The Thesis Committee for Trevor Hollis Diehl
Certifies that this is the approved version of the following thesis:

Toward a Tabloid Press: The Impact of News Aggregation on Content in 12 US News Websites

APPROVED BY
SUPERVISING COMMITTEE:

Supervisor:

Regina G. Lawrence

Homero Gil de Zúñiga.

Toward a Tabloid Press: The Impact of News Aggregation on Content in 12 US News Websites

by

Trevor Hollis Diehl, B.A.

Thesis

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Master of Arts

The University of Texas at Austin

May, 2014

Dedication

This project is dedicated to Jane Vance, for her daily encouragement.

Acknowledgements

This study draws heavily on recent newsroom ethnographies. I don't know these authors, but I would not have been able to validate most of my arguments without their insights. These authors include, but are not limited to C. W. Anderson, David Domingo and Pablo Boczkowski.

I would like to thank my thesis readers, Regina Lawrence and Homero Gil de Zúñiga. Dr. Lawrence's Socratic Method fed a valuable cycle of examination, frustration and revelation. Dr. Gil de Zúñiga was, as usual, a constant source of inspiration. Special thanks to David Ryfe at the University of Nevada, Reno, who took the time to push me on a path toward media scholarship. Finally, I owe much in this venture to Dr. Paula Poindexter. Her dedication to the development of her students is second to none.

Abstract

Toward a Tabloid Press: The Impact of News Aggregation on Content in 12 US News Websites

Trevor Hollis Diehl, M.A.

The University of Texas at Austin, 2014

Supervisor: Regina G. Lawrence

News aggregation is a developing form of professional journalism practice, one uniquely adapted to contemporary communication realities. News companies have always gathered content from a variety of sources when producing their products. However, the sheer volume of information, number of participants and speed of consumption online requires news workers to adopt new routines of collecting and disseminating information. These routines, some argue, fundamentally differ from the beat structure of traditional journalism. As recent ethnographic work has found, online news workers might value a sense of audience and newsworthiness over and above norms like objectivity and getting a good story (Anderson, 2013; Agarwal & Barthel, 2013). As economic pressures continue to strain resources and shrink the number of reporters on staff, news aggregation, both as a practice and a digital filtering tool, is becoming a staple of modern newsrooms.

Few researchers have explored the impact of these divergent routines on content. Through a secondary data analysis of the Pew Research Center's 2012 News Coverage Index, this thesis examines the topics and news-drivers in 12 US news websites. The analysis finds that in-house, so-called "original reporting" tends to rely on institutional actors and hard news topics. When stories are aggregated from a third-party source, soft news topics and celebrity stories are preferred. Finally, different professional practices seem to be favored depending on the type of online news organization. The findings suggest scholars, and those interested in journalism education, think of organizational pressures and professional norms as fluid online, particularly when connecting theories of news work to output in terms of content.

Table of Contents

Introduction

News organizations have always gathered, selected and delivered information. According to this definition of journalism, the news online is not much different than it was in print 100 years ago. On the other hand, the nature and sheer volume of information available on the web alters what Tuchman (1973) called the "raw material of news." In response to this environment, some journalists and news organizations have embraced content aggregation, a developing, and often contentious, practice of news gathering (Anderson, 2013; Boczcowski, 2010). Appropriating third-party content is not new to news work. It is a response to long-standing challenges inherent in the profession: infinite raw material and finite resources. However, the unique demand of making the news online requires a shift in traditional approaches to reporting (Agarwal & Barthel, 2013).

Historically, journalists developed routines to "feed the beast" of daily content demands. These routines produce predictable patterns in news output, namely a reliance on public officials as sources (Cook, 1998; Fishman, 1980; Sigal, 1973; Tuchman, 1973). Since the 1980's, scholars have also noted a decline in public affairs programming and an emphasis on "soft" topics in newspapers and television (Baum, 2002, 2007; Hamilton, 2004; Patterson, 2000). Yet very little scholarship has addressed the impact of information gathering routines online and the reliance on official sources or soft news topics. Employing the Pew Research Center's 2012 News Coverage Index, this study examines the implications of news aggregation as routinized professional practice within 3,102 online stories. The following sections address connections between routines, organizations and news outputs. Second, this study takes a closer look at the various definitions and applications of aggregation as news work. Finally, through a secondary

data analysis, this study explores sourcing patterns and topic selection across 12 major online news websites. The findings suggest that third-party content appropriation leads to less reliance on officials and a preference for soft news and celebrity-driven stories.

