their absence on Google News will lead to the exclusion from Google Web Search accordingly. Italian publishers were concerned that given Google's over-90 percent market share in Italy, the link between Google News and Google Web Search, for example, through the design of algorithms, could negatively affect the ranking, presentation, and online accessibility of publishers' online content. In response, Google presented several proposals, including introducing a separate crawler for Google News and more sophisticated REP to allow publishers the option to opt out of Google News but remain available in Google Web Search.

FIEG also accused Google News of using deep linking to harm news media's business model. Deep linking, maintained FIEG, direct users to particular news articles

that they are interested without them having to browse the homepage of a given news media's website. Italian publishers considered the homepage the most prominent space on a news website because it carried the news media's branding message and the advertisements on the homepage are usually the most profitable. Google defended its deep linking technology by arguing that the very promise of search engines is to save users time and effort in finding the most relevant information from the information sea on the web. Directing users to the homepage rather than specific news articles would harm the users' interest. This argument was supported by AGCM.

Another concern of FIEG was Google's advertising service AdSense. AdSense is an online advertising service provided by Google, through which Google plays an intermediary role between website owners who produce online content and advertisers who seek online advertising space by matching text and display ads to the content and visitor interest of the given website (Google, n.d.). In intermediating the buying and selling of online advertising space, FIEG complained, Google did not offer publishers ways to know and verify how advertising earning acquired via AdSense was calculated and shared and whether Google's intermediary role was justified. In response, Google Ireland, Google's European headquarter, revised the terms and conditions of AdSense by allowing more disclosure about the percentage of advertising revenue sharing and any decisions on the modification to revenue sharing.

The dispute between FIEG and Google News depended, unlike the lawsuits by Agence France Presse and the Belgian publishers, on the antitrust framework. This approach was due to the lack of a national copyright law updated particularly for the digital context in Italy. Although the Italian antitrust authority accepted pledges made by

Google, AGCM realized in its 2010 annual report to Italian Parliament, "a real imbalance between the value generated by publishing for the internet system as a whole and the proceeds which on line publishers are able to collect" that was not resolved in this case. It also pointed out the supranational nature of this imbalance complicated the relationships and the interests of involved parties (AGCM, 2011, p. 5). Having realized that without a solid legal basis the settlement was only temporary, AGCM submitted a report to the Italian government after closing the proceeding to call for new rules and regulations in order to respond to the new developments manifested in the dispute between FIEG and Google News.

Consequence

In January 2011, AGCM was satisfied with the commitments proposed by Google, and the probe was closed. Later, in July 2016, Google and FIEG reached an agreement, in which Google agreed to invest 12-million Euro over three years in helping FIEG publishers, through the use of Google tools, with distribution solutions and content protection (ANSA, 2016). About the same time, Google was involved in another dispute in Italy in which the Italian tax office, the Italian Revenue Agency, alleged that Google did not pay tax over the course of 2009 to 2013. In May 2017, Google agreed to pay 306 million Euros in back taxes (Reuters, 2017). From *Agence France Presse v. Google Inc.,* to *Google Inc. v. Copiepresse*, and then to the dispute between FIEG and Google, issues addressed in these cases went deeper and deeper, touching upon the fundamentals about Google, including its technology, the technical standards that Google used to support its operation, Google's business model, and Google's products as a network. These issues are vital for Google's existence because these are also the specialized areas that define

Google. The pressure from the news media industry required Google News to defend its legitimacy and to maintain its autonomy. In order to gain more autonomy, Google would rather pay millions of dollars to settle the disputes than be pulled into the regulatory systems that the news media industry advocated. At the same time, Google further strengthened its specialization in relation to news media. As shown in the cases examined above, Google argued against news media from a technological point of view. Even in those settlement deals, Google was promoting its own technologies, tools, and the ways of doing and thinking associated with these technologies and tools. That's because Google knew that a higher degree of specialization would help enhance its own capturing power and keep it from being captured by other social sectors.

United States

Background

In the United State, the Associated Press (AP), the American news agency headquartered in New York, pursued the licensing model for Google News to use its content as early as 2005 (Ali, 2007). Under this early licensing agreement, Google News was able to host AP's news stories and pictures by paying the AP, though financial terms of the use were not disclosed (McCarthy, 2006). The agreement was challenged by the AP in August 2006. AP and Google had to work on "new," "innovative" use of AP content on Google News aiming to "protect our [AP's] intellectual property and provide supplemental revenue to subsidize … news gathering and other services for members" (Kramer, 2006b). A new agreement was signed in August 2007, in which Google introduced a new duplication detection feature to highlight original news stories and the publishers who produce them (Cohen, 2007). Before the 2007 agreement was expired in

January 2010, the negotiation between the two parties came to a standstill. As a result, Google News stopped hosting new AP news content after December 2009 (Tartakoff, 2010). The months-long impasse was resolved in August 2010 when Google and AP agreed to extend the existing licensing agreement that permitted Google News to host AP content (Cohen, 2010). The length of the new deal was not disclosed (Krazit, 2010).

Focal Points of Contention

The ups and downs in the Google-AP relationship were affected by AP's concerns about its business model. AP had long used a business-to-business model, under which the news agency was able to develop a business network through content syndication among its member publishers. In this business model, AP invests in news production, including hiring reporters and maintains news bureaus worldwide[9] in producing original news content. Its revenue comes from content licensing to subscribing news organizations, AP's news production software Electronic News Production System (ENPS), broadcast facilities and editorial support to TV networks (GMS), and other (ap.org, 2018). According to the then-president and CEO of AP, Tom Curley, this business model had produced $700 million in revenue for AP in 2006, when AP and Google News were negotiating their early licensing agreement (Kramer, 2006a). AP complained that Google News did not benefit the AP in the way it promised, given AP's business model.

AP argued that since AP did not have its own consumer website, it was not benefited from the traffic driven by Google News as other publishers might have (Claburn, 2007). Traffic-wise, AP was concerned that Google News only kept users on its

[9] As of 2018, AP has 254 bureaus in 100 countries (ap.org, 2018).

own website rather than directed them to publishers who syndicate AP stories (Ali, 2007). This situation will devalue AP's role in its business network and make publishers less willing to pay for AP stories (Claburn, 2007). AP also criticized Google's business model for establishing a "self-referring network" (Seward, 2009), which allowed Google to promote its own properties and gain advertising revenue. For example, news items on Google News were also shown in Google search and Gmail (Kramer, 2006b). "And all that money has come to them, all 22 billion." said Tom Curley, "Folks, they can share" (Seward, 2009). This cry has been echoed recently in the news industry in the U.S. In 2019, the News Media Alliance, an organization that represents over 2,000 American newspapers, released a study that estimated that Google made $4.7 billion from the news industry in 2018. David Chavern, the president of the alliance stated that news publishers "deserve a cut" of that $4.7 billion (Tracy, 2019, para. 1). While the study's methodology was controversial as the research was questioned by some critics for taking "a random, 2008 comment made at a conference, about Google News' value [i.e. Google's Marissa Mayer estimated in 2008 that Google News brought Google $100 million in revenue], and extrapolate that to come up with a number that's supposed to represent how much revenue that Google has made off the news business," (Helmore, 2019) it provided evidence that the economic tension between Google and the U.S. news industry has not been settled after more than a decade.

To respond to the AP's charges, Google News introduced a new feature called "duplicate detection" in 2007 to identify and remove duplicates caused by syndication and to prioritize original news stories and news sources (Cohen, 2007). On the other hand, Google News also took proactive moves to strengthen its power in the disputes

with American publishers. First, Google News worked on securing and expanding its partnership with publishers worldwide. For example, Google News announced in 2009 that eight member news agencies from the European Pressphoto Agency had joined its Hosted News partners (Cohen, 2009). Google also made sure that publishers that syndicated AP stories still agreed to be listed on Google News despite the dispute between the AP and Google (Sullivan, 2010). In December 2009, Google News discontinued updating AP content on its website. Google News used these proactive actions to leverage its negotiation power against publishers by sending the message that "it can live without the AP should it come to that" (Krazit, 2010).

Commercial Rather Than Legal Approach

From the onset, the AP used a commercial, rather than legal, approach in the AP-Google case. Over the years, the AP had developed several rounds of negotiations with Google. Agreements were phase-based in that each covered a relatively short term, *e.g.* 2005-2007, 2007-2009, etc. Each round of negotiation aimed at specific and advanced demands. These strategies gave the AP the flexibility to adjust its negotiation position based on the development of Google News as a new player in the changing media market. The AP also applied PR strategies of being vocal at public events to set the agenda and shape public discourse.

Both parties strategically used the competitor effect. In addition to Google, AP also negotiated with other search engines and internet companies for better deals and more negotiation power against Google News. After Microsoft introduced its own search engine, Bing.com, in June 2009, AP worked with Microsoft for a partnership under conditions more favorable to the AP, including the priority of original news sources in

search results and the availability of real-time metrics for the AP to track and control the use of its content (Seward, 2009). At the same time, the AP also negotiated with Yahoo and praised Yahoo for having "always recognized the value and importance of original, authoritative news" (Liedtke, 2010). In July 2009, Microsoft and Yahoo announced a search deal, which allowed the synergy of the two companies to rival Google (BBC, 2009). The commercial approaches employed by both parties in the AP-Google case reflected the liberal, free-market style, which encourages open competition free from government intervention. These commercial negotiations were often kept from public view and exhibited a lack of transparency. For example, information about the financial details, specific conditions, and the length of the agreements were kept secret, treated as confidential commercial information.

DMCA and fair use

"We [AP] are not suing them [Google]" said Tom Curley, the then-president and CEO of the AP (Kramer, 2006a). An important factor that prevented the AP from pursuing a legal approach was the U.S. Digital Millennium Copyright Act (DMCA). On various occasions, the AP and its leaders criticized the DMCA as "misguided legal theories" (Sandoval, 2009, para. 4) that "enabled Google and the Google wannabes to do what they are doing" (Seward, 2009, para. 9). Under the DMCA, Robert Thomason, the editor of the Wall Street Journal in 2009, said that, "readers have been socialized – wrongly I believe – that much content should be free. ... And there is no doubt that's in the interest of aggregators like Google who have profited from that mistaken perception" (Sandoval, 2009, para. 3).

In the U.S., the DMCA was signed into law by President Bill Clinton in 1998. As the Congress's attempt to update American copyright law to keep pace with the internet era, this legislation aims to clarify copyright and related liability in the digital context. Section 512 of the DMCA defines the copyright infringement liability of online service providers (OSPs). DMCA provides both narrow and broad definitions of OSPs. A narrowly defined OSP is "an entity offering the transmission, routing, or providing of connections for digital online communications, between or among points specified by a user, of material of the user's choosing, without modification to the content of the material as sent or received," and a broadly-defined OSP is "a provider of online services or network access, or the operator of facilities therefore" (DMCA, 1998). Google News is one of these OSPs by both narrow and broad definition.

According to Section 512, four categories of conduct of OSPs are exempted from online copyright infringement liability:

- Transitory communications where OSPs merely act as "data conduit" to transmit digital information from one point on the network to another at the user's request
- System caching that OSPs retain copies of online materials for a limited time and transmit them at the direction of users
- Storage of information on systems or networks at direction of users
- The use of information location tools, such as hyperlinks, online directories, search engines, and the like.

To pursue these limitations, OSPs must not have the requisite level of knowledge of the infringing activity. OSPs' practice in system caching has to be carried out through an

automatic technical process for the purpose of increasing the accessibility of information requested by users. Online materials used by OSPs must be transmitted at the direction of users. And materials that infringed copyright must be removed by the OSPs upon receiving proper notification of claimed infringement.

Overall, the DMCA provides a series of copyright safe harbors for OSPs in using online materials in their operations, which explains why Google attempted to have American law be the applicable law in the Belgian case. The safe harbors provided by the DMCA are complements to the fair use doctrine of U.S. copyright law. U.S. copyright law does not define "fair use," a concept that originated in common law and was made part of statutory law in 1976, beyond noting that "[T]he fair use of a copyrighted work, including such use by reproduction in copies or phonorecords or by any other means specified by that section, for purposes such as criticism, comment, news reporting, teaching (including multiple copies for classroom use), scholarship, or research, is not an infringement of copyright" (Copyright Act. 17 U.S.C. § 107). U.S. law does, however, outline four factors based on which the courts can judge whether the given use of copyrighted work is a fair use:

- The purpose and character of the use, *e.g.* is the use of the copyrighted work commercial or nonprofit in nature?
- The nature of the copyrighted work, *e.g.* is the work creative?
- The amount of the work used in relation to the copyrighted work as a whole
- The effect of the use on the potential market value of the copyrighted work (Copyright Act. 17 U.S.C. § 107).

In 2006, Google successfully won a lawsuit filed by Blake A. Field for copyright violation in the United States (*Field v. Google Inc*., 2006). In *Field v. Google Inc.*, the United States District Court for the District of Nevada ruled that Google was entitled the protection of the DMCA because Google's cache was carried out through an automated technical process and for the purpose of improving access to information on the internet upon users' request. The court found that Google used industry-standard protocols in its operation for the opt-out option, which was a "good faith" that presupposed fair use. The Court concluded that Google's use of the copyrighted work was a fair use. The decision was primarily based on the first condition of fair use defined in the copyright law that Google's use was transformative, rather than superseding the original use, and that it aimed to benefit the public in accessing information on the internet, which was "socially valuable." (*Field v. Google Inc*., 2006) Even though not specifically against Google News, the case set a precedent for Google to use the DMCA, which reflects the First Amendment's commitment to protecting the freedom of information, to pursue protection of online copyright infringement liability in the United States.

Recent developments

Recently, especially after the 2016 presidential election in the United States, the regulatory climate has seemed to change in the United States, in that there has been a stronger voice calling for the end of digital platforms' self-regulation era. The news industry has played a role in driving this climate change by waging several anti-platform campaigns that have put digital platforms into a series of crises about fake news, data privacy breaches, censorship of different political perspectives, Russian interference in the 2016 election, and the spread of harmful information. Congress has held a number of

hearings on these issues since 2017 (Byers & Fiegerman, 2017; Wong, 2018; Sherr, 2018; Zhou, 2019). The most recent charge filed by the news industry against digital platforms in the United States targeted large platforms' anti-competitive practices. Congress held a hearing on June 10, 2019, to investigate these platforms' dominance in digital markets and their influence on America's news media industry (Mui, 2019). On March 7, 2018, U.S. Rep. David N. Cicilline, a Rhode Island Democrat who was then the ranking minority member of the House Judiciary Committee's Subcommittee on Antitrust, Commercial and Administrative Law, proposed the Journalism Competition and Preservation Act, a bill supported by the News Media Alliance, which aimed to allow a four-year antitrust exemption for "the publishers of online content to collectively negotiate with dominant online platforms regarding the terms on which their content may be distributed" (H.R. 5190, para. 1). Cicilline reintroduced the bill—with co-sponsors Doug Collins, a Georgia Republican, and Mark DeSaulnier, a California Democrat—in April 2019 (H.R. 2054). It was referred to the House Subcommittee on Antitrust, Commercial and Administrative Law in May 2019.

These updates showed that the news industry's institutional power through public and policy agenda-setting and its close connection to the political realm. In response, digital platforms have also increased their lobbying effort. It has been reported that Google, Facebook, and Amazon combined spent more than $65 million in lobbying Congress in 2018 (Ng, 2019). In addition, digital platforms have also attempted to play a bigger role in influencing the political agenda in the United States through, for example, their involvement in election campaigns (Kreiss & McGregor, 2018). By pursuing these avenues, digital platforms' efforts to differentiating themselves from news media for a

higher degree of autonomy has depended on not only specialized technologies, like those Google used in the early lawsuits, but also political power.

UK

Background

While seeking expanded partnerships with publishers in the UK, Google came under fire from Rupert Murdoch, founder and Chairman/CEO of News Corp, who criticized Google News and other news aggregators as "parasite" and "content kleptomaniac" (Rushton, 2012). The tension between Google and News Corp. culminated in 2010 when News International, News Corp's UK subsidiary, blocked Google and Google News from indexing content of its newspapers *The Times* and *The Sunday Times*. This decision was reversed later and these newspapers' content was back on Google and Google News (Zara, 2012). Nevertheless, the dispute between News Corp's Britain business and Google continued.

Discursive Strategies

Murdoch and his News Corp. have been severely critical of news aggregators and search engines for "stealing" news media's content (Shafer, 2009), but unlike publishers in other European countries who sought copyright protection, Murdoch didn't pursue litigation. He focused his charge on the idea of quality: quality journalism and quality traffic. Murdoch asserted that Google took a free ride on the news media's content, which produced profit for Google but threatened the economic basis of quality journalism, which made news media unable to afford expensive resources, such as international bureaus. Furthermore, while Google claimed to send traffic to news media's websites, such traffic, Murdoch and his supporters argued, was the "least valuable" traffic, as it

only brought news sites users who "more often than not read one article and then leave the site," rather than loyal readers who have high advertising value. "We'd rather have fewer people coming to our website but paying." Murdoch said in his interview with Sky News Australia (Interview with David Speers, 2009).

Murdoch's anti-Google argument was that quality journalism costs money. Although also related to economic interests, Murdoch's charges had strong discursive power because the idea of "quality journalism" was often associated with democracy and the public good. Murdoch adeptly deployed publicity techniques to gain discursive power over Google. He used public presences, whether media interviews, Federal Communications Commission hearings, industry and academic events, and social media discussions, to portrait Google as engaged in the "theft" of digital news content and a threat to quality journalism (Sarno, 2009). This tactic banked on the supposition that Google might care far more about its public perception than the content and the very limited traffic that News Corp. could bring in (Swisher, 2009). In addition, like the AP in the U.S., News Corp. also leveraged the competition on the search and news aggregation market by strategically involving different players for different purposes, the strategy the AP employed in the United States, including its attempts with Microsoft (Swisher, 2009) and "second-tier" search engines, such as ask.com (Franzen, 2009).

Experiments

In addition to discursive strategies, News Corp. UK took actions that involved various experiments and advocacy strategies. In 2009, News Corp. and its Britain arm, News International, launched Project Alesia (Swisher, 2009-b), News Corp's attempt to create its own online news aggregation service for News Corp's core print titles and

third-party producers across all digital platforms. The project recruited more than 100 people in Britain with about $32.3 million investment from News Corp. (AFP, 2010). The project was called off in 2010 as a result of failing to attract a "critical mass" of publishers who were willing to participate (AFP, 2010). Project Alesia represented News Corp's acknowledgement of online news aggregation as a useful digital distribution method, but that News Corp. would prefer its own news aggregation service against Google News. This action attempted to build up boundaries between "old" and "new" media. Project Alesia aimed at a walled garden for vertical integration, which was not compatible with the open and free nature of the web, while these ideas have been integrated in Google's public discourse to legitimize its service.

Following the example of the Wall Street Journal, an American newspaper owned by News Corp, in June 2010, News Corp. started a paywall experiment in the UK with *The Times* and *Sunday Times*, making these the first newspapers in Britain to apply the paywall model. Users of www.thetimes.co.uk and www.thesundaytimes.co.uk without subscriptions were charged £1 for 24-hour access or £2 for a week's subscription (Saltmarsh, 2010). As the paywall went up, content of these newspapers was no longer available on Google News and Google search (Fallows, 2010). News Corp. reported 105,000 new paying subscribers during the experiment, but its online readership in UK declined significantly. Within a few months, *The Times* website had a 62% drop in online readership and a 90% plummet in pageviews (Schonfeld, 2010). In 2012, content of *The Times* and *Sunday Times* returned to Google News and Google search. It was reported the newspapers reversed their decision because of the fear of "losing their influence because they do not appear in the search engine's rankings" (Rushton, 2012, in teser). The News

Corp's paywall experiment continued and reached "the most successful year" in 2018 when *The Times* and *Sunday Times* had a combined 500,000 subscribers and their digital subscriptions outnumbered print for the first time since the paywall model was launched in 2010 (Tobitt, 2018, para. 2).

Observers pointed out the failure of a series of digital attempts carried out by News Corp. was fundamentally due to its displaced perception that the internet was "mainly a *broadcast* medium, rather than a *communications* medium" (Masnick, 2010, para. 2). Others acknowledged News Corp's digital efforts that paved the way for news media to learn through trial and error. "The three most important things any newspaper can do now are experiment, experiment, and experiment. … Rupert Murdoch is doing just that" (Fallows, 2010, para. 4). These actions also pressured Google News to respond. For example, Google News revised its policy on subscription content in 2009. The previous "First Click Free" policy, which allowed users to access the first page of publishers' subscription content for no charge and for an unlimited time, was replaced by a new policy that limited the free access to paid news to five articles (Goldman, 2009). While whether digital subscription is a good way for news media's digital sustainability remains to be seen, major digital platforms have joined the effort to promote news subscription model. For example, in 2017, Facebook worked with European and American news organizations to launch a test that aimed to support news subscription models in Facebook's Instant Articles, a tool designed for publishers to distribute fast, interactive articles to their readers within the Facebook mobile app and Messenger (Brown, Hardiman, & Salari, 2017). Apple also launched its own subscription service, Apple News+, available in the U.S. and Canada though it focused in 2019 more on

magazines than news, with only three major newspapers included in its service at that time, the *Wall Street Journal*, *Los Angeles Times,* and *Toronto Star* (Apple launches Apple News+, 2019).

Hard-core conservatives, Murdoch and his News Corp. have become advocates of government intervention in platforms in recent years. In 2018, they called for an "algorithm review board" to strongly regulate companies like Google and Facebook (Schwartz, 2018). News Corp. also played a leading role in advocating collective action in the news industry to challenge Google and similar companies (Barnett, 2009). In 2016, it was reported that a consortium of major UK publishers were united against Google and Facebook by pooling their advertising, audience, and data resources to benefit publishers in the competition with their "common enemy" (Elder, 2016).

China

Background

In September 2004, Google News launched Google News in Simplified Chinese in China. Google News China (news.google.cn), which was specifically for mainland Chinese users, differed from the Chinese edition of news.google.com that is available to the world's internet users, with uncensored Chinese language news results. Google News China provided filtered news for mainland Chinese users. It did not link to news sources inaccessible from within China, nor did it return search results "sensitive for political or other reasons" (googleblog, 2004). In November 2004, it was reported that Google News in different languages was banned in mainland China, which left Google News China the only choice for mainland Chinese users to use Google's news aggregation service (Harris, 2004). In 2008, Google News China was among the websites that were

temporarily blocked during riots in Tibet (Nystedt, 2008). In January 2010, as Google China (google.cn), which also applied a self-censorship approach, announced that it was no longer willing to censor its search results in light of a cyberattack (Branigan, 2010), Google entered difficult negotiations with the Chinese government. After a two-month standoff, in March 2010, Google.cn announced that it would shut down its service in mainland China and redirect mainland Chinese users to Google's Hong Kong-based service (news.google.com.hk), which provided uncensored search results (Metz, 2010). Afterward, Google News China withdrew from mainland China, along with other Google services. Since then, Google has maintained limited presence in mainland China until 2017-2018 when Google secretly planned to launch a censored version of its search engine in China through its "dragonfly" project as an attempt to reenter the Chinese market (Gallagher, 2018).

Legal Basis and Focal Points of Contention

Google's business in China was controversial both in China and in the United States. The nature of the Chinese case was complex, as there was no formal litigation involved, yet the case had the strongest consequence: the shutdown of Google's presence in a country. In the Chinese dispute, several legal frameworks were pursed by different parties, including cybersecurity and WTO rules. Moral charges were also raised against Google in the United States.

Cybersecurity

Even though there was no formal litigation, the Information Office of the State Council of the People's Republic of China published a white paper in June 2010, three months after Google shut down its main service in mainland China, which pointed to

some areas closely related to the incident. Among them, cybersecurity was a major concern (Condliffe, 2016). Compared to the U.S., where internet intermediaries enjoy broad safe-harbor provisions based on the content-conduit distinction, and most of the European Union nations, where these companies are offered "conditioned liability," China tends to require "strict liability" on internet companies (MacKinnon, Hickok, & Lim, 2015). A fundamental assumption of the strict regulation is that the Chinese government deems the internet to be the "natural expansion" or "a new dimension" of its national sovereignty. Based on this assumption, the Chinese government has the "unquestionable authority over the Internet" (Liu, 2012, p. 53).

The safety of China's cybersecurity is an overarching principle throughout China's internet laws and regulations. In China, in order to get an Internet Content Provider (ICP) license, internet companies that run businesses in China are required to comply with the Measures for the Administration of Internet Information Services (MAIIS), which, issued by the State Council on September 20, 2000, guides China's ICP licensing system. The measures stipulate that ICPs shall not produce or disseminate any information that "jeopardizing the security of the nation, divulging state secrets, subverting state power, or jeopardizing the integrity of the nation's unity." (Measures for the Administration of Internet Information Services, 2000, Article 15) ICPs that produce and disseminate information that harms China's cybersecurity are illegal and subject to China's Criminal Law. The block of Google News during the Tibet riots in 2008 was an example of such "jeopardizing" activities. While there was no formal litigation filed that was directly related to Google's shutdown in China, Chinese mainstream media generally addressed the case as Google having failed to comply with Chinese laws (Xinhua, 2010).

Censorship and WTO rules

Unlike Chinese media, Western media dominantly reported that censorship in China caused Google's shutdown. On January 21, 2010, when Google had intense negotiations with the Chinese government, then-U.S. Secretary of State Hillary Clinton said in a speech on internet freedom that, "censorship should not be in any way accepted by any company from anywhere." She urged organizations worldwide to address "threats to internet freedom around the world" and maintained that, "the private sector has a shared responsibility to help safeguard free expression" (Clinton, 2010, p. 56).

In the wake of Google's shutdown, Google, the American government, and top U.S. trade officials sought to use WTO rules to charge China with using censorship to discriminate against foreign companies (Palmer, 2010). On the Chinese side, Zheng Zhihai, the deputy director and secretary general of the China Society for World Trade Organization Studies, argued that the WTO entitles its member states the lawful right to supervise and censor internet content by acknowledging openness and supervision as two inseparable aspects of market access. Zheng also argued that China applies the same internet supervision to both domestic and foreign companies, without discriminating against international companies, including Google (Zheng, 2010). Given the long process involved in the WTO system of hearings and appeals and the uncertainties in the process and in the results of the WTO ruling, a WTO complaint, observers held, would serve as only "a negotiating pawn [rather] than a formal legal action" for the United States to pressure Beijing to make concessions in the Google-China dispute in particular and in the U.S.-China relationship in general (Buckley, 2010).

"Do no evil"

The self-censorship approach that Google applied in China was more controversial in the United States than in China. Google's self-censorship was widely criticized in the United States as contradicting Google's "do no evil" motto (Wired, 2004). As a company that aims at the international market, Google added a specific mission to its fundamental commitments: "be responsive to local conditions" (Schrage, 2006). In the case of China's MAIIS, these local conditions included having a local server and maintaining related information records for 60 days and making them available to all relevant government agencies examining them pursuant to law in order to get China's ICP license (Measures for the Administration of Internet Information Services, 2000). Additionally, ICPs also have to follow self-disciplinary regulations in China to establish internet security management systems and technical measures to prevent illegal information. These measures are reinforced in the Cybersecurity Law issued in 2016 (NPC, 2016).

Inside the United States, Google, as well as a few other American technology companies, was severely criticized by the U.S. Congress. Critiques largely focused on ethical and moral issues, especially whether Google harmed human rights and humanity through self-censorship (Gunther, 2006). In testimony before the Congress in 2006, Google's Elliot Schrage defended Google's Chinese business decision as "a meaningful – though imperfect – contribution to the overall expansion of access to information in China" (Schrage, 2006, p. 1). By defining itself as a global promoter of information accessibility, Google argued that, "On balance we believe that having a service with links that work and omits a fractional number is better than having a service that is not available at all" (googleblog, 2004). With this argument Google attempted to justify the

objective of “expanding access to information worldwide” as a legitimate mission that provided an ethical and moral ground for the dispute between Google and China in order to answer the public doubt about Google’s “do no evil” commitment.

In 2018, Google’s attempt to return to the lucrative Chinese market by reintroducing a censored search engine for China, known as Dragonfly, encountered protests from Google employees and human rights groups worldwide. As a result, the Dragonfly plan was shelved (Gallagher, 2019). Unlike the Western disputes examined in this chapter, the dispute about Google in China did not address any economic issues; instead, it had strong political and ideological complexion, which was shaped by the US-China relationship. Such influence was evident in Google’s role in the 2018 US-China trade war when Google decided to comply with the American government’s ban on Huawei, China’s telecomm company (Moon, 2019). The Google dispute in China shows that digital platforms that aim at global media and technology markets are subject to influences not only from professional, technological, and economic domains, but also negotiations driven by international relations—as well as moral and ethical constraints required by different parties domestically and internationally. These issues, as some U.S. tech companies have realized, “are larger than any one company or any one industry" (Gunther, 2006).

Germany

Background

In 2012, a new law, also known as Ancillary Copyright Law for Press Publishers (Leistungsschutzrecht für Presseverleger, or LSR) was drafted in Germany. The law particularly targeted search engines and online news aggregators, preventing these online

services from using German publishers' news content without paying a fee. After intense debate, compromises among stakeholders, and at least three drafts, the law was passed by the German Bundestag and went into effect on August 1, 2013. Google refused to recognize the applicability of the law and the payment obligation. In June 2014, VG Media, the collective that represented over 230 German publishers, sued Google for copyright violation under LSR. In October 2014, Google decided to only display links and headlines for those German publishers who did not opt in Google's news aggregation service (AFP, 2014). Shortly afterward, all Germany publishers represented by VG Media gave Google permission to host their content for free due to the "major economic pressure" of losing traffic (Hebbard, 2014). VG Media has filed several civil complaints against Google since then to enforce the LSR (Lauer, 2016). In May 2017, the Berlin Regional Court ruled that the lawsuit between VG Media and Google was "at least partially justified" but the court needed to examine whether the German Federal Government violated a European Union Directive by failing to notify the EU about the law in 2013 (Dwyer, 2017).

LSR and Controversies

The LSR was drafted as an addendum to Germany's Copyright Act of 9, September 1965, on the right of press publishers over their content, its transferability, the duration and limits of the right, and the interests of the author (Federal Law Gazette, 2013). The new law granted German press publishers a one-year exclusive right over their content after it was made publicly available online for commercial purposes. The addressees of the LSR are defined as search engines and news aggregators. The law states that the public accessibility of press products is permissible but not for "commercial

providers of search engines or commercial providers of services that condition the content accordingly." The latter, according to the European Publishers Council, referred to news aggregators (EPC, 2013). The stipulations contained ambiguities. For example, the law permitted the free use of "single words or very small text excerpts" but did not specify the length permitted, especially in the case of a news snippet. The law did not specify how publishers should be compensated for copyrighted works, whether through contract, license, tax, or other form of remuneration.

Google and opponents of the laws questioned whether the law was necessary in the first place, given that technical tools, *i.e.,* metatags and robots.txt protocols, were available for publishers to opt out of Google News or choose not to include certain elements of their news content, such as the snippet, in Google News results (Oberbeck, 2012). Publishers who supported the law insisted that a legal basis was urgently needed to differentiate undesirable and desired uses of their content (Essers, 2013). Fundamentally, this was a war of control between new and incumbent social sectors. While the technical standards may be effective in certain ways, they were produced and advocated by the technology community—the new comer in the media ecosystem. The legal solution, however, was lobbied for by the big players of the German media industry and legislative powers (Sterling, 2012)—the incumbents. This was a question about who was following whose rules, hence, who was in control. With controversies, fights, and compromises, and without issuing a notification to the European Commission (although the EU Directive 2015/1535 requires EU member states to notify the EU of any draft technical regulation prior to its adoption), in March 2014, the LSR was passed 294-243 in the parliament, supported by Germany's ruling coalition parties—the centre-left CDU

(the Christlich-Demokratische Union Deutschlands, or Christian Democratic Union of Germany) and the liberal party FDP (the Freie Demokratische Partei or Free Democratic Party) (Meyer, 2013).

Aftermath

In the aftermath of the passing of the Ancillary Copyright for Press Publishers, Google tactically used technical methods to bypass the payment obligation required by the Ancillary Copyright law. In 2013, Google introduced a new confirmation system, which required publishers to submit verification information for inclusion (Rabenstein, 2013). Unlike the previous approach, in which Google allowed news publishers to opt out of its services by using robots.txt and metatags, the new confirmation system was based on the opt-in system. This change from opt-out to opt-in was a defiance of the control power that news publishers attempted to seek. It effectively reinforced Google's technological control in its relationship with German media. Then, in October 2014, Google simply removed the snippets and images with only headlines and links available for German publishers that did not opt in (Axel Springer concludes its data documentation, 2014). One month later, after seeing the drop of the online traffic, publishers represented by VG Media decided to give Google permission to use their content for free but criticized Google for "extortion" (Hebbard, 2014).

Axel Springer, one of Europe's largest media companies and publisher of Bild, Germany's biggest mass circulation newspaper, saw a 40 percent traffic drop from Google and an 80 percent drop from Google News (Becker, 2014). Under economic pressure and in fear of losing online relevance, Axel Springer gave its permission for Google to use its content for free. Axel Springer identified this decision as "the most

successful failure," and said "we now know very precisely just how far-reaching the consequences of the discrimination are, as well as the real effects of Google's market power and how Google punishes everyone who exercises a right that has been granted to them by the German Bundestag" (Axel Springer concludes its data documentation, 2014, para. 4). In the cases examined in this chapter, Germany was the first country that passed new legislation to manage the disputes between Google and news organizations. However, due to the law's ambiguity, Google was able to bypass the law through technological tactics, which significantly challenged the effectiveness of legal intervention.

Spain

Background

In 2014, Spain reformed its copyright law, the Ley de Propiedad Intelectual (LPI), in which online news aggregators were required to pay Spanish news publishers unwaivalbe compensation to use their news content. The new law took effect on January 1, 2015. The law is also called the "Google Tax" or "Canon AEDE" as it was considered to specifically target Google News and was a product of the lobbying of the Association of Spanish Daily Publishers—in Spanish the Asociación de Editores de Diarios Españoles (AEDE)—which represents large news publishers in Spain. On December 16, 2014, two weeks before the official implementation of the new LPI, Google News shut down its service in Spain.

New LPI Article 32.2

The new LPI, passed in 2014, introduced amendments to Spain's copyright law. In the new law, section two of Article 32 was modified and sections 3, 4, and 5 were

added. Article 32.2 was most controversial and resulted in the law being dubbed as a "Google Tax" or "Cannon AEDE" as it specifically targeted the news aggregator. Section two of Article 32 allowed online aggregators to use "non-significant fragments" of original content for information, public opinion, and entertainment purposes. In exchange, it entitled the editor or, if applicable, other rightholders, the right of fair compensation for such uses. The right, according to the new law, was inalienable and would be executed by collecting societies, an organized body licensing and managing copyrighted works on behave of copyright owners (BOE-A-2014-11404). Spain's Canon AEDE was widely read as following the precedent of the LSR in Germany, but the Spanish law was distinguished from the LSR by its arbitrary nature. By setting up the inalienable right, Canon AEDE defined a "requirement" for publishers rather than a "privilege" (Gagne, 2015). This mandatory right was also considered a lesson learned from the German case, in which, since the compensation was not mandatory, Google was able to use technical tactics to circumvent the law (Masnick, 2015).

In the second paragraph of Article 32.2, search engines were excluded from the obligation defined above, with three specific conditions:

- The use of content was not for the search engine's own commercial purpose
- The content was used to respond users' queries
- A link to the original content is available (BOE-A-2014-11404).

The clear discrimination made between online aggregators and search engines was a unique characteristic of the Spanish case. Spain may have made this distinction based on its judgment about previous cases. In the Belgian case, for example, Google argued that Google News was merely a search engine as part of an attempt to invoke a

copyright exception enjoyed by search engines. The German case also showed that publishers did not want to be excluded from Google search, given the dominance of Google on the search market, which determines, in part, publishers' online relevance and revenue. Such distinction between online aggregator and search engine defined in Canon AEDE allowed Google to tactically respond to the law without having to lose its influence on the Spanish news market, as discussed next.

Google News Shutdown

Before Canon AEDE came into effect on January 1, 2015, Richard Gingras, then the head of Google News, announced on December 11, 2014, that Google was going to close Google News in Spain on December 16, 2014. Spanish publishers would be removed from Google News outside Spain as well (Gingras, 2014). On December 16, a visit to Google News Spain (news.google.es) only showed a closure message (Sterling, 2014). Nevertheless, due to the distinction between the news aggregator and the search engine addressed in Canon AEDE, observers found that after the shutdown, news from Spanish publishers remained available on Google search in two forms: news box and the "news" tag (Sterling, 2014). News box is a segment embedded in Google web search results. When there are news stories relevant to a searcher's query, these news stories will be included in Google web search results and returned to the searcher in a news box. The other option to get Spanish publishers' news on Google despite the shutdown was to get news through the "news" tab on the Google web search page right below the search window after the searcher submitted a search query from Google homepage. These are the features that Google defined as "Google Universal Search," a strategy that Google uses to integrate search results from its various properties, such as news, maps, images,

and videos, to Google web search results (Fox, 2007). By showing news items in Google universal search, news articles become particles of the web search results, which are exempted from copyright obligation in Spain. By employing this technological tactic, Google once again successfully circumvented Canon AEDE to keep Spanish news in its core business.

Impact of Google News Shutdown

Studies on the impact that the Google News shutdown had on the Spanish media market revealed mixed results. For example, while Chartbeat, an American web analytics company, recorded a 10 percent to 15 percent traffic drop with 50 large Spanish news publishers within hours after Google News was shutdown, AEDE claimed that the association would not worry much, considering that Google News contributed only about 5 percent of the overall traffic for its members (Wohlsen, 2014). Other analyses found that while external traffic, e.g. traffic brought by search engines, dropped sharply, internal traffic actually rose, indicating that Spanish readers who were previously directed by Google News navigated to Spanish news websites themselves once Google News was not anymore available (Ingram, 2014). Others reported that the Google shutdown in Spain had a negative impact on Spanish publishers, but it was not as negative as in Germany (Olmedo & Caballero, 2015). These authors also pointed out that the traffic decline also reflected the overall decrease of the internet penetration between 2014 and 2015 in Spain.

The new LPI law encountered strong challenge from groups such as the Asociación Española de Editoriales de Publicaciones Periódicas (AEEPP), a large association of Spanish periodicals. AEEPP expressed concerns about the negative effects

of the new LPI law, including market concentration, restriction of innovation, and legal uncertainty (AEEPP, 2015). In response to the AEEPP study, some scholars recognized the new LPI law as a positive influence. Gagne (2015) pointed out that the passing of Canon AEDE was the first time "a nation within the E.U. provided a voice for its domestic publishers by way of tangible regulations" (p. 233). The law, the author argued, increased the incentive of Spanish publishers for innovation in their competition with each other for more traffic.

The New European Union Copyright Directive

This analysis reveals that European countries used different national laws to deal with disputes about Google. To avoid a fragmented internal market in Europe, on March 26, 2019, the European Parliament approved, 348-274, a new Copyright Directive that aimed at a "harmonized legal protection for press publications in respect of online uses by information society service providers" (P8_TA-PROV(2019)0231). Article 11 of the new EUCD directly targeted large news aggregators, granting news publishers at the EU level the unwaivable right to have compensation, in the form of a license fee, when news aggregators use their news content. The new EUCD adopted lessons learned from previous European cases—for example, the mandatory right for compensation in Spain's Canon AEDE and the non-discrimination between search engine and news aggregator in Germany's Ancillary Copyright Law—to avoid Google using technical tactics to bypass regulatory measures. As to the key online news elements made available on Google News that concerned European publishers—including news headlines, images, and snippets—the new EUCD only exempted the use of hyperlink and "individual words or very short

extracts," the latter leaves room for EU countries to interpret the law. Overall, the new EUCD reflects a tendency toward stricter regulation over news aggregators.

In early 2019, Google conducted an internal experiment to test what its news search result pages would look like if complying with the updated EU Copyright Directive. The result was a "denuded" news page that presented only the hyperlink and part of the news sources' names, as shown in Figure 4-1, with no news headlines, snippets, or images (Sterling, 2019). How would Google respond to these legal reforms? Would Google be able to work out new technical tactics to bypass? And how the recent legal developments in Europe would affect the regulation style in the United States and other parts of the world? These questions remain unanswered at the time when this dissertation is written, but they will certainly shape the news ecosystem beyond 2019.

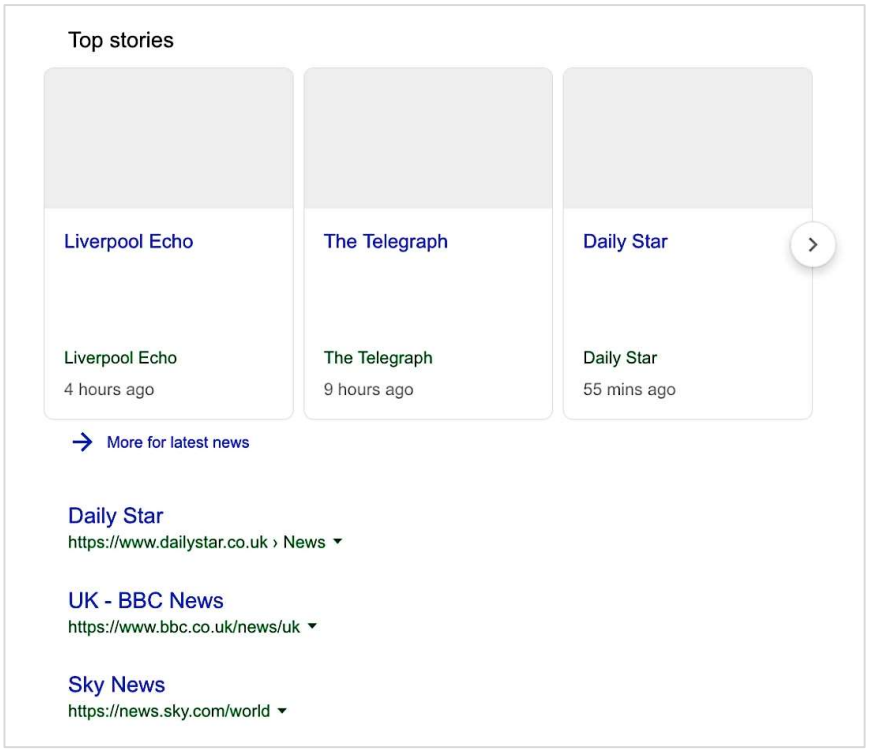

***Figure 4-1*. Google's internal experiment tested the impact of amended EU Copyright Directive. Source: Sterling, 2019.**

Discussions

This chapter provides a global and comparative view that examines and compares eight disputes about Google across Europe, America, and Asia. This analysis demonstrates that stakeholders in different media and political cultures depended on distinct approaches, from litigation, lobbying, and commercial negotiation to political and diplomatic means, when dealing with their disputes with Google. The stakeholders used different legal frameworks—including copyright, antitrust, cybersecurity and WTO rules—to pursue their interests in economic, political, ideological, and moral realms. These disputes generated different consequences, from settlement with commercial agreement, passage of new laws, to service shutdown. Table 4-1 lists the eight case studies examined in this chapter and their characteristics.

COUNTRY	YEAR	PLANTIFF/ACCUSER	APPROACH	LEGAL FRAMEWORKS	FOCAL DISPUTE	CONSEQUENCE
FRANCE	2005-2007	AFP	Litigation	Copyright	Copyrightability of news	Settlement
BELGIUM	2006-2012	Copiepresse	Litigation	Copyright	Technological issues business model	Settlement
ITALY	2009-2011	FIEG	Official compliant	Antitrust	Abusive market domination Linkage between Google products	Settlement
US	2009-2010	AP	Business negotiation	N/A (DMCA)	Licensing agreement, business model	Google discontinued AP content shortly and agreement reached later
UK	2010-2012	News Corp in UK	Took down content from GN, experiments, advocating	N/A	Paying for quality journalism	Took down content from GN and returned shortly after
CHINA	2010-	Chinese government	Political negotiation, diplomacy	Cybersecurity, WTO	National security, censorship	Shutdown
GERMANY	2012-	VG Media	Litigation	Copyright—Ancillary Copyright Law	Compensation Technical standards	Passage of legislation; VG Media publishers opted-in Google News;
SPAIN	2014	AEDE	Lobbying	Copyright—Canon AEDE	Mandatory compensation	Passing of new law; shutdown

***Table 4-1*: Comparison of the Eight Case Studies and Their Characteristics. Source: author.**

The disputes examined in this chapter span more than a decade. Today, they still have profound impact on the global media, technology, and political landscape, as well as important implications for law and policy-making. These cases also outline how normalization and differentiation trends interacted when Google responded to pressures coming from different stakeholders in the changing media ecosystem.

In the early cases, Google largely depended on its specialized technologies to defend its legitimacy. For example, in *Agence France Presse v. Google Inc.* in 2005, the different perception of the nature of news work between Google and AFP reflected different journalistic traditions in the U.S. and France. How Google perceived news was also defined by its specialized technologies, as news was "datafized" in its algorithmic procedures, which reduced news into fact-based part and parcel of Google's data systems. Later, the power relation started to shift. While in *Google Inc. v. Copiepresse*, Google accused Belgian news media of using their "dominant position" to restrict Google's access to local market, Google became the one accused of holding the dominant market position in the Italian case. As Google grew, news media tended to go deeper to touch upon more fundamental issues about Google's operation. In the following cases, Google's specialization was addressed in different areas, including specialized technologies (such as cache, automation, and deep link), technical standards (such as REP), and its business model supported by technological specialization (such as AdSense and Google's lock-in effect across different products).

During this period, differentiation escalated on both sides. On the one hand, news media sought self-control by refusing Google's technological and economic capture; on the other hand, Google resisted to be integrated into the regulatory frameworks advocated

by news media. In a few cases, Google used its technological and economic power to bypass legal constraints. Over these years, Google made responsive changes in its news aggregation service to answer news media's charges, but its urgent goal was to differentiate itself from the news industry to ensure it was not captured into the news industry's institutional power. Differentiation in this case was one social sector's resisting to be normalized by other social sectors.

In later cases, the media industry showed stronger institutional power in setting public and policy agenda. For example, news media were able to use discursive strategies to shape the public perception of digital platforms and wage anti-platform campaigns to put platforms into deep crisis, as in the cases in the U.S. and the UK; the news industry also had strong lobbying power to push the passage of new legislations aiming to incorporate Google into the regulatory systems preferred by the news industry, as what happened in Germany and Spain; In China, Chinese media were in line with the party-state; even in libertarian media systems, such as the U.S. and the UK, there has been an growing voice calling for stricter regulations over digital platforms and collective actions of news organizations to bargain with these platforms. Unlike the news industry, which is an established social institution that has developed its relationship with other social institutions for a long time, digital platforms as a newer industry is still building such relationships. This relationship with other social sectors is an important pillar of the institutional legitimacy. Under such pressures, Google has also worked on developing its own power in areas in addition to technology specialization, for example, expanding partnership networks, and enhancing political influence, because specialization in

technology alone was not enough to sustain its autonomy in relation to established social institutions like the news industry.

From a global perspective, Google—as a digital platform that has global business—has encountered disputes around the world. These encounters revealed that Google's development is subject to complex influence from different stakeholders in the global media, technology, and political landscape. These disputes displayed national interests, international relations, political, economic, and ideological climates in different contexts, and moral and ethical requirements addressed by different stakeholders. The decision of whether American law or Belgian law should be applied in *Google Inc. v. Copiepresse*, the Italian antitrust authority's call for attention to the supranational nature of the Italian dispute, and the U.S.-China relationship that shaped Google's experience in China all demonstrated the complex dynamics at the global level. It's important to consider these digital platforms in a global context and pay attention to the interaction between international and domestic influences.

So far, this dissertation has examined the origin and early history of Google's news aggregation service, the homepage changes on Google News between 2002 and 2019, and international disputes about Google News. These examinations have demonstrated that Google's specialization in technology has played an important role in Google's normalization and differentiation processes. The next chapter will go behind the scenes to explore Google's news-related technologies and algorithms.

Chapter 5. Google's News-related Technologies and Algorithms

The examination in previous chapters reveals that specialization in technology largely defines Google, especially when Google attempts to distinguish itself from traditional news media. This chapter aims to demystify Google's news-related technologies and algorithms and their evolution trends over the past seventeen years by proposing an innovative research method—patent analysis.

Algorithm

Computer algorithms are "the logical series of steps for organizing and acting on a body of data to quickly achieve a desired outcome" (Gillespie, 2016, p. 19). Today, almost all of the most popular and economically successful internet-based services rely heavily on algorithms of different kinds (Latzer et al., 2016). Algorithm scholars see the technical and the social as the dual role of algorithms (Dörr, 2016), which define algorithm as "a complex sociotechnical assemblage" (Gillespie, 2016, p. 24). To understand the social aspect of algorithm, one needs to understand how it works technically. However, the dynamic and secret nature of algorithms sets challenges for algorithm studies.