Literature Review

Organizational Demands, News Routines and Technology

News routines, sometimes called rituals, are the habits journalists have adopted to supply the daily demand for cheap, timely information. As Shoemaker and Reese define them, routines "are those patterned, routinized, repeated practices and forms that media workers use to do their jobs" (1996, p. 100). It should be no surprise then, that the practices developed for creating news content is a key determinant of what is delivered to the public as news. As the aforementioned authors put it, "routines directly affect the media content that reaches the audience" (p. 100). Routine news decisions are not made in a social vacuum however. They are a product of a complex interaction between the individual reporter and the greater social context in which they operate.

Several theoretical concepts have been conceived by scholars in attempt to understand the balance between individual reporter's choices and the broad social environment. Theories range from political economy of the news (Herman & Chomsky, 1998; McChesney, 2008), to institutional models (Cook, 1998) and organizational and sociological models (Schudson, 1989; Shoemaker & Reese, 1996). In quoting Paul Hirsch, Shoemaker and Reese (1996) advocate for an organizational model because, they argue, similarities among organizational structures outstrip their individual differences. There are "stable, patterned sets of expectations and constraints that are common to most media organizations" regardless of their individual genre (p. 102). A daily newspaper has as much in common with an entertainment outlet in this view. Organizations face similar pressures, and therefore adapt similar routines of content creation. These

"expectations and constraints" limit the type of behavior, and in turn, type of content a media outlet can produce.[1]

The organization restrains media routines according to three considerations: the audience, the demands of the organization itself, and the availability of sources. First, news judgment must have some consideration for what appeals to an audience. Second, organizational constraints are best understood in terms of economics. News companies have finite resources to report on possibly infinite raw "news material." Companies also have to turn a profit, so they must acquire content at a low cost. Third, reporters need an efficient, cost effective source. Sources in the Shoemaker & Reese model are often public officials, who can be relied upon on a daily basis.[2] Developing efficient daily work routines is one way to accomplish these goals (Shoemaker & Reese, 1996). News content primarily becomes a product of routines. For example, Keith (2011) writes that television journalists often move between media companies with little discomfort because they rely so heavily on their news routines. In this example, content is the product of professionalized habits honed to organizational demands that transcend the individual companies.

Individual reporters are bound by these constraints, but still exercise some personal control. In another example, relying on the wire services is more efficient than positioning reporters on a live beat. The editor may make personal decisions about what to pull from the wire each day, but ultimately her story options are limited (Shoemaker & Reese, 1996. P. 114).

Though a helpful heuristic for connecting routines to content, the above model may be accused of being too general to account for variations in news genre. Nor does it

[1] For a full discussion of the model see, Reese, 2001.
[2] Sourcing patterns are addressed below.

account for changes in technology. For example, television and newspaper content may be a product of routine, but it is not clear how the context in which those products are created might impact content. Keith (2011) refers to the context of news production as the "media milieu." In this sense, Keith claims that "milieu" might stand for media organizations in transition. Transition periods allow for the individual reporter to impact routines more than they would in more stable systems. Over time, the new routines borne of transition will cement.

The transitioning organizations approach, in Keith's account[3], still fails to explain the nuances between news genres. It may be more useful to think of social contexts, or "milieu" in Keith's vernacular, in terms of technology. Television and newspapers produce different products because they employ different tools of the trade. Those tools of technology impact content to the extent they influence what counts as news on the one hand, and how it is routinized on the other. In her classic ethnography in newspaper and television environments, Tuchman (1973, p.123) succinctly connects technology to content:

> The degree to which resources must be reallocated to meet practical exigencies and the way reallocation is accomplished depends upon both the event being processed and the medium processing it. The technology used by a specific medium does more than "merely" influence the ways in which resources are allocated. It influences the typification of event-as-news or how that news story is perceived and classified.