Algorithm serves the computational process of problem solving, which involves input, throughput, and output (Lazter et al., 2016). In this process, algorithms play an intermediary role by linking user, data, and applications. Algorithm is dynamic in nature because it is subject to influences from various sources. Market feedback, organizational considerations, technological developments, and social, political influences could all affect the I-T-O environment. As a result, algorithm should not be treated as singular artifacts, but as algorithm systems (Seaver, 2013) and a "network of actions upon

actions" (Goffey, 2008, p. 19). The dynamic nature requires systematic analysis of algorithm over a long time frame.

Since algorithms are designed in computer language and operated in the back-end, they are often considered a black box elusive to outside observers (Diakopoulos, 2014). Given the algorithmic power in intellectual, economic, and political domains, some scholars call for Algorithmic transparency for the purpose of monitoring, checking, criticism, or intervention by interested parties (Diakopoulos & Koliska, 2016). Other scholars argue that a degree of secrecy could prevent algorithm holders from malicious attacks, gaming, and manipulation (Cutts, 2008; Manber, 2008). It also encourages healthy competition between similar online services, and maintains competitive advantages and effective cost-gain management (Granka, 2010; Granados & Gupta, 2013). As the debate of algorithmic transparency goes on, researchers often have to depend on reverse engineering to study algorithms (Diakopoulos, 2014). As this study will illustrate, patent analysis could be used as an innovative method for reverse engineering algorithms.

Patent and Patent Analysis

What Is a Patent

Issued by the United States Patent and Trademark Office, a patent is a legal document that describes an invention and its usage right in order to protect the authorization of the owner of the patent (WIPO, 2008). The patent offers the "grant of a property right" to the inventor who "invents or discovers any new and useful process, machine, manufacture, or composition of matter, or any new and useful improvement thereof" (USPTO). Google holds thousands of patents and patents pending (Regalado,

2013). As a company specialized in technology, Google's inventions primarily relate to technical processes or methods, which disclose rich information about Google's technological development and algorithmic design.

A patent file is a semi-structured document, which follows required format. The format varies in different countries and different offices, but the main content usually requires specification, metadata, and drawing. Specification contains sections that provide written description about the invention, such as the title, abstract, background, summary, description, claims, and more, while metadata are informative details about the invention, including patent number, date, inventor, classification, etc. Drawings are used to illustrate the features of the invention specified in description and claims. Together, detailed explanation of the invention in terms of its features, functionalities, and procedures is provided in the patent file. In this study, such information is significant in understanding the rationales, trends, and specific factors that shape Google's news-related algorithms.

Classification

On the patent file, classification is a major element of the metadata to understand the nature of the invention in terms of the subject to which the invention relates. Patent classification is a system that organizes patent documents into specific technology groupings based on common subject matter. The United States Patent and Trademark Office currently adopts the Cooperative Patent Classification (CPC) system, which defines the subject of the patented invention in a hierarchical model. The classification hierarchy can be specified into different levels, such as section, class, subclass, main group, and subgroup. At the section level, patents are classified into eight (A-H)

categories, covering human necessities, operations and transport, chemistry and metallurgy, textiles, fixed constructions, mechanical engineering, physics, electricity, and emerging cross-sectional technologies (cooperativepatentclassification.org). Then patents can be further classified into more detailed and specified levels. A classification symbol usually consists of different components and is formatted as a combination of alphabetical letters, numbers, and other symbols, such as G06F17/30864. A patent may be assigned multiple classification categories as it may relate to different subject matters. In this study, patent classification is used to analyze the technological areas in which Google invests over time.

Patent Analysis

Patent analysis has been traditionally used in the field of business and technology management. Researchers use patent analysis to examine technology development in a given industry in order to guide R&D planning (e.g., Mogee, 1991; Kim and Bae, 2015; Kruppert, 2017). Patent analysis is also useful to explore tech companies' technical strategies. For example, DeVito (2017) uses public text artifacts, patent files among others, to identify factors that affect Facebook's news feed input. Google's patent files reveal detailed information about the key metrics that Google News uses for news retrieval and news ranking (Filloux, 2013). In the field of media and journalism studies, patent analysis has not been widely used. This study proposes an innovative method that uses patent analysis to explore Google's technological and algorithmic evolution. In particular, this chapter aims to serve two goals:

1). Identifying long-term trends about Google's investment in news-related technology innovations, i.e. in what technological areas that Google has increased or

decreased its investment over the years; what are the trends and counter trends across these years and their implications;

2). Understanding how Google's news-related algorithms work, e.g., what are the particular factors that shape Google's algorithmic decision about news, what does the workflow look like in Google's algorithmic systems, and what are the logics that determine the algorithmic design.

Method in This Chapter

This chapter combines quantitative and qualitative research methods as well as computational approaches. The author searched USPTO patent database by using "Google" as the assignee and "news" as the keyword. The search yielded over 3,000 patent files covering the filing date from 2003-2016, but a large portion of these patents do not particularly focus on news (for example, "the candidate content consisting of news, searches, Web pages, images, blogs, RSS feeds, audio files, videos files, and maps"). After de-duplication and based on relevance test across searching fields, the author identified 171 patent files as the key dataset considering these patents particularly address news. The 171 patent files were downloaded as html files and were parsed by Python Beautiful Soup.

To identify long-term trends and countertrends in Google's news-related technologies and algorithms, first, metadata about relevant information, such as given patent's title, filing date, and classification code, were extracted. Data were stored in a normalized relational database by SQLITE so that the author could use SQL to query the database to get basic statistics, such as the counts of a given classification category and the years it was assigned, etc. The data were output into an Excel spreadsheet and were

analyzed and visualized. These data also allowed the author to identify top classification categories by ranking them based on the proportion of the given category in relation to the total counts of all classification categories. Next, a sample of key patents that contain top classification categories and have "news" in title were identified for in-depth, qualitative content analysis. This step identified relevant elements in Google's algorithmic system. These elements were input in NVivo to analyze the major factors that shape Google's news-related algorithms and how Google's algorithmic systems work.

Google's News-related Technologies

This section uses classification information as an important indicator to examine Google's long-term technological trends in its news-related service. In patent files, classification information describes the nature and subject matter of the invention, which indicates technological areas that Google has invested over the years.

Identifying Top Classification Categories

A total of 306 classification categories were identified from the sampled patent files. Over the examined years, these classification categories were assigned to the sampled patents for 788 times at the subgroup level. The distribution of the classification categories—in terms of the frequency of the given category's assignment—shows a salient long-tail pattern (Figure 5-1). Considering a vast of classification categories were distributed on a very long tail, this study decides to focus on the top classification categories located at the head section as the focus of analysis. To identify top categories, this study decides the weight of each category based on the percentage of its assignments in relation to the total counts of all categories. Eleven top categories were identified, which account for less than 4% of the total classification categories but their weight is

over 32% in terms of the frequency of their assignments. Each of the identified top categories has a weight no lower than 1% (Table 5-1).

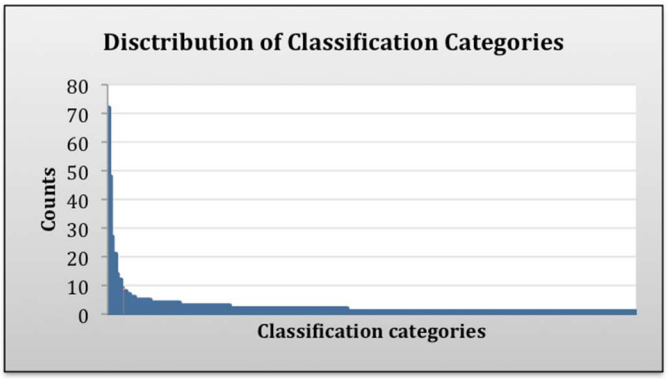

***Figure 5-1*. Distribution of Classification Categories. Source: author.**

Rank by %	Classification	Core Definition	Percentage
1	G06F17/30864	search engine querying	9.1%
2	G06F17/30867	personalization	6.1%
3	G06F17/3053	ranking	3.4%
4	G06F17/30011	document retrieval	2.7%
5	G06Q30/02	advertising, subscription, and commercial behavior	2.7%
6	Y10S707/99933	querying processing in database accessing	1.8%
7	Y10S707/99945	object-oriented database structure processing	1.5%
8	G06F17/3089	website content organization and management	1.5%
9	G06F17/30705	textual data clustering or classification	1.1%
10	G06Q50/01	social networking	1.1%
11	Y10S707/99937	sorting in database or file accessing	1.0%

***Table 5-1*. Top 11 Patent Classification Categories. Source: author.**

Top Classification Categories and Definition

To understand the definition of the classification as detailed as possible, this study analyzes patent classification deeply into the subgroup level. Based on the explanation provided by CPC (the Cooperative Patent Classification, a patent classification system) at different levels of the classification hierarchy, the definition of the top eleven classification categories was summarized as below.

At the main group level, the eleven categories belong to four main groups. G06F17 is the largest main group that contains six of the eleven top categories. Generally speaking, this group relates to electric digital computing or data processing systems or methods that serve "the performance of any automated operation using empirical data in electronic form for classifying, analyzing, monitoring, or carrying out calculations on the data to produce a result or event". The main group addresses information retrieval and database structures.

- 1-G06F17/30864: This subgroup has a specific focus on search engine related techniques, including querying, crawling, and pushing.
- 2-G06F17/30867: This subgroup has a specific focus on personalization.
- 3-G06F17/3053: This subgroup has a specific focus on ranking in the query process with adaptation to user needs.
- 4-G06F17/30011: This subgroup has a specific focus on document retrieval systems.
- 8-G06F17/3089: This subgroup has a specific focus on website content organization and management.
- 9-G06F17/30705: This subgroup has a specific focus on information clustering or classification in the case of unstructured textual data.

The second largest main group is Y10S707. The Y section is used for tagging new technological developments, usually those cross-sectional technologies. This group relates to data processing, with an emphasis on database and file management or data structures. Three categories fall into this main group:

- 6-Y10S707/99933: This subgroup focuses on query processing, or searching, in the process of database or file accessing.
- 7-Y10S707/99945: This subgroup has a focus on object-oriented database structure processing related to database schema or data structure.
- 11-Y10S707/99937: This subgroup has a focus on sorting in the process of database or file accessing.

The main group G06Q30 relates to data processing systems or processes specially adapted for commercial purposes, such as billing, commerce or marketing. One category belongs to this main group:

- 5-G06Q30/02: This subgroup covers data processing systems or processes specially adapted for marketing, such as online advertising, location-based advertising, profiling in connection with buying or selling of goods or services, including paid content, etc.

The main group G06Q50 relates to systems or methods specially adapted for specific business sectors with administrative, commercial, financial, managerial, supervisory or forecasting purposes, especially systems or methods that involve significant data processing operations, i.e. data processing operations that need to be carried out by a technological, e.g. computing, system or device. One classification category falls into this main group.

- 10-G06Q50/01: This subgroup has a specific focus on social networking.

The top 11 classification categories indicate some important technological areas that Google has invested over the years. It's pretty clear that these technologies areas have nothing to do with technologies traditional news media use in their news business,

but they define the technological specialization of Google, on which Google depends to distinguish itself from other social sectors. This analysis therefore provides an opportunity to explore Google's technological specialization that serves its differentiation process as the means. The following section will look at long-term trends and countertrends as well as individual classification categories that show important patterns that could help understand Google's news-related technological evolution.

Long-Term Trends and Countertrends

Figure 5-2 visualizes the top classification categories' long-term trajectories. On the X-axis are the years the identified categories were assigned; the Y-axis shows the number of times each category was assigned to the patents over these years (Figure 5-2). Overall, the chart shows three general peaks in terms of Google's investment in technology innovation in the top classification categories: one peak is around 2006; a small peak around 2009; and a significant surge around 2012.

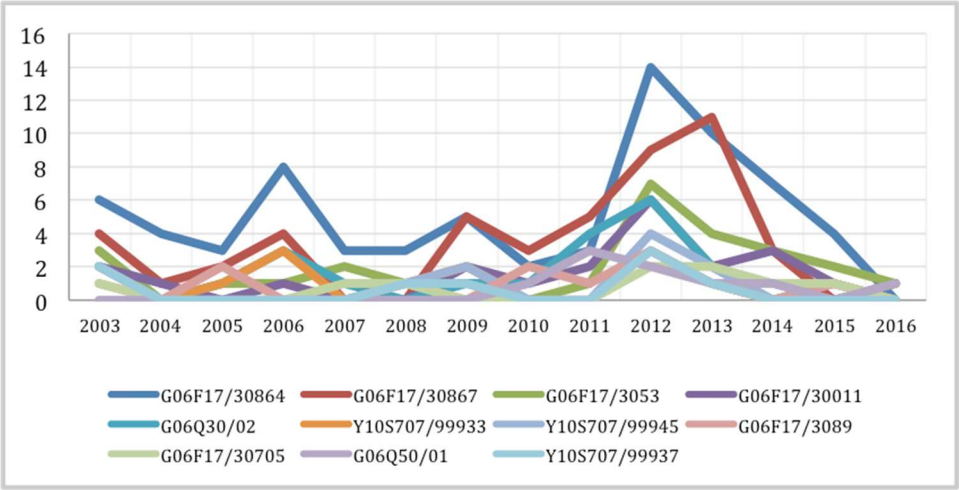

***Figure 5-2*. Long-term Trajectories of Top Classification Categories. Source: author.**

The analysis reveals almost all top classification categories, except G06Q5/01 (social networking), started in 2003. No data before 2003 were recorded, which indicates that news was not a top priority in Google's technology investment before then. After all, Google did not introduce any service or product that specifically addressed news before 2002, when Google News was launched in its Beta version. In 2006, Google News graduated from Beta. The official release of Google News in 2006 correlates with the overall peak of the top classification categories in 2006. In that year, all categories, except G06F17/3089, went up or stayed stable compared to the previous year. G06F17/3089 is a category about web content organization and management. As shown in Chapter 3, the content organization and management on Google News website did not change much since 2002 until the major redesign in 2010. When ranked by long-term growth rate, this category is also the most stable category among the top 11 classification categories in the long run (see Table 5-2).

There is a small peak around 2009. Most categories had a mild increase or stayed at the same level except, again, G06F17/3089, and another category G06F17/30705, which is about textual data clustering and classification. If look at this category over the long run, it was quiet during the years between 2009-2011 but has gained considerable growth in the long term (Table 5-2). Compared to all other categories, G06F17/30867 (personalization) has a sharpest increase in 2009 compared to the previous year. It remained strong in the following couple of years that it even went higher than the top 1 category G06F17/30864 (search). The trend of G06F17/30867 (personalization) around these years makes sense as "News for you", a feature about news personalization, was introduced as the "new heart" of the redesigned Google News in 2010 (see Chapter 3).

The overall increase around 2009 may have prepared for Google News' redesign in 2010, which also explains the rise of G06F17/3089 (content organization) in 2010.

Another significant peak happened in 2012, when almost all categories went up greatly, except G06Q50/01 (social networking), which had an earlier increase in 2011 but went down in 2012. This peak is much higher than other observed peaks, which may be related to the overall increase of the patents that Google owns since 2011. It also correlates with another important event on Google News history—the year 2012 was the ten-year anniversary of Google News. On its official blog, Google disclosed that as its news service became more internationally, its engineers had improved the technology overall, especially in areas such as freshness, grouping, ranking, personalization, and infrastructure (Bharat, 2012).

Personalization

A few categories stand out with unique characteristics when analyzing Google's long-term trends and countertrends in its news-related technologies. Among the top eleven classification categories, G06F17/30864 (search engine technologies) and G06F17/30867 (personalization) have been the leading categories throughout the years of the history of Google News, indicating that Google has been investing heavily in these areas. The category G06F17/30864 was ranked number one on the top list, accounting for over 9% of the total counts of the classification assignments. The dominance of G06F17/30864 indicates that search technology is the key technology applied in Google's news-related services. As shown in Chapter 2, in its early history, Google News was born out of Google search; and it was even introduced as "Google News Search" in its early days. Personalization was one of the categories that started early and stayed

strong. It was the second important category at its introduction in 2003. It has been paralleling the number one category (G06F17/30864) since then, and has even gone above it during 2009-2011 and again in 2013.

G06F17/30864 dropped sharply since 2012, while G06F17/308647, after dropping during 2013-2015, went up again since 2015. This study also calculated the growth rate over years for each top category. The long-term growth rate was then calculated by trendline. The top eleven classification categories were ranked again by weighted growth in Table 5-2. When the weighted growth over years was considered, personalization-related technology (G06F17/30867) becomes the top one category with a growth rate that is much higher than that of search engine technologies (G06F17/30864). These trends could mean that after years of development, Google's search engine related technologies are getting stable. Their innovation and investment are slowing down in recent years, compared to personalization as a newer, rising area. These trends indeed show Google's heavy investment on personalization technology all along. Google's most recent introduction of its AI-driven news service in 2018 continues this trend. Such trend is also reflected in Figure 5-3, which shows a rising trend of the personalization category after 2015 (Figure 5-3).

Rank by Weighted Growth	Rank by %	Classification	Core Definition	Weighted Growth Over Years
1	2	G06F17/30867	personalization	0.85%
2	1	G06F17/30864	search engine querying	0.55%
3	3	G06F17/3053	ranking	0.48%
4	4	G06F17/30011	document retrieval	0.27%
5	10	G06Q50/01	social networking	0.22%
6	9	G06F17/30705	textual data clustering or classification	0.11%
7	7	Y10S707/99945	object-oriented database structure processing	0.06%
8	8	G06F17/3089	website content organization and management	0.02%
9	11	Y10S707/99937	sorting in database or file accessing	-0.03%
10	5	G06Q30/02	advertising, subscription, and commercial behavior	-0.05%
11	6	Y10S707/99933	querying processing in database accessing	-0.09%

***Table 5-2*. Top 11 Classification Categories and Weighted Growth Rate Over Years. Source: author.**

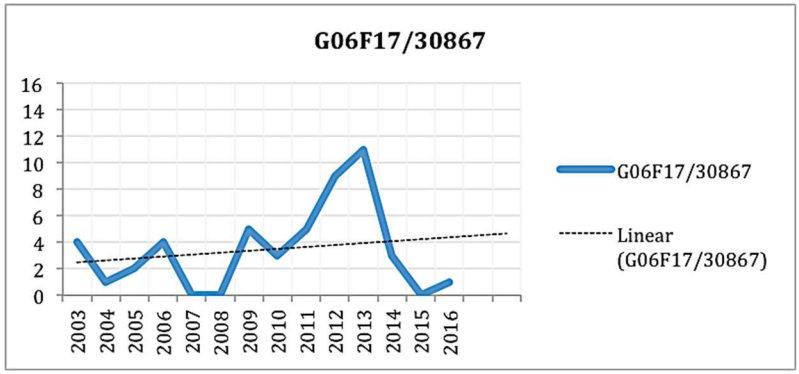

***Figure 5-3*. Trajectory and Trendline of G06F17/30867-Personalization. Source: author.**

Social networking

Among the top classification categories, G06Q50/01 (social networking) is the only one that had a very late start. In the sample patents, this category did not appear in the Google's news-related patents until 2010, a time that witnessed Google's several attempts and adjustments on social networking in response to the rise of social media platforms such as Facebook and Twitter. These attempts included Google Buzz introduced in 2010 and shut down in 2011, and Google Wave launched in 2009 and

ended in 2010, and more. In 2011, Google's focus on social media shifted to Google+, a social network that is owned and operated by Google for online sharing and networking. On Google News, social networking-related features were first introduced on its homepage in 2010 (See Chapter 3 for detailed discussion). Google's attempts on social networking is believed not as successful as its efforts in other areas (Bastone, 2018), as shown in this study that G06Q50/01 (social networking) only accounts for a little bit more than 1% of all patent classifications in the sampled patent files, and was ranked the 10th in the identified top 11 classification categories.

On the other hand, among the top 11 categories, personalization (G06F17/30867) and social networking (G06Q50/01) are the only two areas that show a rising trend after 2015 (Figure 5-3 and Figure 5-4). Overall, social networking-related technology has a high growth rate over the years, which increases the rank of the category from number 10 (by percentage) to number five (by growth, see Table 5-2). In this regard, social networking seems to be a potential area in the case of Google and its news-related technologies. It's a critical moment for Google to decide whether it should increase or decrease its investment on social networking technology and its application in news. With the unsolved fake news problem on social media, innovation in this area could be both a hot demand and a hot potato.

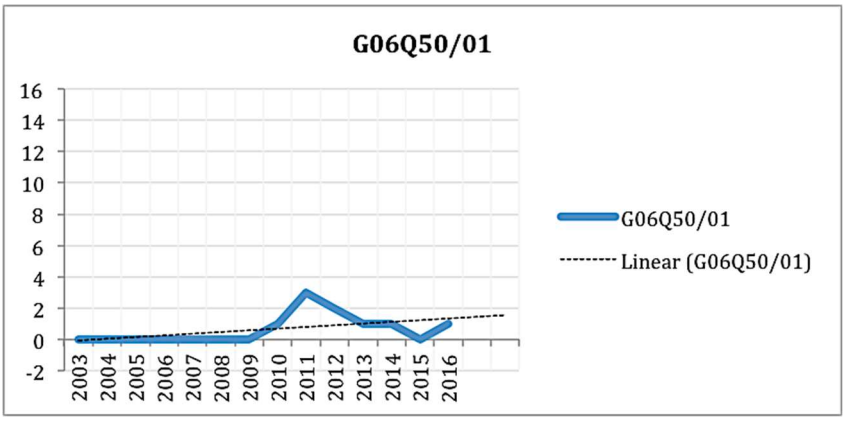

Figure 5-4. Trajectory and Trendline of G06Q50/01-Social Networking. Source: author.

Other important categories

When the identified top 11 categories were considered by percentage (Table 5-1) and the long-term growth (Table 5-2) together, the category of ranking (G06F17/3053) and document retrieval (G06F17/30011) are also strong areas, ranked top three and top four in both ranking criteria (by percentage and by long-term growth). As shown in previous chapters, ranking is historically a core technology for Google search engine and Google News, which provides users the information ranked by certain orders, such as relevance, importance, freshness, etc. In the early history of Google News, this technology differentiated Google from other web directories, and later it remained the critical technology supporting Google's algorithms, e.g., PageRank. The overall importance and high growth rate of document retrieval (G06F17/30011) is also understandable considering news is traditionally managed as text and innovation in texts is relatively cheaper than other forms of content, such as image and audio-video. Comparatively, the category G06F17/3089 for website content organization and management is the most stable category, ranked top 8 by both ranking criteria and the extent of fluctuation over years is close to zero. As the analysis in Chapter 3 shows,

content on Google News website was mostly organized by news sections throughout its history. Significant resign of the website homepage only occurs every 7-8 years averagely. While minor changes were more frequent, these changes did not primarily affect the basic structure in terms of content organization and management.

G06Q30/02 (advertising, subscription) and Y10S707/99933 (querying processing in database accessing) are the only categories that have significant drop over time. Particularly, the rank of the former dropped from number 5 (by percentage) to number 10 (by growth) and the latter dropped from number 6 (by percentage) to number 11 (by growth). The drop of G06Q30/02 is perhaps the most surprising finding here. This classification category contains technologies about commercial activities, such as online advertising, location-based advertising, and using user profile data to match buying and selling demand, etc. If these findings were read in connection with the findings of Chapter 4 about the international disputes, one would be surprised to find that while Google's business model largely driven by its advertising technologies was one of the key concerns of the traditional media industry, Google's investment in this technological area dropped significantly over the past fifteen years. This category also addresses users' commercial behavior, such as subscription to paid content behind the paywall. While paid content also concerns some news organizations in terms of their business model in the digital age, Google News' investment in this area is dropping. On the other hand, personalization as the most important, and still rising, area seems not to get enough attention from the traditional media industry, at least not reflected in the international disputes examined in Chapter 4. It would be expected that the tensions and negotiations between digital platforms and traditional media would soon be shift from older

generation of technology such as advertising to newer generation of technology such as personalization.

Google's News-related Algorithms

To identify the key determining factors that shape Google's algorithmic design and to understand how Google's news-related algorithms work, this study identified sixteen key patent files to conduct in-depth textual analysis. These files also meet two other conditions: 1). They all contain identified top classification categories. And 2). They all have the keyword "news" in title to make sure the high level of relevance to this analysis. I then used NVivo to analyze the key patent files for themes and patterns. The sixteen patent files are:

- Systems and methods for improving the ranking of news articles
- Personalizing aggregated news content
- Generating a news timeline
- Identification and ranking of news stories of interest
- Translated news
- Systems and methods for monetizing subscription and archival news content
- News topic-interest-based recommendations twiddling
- Systems and methods for personalizing aggregated news content
- Systems and methods for providing news alerts
- Discovery of news-related content
- Methods and apparatus for clustering news online content based on content freshness and quality of content source
- Image selection for news search

- Relevance determination and summary generation for news objects
- Labeling events in historic news
- Systems and methods for implementing a news round table
- Creating a customized news collection based on social networking information

Key Determining Factors

Google News' algorithms are multi-layered, interlinked systems that connect input, throughput, and output (Latzer, Hollnbucher, & Saurwein, 2016). While the throughput part is in the black box, the output and input ends can be observed. The homepage changes observed from Google News website reveal rich information about the output, this section, based on the in-depth textual analysis of the 16 key patent files, looks into the input side and explores the throughput. The rationale is when findings about output and input are put together, it discloses clues that are useful to understand the logics, determining factors, and decision-making behind Google algorithms.

In Google's algorithmic systems, the input involves the process of datafication, in which "many aspects of the world" are rendered into data, mostly in quantified manner (Cukier & Mayer-Schoenberger, 2013, p. 29). In the case of Google News, the input relates to news-related information and the process that this information is datafied. The input data come from four major data sources: 1). news source-related information; 2). news content-related information; 3). back-end machine language; and 4). user information.

News source

Source quality is an important weighting factor in Google News' algorithm systems. Therefore, news source-related information is collected and processed. This patent study

finds that source quality plays a key role in determining news stories' ranking in Google's algorithmic systems. Google defines over a dozen of metrics to rank news (Curtiss, Bharat, and Schmitt, 2009). These metrics primarily address news source quality, including:

- the number of articles produced by the news source during a given time period;
- the average length of the articles produced by the news source;
- the importance of coverage by the news source, e.g. the capability that the news source produces news stories on important topics, or news stories that have wide influence;
- breaking news score, i.e. how quickly the news source can respond to breaking news, e.g. publishing a story soon after an important event has occurred;
- usage pattern, such as the news site's click rate and popularity;
- news source reputation, or the human opinion of the news source, such as how users, public, and other professional agencies evaluate the given news source, e.g. the number of Pulitzer prizes the newspapers have won;
- circulation statistics of the news source;
- staff size of the news source;
- the number of news bureaus associated with the news source;
- originality, i.e. the capability that the news source produced original content, e.g. the number of original named entities the news source produces within a news cluster;
- breadth, the range of topics the news source produces;

- international influence of the news source, e.g. international online traffic of the news source's website;
- writing style, e.g. spelling, grammar, reading levels;

Additional factors include the age of the news source, the quality of the news source's hub pages, and the genre of the news source, such as commentary, interactive, blog, modifiable, or amateur (Bharat, Stoll, and Mayer, 2017).

A group of these metrics focus on news source's organizational characteristics, such as the circulation, staff size, and the number of news bureaus; another group of metrics consider the professionalism of the news source, including the productivity of the news source, such as the amount and length of news articles being produced; the quality of the news articles that the news source produces, such as the originality, breadth, and writing style; other professional capabilities, such as capability of covering breaking news and winning professional prizes; and digital capabilities, such as the news source's online presence and online influence.

Many of these metrics reflect existing journalistic norms and practices—from something as small as the grammatical quality to something as big as news organizations' reputation. But Google only selectively adopted journalistic ideas into its algorithms given that only those things that can be *algorithmically* processed are selected as the ranking metrics. For example, breaking news score is considered because this factor can be quantified by calculating the length of time between when the event occurred and when the related news was published. Similarly, originality and breadth are selected because these factors can be determined by algorithmically identifying the duplication and calculating the story size depending on named entities. However, objectivity as an

important journalistic value is not selected as a ranking metric because it cannot be datafized. This process of datafication transformed certain aspects of journalism into quantified variables and reduced news into algorithmically recognizable raw data.

News content-related information

In addition to news source-related information, data about news content are also input and processed in Google News algorithmic systems. This content-related information covers news texts, images, audios, and videos. Once collected, features are identified through various pattern recognition techniques that could analyze different types of news content, such as character recognition, facial recognition, and voice recognition. Features mainly concern three types of information: temporal, geographic, and topical. Temporal information includes when the news event happened and when the related news article is published. Geographic information relates to places and locations involved in the news story. Topical information is extracted from news articles including information about the event, people, and issues covered by the news report. Clearly, these features cover at least four of the Ws in the five Ws that are basics to news: the when, where, who, and what, but the "why", once again, is hard to be processed algorithmically.

In Google's algorithmic systems, news articles are processed individually and in aggregate. For example, related news articles are clustered by parsing each piece's headline, determining the collection of words that are most representative of the news document, and calculating the amount and identifying the position of the keywords in the article. In addition, the centroid of the cluster is calculated using mathematic techniques so that each article's relevance score can be computed in relation to the whole cluster. For example, the centroid is computed by averaging individual term vectors—a weighted set

of terms—from the documents contained by the cluster (Bharat, Curtiss, and Schmitt, 2016). Algorithms could also calculate how many times the words in the news headline appear in the body of the documents in the cluster. These approaches highly datafize news content, converting qualitative content into fragmented, quantified data that can be processed by machine. This is the machine way, or "Google way" to handle news. "Once we datafy things," Cukier and Mayer-Schoenberger warned, "we can transform their purpose and turn the information into new forms of value." (2013, p. 35)

Back-end machine language

Google News algorithms also largely extract information from back-end machine language, such as HTML codes and metadata, which can be easily recognized and processed by computational algorithms. Temporal, geographic, and topical information can be extracted from timestamp, HTML codes, and metadata. Other key elements can also be identified automatically, such as headline, keywords, and byline, as well as the news website's layout, the position of the given news article on the webpage, and the characteristics of hyperlink, images, and more.

To respond to such techniques, some newsrooms have adopted search engine optimization (SEO) systems that aim to increase their content's visibility and ranking in search engines' algorithm systems (Giomelakis & Veglis, 2015; Dick, 2011). SEO practices include strategically using keywords in the main body of the content, URL, and kickers, building inter-site links, having different versions of headline for the same news story, and more (Giomelakis & Veglis, 2015). These newsrooms have either hired outside SEO specialists, or set up in-house SEO teams, or used both to promote their online presence (Dick, 2011). By adopting SEO techniques, the meaning of news production

expanded since newsrooms have to pay attention to the production of both front-end content and back-end information.

Meanwhile, concerns arose in other newsrooms that considered SEOs to be profit-driven and harm journalistic autonomy. Practically, they were also concerned that SEO could make news dry and cookie cutter-like. Take headline writing for example, "Novak Djokovic beats Roger Federer in epic Wimbledon 2014 men's final" is considered SEO-friendly while "Djoko did it again!" is non-SEO friendly (Giomelakis & Veglis, 2015, p. 26). Under the SEO rules, "Gotcha," the famous headline published in London's tabloid newspaper *The Sun* in 1982, referring an Argentine ship, the General Belgrano, being sunk by the British Navy during the Falklands war, has little chance to rank high in search engines algorithms (Richmond, 2008). Defenders of SEO argue that SEO rules actually support the traditional journalistic formula, such as the five Ws and the inverted pyramid structure, that gets news "straight to the point" (p. 54). After all, argued Richmond, journalists are writing for being read not for showing audience how "clever" they are. This debate over SEOs reminds us *Agence France Presse v. Google Inc.*, in which the two parties had stark different understanding about the nature of journalistic work (See Chapter 4). As shown in this section, how computational algorithms read news determines how Google saw news as uncopyrightable facts, opposite from AFP's perception of news as creative, copyrightable work.

User information

In Google News algorithm systems, another very important strand of information is user information. User data are collected from a range of sources, including users' personal information associated with their Google account, such as name, email address,

and demographic information; users' search history, including search queries, selections, preferences, and click data; users' social network data, including their contacts and the interactions among contacts, and their behavior on social media such as posting, sharing, endorsing, and commenting; and users' information stored by cookies, e.g. device information, time spent on a certain item, IP address, and more.

User data are collected and processed at individual and aggregate levels. For example, one user may be considered with users who use same device, have same location, and speak same language. News relevance may be decided based on an individual user's interest as well as the broader communities' interest. User data are also considered in a dynamic way. For example, a user's current interest is compared to the user's past interest to find long-term interest patterns, which is used to predict the user's future interest. In the algorithm systems, users' old interest may be demoted in ranking while users' current interest may have a higher ranking score, data associated to users' old interest are not deleted from the systems all together.

How Google obtains data input from back-end machine languages and user information is very different from journalistic practices in the news industry. While the former is technical, the latter reflects different relationship with the audience. In newsrooms, the news media are agenda setters deciding what's newsworthy and what's something that the public should know. Even in the digital environment, this kind of top-down media-audience relationship dose not change much. In Google's algorithmic systems, users' requests are often the starting point of algorithmic processing. While the Google way commercializes users' needs, it also attempts to understand users as detailed as it can. Such understanding is achieved through the process of datafication, in which

users' needs are converted into data that could be processed by algorithm. News media, on the other hand, in the tradition of journalism, understand their audience by talking to them. In the digital era, how to understand audience is another dimension of the competition between digital platforms and news media. While digital platforms try to understand users through statistics and deep contextualization of user data, news media need to strengthen their human tradition—going into local communities and making people connections.

Google's use of user data is also constrained by legal and ethical concerns about privacy and data protection. In May 2018 the General Data Protection Regulation (GDPR) went into effect in the EU. The GDPR imposes obligation on data controllers and processors who "process personal data on behalf of another body" (General Data Protection Regulation, 2017). The GDPR defines data protection as people's fundamental right and urges data organizations to commit to data protection by design and by default. The regulation stipulates detailed principles relating to the process of personal data, including data controllers' obligations and data subject's rights. Under the GDPR, data companies should inform the data subjects about the nature, scope, context, risk, and purpose of the data processing in a clear and transparent way. It also grants data subjects the right to consent, access, rectify, and erase their data (GDPR, 2017). In January 2019, the National Commission on Informatics and Liberty (CNIL, Commission nationale de l'informatique et des libertés), France's data protection authority, issued a $57 million fine to Google for GDPR violation (Price, 2019). Google intended to appeal the fine (Cerulus, 2019). In the U.S. Google announced in May 2019 that the company rolled out

auto-delete controls that could automatically delete users' location history and activity data on an ongoing basis (Monsees, 2019).

Workflow

Google News algorithms are a complex network of systems. In Google's patent files, these systems consist of a wide range of components, modules, or units (see Figure 5-5 for an example). The general workflow is a structured process that starts from data collection, which can be done through automatic crawling or the submission of sources. Once different types of data, e.g., news source-related information, news content-related information, back-end machine language, and user data, are collected, features will be identified and labeled. Features are predetermined to define characteristics of different data type. Collected data then will be classified into categories, such as source, genre, and topic. These categories can be refined into levels of sub-categories as needed. Classified data are stored in individual servers, repositories, or databases. Data can be searched, queried, and retrieved within the server/repository/database where they are stored and processed across different data storages and systems (see Figure 5-6 for an example). Data categories are usually scored based on different models and criteria. A wide variety of scoring systems are responsible for different scoring purposes, such as freshness score, relevance score, interest score, and source score. The given score may be an absolute score or a normalized one to quantify pre-defined features. Scored items and their scores are stored respectively. Scored categories are also networked as multiple scores are often taken into consideration when ranking decision needs to be made. Based on a query, candidate items are determined by the combination of relevant scores through appropriate

formulas. A final or overall score is calculated and based on which candidate items will be ranked and presented to users by the ranking order (see Figure 5-7 for workflow).

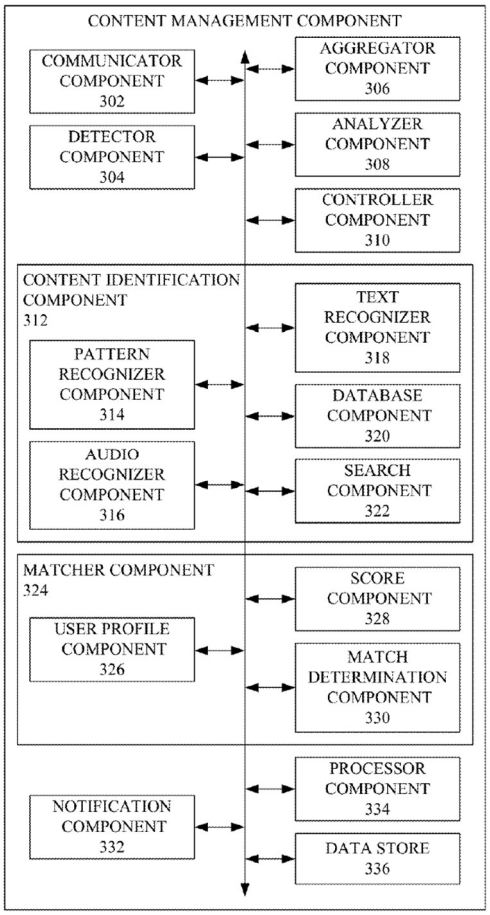

***Figure 5-5*. Example of Components. Source: Varadarajan et al., 2016.**

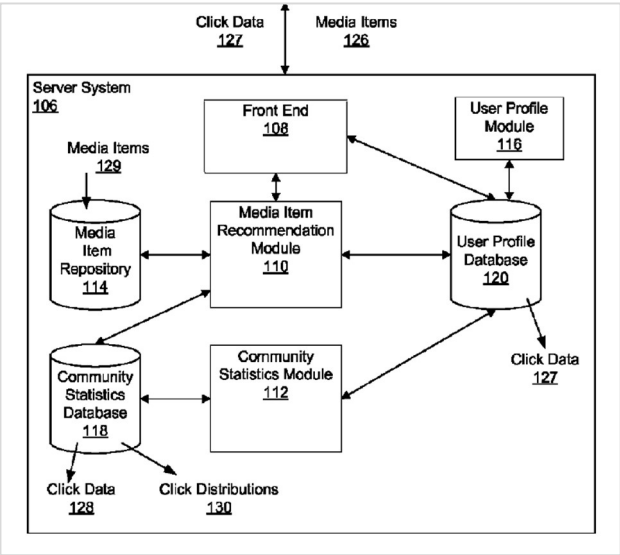

***Figure 5-6*. Example of the Network Between Modules and Data Storages. Source: Dolan and Liu, 2012.**

The actual workflow is much more complex and nonlinear. Processes described above often overlap. Take classification for example, while labeled data are classified and stored respectively, algorithmic decisions are made across different types of classification boundaries, in terms of both software and hardware, to make sense of the relationships between features (Figure 5-6). Each media item can be associated with multiple categories depending on topic, location, media type, source, and more. The multiple taxonomies allow a news item to have relationships with various factors. Like a huge, complex machine, the interconnected classification systems form a network, in which systems and data storages that support the algorithmic processes are pipelined in a relational way.

Algorithmic Logics: Two Types of Sense-making

Sense-making—the machine way

When Google News was first introduced to the public in 2002, the slogan said, "Get your news the Google way". So, what is the "Google way"? In a searching environment, when a user inputs a search query with a couple of discreet features, i.e. search keywords such as "Olympics", "drug scandal", "CNN", what the user actually does is to define some relationships, for example, a topic (drug scandal) in relation to an event (Olympics) and in relation to a news source (CNN). Based on the user's request Google needs to make judgment what relationships the user is looking for, how to locate these relationships in its databases, and how to present these relationships.

In the environment of online news consumption, however, users' actions are hard to interpret than those in searching. Relationship sense-making is often not as straightforward. For example, when a user clicks on a news story, what do these actions tell about what relationships the user is actually looking? Was the user interested in the topic, source, genre, or something else? To make this judgment, Google algorithms have to contextualize information from different, and as many as possible, data sources. By contextualizing, it means data cannot be treated in a static way but a relational way. Algorithms have to make decision by going beyond the time and place when/where the data are collected or when/where the behavior behind the data occurred. One individual piece of information has to be considered in relation to other individual pieces of information. For instance, data collected in one crawl may be compared to those in the subsequent crawl; user's current interest may be compared to the user's past interest; an individual user's information may be considered in relation to the broader communities that this user has connect with, etc. For each small task, there could be a huge web of relationships behind it that algorithms need to make sense through contextualization. As

tasks become more and more complex in the digital environment, rule-based algorithms show limitations. This kind of contextualized relationship sense-making now depends on AI and machine learning technologies that Google News focuses on recently. Using algorithms to make sense of data and relationships of data is a machine way of sense making. This is a process of datafication, however complex the process is. Throughout the algorithm systems, the machine way always parallels with the human way of sense making.

Sense-making—the human way

Patent analysis reveals that Google's algorithmic processes are mostly automatic, but human factor plays an important role at various points. First, many criteria that are applied in Google's algorithmic decision-making processes, whether in clustering, scoring, or ranking, are defined by human. For example, the aforementioned metrics that are used to rank news articles are defined by human. In the workflow described above, many decisions are made based on pre-defined thresholds or hyper-parameters. In one example, relevance threshold may be predetermined by an administrator. Humans are also involved into reviewing, labeling, and evaluating processes, e.g. human evaluators are needed to review news sources, articles, and webpages to evaluate source and document quality. Human users can also read news articles and write brief description as news summary. People involved in these human roles include evaluators, experts, administrators, employee-users, and programmers. In addition, general users and Google News users are sometimes polled or recruited for user studies to help make algorithmic decisions. These human decisions are integrated into Google's algorithm design.

There is another layer of human intervention that is introduced on the top of the computational algorithms. As shown in Chapter 3, the introduction of many features on Google News website is not algorithm-driven but decided by human. In the case of Local, Spotlight (long-form, in-depth feature stories), Fact check, and some special sections, news content displayed in these sections may have little chance to stand out in Google's algorithmic systems, based on the ranking criteria described earlier. It was human decisions that pushed such content to be presented on Google News homepage. Many of these human decisions were made to respond to the pressure coming from the news industry as shown in Chapter 3. These decisions made by human have been superimposed on algorithmic systems.

In early days, Google largely depended on rule-based algorithms that arbitrary rules were defined by human experts. In 2018 when the "new" Google News was reintroduced to the public, it was described as an AI-powered service. Algorithms today involve more machine-learning technologies, in which the system is fed with a great amount of raw data (Figure 5-7). The machine can learn from the training data through supervised machine learning, or process un-labeled data and identify patterns, associated rules, and relationships on its own through unsupervised machine learning. In the more advanced reinforcement learning model, algorithms are even more powerful that they can process data and make decisions in real time. As technology enters the artificial intelligence era, rule-based algorithms may phase out into AI algorithms, but the logics and decision-making that ruled the pre-AI time will still be at play. The new Google News, said Google, is "AI meets human intelligence". Indeed, the "Google way" has always been an integration of the machine way and the human way.

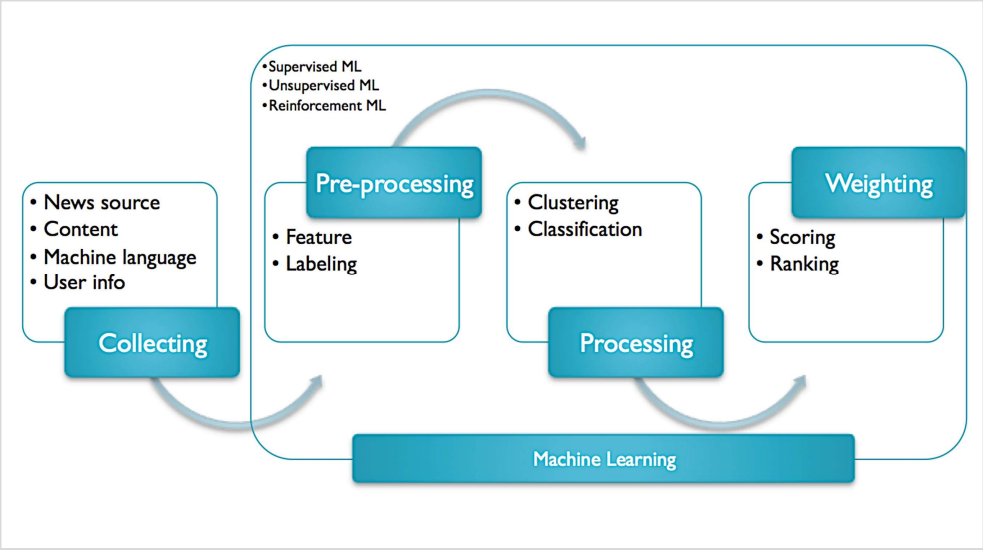

***Figure 5-7*. Workflow in Google's Algorithmic Systems. Source: author.**

Discussions

This chapter provides patent analysis that aims to understand Google's news-related technologies and algorithms. As shown in previous chapters, specialization in technology largely defines Google. In the process of Google's differentiation, technological specialization has also become the means to serve the end of autonomy. Patent analysis in this chapter offers an opportunity to systematically study Google's specialization in technology.

The top patent classification categories identified in this chapter reveal key technological areas that Google has invested. These technologies clearly differentiate Google from news media. However, Google normalizes certain journalistic ideas and resources and integrates them into its algorithmic routines, for example, its ranking algorithm heavily depends on news source's organizational characteristics and professionalism-related metrics; different types of news content also serves as one of the main data sources in Google's algorithmic systems. Such adoption is very selective and

generates new norms, that is, Google only adopts journalistic ideas that could be algorithmically processed to serve a process of datafication. During this process, news is reduced into digital data. As a result, the value of news that cannot be datafized is lost, new norms are encouraged, such as SEO rules adopted in some newsrooms.

How Google and news media understand audience is another dimension that differentiates them from one another. As Google evolves into the AI era, Google could have more precise yet commercialized understanding of users through AI technology-supported contextualization of user data. While Google deepens its differentiation in this area, its specialization in user understanding is constrained by legal and ethical limitations. News media should take a different path by strengthening journalism's tradition in humanity to make human connection with audience and communities that they serve. The differentiation in audience understanding and relationship, the author believes, is the real battlefield for digital platforms and news media as two social institutions to claim their legitimacy and autonomy.

Chapter 6. A Growing Institutional Power and N-D-N—A New Theoretical Model

At the end of the second decade of the 21st century, Google's ambition in news remained strong. Beyond Google News, Google had introduced several projects that aimed to support its news business, including the Google News Initiative (GNI). Simultaneously, regulatory pressures kept growing—especially in Europe—in ways that resurfaced unsolved issues left from early disputes around Google News. Is it possible to discern the future from the past, understanding how these updates may affect the future of Google's news business and the future of the news industry? This final chapter first provides an overview of GNI and then proposes an N-D-N model to conclude the dissertation.

Google News Initiative

As of this writing, GNI was an umbrella initiative of Google that integrates earlier news-related projects – such as Google News, Google Digital News Initiative, and Google News Lab – with newer initiatives. In March 2018, Google announced that GNI would invest $300 million over three years aiming at "building a stronger future for news" (Schindler, 2018; Wang, 2018). Compared to earlier, discrete news-related efforts, the launch of GNI signified that Google had brought forth a more systematic and strategic plan for the future development of its news business.

To understand GNI's influence on the future of the news industry, this section provides an overview of GNI in three parts. The first part examines GNI's website and its manifestos—the discourses that technology companies use on their websites to outline "the goals and processes" (Carlson & Usher, 2016, p. 567) and to "clearly explain their unique contributions to an existing industry" (p. 565)—to understand how GNI works,

how it introduces itself to the public, and the implications of GNI's actions, objectives, and interpretations. Part 2 examines the 113 blog posts published on GNI's official blog site as of January 31, 2019. These blog posts provided additional information not included on the GNI website, focusing especially how Google has changed its news focus since 2016. The third part looks at Google's Digital News Innovation Fund (DNI Fund), an important part of GNI that has a longer history than GNI itself. DNI Fund is a Europe-focused program that provides funding for selected European news organizations and startups that proposed projects regarding digital news innovation. This section of the chapter explores the areas Gogle has funded in the last few years of the 2010s and their implications for the news industry. Together, this section explores how Google attempts to shape the future of the news industry through a series of initiatives associated with GNI.

An Overview of GNI

In early 2020, the GNI website displayed a strong rhetoric of collaboration, using language such as "Gone are the days when news organizations – or tech companies – can 'go it alone'" and "We believe in spreading knowledge to make life better for everyone. It's at the heart of Google's mission. It's the mission of publishers and journalists. Put simply, our futures are tied together" (About, n.d.) Indeed, observers saw GNI as Google's effort to "sweeten" its relationship with the news industry (Wang, 2018), which many perceived to have been hurt due to the disputes between news media and Google News' aggregation.