This passage is quoted at length because it captures the influence of technology on content. Tuchman's contribution to contemporary media scholarship is unique here. Three conclusions are drawn from Tuchman for the purposes of this study: news output is contingent upon the *type* of event being covered, the *resources* available within the

[3] Keith draws on the classic work of Kurt Lewin, see Lewin, K. (1951). Field theory in social science.

organization and the *technology* employed.[4] In this approach, news organizations allocate resources to cast a "news net" (in the form of beat reporters) to cover news events (Tuchman, 1978).

Tuchman emphasizes organizational and technological constraints in informing the decisions of individual reporters. However, the basic problem of control over infinite "raw news material" and the demands of creating a news product for an audience extend beyond individual new decisions. News media products are created in a multi-dimensional, complex society with several competing pressures and conditions. Furthermore, the economic and technological change of the past two decades undermines a definitive account of the news making process. Bennett (2004) offers a four-part typology for a more complete understanding of underling factors that influence news content: the reporter's news sense, the organizational routines, the economic constraints and the technology used for gathering and transmitting information. Bennett offers six factors that influence news construction under each categorical type. The full exposition and application of this model is beyond the scope of this paper. The key for this study is to note contemporary changes in the composition of the model.

Bennett notes that for much of twentieth century, a reporter-driven, organizationally bound hybrid order dominated. The result was news content determined by news beats centered on officials, notions of objectivity and the role of reporter as watchdog.[5] Since the 1980s, the "reporter-organization" influences on news construction have been upset by technological advancement, increased competition, and the

---

[4] Tuchman refers to technology as the news medium. Tuchman emphasizes the impact of technology on the size of an organization, labor inputs to create a story and how stories break (p. 119).

[5] For more, see Bennett, (2004). Gatekeeping and press-government relations: A multigated model of news construction. *Handbook of political communication research*, 283-314.

encroachment of business and marketing concerns on news programming. The result is what Bennett calls the rise of the "entertainment-technology hybrid order." Bennett then poses the question: Does the rise of a new order represent a sea change in gate-keeping (or news production practices), or are we witnessing a hybrid model where traditional news making is simply supplemented, but not replaced? (p. 302). The question might inform our investigation of online news routines as antecedent to news outputs. But first, the answer requires a more thorough discussion of current features of the news environment.

This paper will argue that the organizational approach to media study might be improved with a more thorough account of developing news practices, enabled by technology. The same is called for in consideration of organizations themselves. Tuchman's (1973) observation about the differences between print and television newsrooms was of little consequence in an age dominated by a few, consistent media organization types. Accordingly, Shoemaker and Reese could note that similarities among organizational structures outstrip their individual differences. The contemporary media environment calls that assumption into question. News organizations have asymmetrically adapted to online pressures, both in terms of practice and the content they produce.

The Hyper-Real News Beat

For much of the twentieth century, media organizations were products of what Castells (2009) calls the Fordist Industrial Age. As Shoemaker and Reese (1996) point out, newsroom routines mimic a "news factory" complete with division of labor and an "assembly line" (p.103). It stands to reason that newsrooms resemble the greater economic paradigms they were created in. In the news factory, individual agency is often

absorbed by the industrialized process. Network communication technologies, primarily the internet, alter the roles of the individual in information production. As newsrooms move online they must alter their routines to fit these new circumstances. The following explores the challenges to news work online.

A key attribute of the industrial age was the high cost of knowledge production (Benkler, 2006). This raises barriers to information production. In turn, industrial era organizations are able to monopolize content creation. Network communication technologies, like the internet, fundamentally challenge industrial organizational structures by lowering barriers to production. Contemporary communication technologies allow users to create their own content and distribute it through public networks. O'Reilly (2005) defines the digital media environment as "a continually-updated service that gets better the more people use it, consuming and remixing data from multiple sources, including individual users….[through] an 'architecture of participation.'" The sheer number of participants and the ease with which they can create content has led to a condition of what one scholar calls "information abundance" (Bimber, 2003).