The collaborative relationship was developed through a vast partnership network at institutional level that brought together a wide range of stakeholders in the news ecosystem, including:

- News media in the United States and beyond, such as The Washington Post, The New York Times, The Guardian, Le Monde, South China Morning Post
- Digital native media that have news-related services such as Slate, Reddit, Gizmodo
- Non-profit organizations that have their own associated networks, such as the Local Media Consortium, which represents more than 70 local media companies in the 50 U.S. states and Puerto Rico; the Association of Magazine Media with 150 domestic, international, and members, representing more than 500 magazine media brands; the European Journalism Center, which has hundreds of media partners across Europe; the World Association of Newspaper and News Publishers, a global network of 3,000 news publishing and technology companies and 80 member publisher associations representing 18,000 publications in 120 countries
- Research organizations, such as the Reuters Institute for the Study of Journalism at Oxford University
- Membership organizations, such as American Society of News Editors, founded as professional group for top newspaper editors but expanded to all journalism executives in 2009 when the organization removed “Newspaper” from its name and substituted “News” (About us, n.d.). GNI’s partnership with these

organizations took various forms, including financial and technological support, research collaboration, education, and membership.

Another important approach that GNI used to expand its partnership network was to offer various programs in the form of training workshops, fellowships, and awards, such as Society of Professional Journalists, GNI Fellowship, etc. GNI also developed a university network of over 200 universities around the world and worked with journalism schools of participant universities to provide journalism educators with training. As of 2019, these training, fellowship, and university networks covered America, Europe, Asia, Africa, and Australia. GNI had annual application and sign-up systems to recruit participants; it also offers online courses and recommendations based on participants' jobs, skill levels, and available time (see Training, n.d.).

As part of its network building, GNI normalized the news industry's economic concerns and provided technology-driven solutions. GNI classified "the needs of news organizations" and "industry challenges" into four categories: distribution and audience engagement, revenue growth, digital business transformation through data and infrastructure management, and new ways of storytelling. For each category, GNI offered a wide range of Google products as solutions, from YouTube and Accelerated Mobile Pages (AMP), to Google Trends, Google Earth, and more.

GNI offered training and courses revolving around Google's tools and products as well. On GNI's website, signed-in users could find as many as 25 Google tools and lessons corresponding to Google products. Case studies about how partner organizations applied Google products in their practices were mapped onto the four types of industry challenges as examples of best practice to solve these challenges, including storytelling,

monetization, distribution & engagement, and data and infrastructure (see Case Studies, n.d.). By defining industry challenges facing news media as techno-economic problems and offers Google's own technological solutions. By doing so, GNI builds up a media-tech network in the name of "help[ing] journalism thrive in the digital age." (Case studies, n.d.)

GNI Blog

Reflections on GNI were also recorded on the initiative's official blog, which had 113 blog posts as of January 31, 2019. An examination of these blog posts revealed similar observations made on GNI website. In addition, it also demonstrates the change of focus in GNI's short history. The largest number of the 113 posts (43%) were about Google technologies and services, concentrating on Google's artificial intelligence (AI) technology, Google Trends and data analysis-related tools, Google's advertising and subscription services, VR/AR technologies, and Google's Cloud service. The second-largest category (31%) of posts covered GNI's initiatives, including fellowships, awards, and training. Posts about the DNI Fund made up the largest sub-category in this group (29% of the category). Other posts in this category covered Google's influence in Asia (11% of this category) and Africa (11% of this category). The third-largest category of blog posts focused on the applications of Google technologies and tools. In this category, 8% of the blog posts addressed the applications of these Google technologies in fact checking, 6% were about how Google technologies were applied in covering elections in different parts of the world, and 5% of the blog posts focused on local news.

Considered by year, the blog posts could be divided into three periods: 2016 and earlier, 2017, and 2018 and after. Comparison of the posts in these three periods reveals

changes in Google's focus on news-related business. In terms of the Google technologies, in 2016 and earlier, there were blog posts about AMP (Accelerated Mobile Pages Project) and Google Trends. In 2017 more attention was paid to AI technologies, such as machine learning tools and smart assistant. In 2018 and 2019, GNI promoted more diverse technologies on its official blog, covering multimedia, subscription, cloud, analytics, mobile, and video. A focus on fact checking, local news, and elections appeared after 2016. Fact checking grew from less than 4% in 2016 or earlier to 11% in 2017. It fell to 8% in 2018 but remained a trending area on GNI blog. The local category grew from less than 4% in and before 2016 to over 4% in 2017 and to about 8% in 2018. Election-related posts grew from less than 4% in and before 2016 to more than 4% in 2017 and to 10% in 2018. The growth of this category indicated Google's growing interest in promoting its technological influence in the political realm.

DNI Fund

In the wake of the disputes between European news publishers and Google News, in 2015, Google launched the Digital News Initiative, a program that aimed to "support high-quality journalism in Europe through technology and innovation." (Verney, A. 2015, Editor's note) This initiative was also a gesture to improve Google's relationship with the European news industry, according to Carlo D'Asaro Biondo, Google's president of strategic partnerships, Europe. When he announced the launch of the DNI in 2015 in London, "I firmly believe that Google has always wanted to be a friend and partner to the news industry, but I also accept we've made some mistakes along the way" (Verney, 2015, para. 14).

The DNI Fund, which was part of this initiative, invested 150 million Euros from 2016-2019 to fund projects for digital news innovation in Europe, categorized into three types: large, medium, and prototype. The DNI Fund, a competition-based program, was open to European publishers of all sizes, with the winning projects are selected by the Google Project Team and the DNI Fund Council, which consisted of representatives from Google, the European news industry, and academia. From 2016 to 2019, the DNI Fund had six rounds of competitions that funded 662 projects covering more than 30 European countries with funding of more than 140 million euros (Table 6-1).

Round	Time	No. of projects	Amount of funding (€)	No. of countries
1	Feb-16	128	27M	23
2	Nov-16	124	24M	25
3	Jul-17	107	22M	27
4	Dec-17	102	21M	29
5	Jul-18	98	21M	28
6	Mar-19	102	25M	23

***Table 6-1*. DNI Fund Projects Round 1-6. Source: Google DNI, compiled by author.**

In 2016, the first year of the DNI Fund, which covered Round 1 and Round 2, the top ten topics addressed in the winning projects were multimedia (8.3% of the winning projects), analytics and research (6.3%), payment models (6.3%), data journalism (5.9%), AI technologies (5.5%), visualization (5.5%), user-generated content (5.5%), niche editorial products (5.1%), investigative journalism (4.7%), and personalization (4.3%) (Google Digital News Initiative, 2017). Compared with the first round, Round 2 projects showed evident growth in two categories: the intelligence category and the distribution and circulation category. The former grew from less than 12% in Round 1 to over 20% in Round 2, and the latter rose from 4% in Round 1 to 9% in Round 2. The biggest drop

came from the social and community category. Projects in this category decreased from more than 20% in Round 1 to 9% in Round 2.

The year 2017 covered Rounds 3 and 4 of the DNI Fund. In Round 3, the focus on artificial intelligence continued to increase, with about 23% more applications in this area than Round 2. Fact checking, which was not among the top ten topics in Rounds 1 and 2, stood out in Round 3, making up 29% more of the applications. Additionally, investigative journalism, one of the top ten topics in the previous year, grew in Round 3 (up 20%). Immersive technologies, such as virtual reality and augmented reality, increased largely in this round. (See Blecher, 2017.) Round 4 saw a continued growth in AI technologies. What was new about this round was the rise of the application of AI technologies in the exploration of new business models, such as using machine learning technology for subscription and improving content visibility through personalization. Related categories, such as analytics, audience development, and new advertising models, were strong in Round 4 as well (DNI, 2017).

In Round 5 projects announced in 2018, a focus on using new technologies to help news organizations with monetization and diversifying revenue streams was still the trend. Audio technologies, smart assistants, mobility-focused applications were widely addressed by Round 5 projects. New possibilities for existing business models were explored further, for example, using personalization technologies for dynamic pricing systems or subscriptions. Automation and AI technologies also remained important topics in Round 5 (Blecher, 2018).

In 2019, Round 6 winning projects were announced. In this round, Google asked large and medium project applicants to focus on “one of the most pressing issues

identified by the news ecosystem: the diversification of revenue streams" (Blecher, 2019). In addition, projects about artificial intelligence and machine learning technologies continued to be a top technology focus. The applications of these technologies concentrated on opportunities driving subscriptions, creating new payment models, and finding solutions to minimize churn (subscribers who leave the service during a given time period). Right after Round 6 was closed in March 2019, Google announced the launch of the Google News Innovation Challenge in Europe to support local news. In May 2019, Google also for the first time launched the GNI Innovation Challenge in North America, which aimed to fund selected projects up to $300K to support local news (Shaw, 2019).

DNI's 2018 report identified four main areas that were funded in 2016 and 2017: using fact checking-related technologies to combat misinformation (about 7% of the funded projects), innovations that support local and small publishers in the digital space (11%), boosting digital revenues (11%), and exploring new technologies, especially technologies that enhanced automation in the news workflow, content, and user experience (23%) (DNI, 2018). In 2018 and 2019, fact check-related projects decreased noticeably in Europe. Google's investment in fact checking had different patterns in Europe-focused DNI Fund projects and the GNI's general initiatives, e.g. since 2018 the former withdraws much faster than the latter in funding fact checking related projects. This differentiation indicates that fact checking is a time- and region-specific issue, with a higher level of concerns about "fake news" in the U.S. than in Europe overall, as shown in the 2018 Reuters Institute report (Newman et al., 2018). In addition, the differentiation

indicates that the concern with fact checking is stronger at certain times, such as during election seasons, than other times.

Since 2018, AI technology-related projects have become a top focus. Google has been training the news industry to apply these technologies in various fields, including content production and distribution, dynamic pricing systems, and audience engagement. These technologies have pushed the future of news in a data-driven direction, as all these technologies require data to function. If data journalism becomes a new norm in the news industry, Google's technological control will only grow stronger. Projects that concerned local news and local media kept going up in Round 5 (13%), and the area was further encouraged in 2019. Since the winning projects were selected based on Google's criteria, the composition of the DNI Fund projects reflects areas that Google encourages and invests in.

This section of the chapter about GNI, the GNI blog, and the DNI Fund reveals that Google has systematically involved itself in the news media's future development through financial means (such as funding for projects it selects), technological means (such as the cross-promoting its own products), and institutional means (such as network building and the investment in election-related projects in different parts of the world). Google has realized that technological specialization alone is not enough to differentiate itself as an autonomous social sector from established social institutions, such as the news industry. To gain a higher level of autonomy, Google must extend its economic and technological influence into socio-political realm. Before it is ready to become a differentiated social unit, it seeks power negotiation with the news industry through various means.

Conclusion: The N-D-N Theoretical Model

Drawing on the theoretical frameworks of normalization and differentiation, this dissertation provides an in-depth analysis of Google's news aggregation service in terms of the origin and early history of Google News, the structural, visual, and functional evolution on Google News homepage design, international disputes between Google and news media, Google's news related technologies and algorithms, and Google's systematic initiatives in the news area and their influence on the future of the news industry. When observations from different chapters are put together and mapped onto a timeline as shown in Figure 6-1, it becomes clear that normalization and differentiation trends have shaped the development of Google's news business and its relationship with the news industry over the past two decades (Figure 6-1).

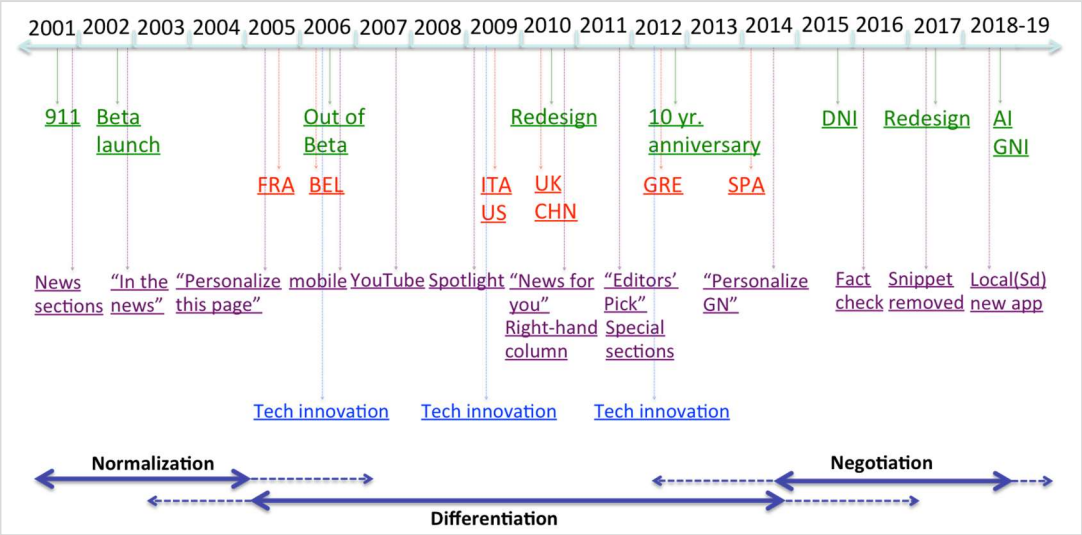

***Figure 6-1*. The N-D-N Model—Normalization and Differentiation in Google's Development in the News Area. Source: author.**

At the top of Figure 6-1 is the timeline that shows the years this study examines. Below the timeline are the key observations discussed in each chapter in this dissertation. These observations are marked with different colors. Green text indicates important

historical moments in the history of Google's news aggregation service and its news business. Red text represents countries involved in the international disputes surrounding Google and Google News. Purple text shows major features or changes observed on the Google News homepage over the past two decades. Blue text illustrates the peak moments when Google invested heavily in news-related technological innovation.

The information in Figure 6-1 provides an opportunity to examine the development of Google's news endeavors from different perspectives. Based on this examination, this dissertation proposes a new theoretical model, the N-D-N model, to conceptualize Google's development in the news arena in relation to the news industry. The N-D-N model divides Google's news initiatives into three historical stages: normalization, differentiation, and negotiation. This model builds on existing theoretical frameworks of normalization and differentiation to understand social changes and the interrelationship between "old" and "new" media sectors in the 21st century media-tech ecosystem.

Based on the review of existing studies, this dissertation, in Chapter 1, identifies five characteristics involved in the social processes of normalization and differentiation: adoption and appropriation in the process of normalization; specialization and autonomy in the process of differentiation; and negotiation in the interaction between normalization and differentiation. Normalization involves adoption and appropriation. In this process, a social sector tends to accept certain existing ideas, practices, and norms that other social sectors establish in their social routines and integrate them into the adopting sector's own routines. When normalization progresses, adoption becomes more selective in certain cases. The adopting party transforms certain aspects of the adopted ideas, practices, and

norms to serve new purposes or it marries adopted ideas, practices, and norms to new ones. As new ideas, practices, and norms take hold in the adopting party's social routines, specialization develops in one or more areas, which distinguishes the given social sector from other social sectors in terms of their respective social functions and roles. As specialization develops, the demand for autonomous social status grows, as specialization allows the given party unique resources and power that could sustain it to function independently from other social sectors. Specialization and the demand for autonomy contribute to a force of differentiation with the former as the means while the latter the end.

The social processes, however, do not develop in a neat, linear way. Often, normalization and differentiation interact in a dialectical fashion as these processes involve negotiations between social sectors and internal and external influences. As the dotted lines in Figure 6-1 indicate, the boundaries of stages are not clear-cut. In transitional periods, stages overlap, with certain trends more dominant than others during a given period of time. This is a common feature of historical periodization in all fields.

Normalization Stage

Normalization often occurs when a social sector acknowledges existing practices or norms because its own specialization is not yet clearly and strongly established. As shown in this study when Google started its news aggregation attempt after the 911 attacks and its series of quick moves afterwards, the decision of normalization usually comes with a timing condition when the given social sector has the need for certain undertaking, for social and/or economic considerations, but has no time to start that undertaking in a substantial way without relying on existing resources or other

advantages. In the early years of Google, when news was not treated differently from other types of information, the goal of the young Google was to organize the information on the web. The early history of Google in adopting the directory model while experimenting with the algorithm model indicates that Google's specialization had not yet taken shape in any specific area, including its resources, identity, and legitimacy. The attacks of September 11, 2001 gave Google an opportunity to realize news as a particular type of information with both social and economic value. In the aftermath of the 9/11 attacks, news paralysis and news demand were both salient. It was against the backdrop of this national emergency that Google adopted the news media's social role in informing the public. For example, it provided users with information and access to public resources to meet the public's critical information needs. This role was different from Google's traditional role as a pure search engine using a user-activated model in which it only returns information per users' request, rather than actively informing the public.

Google's ambition in news did not fade after 9/11. Instead, it became a formalized goal when Google launched Google News in 2002. Through its news aggregation service, Google normalized news business into its search business, and news became both the supply and the final product of Google News. To start, Google News adopted news content in a more systematic way (e.g. automatically crawling thousands of news sources worldwide). It also integrated common journalistic practices into its website structure and design, for example, using standard newspaper sections to structure its website. Existing journalistic ideas were also adopted into Google's news-related algorithmic systems, from the news media's organizational characteristics to professional criteria. In fact, Google founders' early statement made it very clear that Google adopted ideas of the

news industry in many other areas, including Google's corporate structure, its interest in objectivity, and how it justified its advertising business (Page, 2004).

In the meantime, the "Google way" of handling news introduced new ideas and practices. For example, Google's news aggregation service was automatic without human editors' intervention; it aggregated news at the individual story level and broke the boundary between particular news organizations' content. As a result, news was de-bundled, which to a certain extent invalidated the news media's role in agenda setting and gatekeeping. Google then used its own method of re-bundling news, for example, by clustering news around named entities and labels that defined news types. It could be argued that adoption was gradually taken over by appropriation at this time. Overall, normalization drove the development of Google's news business in this phase, which lasted until about the mid-2000s, since Google's specialization was still developing and had not reached a degree of technological capture as it did later on when its technological power was able to define the digital infrastructure. Furthermore, the demand for a status independent from the news industry was not strong then. After all, Google depended on the news media for content supply; it also needed time to legitimize the new ideas and practices that it introduced through its news aggregation service.

Differentiation Stage

The normalization stage that characterized the early history of Google's news aggregation service transformed into a long differentiation phase, a period of time that featured high degree of specialization and a strong demand for autonomy, with the former as the internal influence and the latter the external influence. From the mid-2000s to early 2010s, Google heavily invested in its technological innovation in the news area, as shown

in Figure 6-1 and in the more detailed analysis in Chapter 5. The technological developments in those years focused on areas, such as personalization and search-related technologies, that clearly differentiated Google from traditional news media. As Google's technological and economic power grew, Google encountered strong resistance from the news industry. From 2005 to 2014, Google and its news aggregation service were involved in a series of disputes in different parts of the world. News organizations that made charges against Google targeted its specialized technologies as well as the ideas, norms, and practices supported by these technologies, including automated techniques in crawling and caching news, industry technical standards, algorithmic interlink, and Google's business model and competition position. In these disputes, news publishers attempted to use existing legal frameworks to define Google's technologies and practices. In response, Google defended itself mostly from technological perspective as well, trying to use its technological specialization to justify its operations and resist being incorporated into the regulatory frameworks advocated by the news industry. The external pressure from the news industry gave Google a strong sense of crisis as publishers' charges touched upon fundamental issues concerned Google's bottom line and lifeline. The past experience of other large tech companies, such as Microsoft, told Google that this was a critical moment (Novet & Elias, 2019). This strong sense of crisis led to a high degree of demand for autonomy.

During the years that Google was involved in international disputes, it made changes to its news aggregation service. A variety of new features—Spotlight, Editors' Picks, fact check, and Local—were introduced on Google News to respond to the news media's challenges. While these practices looked like adoption, the analysis in Chapter 3

indicates they were responsive in nature, because these changes were made by Google as a response to pressure from the news industry. These responsive changes reflected Google's limited level of autonomy; compared to the established news industry, Google represented a young social sector. While Google had technological and economic power, the news media showed strong institutional power in these cases, e.g., through lobbying, journalistic legitimacy, and public discourse, etc., which threatened the legitimacy of Google's technologies and practices. As discussed earlier in this dissertation, organizational legitimacy is significant for young organizations to survive their early stages, which have a high death rate. The demand for autonomy became urgent under such pressure.

While Google did successfully settle a few early cases by drawing on its commercial and technological power, such settlements were temporary. In different media markets, Google was hit by a strong backlash. In Germany and Spain, news publishers used their lobbying power to push the passage of new laws that targeted Google. Even in the U.S. and the UK, countries with liberal traditions, there have been growing calls for the end of the self-regulation era for large tech companies like Google. At then end of 2019, the U.S. Justice Department was planning an antitrust investigation on Google; and **U.S.** Sen. Elizabeth Warren of Massachussets, one of the candidates for the 2020 Democratic presidential nomination, urged regulators to pass legislation to break up large tech platforms (Novet & Elias, 2019). Furthermore, the news industry also used its institutional power to wage a series of anti-platform campaigns to influence public discourse and policy agenda (see Chapter 4).

Although Google was able to bypass some legal constraints via technological tactics, as shown in Chapter 4 of this dissertation, and simply refused in 2019 to pay publishers in France, the first country that transposed the European Union Copyright Directive (EUCD) into national law (Willsher, 2019), there was a sense in late 2019 that regulatory measures were getting stricter. The news industry has played a role in this climate change. The recent reforms in copyright, antitrust, privacy, and data protection regulations in different parts of the world seemed to suggest that although Google's attempt at differentiation was intense, the outcome so far has not effectively eased Google's struggles with the news industry, a powerful, established social institution worldwide, despite declines in some news sectors, such as newspapers, in many countries. From the mid-2000s to mid-2010s, Google's differentiation process was driven internally by a high degree of specialization and externally by the pressure from other social sectors that resulted in the demand for autonomy. At this stage, differentiation played a bigger role than normalization, although the two processes were not completely broken away from each other, as shown by the dotted lines in Figure 6-1.

Negotiation Stage

Negotiation is a result of the deliberate studying and weighing of a situation. It involves the given social sector's judgment of the broader context, its understanding of other social sectors' leverage, and the self-assessment of its own power position relative to other social sectors. In the case of Google, the negotiation phase happened after the normalization and the differentiation phases, since that was when Google was able to examine all the factors above.

As far as the broader context is concerned, the interrelationship between traditional news media and digital platforms had evolved by the time Google News entered the negotiation phase. News media had passed the stage of early adoption of digital technology and entered a new stage that sought a closer relationship with tech companies. Entering the second half of the 2010s, such trends became clear. For example, a study by the Tow Center for Digital Journalism at Columbia University's Graduate School of Journalism showed that in 2016 "publishers spoke about platforms with more detachment than they do today [2018]: these were distribution channels to put content in front of audiences. The sentiment worked both ways." (Bell, 2018, para. 47) Two years later, the research team found that "both platforms and publishers repeatedly used words like 'partner' and 'partnership' to describe their increasingly close relationship" (para. 48). This study also found that the news media had "wrestle[d] back a degree of control" (para. 136) through years of trying out various technologies and dealing with their relationship with digital platforms. For example, the news media knew better what technological features rolled out by digital platforms do or do not work in serving their news products or editorial purposes and therefore gained more negotiation power.

Under such circumstances, Google stressed a collaborative relationship with the news industry through its public discourse and efforts it made to answer the news industry's criticisms and concerns, such as applying Google's technologies to address fake news, local journalism, and business model issues; removing news snippets from Google News; and making it technologically easier to direct users to news media outlet websites. Google also realized that technological power alone is not enough to sustain an

autonomous social status relative to other established social institutions inside and outside the U.S. As a result, Google aimed to grow its institutional power more systematically through a series of initiatives worldwide. On the one hand, it continued to develop and promote its technologies and products through training and funding opportunities and an expanding partnership network; on the other hand, Google attempted to enhance its political influence, e.g., by playing a growing role in elections in different parts of the world through GNI. Google also joined other large digital platforms to increase lobbying spending and strengthen its influence during elections. Google's experience in China and other parts of the world also helped it realize the complex political and ideological contexts in the global media-tech market. Moving forward, Google's – and other large tech companies' – technological and economic development will continue to be subject to these influences domestically and globally. As realized by Google's founders in its early history, "We want Google to become an important and significant institution. That takes time, stability and independence." (Page, 2004, para. 24) Given the complex dynamics and power relations examined above, a long negotiation phase is expected before Google can achieve the goal of its founders.