The new reality of information abundance is an accelerated version of an old problem. As Tuchman (1973) pointed out two decades before the invention of the *World Wide Web*, journalists must solve the problem of finite resources and infinite "raw news material." During the latter half of the previous century, the raw news material was grounded in the physical world. The routines developed in response to the variations in possible events in that world (raw news material) were an attempt by journalists to exercise some control over the environment (Tuchman, 1973). The internet adds yet another layer of complexity, forcing journalists to adapt new means of control over the information environment. Two features of the new media environment are important for this discussion: abundance and hyper-reality.

Abundance refers to the "diverse, extensive and fluid nature" of the content and form of new media messages (Williams & Carpini, 2011). The media content online is as diverse as the individuals who produce it. Contemporary newsrooms must negotiate an information environment that blurs long-held distinctions between "type of media and genre, producers and consumers, mass and interpersonal communication, and public affairs and popular culture" (p. 85). Using *Yahoo News* as an example, Williams & Carpini show how one major news website departs from industrial era publications. The news website includes a wide, unstructured range of "topics, sources and points of view." The content of these websites depends in large part on new types of reporting tools (p.85-86).

Hyper-reality is the second feature of the new media environment that journalists must grapple with. Citing Dylan and Katz, Williams and Carpini (2011) define hyper-reality as a place where there is "no clear distinction between a 'real' event and its mediated representation" (p. 140). This environment challenges the epistemic nature of knowledge production in society; notions of the "true" reality are based, at least in part, by the mediated constructions of events and characters in the 'real' world. The internet itself is, in certain respects, a mass web of mediated constructions of reality.

Abundance and hyper-reality have implications for online journalism for several reasons. First, online reporting challenges news workers to wade through and filter an abundance of mediated messages on the internet. The "raw news material" Tuchman (1973) referred to is now of two stripes: the analog 'real' world and the digital, hyper-real world. This distinction has led to an internal split in the profession of journalism. Traditional, so-called "original reporters" lay claim to certain reporting practices based in the analog world. News aggregators on the other hand, tend to operate in the hyper-real, and often treat the internet itself as raw material for news (Anderson, 2013). Second, as

more of social life takes place in hyper-real environments, it becomes incumbent upon reporters to make editorial decisions about what counts as news. Yet many have found that journalistic routines of the analog, industrial era are not equipped to manage the features of the new media environment (Brannon, 2008; Klinenberg, 2005; Ryfe, 2012).

News Routines Online

Scholarship concerned with online news and professional practice is diverse.[6] The overall condition, as Deuze and Marjoribanks (2009) put it, is that journalists in today's newsrooms "are expected to do more with less time, fewer resources and fewer colleagues" (p.555). In Mitchelstein and Boczkowski's (2009) summary of online scholarship, they note that technological appropriation and speed are altering information gathering and distribution routines.

On the whole, journalists seemed to have been eager to adopt the internet as a research and reporting tool, but reluctant to abandon traditional conventions of news value (O'Sullivan & Heinonen, 2008). As O'Sullivan and Heinonen (2008) conclude in their survey of 239 journalists, "newspaper journalists want to stay newspaper journalists" (p. 368). Pavlik (2000), also drawing on survey data, notes that when reporters incorporate more multimedia content online, the finished product often eschews the inverted pyramid style narrative. This may create tension in the newsroom. For example, Cawley's (2008) ethnography of a transitioning Irish newspaper revealed an inherent tension in the values internet technologies offer and those of traditional reporting. Online reporters were "trying to escape the company's print traditions (in immediacy of publication and experiments with audio and video) while also being bound to them (in text…)" (p. 51). That tension often results in news managers reining-in web

[6] For a full review see Lewis, 2012 and also Mitchelstein & Boczkowski, 2009.

practices to more closely mimic print products. This was the case in Argentina, where Garcia (2008) saw the development of new routines online actively curtailed by the print news editors. In other cases, technology was considered a barrier to quality content production, because reporters were unable to develop routines adequate for efficient online reporting (Brannon, 2008; Ryfe, 2012).