Implications

This dissertation was written against the backdrop of an ongoing conversation about digital platform governance. A fundamental question related to this topic is whether digital platforms are, by nature, media companies or tech companies (Napoli & Caplan, 2017). This question matters because media and tech sectors are subject to different regulatory standards in the U.S. Napoli (2019) pointed out that under the existing media regulation principles, the media sector, especially the electronic media

sector, is required to abide by public interest obligations while the same public interest framework does not apply to the tech sector so far. Instead, tech companies tend to be protected by such internet laws as Section 230 of the Communications Decency Act, which stipulates that "interactive computer service providers" shall not be treated as "the publisher or speaker of any information provided by another information content provider" (47 U.S.C. § 230). The outcome of the ongoing debate about whether and how to regulate digital platforms will have a far-reaching impact on the future of journalism and democracy. The analysis of Google's development in the news area and the N-D-N model proposed based on this analysis will have important implications for law and policymaking regarding digital platform governance and the future development of the news industry in three areas.

The first is related to the hybridity and its implications for law and policymaking. The interaction between normalization and differentiation indicates that Google's news business has a hybrid nature, mixing characteristics of both media and tech sectors. During Google's evolution, the hybridity placed emphasis on different aspects in different phases in the N-D-N model. For example, in the normalization phase, Google News was more of a *media-tech* enterprise, while in the differentiation phase it presented more *tech-media* characteristics. In the negotiation phase, Google pursued a balance between influences from both sectors. This hybridity and its dynamic nature called for more attention to innovative regulatory approaches as an either/or (i.e., either media or tech) or a one-size-fits-all approach in law and policymaking about platform regulation may not be able to effectively address the complexities involved in the processes of normalization, differentiation, and negotiation. To better understand these complexities,

one needs to examine the internal and external influences that drive the decision-making at different stages in the N-D-N model, the resulting practices, and their effects. In addition to understanding individual actors, such as Google, it is important for lawmakers and policy crafters to have a vision about the media and tech ecosystem.

The second area of important implications for the N-D-N model involves a timing issue in law and policymaking. The N-D-N model marks Google's development in the news area over the past two decades into three stages. While the *order* of these stages may vary for digital platforms with different evolutionary trajectories, these processes are important components for understanding the interrelationship between "old" and "new" social sectors in the changing media and tech environment. The N-D-N model provides lawmakers and policymakers with a tool for thinking about when or at what stage is the right time to intervene? Should it be at the normalization stage, the differentiation stage, or the negotiation stage? Since the normalization stage involves the adoption of existing practices and norms, the urgency for regulatory intervention is weak at this time, as these adopted practices and norms have already been addressed by existing laws and policies. But new issues may emerge during the appropriation process, which signifies areas for regulatory reforms. At the differentiation stage, regulators need to take into consideration the relationship between new and incumbent social sectors as well as their respective interests. By the time the negotiation stage starts, the big picture is clearer in terms of the trends in the broader context as well as involved parties' power positions and expectations. Since the negotiation stage often indicates a close match between the competing parties, external intervention may be needed at this point to break the standoff. An important question that lawmakers and policymakers need to address is what kind of

approach—restrictive or non-restrictive—is more proper at any stage? Different approaches may have different effects in shaping a given sector's development, in balancing innovation and responsibility, and in defining the overall influence of the sector on society.

The N-D-N model's third area of implication for law and policy involves its implication for the future development of the news industry. The debate about whether digital platforms are media or tech companies is, in fact, an indicator of the increasingly complex and constantly changing media and tech environment, in which the boundary between the two sectors is being negotiated. In this environment, each sector needs to act in relation to the other sector. This two-way shaping between the two sectors will determine where they go in the next decade. Journalism plays a critical role in democracy. In the digital era, its future development requires the grasp of both itself and other sectors that share today's news ecosystem. The N-D-N model was used in this dissertation to explain how and under what circumstance Google defined its news strategies differently. It also demonstrated the areas in which Google adopted existing journalistic practices and ideas and the areas in which Google introduced completely different practices and ideas when dealing with news. The analysis of Google's news strategies and the N-D-N model reveal the respective strengths and limits of the media and tech sectors. While Google had strong technological power, it lacked the news media's legitimacy, which was rooted in journalistic traditions. For example, when Google pushed technology and datafication, journalism's relationship with real people and real communities becomes invaluable and powerful. The N-D-N model is thus useful not only for scholarly theorizing but also a practical tool for the news industry to

strategically plan its future development and to decide its own processes of normalization, differentiation, and negotiation.

Limitations and Future Studies

This dissertation provides an in-depth, systematic analysis of Google's endeavors in the news area. It cannot, of course, tell us everything we might like to know about Google News. It might, for example, be useful to conduct interviews with Google employees in future studies to get first-hand accounts to enrich this analysis. This study, however, has provided a more comprehensive and multi-faceted examination of the world's largest news aggregator than had previously been conducted.

The study proposes the N-D-N model based on an analysis of Google and its news aggregation service over the past two decades. To what extent this model can be generalized to other digital platforms that have distinct historical backgrounds, evolutionary trajectories, and contingencies must be explored through future research. It is also important to keep in mind that when media and technology advance from one generation to another, the media and technology systems evolve through a dynamic process, in which individual cases are, "at any given time, at different stages in their development" (Napoli, 1998, p. 318). For example, observing in the year 2020, while it looks like print is ancient, broadcasting is mature, and digital media are rising, these media forms are at different stages in their respective development. They may have experienced their own normalization, differentiation, and normalization at different times. Therefore, when testing the N-D-N model in future studies, researchers should be alert to this dynamic nature of media evolution. In fact, the N-D-N model contributes to a better

understanding of this dynamics as it addresses the interactions between the "old" and "new" media sectors.

It is worth noting that the N-D-N model is based on an approach that divides history into periods. Some historians considered period divisions "arbitrary constructs" (Blackbourn, 2012, p. 301) that "flatten up the diversity" of "phenomena that appear in any particular time" (Kaufmann, 2010, p. 3). They argue that discrepancies across time and space and cross-cultural interactions in the global context could complicate efforts at periodization and therefore affect their effectiveness (e.g., Bentley, 1996). Thus, deep understanding of the nuances and complexities in different contexts and careful examination of the relationship between uniqueness and generalization will be required when applying the N-D-N model in future studies. Variations should be allowed and the variations themselves deserve great scholarly attention. The N-D-N model is nonetheless valuable because, as many historians have pointed out, it's almost impossible to advance any evidence-based interpretation of history without resorting to seeing certain eras as indicative of certain tendencies (Green, 1995; White, 1987).

As the world enters the third decade of the 21st century, this dissertation presents an endeavor in exploring such tendencies in order to gain a comprehensive understanding of the patterns of continuity and change as well as the driving forces, power relations, and value judgments involved in the processes that shape the media and technology environment, in particular, and our society in general.

References

1:05CV00546-GK (2007). *Agence France-Presse v. Google, Inc*. In the United States District Court For the District of Columbia.

1:05CV00546-GK (2006). *Agence France-Presse v. Google, Inc*. Report.

1:05CV00546-GK (2005). *Agence France-Presse v. Google, Inc*. Google's Reply to AFP's Opposition to Partial Summary Judgment Dismissing Count II (Headlines) For Lack Of Protectable Subject Matter.

47 U.S.C. § 230. Section 230 of the Communications Decency Act (CDA) of 1996.

About news search (2002). Google.com. https://web.archive.org/web/20021013214112/http://news.google.com:80/help/about_news_search.html

About the Google Directory (2000). Google.com. https://web.archive.org/web/20000620000041/http://directory.google.com:80/dirhelp.html

About the Open Directory Project. (2009, September 1). Facebook. https://www.facebook.com/notes/nico-associates/about-the-open-directory-project/127986343276/

About (n.d.). Google News Initiative, https://newsinitiative.withgoogle.com/about

About Us. (n.d.) ANSE.org. https://members.newsleaders.org/about-us

AEEPP (2015, July 9). Economic report on the impact of the New Article 32.2 of the LPI. Aeepp.com. https://www.aeepp.com/noticia/2272/actividades/informe-economico-del-impacto-del-nuevo-articulo-32.2-de-la-lpi-nera-para-la-aeepp.html

AFP. (2010, October 23). Paid content takes hit as News Corp drops $32m Project Alesia "digital newsstand." News.com.au. https://www.news.com.au/breaking-news/paid-content-takes-hit-as-news-corp-drops-32m-project-alesia-digital-newsstand/news-story/b547dbdef6f14cc9ca12cf798d2b4e7c

AFP (2014. October 1). Google removes news snippets in Germany legal fight. Phys.org. https://phys.org/news/2014-10-google-news-snippets-germany-legal.html

AGCM (2011). AGCM annual report. Agcm.it. http://www.agcm.it/en/component/joomdoc/annual-reports/AnnualReport2010_Catrical.pdf/download.html

Agence France Press v. Google Inc, 1:05CV00546, D.D.C. (2005)

Alexander, J. C., & Colomy, P. B. (1990). *Differentiation theory and social change: Comparative and historical perspectives*. Columbia University Press.

Alexander, J. C. (1990). Introduction: Differentiation theory: Problems and prospects. In J. C. Alexander, & P. B. Colomy (Eds.), *Differentiation theory and social change: Comparative and historical perspectives* (pp. 1-15). Columbia University Press.

Ali, R. (2007, August 31). Google News starts publishing full stories; AP, AFP, PA and Canadian Press; Still no ads. Gigaom.com. https://gigaom.com/2007/08/31/419-google-starts-publishing-full-stories-on-its-news-section-ap-afp-pa-and/

Anderson, B. (2006). *Imagined communities: Reflections on the origin and spread of nationalism*. Verso Books.

Anderson, C. W. (2013). What aggregators do: Towards a networked concept of journalistic expertise in the digital age. *Journalism*, 14(8), 1008-1023.

Anderson, C. W. (2011). Between creative and quantified audiences: web metrics and changing patterns of newswork in local US newsrooms. *Journalism, 12*(5), 550–566.

ANSA (2016, June 7). FIEG-Google deal to boost media sector (2). Ansa.it. http://www.ansa.it/english/news/science_tecnology/2016/06/07/fieg-google-deal-to-boost-media-sector-2_fd3880f1-19b9-4a88-a49a-42b29e602ec0.html

Antitrust: Consumer welfare at heart of Commission fight against abuses by dominant undertakings. (2008). Europa.eu. https://ec.europa.eu/commission/presscorner/detail/en/IP_08_1877

Ap.org (2018). AP by the numbers. Ap.org. https://www.ap.org/about/annual-report/2018/ap-by-the-numbers

Apple launches Apple News+, an immersive magazine and news reading experience. (2019, March 25). Apple.com. https://www.apple.com/newsroom/2019/03/apple-launches-apple-news-plus-an-immersive-magazine-and-news-reading-experience/

Arthur, B. W. (2009). *The nature of technology: What it is and how it evolves*. Free Press.

Associated Press (2009, August 27). Antitrust watchdog probes Google Italy. AP. https://phys.org/news/2009-08-antitrust-watchdog-probes-google-italy.html#nRlv

Athey, S., & Mobius, M. (2012). *The impact of news aggregators on internet news consumption: The case of localization* (pp. 1-36). working paper.

Athey, S., Mobius, M. M., & Pál, J. (2017). The impact of aggregators on Internet news consumption. Stanford University Graduate School of Business Research Paper No. 17-8.

Auchard, E. (2007, April 7). AFP, Google News settle lawsuit over Google News. Reuters.com. https://www.reuters.com/article/us-google-afp/afp-google-news-settle-lawsuit-over-google-news-idUSN0728115420070407

Axel Springer concludes its data documentation: Major losses resulting from downgraded search notices on Google. (2014, November 5). Axelspringer.com. https://www.axelspringer.com/en/presse/Axel-Springer-concludes-its-data-documentation-Major-losses-resulting-from-downgraded-search-notices-on-Google_22070687.html

Bakker, P. (2012). Aggregation, content farms and Huffinization: The rise of low-pay and no-pay journalism. *Journalism Practice*, *6*(5-6), 627-637.

Baram, M. (2017). Facebook finally admits it's a giant media company — almost. *Fast Company* (11 January). https://www.fastcompany.com/3067148/facebook-finally-admits-its-a-giant-media-company-almost

Barnett, E. (2009, November 13). Rupert Murdoch has warned that his paywall plans could be delayed. Telegraph.co.uk. https://www.telegraph.co.uk/finance/newsbysector/mediatechnologyandtelecoms/digital-media/6559694/Rupert-Murdoch-to-remove-News-Corps-content-from-Google-in-months.html

Bastone, N. (2018, November 7). Eric Schmidt takes the blame for Google's social networking failures: "I suspect we didn't fully understand how to do it." Businessinsider.com. https://www.businessinsider.com/eric-schmidt-takes-responsibility-for-google-social-networking-problems-2018-11

BBC (2009). Microsoft and Yahoo seal web deal. News.bbc.co.uk. http://news.bbc.co.uk/2/hi/business/8174763.stm

Becker, A. (2014, October 30). German publishers vs. Google. Dw.com. https://www.dw.com/en/german-publishers-vs-google/a-18030444

Becker, M. C., Lazaric, N., Nelson, R. R., & Winter, S. G. (2005). Applying organizational routines in understanding organizational change. *Industrial and corporate change*, *14*(5), 775-791.

Beckmann, C. (2010, July 15). Google News changes reflect your feedback. Google Blog. https://news.googleblog.com/2010/07/google-news-changes-reflect-your.html

Bell, E.; & Albright, J. (2018). Friend and foe: the platform press at the heart of journalism. Columbia Journalism Review. https://www.cjr.org/tow_center_reports/the-platform-press-at-the-heart-of-journalism.php/

Bell, E., & Owen, T. (2017). The platform press: How silicon valley reengineered journalism. Columbia.edu. https://academiccommons.columbia.edu/catalog/ac:15dv41ns27

Bell, E. (2018, June 14). Friend and foe: The platform press at the heart of journalism. cjr.org. https://www.cjr.org/tow_center_reports/the-platform-press-at-the-heart-of-journalism.php

Bemis, A. H. (2001, September 12). Remains of the day. Laweekly.com. https://www.laweekly.com/news/remains-of-the-day-2133806

Benson, R. (2013). *Shaping Immigration News*. Cambridge University Press.

Benson, R. (2005). Mapping field variation: Journalism in France and the United States. In R. Benson, & E. Neveu (Eds). *Bourdieu and the journalistic field* (pp. 85-112). Polity Press.

Bentley, J. H. (1996). Cross-cultural interaction and periodization in world history. *American Historical Review*, *101*(3), 749-770.

Berne Convention (1979). Berne Convention for the Protection of Literary and Artistic Works. https://wipolex.wipo.int/en/text/283693

Bharat, K. (2012, September 22). Google News turns 10. Google Blog. https://news.googleblog.com/2012/09/google-news-turns-10.html

Bharat, K., Curtiss, M., & Schmitt, M. (2016). *U.S. Patent No. 9,361,369*. Washington, DC: U.S. Patent and Trademark Office.

Bharat, K., Stoll, N., & Mayer, M. (2017). *U.S. Patent No. 9,760,629*. Washington, D.C.: U.S. Patent and Trademark Office.

Blackbourn, D. (2012). "The horologe of time": Periodization in history. *PMLA*, *127*(2), 301-307.

Blanco, C. F., Sarasa, R. G., & Sanclemente, C. O. (2010). Effects of visual and textual information in online product presentations: looking for the best combination in website design. *European Journal of Information Systems*, *19*(6), 668-686.

Blecher, L. (2019, March 21). Digital News Innovation Fund: three years in, and 662 total projects supported. Google Blog. https://www.blog.google/around-the-globe/google-europe/digital-news-innovation-fund-three-years-and-662-total-projects-supported/

Blecher, L. (2018, July 24). Digital News Innovation Fund: More than €21m of funding to reach €115m to date for innovation in news. Google Blog.

https://www.blog.google/around-the-globe/google-europe/digital-news-innovation-fund-more-21m-funding-reach-115m-date-innovation-news/

Blecher, L. (2017, July 6). More than EUR 21 million in funding from Round 3 of the Digital News Initiative Fund. Google Blog. https://www.blog.google/around-the-globe/google-europe/21-million-funding-round-3-digital-news-initiative-fund/

Blythe, S. E. (2006). The U.S. Digital Millennium Copyright Act and the E.U. Copyright Directive: Comparative impact on fair use rights, 8 Tul. J. Tech. & Intell. Prop. 111 (2006)

BOE-A-2014-11404. (2014). Ley 21/2014, de 4 de noviembre, por la que se modifica el texto refundido de la Ley de Propiedad Intelectual, aprobado por Real Decreto Legislativo 1/1996, de 12 de abril, y la Ley 1/2000, de 7 de enero, de Enjuiciamiento Civil. https://www.boe.es/buscar/doc.php?id=BOE-A-2014-11404

Bogust, I. (2016, August 3). Facebook is not a technology company. Atlantic. http://www.theatlantic.com/technology/archive/2016/08/facebook-is-not-a-technology-company/494183/

Bourdieu, P. (1998). *On Television*. Translated by P. P. Ferguson. The New Press.

Bourdieu, P. (1993). *The field of cultural production: Essays on art and literature*. University Press.

Branigan, T. (2010, January 12). Google to end censorship in China over cyber attacks. Theguardian.com. https://www.theguardian.com/technology/2010/jan/12/google-china-ends-censorship

Bright, J., & Nicholls, T. (2014). The life and death of political news: Measuring the impact of the audience agenda using online data. *Social Science Computer Review*, 32 (2), 170–181.

Brown, C., Hardiman, A., & Salari, S. (2017, October 19). Testing subscriptions support in instant articles. Facebook. https://facebookjournalismproject.com/article/testing-subscriptions-support-in-instant-articles/

Buckley, C. (2010, March 10). Q+A: Is there a WTO case against Chinese Internet censorship? Reuters.com. https://www.reuters.com/article/us-china-usa-internet/qa-is-there-a-wto-case-against-chinese-internet-censorship-idUSTRE6290T120100310

Byers, D., & Fiegerman, S. (2017, November 1). Facebook, Twitter, Google face aggressive questioning at Senate hearing. CNN.com. https://money.cnn.com/2017/11/01/media/facebook-twitter-google-russia-senate-house-intelligence-committees/index.html

Byford, S. (2012, December 13). Google finally settles with Belgian publishers over content scraping lawsuit. Theverge.com. https://www.theverge.com/2012/12/13/3764692/google-copyright-lawsuit-settlement-belgium

Cain, Á. (2018, May 4). 9 famous predictions by Nostradamus some people say foresaw the future. Businessinsider.com. https://www.businessinsider.com/predictions-of-nostradamus-2011-12

Calzada, J., & Gil, R. (2016). What do news aggregators do? Evidence from Google News in Spain and Germany. The Social Science Research Network Electronic Paper Collection 2016.

Carlson, M. (2007). Order versus access: News search engines and the challenge to traditional journalistic roles. *Media, Culture & Society*, *29*(6), 1014-1030.

Carlson, M., & Lewis, S. C. (Eds.). (2015). *Boundaries of journalism: Professionalism, practices and participation*. Routledge.

Carlson, M., & Usher, N. (2016). News startups as agents of innovation: For-profit digital news startup manifestos as metajournalistic discourse. *Digital Journalism* 4 (5), 563–581.

Carlson, M. (2015). Introduction: The many boundaries of journalism. In M. Carlson, & S. C. Lewis (Eds.). *Boundaries of journalism: Professionalism, practices and participation* (pp. 1-18). Routledge.

Carlson, S. (2011, May 18). Rededicating the Newseum's Journalists Memorial. Google Blog. https://news.googleblog.com/2011/05/rededicating-newseums-journalists.html

Campbell, C. (2018, January 18). Baidu's Robin Li is helping China win the 21st Century. Time.com. http://time.com/5107485/baidus-robin-li-helping-china-win-21st-century/

Case studies. (n.d.) Google News Initiative, https://newsinitiative.withgoogle.com/success-stories/

Catone, J. (2007, July 10). 5 News Aggregation Methods Compared. ReadWriteWeb. https://readwrite.com/2007/07/10/news_aggregation_methods/

Cerulus, L. (2019, January 23). Google to appeal €50 million GDPR fine. Politico.eu. https://www.politico.eu/article/google-appeals-e50-million-gdpr-fine/

Chavern, D. (2018, August 21). It's time to rein in the tech platforms' anticompetitive behavior. News Media Alliance. https://www.newsmediaalliance.org/ceo-time-to-rein-in-platforms/

Chiou, L., & Tucker, C. (2011). Copyright, digitization, and aggregation. *Digitization, and Aggregation (December 13, 2011). NET Institute Working Paper, 11-18*.

Cho, J., Roy, S., & Adams, R. E. (2005, June). Page quality: In search of an unbiased web ranking. In Proceedings of the 2005 ACM SIGMOD international conference on Management of data (pp. 551-562).

Cipriani, J. (2018, May 16). The new Google News app has arrived: Here's what you need to know. Cnet.com. https://www.cnet.com/how-to/new-google-news-app-what-you-need-know/

Claburn, T. (2007, August 31). Google News becomes a publisher. Informationweek.com. https://www.informationweek.com/google-news-becomes-a-publisher/d/d-id/1058737

ClaimReview. (n.d.). schema.org. https://schema.org/ClaimReview

Clement, J. (2019, September 9). Global market share of search engines 2010-2019. statista.com. https://www.statista.com/statistics/216573/worldwide-market-share-of-search-engines/

Clinton, H. (2010). U.S. Department Of State remarks by Secretary of State Hillary Rodham Clinton on internet freedom. State.gov.nhttps://2009-2017.state.gov/secretary/20092013clinton/rm/2010/01/135519.htm

Coddington, M. (2019). *Aggregating the News: Secondhand Knowledge and the Erosion of Journalistic Authority*. Columbia University Press.

Coddington, M. (2015). Clarifying journalism's quantitative turn: A typology for evaluating data journalism, computational journalism, and computer-assisted reporting. *Digital Journalism. 3*(3), 331–348.

Coddington, M. (2014). Normalizing the hyperlink: How bloggers, professional journalists, and institutions shape linking values. *Digital Journalism*, *2*(2), 140-155. https://doi.org/10.1080/21670811.2013.785813

Cohen, J. (2007, August 31). Original stories, from the source. Google Blog. https://news.googleblog.com/2007/08/original-stories-from-source.html

Cohen, S., Zampano, G., & Kiviniemi, P. (2009, August 28). Italy Opens Antitrust Probe of Local Google News Service. Wsj.com. https://www.wsj.com/articles/SB125137301333663261

Collins (n.d.). Definition of "mainstream media". Collinsdictionary.com. https://www.collinsdictionary.com/dictionary/english/mainstream-media

Condliffe, J. (2016, November 7). China's new cybersecurity rules may block western innovation. Technologyreview.com. https://www.technologyreview.com/s/602812/chinas-new-cybersecurity-rules-may-block-western-innovation/

Coner, A. (2003). Personalization and customization in financial portals. *Journal of American Academy of Business*, 2, 498–504.

Cook, T. E. (1998). *Governing with the news: The news media as a political institution*. University of Chicago Press.

Cooke, L. (2005). A visual convergence of print, television, and the internet: charting 40 years of design change in news presentation. *New Media & Society*, *7*(1), 22-46.

Copiepresse (2011). Introduction. copiepresse.be. http://www.copiepresse.be/judiciaire.php?classement=02

Copyright Act. 17 U.S.C. § 107, 2016. Limitations on exclusive rights: Fair use

Costa, M., & Silva, M. J. (2010). Understanding the information needs of web archive users. In Proceedings of the 10th International Web Archiving Workshop.

Court of first instance of Brussels. 2006/9099/4. Ref. no 39513. August 3, 2006

Crawford, W. (2002, February 1). Turn on the radio. Freelibrary.com. https://www.thefreelibrary.com/_/print/PrintArticle.aspx?id=82103779

Cukier, K, & Mayer-Schoenberger, V. (2013). The Rise of Big Data. *Foreign Affairs* (May/June), 28–40.

Curtiss, M., Bharat, K. A., & Schmitt, M. (2016). *U.S. Patent Application No. 14/696,931*.

Cutts, M. (2008, June 28). Using data to fight webspam. Google Blog post. http://googleblog.blogspot.com/2008/06/using-data-to-fight-webspam.html

Das, A. S., Datar, M., Garg, A., & Rajaram, S. (2007, May). Google news personalization: scalable online collaborative filtering. In *Proceedings of the 16th international conference on World Wide Web* (pp. 271-280). ACM.

DeVito, M. A. (2017). From editors to algorithms: A values-based approach to understanding story selection in the Facebook news feed. *Digital Journalism*, *5*(6), 753-773.

Diakopoulos, N. (2014). Algorithmic accountability reporting: On the investigation of black boxes. Tow Center for Digital Journalism. http://towcenter.org/wp-content/uploads/2014/02/78524_Tow-Center-Report-WEB-1.pdf

Diakopoulos, N., & Koliska, M. (2016). Algorithmic transparency in the news media. *Digital Journalism*, 1-20.

Dick, M. (2011). Search engine optimisation in UK news production. *Journalism Practice*, 5(4), 462-477.

Dimmick, J., & Rothenbuhler, E. (1984). The theory of the niche: Quantifying competition among media industries. *Journal of Communication*, *34*(1), 103-119. https://doi.org/10.1111/j.1460-2466.1984.tb02988.x

Dimmick, J., Chen, Y., & Li, Z. (2004). Competition between the Internet and traditional news media: The gratification-opportunities niche dimension. *The Journal of Media Economics*, *17*(1), 19-33.

Directive 2001/29/EC of the European Parliament and of the Council of 22 May 2001 on the harmonisation of certain aspects of copyright and related rights in the information society. The European Parliament and the Council of the European Union.

DMCA (1998). The Digital Millennium Copyright Act of 1998.

DNI (2017). Announcing successful recipients of Round 4 funding. Google Blog. https://digitalnewsinitiative.com/news/announcing-successful-recipients-round-4-funding/

DNI (2018). Elevating quality journalism: Digital News Innovation Fund report 2018. Google Blog. https://newsinitiative.withgoogle.com/dnifund/report/

Dolan, P.; & Liu, J. (2012). *U.S. Patent No. 8,229,873 B1*. Washington, DC: U.S. Patent and Trademark Office.

Domingo, D., & Paterson, C. (2011). *Making online news, vol. 2: Newsroom ethnographies in the second decade of Internet journalism*. Peter Lang.

Dörr, K. N. (2016). Mapping the field of Algorithmic Journalism, *Digital Journalism*, 4:6, 700-722, DOI: 10.1080/21670811.2015.1096748

Duke, J. (2018, February 26). Facebook, Google to be probed over role in creating "filter bubbles." Smh.com.au. https://www.smh.com.au/business/companies/facebook-google-to-be-probed-over-role-in-creating-filter-bubbles-20180226-p4z1pj.html

Dwyer, C. (2017, May 10). German publishers' lawsuit against Google threatens to backfire. Npr.org. https://www.npr.org/sections/thetwo-way/2017/05/10/527800498/german-publishers-lawsuit-against-google-threatens-to-backfire

Edmond, R. (2015, July 12). Mid-year report: The newspaper industry's billion dollar challenge. Poynter. http://www.poynter.org/2015/mid-year-report-the-newspaper-industrys-billion-dollar-challenge/353911/

Elder, R. (2016, August 9). Publishers in the UK are uniting against Google and Facebook. Businessinsider.com. http://www.businessinsider.com/uk-publishers-unite-to-take-on-google-and-facebook-2016-8

EPC (2013, March 1). Europe's publishers welcome new German law to force content aggregators and search engines to recognise copyright. European Publishers Council. http://www.editorandpublisher.com/news/europe-s-publishers-welcome-new-german-law-to-force-content-aggregators-and-search-engines-to-recognise-copyright/

Essers, L. (2012, December 13). Belgian news publishers settle copyright dispute with Google. Cio.com.au.

https://www.cio.com.au/article/444548/belgian_news_publishers_settle_copyright_dispute_google/

Evolution of search. (2011, November 27). Youtube.com. https://www.youtube.com/watch?feature=player_embedded&v=mTBShTwCnD4&t=122

Facebook News & the current wave of news aggregators. (2019, November 9). Medium.com. https://medium.com/6pages/facebook-news-the-current-wave-of-news-aggregators-e522795fbcc

Fact Check. (n.d.). Google. https://developers.google.com/search/docs/data-types/factcheck

Fallows, J. (2010, July 16) On Rupert Murdoch's Times Paywall. theatlantic.com. https://www.theatlantic.com/technology/archive/2010/07/on-rupert-murdochs-times-paywall/59878/

Federal Law Gazette. (2013). Law on Copyright and Related Rights. https://www.gesetze-im-internet.de/englisch_urhg/englisch_urhg.html

Feldman, M. S., & Rafaeli, A. (2002). Organizational routines as sources of connections and understandings. *Journal of Management Studies*, *39*(3), 309-331.

FIEG/Google (2010). A420-FIEG- Federazione Italiana Editori Giornali/Google. No. 21959, December 22, 2010.

Field v. Google Inc. CV-S-04-0413-RCJ-LJL. United States District Court, D. Nevada. (2006).

Filloux, F. (2013, February 25). Google News: the secret sauce. Theguardian.com. https://www.theguardian.com/technology/2013/feb/25/1

Franzen, C. (2009, November 9). What lies behind Murdoch's move to block Google? Theatlantic.com. https://www.theatlantic.com/entertainment/archive/2009/11/what-lies-behind-murdoch-s-move-to-block-google/347627/

Fortt, J. (2008, July 23). What's Google News worth? $100 million. Fortune.com. http://fortune.com/2008/07/22/whats-google-news-worth-100-million/

Fox, V. (2007, May 17). Taking advantage of universal search. Google Blog. https://webmasters.googleblog.com/2007/05/taking-advantage-of-universal-search.html

Gagne, C. (2015). Canon AEDE: Publishers' protections from digital reproductions of works by search engines under european copyright law. Temp. Int'l & Comp. LJ, 29, 203.

Gallagher, R. (2018, August 1). Google plans to launch censored search engine in china, leaked documents reveal. Theintercept.com. https://theintercept.com/2018/08/01/google-china-search-engine-censorship/

Gallagher, R. (2019, March 4). Google employees uncover ongoing work on censored China search. Theintercept.com. https://theintercept.com/2019/03/04/google-ongoing-project-dragonfly/

Garber, M. (2010, June 10). Google News experiments with human control, promotes a new serendipity with Editors' Picks. Niemanlab.org. http://www.niemanlab.org/2010/06/google-news-experiments-with-human-control-promotes-a-new-serendipity-with-editors-pick/

Garber, M. (2011, April 8). Google News founder Krishna Bharat: For news consumers, "the whole experience is what counts," Niemanlab.org. http://www.niemanlab.org/2011/04/google-news-founder-krishna-bharat-for-news-consumers-the-whole-experience-is-what-counts/

Garcia, M. (2012, September 9). USA TODAY turns 30: Part 1. Garciamedia.com. https://www.garciamedia.com/blog/usa_today_turns_30-part_1-looking_into_the_attic_for_those_early_sketches/

Geerts, T. (2012, December 12). Partnering with Belgian news publishers. Google Blog. https://europe.googleblog.com/2012/12/partnering-with-belgian-news-publishers.html

General Data Protection Regulation: a guide to assist processors. (2017). Cnil.fr. https://www.cnil.fr/en/general-data-protection-regulation-guide-assist-processors

George, L. M., & Hogendorn, C. (2012). Aggregators, search and the economics of new media institutions. *Information Economics and Policy*, *24*(1), 40-51. https://doi.org/10.1016/j.infoecopol.2012.01.005

Glaser, M. (2010, February 4). Google News to publishers: Let's make love not war. Mediashift.org. http://mediashift.org/2010/02/google-news-to-publishers-lets-make-love-not-war035/

Gillespie, T. (2017). Governance of and by platforms. *Sage handbook of social media.* Sage.

Gillespie, T. (2016). Algorithm. In B. Peters (Ed). *Digital keywords: A vocabulary of information society and culture* (pp. 18-30). Princeton University Press.

Gillespie, T. (2010). The politics of 'platforms'. *New media & society*, *12*(3), 347-364.

Gingras, R. (2014, February 24). Helping Local News Thrive. Google Blog. https://news.googleblog.com/2014/02/helping-local-news-thrive.html

Gingras, R. (2016, October 13). Labeling fact-check articles in Google News. Google Blog. https://blog.google/outreach-initiatives/google-news-initiative/labeling-fact-check-articles-google-news/

Giomelakis, D., & Veglis, A. (2015). Employing search engine optimization techniques in online news articles. *Studies in Media and Communication*, 3(1), 22-33.

Goffey, A. (2008). Algorithm. In M. Fuller (Ed.), *Software studies: A lexicon* (pp. 15-20). Cambridge, MA: MIT Press.

Goldman, E. (2005). Search engine bias and the demise of search engine utopianism. Yale JL & Tech., 8, 188.

Goldman, D. (2009, January 9). Worst year for jobs since '45. CNN.com. https://money.cnn.com/2009/01/09/news/economy/jobs_december/

Google (n.d.). How AdSense Works. Google.com. https://support.google.com/adsense/answer/6242051?hl=en

Google (n.d.). What does each label mean? Google.com. https://support.google.com/news/publisher-center/answer/4582731

Google (n.d). Google News Publisher Center overview. Google.com. https://support.google.com/news/publisher-center/answer/9606538?hl=en-GB&visit_id=637183293408216825-2342163076&rd=1

Google. (2019). Everyone benefits from access to information and creativity online. Google.com. https://www.google.com/intl/en/togetherforcopyright/

Google. (2018, May 15). Google SearchLiaison. https://twitter.com/searchliaison/status/996488444883828736

Google. (2009, December 8). Exploring a new, more dynamic way of reading news with Living Stories. Google Blog. http://googleblog.blogspot.com:80/2009/12/exploring-new-more-dynamic-way-of.html

Google. (2006, February 28). Google News for mobile devices. Google Blog. http://googlepress.blogspot.com/2006/02/google-news-for-mobile-devices_28.html

Google and France reach landmark agreement. (2013, February 2). France24.com. http://www.france24.com/en/20130201-google-france-reach-landmark-agreement

Google Inc v. Copiepresse, No. 2007/AR/1730 (2011)

Googleblog. (2004, September 27). China, Google News and source inclusion. Google Blog. https://googleblog.blogspot.com/2004/09/china-google-news-and-source-inclusion.html

Google Digital News Initiative. (2017). Stimulate engage innovate: Digital News Initiative Innovation Fund. Google Blog. https://newsinitiative.withgoogle.com/dnifund/insights/2016-2017-dni-innovation-fund-report-available-now-download/

Google SearchLiaison. (2018). Twitter.com. https://twitter.com/searchliaison/status/996488444883828736?lang=en

Görke, A., & Scholl, A. (2006). Niklas Luhmann's theory of social systems and journalism research. *Journalism Studies*, *7*(4), 644-655.

Granados, N.; & Gupta, A. (2013). Transparency strategy: Competing with information in a digital world. *MIS Quarterly: Management Information Systems* 37 (2): 637–641.

Granka, L. A. (2010). The politics of search: A decade retrospective. The Information Society, 26(5), 364-374.

Green, W. (1995). Periodizing World History. *History and Theory*, *34*(2), 99-111.

Groklaw. (2008, October 16). An update on the Copiepresse/Google Dispute—ACAP enters the picture. Groklaw.net. http://www.groklaw.net/articlebasic.php?story=20081016024354825

Gunther, M. (2006, February 16). Tech execs get grilled over China business. Cnn.com. https://money.cnn.com/2006/02/15/news/international/pluggedin_fortune/index.htm

Hågvar, Y. B. (2012). Labelling journalism: The discourse of sectional paratexts in print and online newspapers. *Nordicom Review*, *33*(2), 27-42.

Haim, M., Graefe, A., & Brosius, H. B. (2018). Burst of the filter bubble? Effects of personalization on the diversity of Google News. *Digital Journalism*, *6*(3), 330-343.

Halavais, A. (2002). Part 3. The rise of do-it-yourself journalism after September 11. In: One Year Later: September 11 and the Internet. Pewinternet.org. http://www.pewinternet.org/2002/09/05/one-year-later-september-11-and-the-internet/

Hallin, D. C., & Mancini, P. (2004). *Comparing media systems: Three models of media and politics*. Cambridge university press.

Hallin, D. C. (2005). Field theory, differentiation theory, and comparative media research. In R. Benson, & E. Neveu (Eds). *Bourdieu and the journalistic field* (pp. 224-243). Polity Press.

Hallin, D. C., & Mancini, P. (Eds.). (2012). *Comparing media systems beyond the Western world*. Cambridge University Press.

Hammonds, K. (2003, March 31). How Google grows… and grows… and grows. Fastcompany.com. https://www.fastcompany.com/46495/how-google-growsand-growsand-grows

Hammond, P. (2015). From Computer-Assisted to Data-Driven: Journalism and Big Data. *Journalism*, *18*(4), 408-424. https://doi.org/10.1177/1464884915620205

Hargittai, E. (2007). The social, political, economic, and cultural dimensions of search engines: an introduction. *Journal of Computer-Mediated Communication, 12*(3), 769-777.

Harris, W. (2004, December 1). China bans Google News. Theinquirer.net. https://www.theinquirer.net/inquirer/news/1001014/china-bans-google-news

Hebbard, D. B. (2014, October 23). German publishers "bow to pressure," will allow Google to display search result snippets. Talingnewmedia.com. http://www.talkingnewmedia.com/2014/10/23/german-publishers-bow-to-pressure-will-allow-google-to-display-search-result-snippets/

Helmore, E. (2019, June 10). Google made $4.7bn from news sites in 2018, study claims. Theguardian.com. https://www.theguardian.com/technology/2019/jun/10/google-news-revenue-2018-new-study

Holzmann, H., Nejdl, W., & Anand, A. (2017, June). Exploring web archives through temporal anchor texts. In *Proceedings of the 2017 ACM on Web Science Conference* (pp. 289-298). ACM.

Hoffmann, R., & Krauss, K. (2004, October). A critical evaluation of literature on visual aesthetics for the web. In *Proceedings of the 2004 annual research conference of the South African institute of computer scientists and information technologists on IT research in developing countries* (pp. 205-209). South African Institute for Computer Scientists and Information Technologists.

H.R.5190. (2018, March 7). Journalism Competition and Preservation Act of 2018. https://www.congress.gov/bill/115th-congress/house-bill/5190/text

H.R.2054. (2019, April 3). Journalism Competition and Preservation Act of 2019. https://www.congress.gov/bill/116th-congress/house-bill/2054/text?q=%7B%22search%22%3A%5B%22Hr+2054%22%5D%7D&r=1

Huang, J. S., Yang, M. J., & Chyi, H. I. (2013). Friend or foe? Examining the relationship between news portals and newspaper sites in Taiwan. *Chinese Journal of Communication*, *6*(1), 103-119.

Ingram, M. (2014, December 16). External traffic to Spanish news sites plummets after Google move. Gigaom.com. https://gigaom.com/2014/12/16/traffic-to-spanish-news-publishers-plummets-after-google-move/

Interview with David Speers (2009). Sky News Australia. https://www.youtube.com/watch?v=M7GkJqRv3BI

Introna, L., & Nissenbaum, H. (2000). Shaping the Web: Why the politics of search engines matters. *The Information Society, 16*(3),1–17.

Isbell, K. (2010). The rise of the news aggregator: Legal implications and best practices. Citizen Media Law Project. Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1670339

Jenkins, H. (2006). *Convergence culture: Where old and new media collide*. NYU press.

Jeon, D. S., & Nasr, N. (2016). News Aggregators and Competition among Newspapers on the Internet. *American Economic Journal: Microeconomics*, *8*(4), 91-114.

Jiang, M. (2014). The business and politics of search engines: A comparative study of Baidu and Google's search results of Internet events in China. *new media & society*, *16*(2), 212-233.

Johnson, B. (2009, November 9). Murdoch could block Google searches entirely. Theguardian.com. https://www.theguardian.com/media/2009/nov/09/murdoch-google

Kaufmann, T. D. (2010). Periodization and its discontents. *Journal of Art Historiography*, *2*(2), 1-6.

Keith, S. (2007). Searching for news headlines: Connections between unresolved hyperlinking issues and a new battle over copyright online. In M. Machill, & M. Beiler (Eds.) *Die wachsende Macht von Suchmaschinen im Internet /The power of search engines* (pp. 202-219). Herbert von Halem.

Key statistics. (2018). Agence France Press annual report. https://annualreport.afp.com/en/

Kim, G., & Bae, J. (2017). A novel approach to forecast promising technology through patent analysis. *Technological Forecasting and Social Change*, 117, 228-237.

Kiss, J. (2010, November 9). My Taptu app takes the visual news aggregator to mobile. Theguardian.com. https://www.theguardian.com/technology/pda/2010/nov/09/taptu-app-aggregator

Klett, F. (2002). Visual communication in web-based learning environments. *Educational Technology & Society*, *5*(4), 38-48.

Kosslyn, J., &Yu, C. (2017, April 7). Fact Check now available in Google Search and News around the world. Google Blog. https://www.blog.google/products/search/fact-check-now-available-google-search-and-news-around-world/

Kramer, S. (2003, September 25). Google News creator watches portal quiet critics with 'best news' webby. Ojr.org. http://www.ojr.org/ojr/kramer/1064449044.php

Kramer, S. (2006a, April 25). AP's Curley, Google's Mayer talk aggregation — carefully; AP-Google Deal? Gigaom.com. https://gigaom.com/2006/04/25/nab-aps-curley-googles-mayer-talk-aggregation-carefully-ap-google-deal/

Kramer, S. (2006b, August 2). Parsing & breaking down the AP-Google agreement. https://gigaom.com/2006/08/02/parsing-breaking-down-the-ap-google-agreement/

Krazit, T. (2010, August 30). Google, AP reach deal for Google News content. Cnet.com. https://www.cnet.com/news/google-ap-reach-deal-for-google-news-content/

Kreiss, D., & McGregor, S. C. (2018). Technology firms shape political communication: The work of Microsoft, Facebook, Twitter, and Google with campaigns during the 2016 US presidential cycle. *Political Communication*, *35*(2), 155-177.

Kruppert, R. (2017). Dispute the patent, short the stock: Empirical analysis of a new hedge fund strategy. *International Review of Law and Economics*, 50, 25-35.

Lasar, M. (2011, February 16). La Taxe Google is back, this time to help French ISPs. Arstechnica.com. https://arstechnica.com/tech-policy/2011/02/la-taxe-google-is-back-this-time-to-help-french-isps/

Lasorsa, D. L., Lewis, S. C., & Holton, A. E. (2012). Normalizing Twitter: Journalism practice in an emerging communication space. *Journalism studies*, *13*(1), 19-36.

Latzer, M., Hollnbuchner, K., Just, N., & Saurwein, F. (2016). The economics of algorithmic selection on the Internet. *Handbook on the Economics of the Internet*, 395-425. http://dx.*doi*.org/10.4337/9780857939852.00028

Lauer, K. (2016, January 5). German publishers have filed complaint against Google: VG Media. Reuters.com. https://www.reuters.com/article/us-google-media-germany/german-publishers-have-filed-complaint-against-google-vg-media-idUSKBN0UJ1KF20160105

Lee, A. M., & Chyi, H. I. (2015). The rise of online news aggregators: Consumption and competition. *International Journal on Media Management*, *17*(1), 3-24.

Lee, J., Kim, J., & Choe, D., (2001). Homepages with Emotions: Quantitative Relations Between Homepage Design Factors and Emotions. Human Computer Interaction Lab, Yonsei University, Korea.

Lenssen, P. (2006, January 21). Google News adds most popular, recommended. Blogoscoped.com. http://blogoscoped.com/archive/2006-01-21-n38.html

Lewis, S. C.; & Usher, N. (2013). Open source and journalism: Toward new frameworks for imagining news innovation. *Media, Culture & Society, 35* (5), 602–619.

Li, K. (2019, January 15). Facebook to invest $300 million to help local news survive. Reuters.com. https://www.reuters.com/article/us-facebook-news/facebook-to-invest-300-million-to-help-local-news-survive-idUSKCN1P9163

Liedtke, M. (2010, February 1). The Associated Press Strikes Deal with Yahoo. editorandpublisher.com. https://www.editorandpublisher.com/news/the-associated-press-strikes-deal-with-yahoo/

Liu, Y. (2012). Rise of China and global Internet governance. *China Media Research*, 8(2).

Lowrey, W. (2011). Institutionalism, news organizations and innovation. *Journalism Studies, 12*(1), 64-79. http://dx.*doi*.org/10.1080/1461670X.2010.511954

Lowrey, W., Parrott, S., & Meade, T. (2011). When blogs become organizations. *Journalism*, *12*(3), 243-259.

Luhmann, N. (1997). Globalization or world society: how to conceive of modern society? *International Review of Sociology*, *7*(1), 67-79. http://dx.*doi*.org/10.1080/03906701.1997.9971223

Ma, O. (2009, June 28). A call to news publishers: How to share your video. Google Blog. https://news.googleblog.com/2009/06/call-to-news-publishers-how-to-share.html

MacKinnon, R., Hickok, E., Bar, A., & Lim, H. (2015). Fostering freedom online: The role of internet intermediaries. *Other Publications from the Center for Global Communication Studies.* https://repository.upenn.edu/cgcs_publications/21

Manber, U. (2008, May 20). Introduction to Google search quality. Official Google Blog. http://googleblog. blogspot.com/2008/05/introduction-to-google-search-quality.html

Manovich, L. (2013). Software takes command. Bloomsbury Academic.

Margolis, M., & Resnick, D. (2000). *Politics as usual* (Vol. 6). Sage Publications.
Masnick, M. (2010, December 21). A Look At Rupert Murdoch's History Of Internet Failures. Techdirt.com. https://www.techdirt.com/articles/20101219/22155912331/look-rupert-murdochs-history-internet-failures.shtml
McCarthy, C. (2006, August 4). Google reveals payment deal with AP. https://www.cnet.com/uk/news/google-reveals-payment-deal-with-ap/
Measures for the Administration of Internet Information Services (2000). State Council of the People's Republic of China. http://www.gd.xinhuanet.com/web/gdga/jian/jian34.htm
Mehta, Y. (2011, August 4). Google News highlights unique content with Editors' Picks. https://news.googleblog.com/2011/08/google-news-highlights-unique-content.html
MEMO/08/761 (2008). Antitrust: Guidance on Commission enforcement priorities in applying Article 82 to exclusionary conduct by dominant firms – frequently asked questions. European Commission. https://ec.europa.eu/commission/presscorner/detail/en/MEMO_08_761
Metz, C. (2010, March 22). Google redirects China to uncensored Hong Kong servers. The Register. https://www.theregister.co.uk/2010/03/22/google_redirects_china_to_hong_kon
Meyer, D. (2013, March 1). German parliament passes 'Google tax' law, forcing royalty payments for news snippets. https://gigaom.com/2013/03/01/german-parliament-passes-google-tax-law-forcing-royalty-payments-for-news-snippets/
Michielsen, P. (2006, November 20). Belgian press association sued Google News. Twobirds.com. https://www.twobirds.com/en/news/articles/2006/belgian-press-association-sued-google-news
Miller, R. (2001). From niche site to news portal. *Online Journalism Review*, *14*, 2001.
Mogee, M. (1991). Using patent data for technology analysis and planning. *Research-Technology Management*, 34, 43–49.
Molyneux, L., & Mourão, R. R. (2019). Political journalists' normalization of twitter: Interaction and new affordances. *Journalism Studies*, *20*(2), 248-266. https://doi.org/10.1080/1461670X.2017.1370978
Monsees, D. (2019, May 1). Introducing auto-delete controls for your Location History and activity data. Google Blog. https://www.blog.google/technology/safety-security/automatically-delete-data/?utm_source=feedburner&utm_medium=feed&utm_campaign=Feed%3A+InsideSearch+%28Search+%7C+Google+Blog%29
Moon, A. (2019, May 19). Exclusive: Google suspends some business with Huawei after Trump blacklist – source. Reuters.com. https://www.reuters.com/article/us-huawei-tech-alphabet-exclusive/exclusive-google-suspends-some-business-with-huawei-after-trump-blacklist-source-idUSKCN1SP0NB
Morehead, J. (2016, May 9). Putting a spotlight on local news sources. https://blog.google/outreach-initiatives/google-news-initiative/putting-spotlight-on-local-news-sources/

Mortaigne, V. (2010, July 1). Zelnik report advocates a "Google tax". Lemonade.fr. https://www.lemonde.fr/technologies/article/2010/01/07/le-rapport-zelnik-prone-une-taxe-google_1288659_651865.html

Mui, Y. (2019, June 10). Congress takes on Big Tech in hearing on anti-competitive behavior among digital giants. Cnbc.com. https://www.cnbc.com/2019/06/10/congressional-hearing-takes-on-antitrust-behavior-among-digital-giants.html

Napoli, P. (2019, October 4). What would Facebook regulation look like? Start with the FCC. Wired.com. https://www.wired.com/story/what-would-facebook-regulation-look-like-start-with-the-fcc/

Napoli, P. M. (2015). Social media and the public interest: Governance of news platforms in the realm of individual and algorithmic gatekeepers. *Telecommunications Policy*, *39*(9), 751-760.