The shift in the speed of online content production also places pressures on news routines. For Boczkowski (2004), the pace of information production online requires digital reporters to adopt speed as a professional norm. In turn, news organizations must re-align "editorial strategies; work processes and production and distribution technologies to cater to this dominant temporal and special patterning of news consumption" (p. 3). Quandt (2008) echoed these findings in his ethnography of five German newsrooms. Pressures to get the story out quicker were the only notable differences in professional routines. Overall, he summarized, the promise of the individual actor transcending organizational constraints (as Keith predicted in her critique of the Shoemaker and Reese model) wasn't observed. Instead, he suggested that the norms of print media still dominate online news.

The key take-away of the recent work in online newsrooms for the purposes of this study is best articulated in Domingo's insight: "each context (each company, in this case) develops concrete strategies, definitions, tools, routines and roles that can only be explained by deep analysis of the actors and material conditions in each environment" (2008, p. 124-125). Domingo's call for more ethnographic work could also be interpreted as a signal that organizations have not uniformly adapted to their online environments. There does seem to be some consensus that print organizations are less open to adapt to new technologies, either through editorial oversight or an inability to adapt efficient routines. Domingo also implies that news online is shaped, at least in part, by variations

in the application of reporting tools employed. Following Truchman (1973) we also see that different types of organizations appropriate resources differently according to the news medium. Finally, echoing Quandt (2008b), the parent company often determines what makes it to the website. It stands to reason then, that the parent news organization and the medium shape, at least in part, the means of content collection. Few empirical studies have explored this relationship. One overarching research question this paper addresses is *how is professional practice different across various news organization types?*

DEFINING A NEW AREA OF NEWS WORK

One attribute of the networked information ecology is the ability to transfer and recycle content with ease. This attribute is best illustrated by an internet culture maxim known as "the 1% rule." In many online contexts, the saying goes, roughly 1% of the population actually create content, 10% interact through comments or sharing, and 89% simply watch from the sidelines (Arthur, 2006). Early studies of news practice online referred to recycling the content of others as "shoveling." Journalists tended to look down on the practice and see it as a threat to content quality and journalistic integrity (O'Sullivan & Heinonen, 2008). One incarnation of this practice that has gained traction in reporting online is the news aggregator.

At first glance, news aggregation is the process of collecting content from around the web and displaying it on one website (Isbell, 2010). In the field of news media and journalism scholarship, aggregation may be used in at least three related, but separate ways. It may refer to computer applications built to gather content, the news work of web producers who bundle content from other sources, or a web-based media company that

relies on the content of others. All three definitions incorporate the notion of collecting material from third-party sources to create a news product.

News aggregators as computer applications filter and disseminate information from around the web and post it on one web page. In a sense, these tools are the digital versions of a news beat. Instead of relying on "shoe leather" and face-face interactions with humans, the programs rely on the zeros and ones of digital programming languages to gather information on the web. These applications, also called a "news feed," require little human intervention. They also cost much less than human employees (Grueskin, Seave & Graves, 2011).

Aggregation is also at the center of a shift in the so-called "jurisdictional" definition of what it means to be a journalist (Anderson, 2013b; Lewis, 2012). In his ethnography work of Philadelphia online metro papers, Anderson (2013b) witnesses a division of labor in the newsroom unique to the demands of the internet. A new type of news worker, separate from the reporters and editors, coordinates content on the web. Anderson calls these producers "second-level news workers" (p.70). Their primary task is to link, bundle, rank and post content according to "rapidly changing sense of their importance, popularity and newsworthiness" (p. 70).

Second-level News Processes

Anderson's insights relate to this discussion in two important ways. First, his conception of "second-level" news processes can be used to understand how the news is made online. On a basic level, the work of the aggregator isn't much different than traditional news making. News companies have always synthesized material from editorial staff, reporters, wire, and freelancers (Grueskin, Seave & Graves, 2011). Anderson makes a distinction between bundling a news product for an audience on one

hand, which is nothing new for news companies, and turning news facts into stories on the other. The latter, what he calls first-level news processes, refers to the "traditional" or "original" reporting practices of professional journalists.