Napoli, P. M. (1998). Evolutionary theories of media institutions and their responses to new technologies. *Communication theory: A reader*, 315-329. Kendall/Hunt.

Napoli, P., & Caplan, R. (2017). Why media companies insist they're not media companies, why they're wrong, and why it matters. *First Monday*, *22*(5). https://doi.org/10.5210/fm.v22i5.7051

Nechushtai, E., & Lewis, S. C. (2019). What kind of news gatekeepers do we want machines to be? Filter bubbles, fragmentation, and the normative dimensions of algorithmic recommendations. *Computers in Human Behavior*, *90*, 298-307.

Newman, N., Fletcher, R., Kalogeropoulos, A., Levy, D. A. L., & Nielsen, R. K. (2018). Reuters Institute digital news report 2018. Digitalnewsreport.org. http://media.digitalnewsreport.org/wp-content/uploads/2018/06/digital-news-report-2018.pdf?x89475

News Media Alliance. (2019). Legislation to protect local news. Newsmediaalliance.org. https://www.newsmediaalliance.org/protect-local-news-video/

Ng, A. (2019, January 23). Google, Facebook and Amazon spending more than ever lobbying Congress. cnet.com. https://www.cnet.com/news/google-facebook-amazon-spending-more-than-ever-lobbying-congress/

Novet, J; & Elias, J. (2019, May 31). Justice Department is reportedly preparing antitrust probe of Google. Cnbc.com. https://www.cnbc.com/2019/05/31/doj-preparing-antitrust-probe-of-google---dow-jones.html

NPC (2016). Cybersecurity Law of the People's Republic of China. The National People's Congress of the People's Republic of China. Npc.gov.cn. http://www.npc.gov.cn/npc/xinwen/2016-11/07/content_2001605.htm

Nystedt, D. (2008, March 17). Update: Google News, YouTube blocked in China amid Tibet riots. Infoworld.com. http://www.infoworld.com/article/2651150/security/update--google-news--youtube-blocked-in-china-amid-tibet-riots.html

Oberbeck, K. (2012, August 21). A law to the detriment of all. https://plus.google.com/+KayOberbeck

Oliver, A. (2017, March 27). What Broadcasting Budget Cuts Means for Local Journalism. duke.edu. http://dewitt.sanford.duke.edu/broadcasting-budget-cuts-means-local-journalism/

Olmedo, p. M., and Caballero, J. L. (2015). A battle without winners, as Spain examines early results of its "Google Tax". Ecija.com. https://ecija.com/sala-de-prensa/a-battle-without-winners-as-spain-examines-early-results-of-its-google-tax-2/

Organize the Web. (n.d.) dmoz.org. https://dmoztools.net/

Owen, L. H. (2017, June 21). News apps are making a comeback. More young Americans are paying for news. 2017 is weird. Nieman Lab. http://www.niemanlab.org/2017/06/news-apps-are-making-a-comeback-more-young-americans-are-paying-for-news-2017-is-weird/

P8_TA-PROV(2019)0231. (2019). Copyright in the Digital Single Market. March 26, 2019. European Parliament. http://www.europarl.europa.eu/doceo/document/TA-8-2019-0231_EN.pdf

Page, L., Brin, S., Motwani, R., & Winograd, T. (1998, January 29). The PageRank citation ranking: Bringing order to the web. Stanford.edu. http://ilpubs.stanford.edu:8090/422/1/1999-66.pdf

Page, L. (2004). 2004 founders' IPO letter. Google.com. https://abc.xyz/investor/founders-letters/2004-ipo-letter/#_ga=2.165626872.610004439.1532311821-929489725.1521479135

Paka, A. (2017, June 27). Redesigning Google News for everyone. Google Blog. https://www.blog.google/outreach-initiatives/google-news-initiative/redesigning-google-news-everyone/

Palmer, D. (2010, March 9). U.S. weighing China Internet censorship case. Reuters.com. https://www.reuters.com/article/us-usa-china-google-wto/u-s-weighing-china-internet-censorship-case-idUSTRE6284YG20100310

Parasie, S, & Dagiral, E. (2013). Data-driven journalism and the public good: "Computer-assisted reporters" and "programmer-journalists" in Chicago. *New Media and Society*, *15*(6), 853–871.

Parfeni, L. (2010, December 1). Google News gets a most shared section. Softpedia.com. https://news.softpedia.com/news/Google-News-Gets-a-Most-Shared-Section-169797.shtml

Paterson, C., & Domingo, D. (2008). *Making online news: The ethnography of new media production*. Peter Lang.

Perez, J. C. (2007, April 6). AFP-Google settlement leaves open questions. Infoworld.com. https://www.infoworld.com/article/2662703/techology-business/afp-google-settlement-leaves-open-questions.html

Pew Research Center. (2019). For local news, Americans embrace digital but still want strong community connection. Journalism.org. https://www.journalism.org/2019/03/26/for-local-news-americans-embrace-digital-but-still-want-strong-community-connection/

Pew Research Center. (2002). One year later: September 11 and the internet. Pewinternet.org. http://www.pewinternet.org/2002/09/05/one-year-later-september-11-and-the-internet/

Pfanner, E. (2013, February 17). French publishers forge deal with Google, breaking ranks with Europe. NYTimes.com. https://www.nytimes.com/2013/02/18/technology/a-first-step-on-continent-for-google-on-use-of-content.html

Pfanner, E. (2011, January 17). A Google worry recedes, for now, as Italy ends investigation into news service. Nytimes.com. https://www.nytimes.com/2011/01/18/technology/18iht-google18.html

Potts, K. (2007). *Web Design and Marketing Solutions for Business Websites*. Apress.

Price, E. (2019, January 21). France fines Google $57 million for GDPR violations. Fortune.com. http://fortune.com/2019/01/21/france-fines-google-57-million-for-gdpr-violations/

Rabenstein, G. (2013, June 21). Google News remains an open platform for all German publishers. Googleblog. https://germany.googleblog.com/2013/06/google-news-bleibt-offene-plattform-fuer-verlage.html

Raj, S. (2019, January 17). Krishna Bharat—The Indian guy who invented Google News. Medium.com. https://medium.com/@sushantraj436/krishna-bharat-the-indian-guy-who-invented-google-news-8b6caf606468

Regalado, A. (2013, November 29). Google's growing patent stockpile. Technologyreview.com. https://www.technologyreview.com/s/521946/googles-growing-patent-stockpile/

Reuters (2017, May 4). Google to pay $334 million to settle Italian tax dispute. Reuters.com. https://www.reuters.com/article/us-google-italy-tax-idUSKBN1801CP

Richmond, S. (2008). How SEO is changing journalism: the going down of the pun. *British Journalism Review*, 19(4), 51-55.

Ricknäs, M. (2008, May 28). Belgian newspapers ask google for $77.5 million in damages. Pcworld.com. https://www.pcworld.com/article/146380/article.html

Rushton, K. (2012, September 26). Rupert Murdoch backs down in war with 'parasite' Google. Telegraph.co.uk. https://www.telegraph.co.uk/finance/newsbysector/mediatechnologyandtelecoms/9566353/Rupert-Murdoch-backs-down-in-war-with-parasite-Google.html

Saltmarsh, M. (2010, March 26). Murdoch finalizes paywall for two British papers. Nytimes.com. https://www.nytimes.com/2010/03/27/business/media/27paper.html

Sandoval, G. (2009, April 7). Wall Street Journal, AP take aim at Google. Cnet.com. https://www.cnet.com/news/wall-street-journal-ap-take-aim-at-google/

Saoji, M. (2014, August 27). Google News & Weather now on the Play store with a new look. Google Blog. https://news.googleblog.com/2014/08/google-news-weather-now-on-play-store.html

Sarno, D. (2009, December 2). Murdoch accuses Google of news 'theft'. Latimes.com. https://www.latimes.com/archives/la-xpm-2009-dec-02-la-fi-news-google2-2009dec02-story.html

Sasseen, J., Olmstead, K., & Mitchell, A. (2013). Digital: By the numbers. The Pew Research Center. http://www.stateofthemedia.org/2013/digital-as-mobile-grows-rapidly-the-pressures-on-news-intensify/digital-by-the-numbers/

Sayer, P. (2009, September, 4). Google faces broad antitrust investigation in Italy. Pcworld.com. https://www.pcworld.com/article/171460/article.html

Schindler, P. (2018, March 20). The Google News Initiative: building a stronger future for news. Google Blog. https://blog.google/outreach-initiatives/google-news-initiative/announcing-google-news-initiative/

Schmidt, E. (2013, February 1), Google creates €60m Digital Publishing Innovation Fund to support transformative French digital publishing initiatives. Google Blog. https://blog.google/topics/journalism-news/google-creates-60m-digital-publishing/

Schonfeld, E. (2010, November 2). The Times UK lost 4 million readers to its paywall experiment. Techcrunch.com. https://techcrunch.com/2010/11/02/times-paywall-4-million-readers/

Schrage, E. (2006). Testimony of Google Inc. http://www.nytimes.com/packages/pdf/business/GoogleStatement.pdf

Schroeder, R., & Kralemann, M. (2005). Journalism ex machina—Google News Germany and its news selection processes. *Journalism studies*, *6*(2), 245-247.

Schudson, M. (1981). Discovering the news: A social history of American newspapers. Basic Books.

Schwartz, B. (2011, July 20). Google directory no longer. Seroudtable.com. https://www.seroundtable.com/google-directory-gone-13731.html

Schwartz, J. (2018, May 18). Murdoch embraces liberal approach to curb Facebook, Google. Politico.com. https://www.politico.com/story/2018/05/18/rupert-murdoch-facebook-google-news-social-media-algorithm-596738

Schweigert, A. (2012, June 14). Towards a Better Definition of Curation in Journalism. Adamschweigert.com. https://adamschweigert.com/towards-a-better-definition-of-curation-in-journalism/

Scolari, C. A. (2018). Media Evolution. In P. M. Napoli (Ed), *Mediated Communication* (pp. 149–168). De Gruyter.

Scott, W. R. (2008). Lords of the dance: Professionals as institutional agents. *Organization Studies*, *29*(2), 219-238.

Seaver, N. (2013). Knowing algorithms. *Media in Transition*, 8 (2013): 1-12.

September 11th Attacks. (n.d.). https://historyday911.weebly.com/largest-attack-on-us-soil.html

Seward, Z. (2009, October 9). What The Associated Press is saying to Google, Microsoft, and Yahoo. Niemanlab.org. http://www.niemanlab.org/2009/10/what-the-associated-press-is-saying-to-google-microsoft-and-yahoo/

Shafer, J. (2009, November 10). Read between the lies. https://slate.com/news-and-politics/2009/11/what-s-behind-rupert-murdoch-s-trash-talking-of-google.html

Shaw, B. (2019, May 28). A challenge to stimulate local news in North America. Google Blog. https://blog.google/outreach-initiatives/google-news-initiative/challenge-stimulate-local-news-north-america/

Shaw, R. (2010). Online newspapers' visual character and perceptions of credibility. Doctoral dissertation. University of Missouri.

Sherman, C. (2000). Humans do it better: Inside the Open Directory Project. Infotoday.com. http://www.infotoday.com/online/OL2000/sherman7.html

Sherr, I. (2018, September 4). Facebook, Twitter and Google have a lot to prove to Congress. Cnet.com. https://www.cnet.com/news/facebook-twitter-and-google-have-a-lot-to-prove-to-congress/

Shiels, M. (2018, September 18). Building Google News for everyone—a retrospective. Google Blog. https://www.blog.google/products/news/building-google-news-everyone-retrospective/

Shoemaker, P. J., & Vos, T. P. (2009). *Gatekeeping theory*. Routledge.

Silcock, B. W., & Keith, S. (2006). Translating the tower of Babel? Issues of definition, language, and culture in converged newsrooms. *Journalism Studies, 7*(4), 610-627. http://dx.*doi*.org/10.1080/14616700600758025

Singer, J. B. (2014). User-generated visibility: Secondary gatekeeping in a shared media space. *New media & society*, *16*(1), 55-73.

Singer, J. B. (2005). The political j-blogger: "Normalizing" a new media form to fit old norms and practices. *Journalism*, 6(2), 173-198. https://doi.org/10.1177/1464884905051009

Singh, J. V., Tucker, D. J., & House, R. J. (1986). Organizational legitimacy and the liability of newness. *Administrative science quarterly*, 31(2), 171-193.

Slashdot.org (2001). World Trade Towers and Pentagon attacked. Slashdot.org. https://slashdot.org/story/01/09/11/1314258/world-trade-towers-and-pentagon-attacked

Slocum, M. (2010, January 6). What qualifies as a Spotlight story on Google News? Here's a few clues. Niemanlab.org. http://www.niemanlab.org/2010/01/what-qualifies-as-a-spotlight-story-on-google-news-heres-a-few-clues/

Smith, B. (2018, October 8). Project Strobe: Protecting your data, improving our third-party APIs, and sunsetting consumer Google+. Google Blog. https://www.blog.google/technology/safety-security/project-strobe/

Smith, G. (2007, March 13). Copiepresse v Google - the Belgian judgment dissected. Twobirds.com. http://www.twobirds.com/en/news/articles/2007/copiepresse-v-google

Smydra, D. (2011, September 24). Recognizing publishers' standout content in Google News. Google Blog. https://news.googleblog.com/2011/09/recognizing-publishers-standout-content.html

Sterling, G. (2019, January 15). EU copyright directive nearing final form as Google tests stripped-down news SERPs. Searchengineland.com. https://searchengineland.com/eu-copyright-directive-nearing-final-form-as-google-tests-stripped-down-news-serps-310494

Sterling, G. (2014, December 16). Google News Still Available In Spain Despite Closure Deadline. Searchengineland.com. https://searchengineland.com/google-news-still-available-spain-despite-closure-deadline-210990

Sterling, G. (2012, August 22). Germany wants to force Google to pay license fees for links. Searchengineland.com. https://searchengineland.com/germany-wants-to-force-google-to-pay-license-fees-for-links-131288

Stinchcombe, A. L. (1987). *Constructing social theories*. University of Chicago Press.

Stier, L. (2018, September 10). Making Google News work on every phone, everywhere. Google Blog. https://www.blog.google/products/news/making-google-news-work-every-phone-everywhere/

Stolt, K. (2010, June 30). Extra! Extra! Google News redesigned to be more customizable and shareable. Google Blog. https://googleblog.blogspot.com/2010/06/extra-extra-google-news-redesigned-to.html

Sullivan, D. (2014, September 26). The Yahoo Directory—once the internet's most important search engine—is to close. Searchengineland.com. https://searchengineland.com/yahoo-directory-close-204370

Sullivan, D. (2010, January 8). Where's AP In Google News? Apparently in limbo, as Contract Running Out. Searchengineland.com. https://searchengineland.com/wheres-ap-in-google-news-33164?utm_source=feedburner&utm_medium=feed&utm_campaign=Feed:+searchengineland+(Search+Engine+Land)&utm_content=Twitter

Sullivan, D. (2007, April 6). AFP & Google Settle Over Google News Copyright Case. Searchengineland.com. https://searchengineland.com/afp-google-settle-over-google-news-copyright-case-10926

Sundar, S. S., & Marathe, S. S. (2010). Personalization versus customization: The importance of agency, privacy, and power usage. *Human Communication Research,* 36(3), 298-322.

Swisher, K. (2009, November 24). What's really behind the rupe-a-dope with Google and Microsoft? Here are five possibilities! Kara.allthingsd.com. http://kara.allthingsd.com/20091124/whats-really-behind-the-rupe-a-dope-with-google-and-microsoft-here-are-five-possibilities/

Swisher, K. (2009b, December 23). Project Alesia: News Corp.'s Roman battle cry–Does that cast Googlers as the Gauls? Alltingsd.com. http://allthingsd.com/20091223/project-alesia-news-corp-s-roman-battle-cry-does-that-cast-googlers-as-the-gauls/

Tandoc, E. C., & Vos, T. P. (2016), The journalist is marketing the news: Social media in the gatekeeping process. *Journalism Practice*, *10*(8), 950-966. https://doi.org/10.1080/17512786.2015.1087811

Tartakoff, J. (2010, January 8). Google News stops hosting new AP content. Gigaom.com. https://gigaom.com/2010/01/08/419-google-stops-hosting-new-ap-content/

Taylor, C. P. (2001, September 17). Web hits rise. Adage.com. https://adage.com/print/53814

Ted Olson on loss and love in the decade since 9/11. (2011, September 11). Washingtonpost.com. https://www.washingtonpost.com/blogs/reliable-source/post/ted-olson-on-loss-and-love-in-the-decade-since-911/2011/09/09/gIQAlqKDLK_blog.html?noredirect=on&utm_term=.e57352cb4577

Tobitt, C. (2018, July 6). Times and Sunday Times hit 500,000 subscribers as digital outnumbers print for first time. Pressgazette.co.uk. https://www.pressgazette.co.uk/times-and-sunday-times-hits-500000-subscribers-as-digital-outnumbers-print-for-first-time/

Toor, A. (2016, October 14). Google News launches fact check feature ahead of Election Day. Theverge.com. http://www.theverge.com/2016/10/14/13283200/google-news-fact-check-election

Tracy, M. (2019, June 9). Google made $4.7 billion from the news industry in 2018, Study says. Nytimes.com. https://www.nytimes.com/2019/06/09/business/media/google-news-industry-antitrust.html

Training (n.d.) Google News Initiative. https://newsinitiative.withgoogle.com/programs/

Tuch, A. N., Presslaber, E. E., StöCklin, M., Opwis, K., & Bargas-Avila, J. A. (2012). The role of visual complexity and prototypicality regarding first impression of

websites: Working towards understanding aesthetic judgments. *International journal of human-computer studies*, *70*(11), 794-811.

USPTO. United States Patent and Trademark Office. https://www.uspto.gov/patents-getting-started/general-information-concerning-patents

Upstill, T. (2018, May 8). The new Google News: AI meets human intelligence. Google Blog. https://www.blog.google/products/news/new-google-news-ai-meets-human-intelligence/

Van Couvering, E. (2011). Navigational media: The political economy of online traffic. *The Political Economics of Media: The Transformation of the Global Media Industries*. Bloomsbury Academic.

Varadarajan, B., Vijayanarasimhan, S., Shetty, S., Kothari, N. D., & Rizzolo, N. D. (2016). U.S. Patent No. 9,384,242 B1. Washington, D.C.: U.S. Patent and Trademark Office.

Verney, A. (2015, April 28). Let's work together to support quality journalism. Google Europe Blog. https://europe.googleblog.com/2015/04/lets-work-together-to-support-quality.html

Vobič, I. (2007). The normalization of the blog in journalism: online newspapers of Slovene traditional media. *Medijska istraživanja*, *13*(2), 59-83.

Vos, T., & Heinderyckx, F. (2015). Revisiting gatekeeping theory during a time of transition. In T. Vos, & F. Heinderyckx (Eds.). *Gatekeeping in transition* (pp. 3024). Routledge.

Vos, T., &, Russell, F. M. (2019). Theorizing journalism's institutional relationships: An elaboration of gatekeeping theory. Journalism Studies, *20*(16), 2331–2348. https://doi.org/10.1080/1461670X.2019.1593882

Wang, Q. (2018c). Mechanisms and implications of news personalization: A case study of Google News. Association for the Advancement of Artificial Intelligence. https://apen18.hans-bredow-institut.de/program/

Wang, Q. (2018b). Media differentiation between Chinese state media and commercial online media: A multi-stakeholder media system in the digital China. *International Communication Research*, *53*(1), 2-24.

Wang, Q. (2018a). Dimensional field theory: The adoption of audience metrics in the journalistic field and cross-field influences. *Digital Journalism*, *6*(4), 472-491. https://doi.org/10.1080/21670811.2017.1397526

Wang, Q. (2017). Participatory journalism in the Chinese context: Understanding journalism as process in China's participatory culture. *Journalism: Theory, Practice and Criticism. 18*(4), 501–517. https://doi.org/10.1177/1464884916636177

Wang, S. (2018, March 20). Google announces a $300M "Google News Initiative." Niemanlab.org. http://www.niemanlab.org/2018/03/google-announces-a-300m-google-news-initiative-though-this-isnt-about-giving-out-grants-directly-to-newsrooms-like-it-does-in-europe/

Watanabe, K. (2013). The western perspective in Yahoo! News and Google News: Quantitative analysis of geographic coverage of online news. *International Communication Gazette, 75*(2), 141-156.

Weaver, D. A., & Bimber, B. (2008). Finding news stories: a comparison of searches using LexisNexis and Google News. *Journalism & Mass Communication Quarterly*, *85*(3), 515-530.

Weber, M. S. (2014, April). Observing the web by understanding the past: archival internet research. In *Proceedings of the 23rd International Conference on World Wide Web* (pp. 1031-1036). ACM.

Weber, M. S., & Monge, P. R. (2014). Industries in turmoil: Driving transformation during periods of disruption. *Communication Research*, *44*(2), 147-176.

Weigle, E. (2010, November 16). Credit where credit is due. Google Blog. https://news.googleblog.com/2010/11/credit-where-credit-is-due.html

Welch, C. (2013, February 1). Google ends standoff with French publishers, commits to €60 million 'innovation' fund. Theverge.com. https://www.theverge.com/2013/2/1/3941854/google-ends-standoff-french-news-sites-digital-publishing-innovation-fund

Whetstone, R. (2006). About the Google News case in Belgium. Google Blog. https://googleblog.blogspot.com/2006/09/about-google-news-case-in-belgium.html

White, D. M. (1950). The "gate keeper": A case study in the selection of news. *Journalism Bulletin*, *27*(4), 383-390.

White, H. L. (1987). A Technological Model of Global History. *The History Teacher*, *20*(4), 497-517.

Wiggins, R. (2001). The Effects of September 11 on the Leading Search Engine. *First Monday*, 7(10). Firstmonday.org. http://firstmonday.org/issues/issue6_10/wiggins/index.html

Willsher, K. (2019, October 17). France accuses Google of flouting EU copyright law meant to help news publishers. Latimes.com. https://www.latimes.com/business/story/2019-10-17/france-accuses-google-ignoring-copyright-law

WIPO. (1996). Summary of the WIPO Copyright Treaty (WCT) (1996). WIPO. https://www.wipo.int/treaties/en/ip/wct/summary_wct.html

WIPO. (2008). WIPO intellectual property handbook: Policy, law and use. Chapter 2: Fields of intellectual property protection. WIPO 2008. P17-160.

Wired (2004, September 25). Google bows to Chinese censorship. Wired.com. https://www.wired.com/2004/09/google-bows-to-chinese-censorship/

Wohlsen, M. (2014, December 16). Spain's Google news shutdown is a silly victory for publishers. Wired.com. https://www.wired.com/2014/12/google-news-shutdown-spain-empty-victory-publishers/

Wojazer, P., & Vinocur, N. (2013, February 1). Google to pay 60 million euros into French media fund. Reuters.com. https://www.reuters.com/article/us-france-google/google-to-pay-60-million-euros-into-french-media-fund-idUSBRE91011Z20130201

Wong, J. C. (2018, April 10). Congress grills Facebook CEO over data misuse – as it happened. Theguardian.com. https://www.theguardian.com/technology/live/2018/apr/10/mark-zuckerberg-testimony-live-congress-facebook-cambridge-analytica

Xinhua (2010, March 11). NPC deputy: International company should comply with local laws. Sina.com.cn. http://tech.sina.com.cn/i/2010-03-11/17553931626.shtml

Zara, C. (2012). Rupert Murdoch vs. Google -- Google wins. Ibtimes.com. http://www.ibtimes.com/rupert-murdoch-vs-google-google-wins-796615

Zettl, H. (2013). *Sight, sound, motion: Applied media aesthetics*. Cengage Learning.

Zhang, L. (2007, August 21). Would you like video with that? Google Blog. https://news.googleblog.com/2007/08/would-you-like-video-with-that.html

Zheng, Z. (2010, March 9). Google incident does not break WTO rules. Chinadaily.com.cn. http://www.chinadaily.com.cn/china/2010-03/09/content_9562138.htm

Zhou, L. (2019, April 9). Republicans and Democrats have completely different priorities on tech. Vox.com. https://www.vox.com/2019/4/9/18300659/facebook-google-congress-hearing-white-nationalism-censorship