In networked environments, the second-level workflow is vastly different from previous eras. As Agarwal & Barthel (2013) put it, aggregation requires constant monitoring of the internet, reliance on news feeds, extreme speed and the ability to appropriate the content of others. These skills are distinct from first-level processes because they require synthesizing the content of others in a hyper-real, information abundant environment. As one news executive explains, "What we do as aggregators isn't about journalism…it's about making sense of the internet" (Anderson, 2013, p. 1018).

Second, these second-level media workers have an elevated sense of their audience. Instead of "getting it right" or writing a complete story, aggregators favor an internal notion of newsworthiness. As a reporter, one former *New York Post* web editor put it, "You need to know how to develop and cultivate sources…As an aggregator, you're number one, going to need to know how to have a sense of story" (p. 1017). Other scholars have also noted the importance of knowledge of the audience in web production work (Agarwal & Barthel 2013).

Web producers operate in a space distinct from those in previous eras. Anderson (2013b) notes that these workers often exists in a quasi-institutionalized role; they embody some sense of journalistic norms, but at the same time are not bound by traditional institutional ways of thinking. Agarwal and Barthel (2013) found that this can lead to ambivalent attitudes toward objectivity. As noted above, these news workers favor audience over intense reporting, speed and fairness over "getting it right" and a personal sense of responsibility over institutional verification.

Aggregation Websites

A third definition of aggregation is used to separate "traditional" news company websites from those that rely primarily on the content of others. Pew Research Center (2006; 2014) commonly uses the term to refer to websites like *Google News, Yahoo News* and *The Huffington Post*. Though often type cast as bundlers of content, the differences between websites in this genre highlight the complexity of the news landscape online. For example, *Google News* relies solely on news feeds, and does not create any original content. Stories are filtered according to criteria set by algorithms, like how recently the story was released or how popular it becomes. News filtering at Google most resembles the first definition of aggregation, a computer application. It is also an inexpensive means to disseminate information because it does not require a staff of news workers. (Grueskin, Seave & Graves, 2011).

According to Hindman (2012), *Yahoo News,* like *Google,* depends heavily on algorithms. Built as applications that select content based on the interest of the individual news consumers, Hindman argues that these so-called content recommender systems have important consequences for the future of news online. In particular, the need for advertising dollars and web traffic will replace human editorial decisions with those of the aggregator program. These tools, Hindman finds, favor the largest news companies with the most resources, concentrate audiences and hurt smaller news operations.

Other news organizations in this category, including *The Huffington Post* and *Yahoo News,* have dedicated content editors responsible for aggregating the news. However, there is variation between these two in terms of source material. *The Huffington Post* employs 187 full-time editorial staff members, and claims to have a total of 575 domestic and international editorial positions (Pew, 2014).[7] According to

[7] It should be curious to note that the media organization did not refer to their staff as journalists.

Anderson (2013), content editors at organizations like *The Huffington Post* rely heavily on news feeders and web monitoring, and spend much of their time creating headlines. *Yahoo News* in contrast, reportedly has a small number of news editors (Grueskin, Seave & Graves, 2011). As of 2011, *ABC* and *Yahoo News* have a partnership of collaboration, where *Yahoo News* will draw from *ABC's* television stories for content.

Aggregation is a cheap, efficient way to filter information on the internet. It enables reporters to routinize the collection and dissemination of a vast range of content on the web. Though collecting material is nothing new for news organizations, doing so on the web introduces new types of workers into the newsroom. Often, these workers are not as bound to traditional conceptions of journalism practice. There is also diversity in terms of the degree to which aggregation is used, both as a computer application and as a form of news work. Theses nuances in organizational adoption of aggregation have yet to be explored. Based on the tendency for print journalists to refrain from drawing explicitly on third-party sources (O'Sullivan & Heinonen, 2008) and the tendency for aggregator websites to rely on the content of others, the following hypothesis are proposed:

H1 Print-based websites will rely less on third-party content

H2 News aggregator websites will favor third-party content

One attribute of news aggregation outlined above, in all forms, is speed and brevity. The outcome tends to be shorter stories. In the case of aggregation as a computer application, rarely more than a headline and a few sentences are posted (Grueskin, Seave & Graves, 2011; Isbell, 2010). In the case of human news workers, Agarwal and Barthel (2013) found that many web producers summarize several stories a day, and often write less than 300 words per story. They also report that some websites feature a mix of long-

form stories, produced by "traditional" reporting, and shorter, aggregated stories. Based on these assumptions, the following hypothesis is proposed:

H3 News aggregator websites will feature shorter stories

It is also unclear if changing habits (and changing conceptions of journalistic norms) signal a shift in the content news aggregators select. Few empirical studies exist that take an account of how reliance on third-party news material might impact content online. The next section will outline a means to categorize news outputs in terms of topics. It also discusses news inputs as a measure of the degree to which institutional actors are favored in stories.

Finally, researchers should be cautious not to generalize when referring to aggregation as a practice or as a primary function of a news company. Depending on the different organizational cultures, level of editorial oversight and personal preference of the web producers, there may be considerable variation in terms of professional practice and content selection among aggregator websites. Therefore, the following research question is posed:

RQ2 How do aggregator websites differ in terms of professional practice, reliance on institutional actors and story length?

Understanding Content Online: Topics and News-Drivers

As the previous section outlined, news workers routinize the collection and dissemination of information in response to a set of environmental and organizational constraints. These routines often produce predictable patterns of content: reliance on official sources and staged news events (Bennett, 1990; Cook, 1998; Fishman, 1980; Shoemaker & Reese, 1996; Sigal, 1973) However, few have explored the relationship between developing news routines and content in the online news media. This study will examine the practice of collecting third-party source material online, relative to more traditional reporting practices.

The following outlines a rationale measuring online news content. Two measures are considered. The first concerns news outputs in terms of topics. The second accounts for news inputs in terms of news actors. The second measure draws from work on sourcing the news (Bennett, 2004; Bennett, Lawrence & Livingston, 2006; Lawrence, 2001; Livingston & Bennett, 2003). This approach is useful here because it directly connects the above discussion on organizational constraints and sourcing patterns (Cook, 1998; Fishman, 1980; Shoemaker and Reese, 1996; Sigal, 1973).

HARD AND SOFT NEWS

Early scholars found that journalists often dealt with variations in news events through internal classifications of content. Hard and soft news were terms developed to demarcate breaking events from less timely material (Tuchman, 1973). Notions of hard and soft news take their roots from journalist's own classification of daily news events. In Tuchman's account, these classifications are preconceptions based on the nature of the story (breaking, planned or predictable future coverage) and the technology of the medium. These classifications help journalists routinize their work, in an attempt to

reduce the variability of information on any given day. Boczkowski (2004) reminds his readers that such distinctions are fluid. Distinctions between hard and soft news are the product of individual news workers adapting to their environment. Furthermore, he found that online journalists have an accelerated concept of hard news. These journalists spend more time writing headlines, and tend to favor quantity and speed.

A separate literature applies a related, but different conception of hard and soft news. Political communication scholars adapted similar language to categorize media content according to its topical proximity to public affairs information (Baum, 2002, 2007; Hamilton, 2004; Patterson, 2000). In this tradition, hard and soft are categories of topical content. These studies appeared in response to increasing competition in media markets, mostly in television programming. The push toward entertainment content as "soft" programming in this view is a product of market forces on media companies (Hamilton, 2004). According to Hamilton, audiences tend to favor entertainment and human interest stories. Since owners tend to supply content based on consumer demand, the result is a lack of public policy and public affairs programming.

Hard news according to Patterson (2000) refers "to coverage of breaking events involving top leaders, major issues, or significant disruption in daily life" (p. 3). In contrast, soft news "eschews discussion of politics and public policy in favor…of celebrity gossip, crime dramas, disasters or other dramatic human interest stories" (Baum, 2002, p. 91).[8] Over time, Patterson argues, news content has shifted away from public affairs programming, replaced collective references with personal ones, and increasingly favors sensationalism and negative political coverage. Patterson also makes a normative claim: soft news breeds news audiences disinterested in both news and politics.

[8] For a full discussion on soft and hard news, particularly crime stories, see Reinemann et al., 2012 and Curran et al., 2009.

Though following the same basic definitional attributes of the categories, Baum (2002; 2007) notes that soft news is not entirely devoid of relevant policy information. For example, he argues that soft news programming often presents policy information in an entertaining way. Consequently, those who are not generally interested in policy issues may find themselves following foreign policy stories when they pop-up in soft programming (2002). As political information reaches these viewers, Baum argues, it may represent a democratization of policy information.

Scant research has explored whether hard or soft news content prevails online. It's also unclear if news workers in online organizations (journalists and web producers alike) are tending to routinize decisions about newsworthiness in one direction or another. Following the observations of Agarwal and Barthel (2013) and Anderson (2013), if web producers favor a sense newsworthiness oriented toward an audience, one might logically infer that editorial decisions lean toward soft news topics, as suggested by Hamilton (2004). In Hamilton's examination of the demand for online content, he found an overwhelming preference for soft news. The demand for soft news outweighs that of hard news by up to ten times according to Hamilton (p. 207). Accordingly, the following hypothesis is gleamed from the above discussion:

H4 Increased reliance on third-party stories will favor soft news topics

NEWS DRIVERS: OFFICIAL SOURCES

Hard and soft news distinctions act as general proxies for understanding the nature of topics covered in the news. These interpretive categories limit our discussion to consideration of public affairs versus entertainment programming. Consideration of who drives the news, that is, who or what creates the impetus of a story is another means for understanding content. Analysis of the sources tells us something about the popular and

political culture, the framing of issues, and the problems addressed in the news. As Cook (1998) points out, the media do not simply mirror the world, they direct us toward particular values and politics (p. 91).

In Gans' (1979/2004) early work on sourcing the news, he noted several imperatives that influence "source consideration" (p.128). In his study of television and magazine news, over 70% of sources were candidates, officials, state and local officials. Several contemporary and classic studies found a similar reliance on government officials (Brown, Bybee, Wearden, &Straughan, 1986; Daley & O'Niell, 1991; Dickson, 1992; Sigal, 1973; all cited in Shoemaker & Reese, 1996 p. 46).

According to Gans, journalists prefer official authority figures because they solve other newsworthiness criteria, particularly timeliness, authority and reliability (p.130-131). Officials need journalists to get their message out, and journalists need legitimate sources of content. The result "resembles a dance" where "more often than not, sources do the leading" (p.116). Cook (1996) gives journalists more credit for content, as they ultimately decide what's fit to print. Regardless of "who leads," officialdom is such a central feature in the news that Cook regards the news as an arm of government itself. Drawing on Gans, Ruggiero (2004) notes that journalists prefer government officials as sources because it helps maintain the objectivity norm, while at the same time meeting organizational and economic pressures of timely, cheap information.

Other scholars have made distinctions between managed and unmanaged news (Lawrence, 2001; Livingston & Bennett, 2004). Managed news refers to press conferences, political rallies and other staged news events. Unmanaged news, or unexpected news events, may or may not include government officials as sources. As Lawrence (2001) notes, when journalists shift coverage away from routine, scheduled news events, both the range of voices and issue frames employed tend to be more

expansive. Some scholars argue that as technology lowers the barriers to communication, events will become more common in daily news coverage (Livingston & Bennett, 2004).

As many scholars have noted, traditional reporting practices tend to rely on official sources for information. These practices are the outcome of years of stability in the news business, where news gathering routines developed in response to organizational, economic and technological constraints of the time (Bennett, 2004). The very notion of "institutional" norms and practices (and the content they produce) are a result of this history. However, as noted above, contemporary news workers operate in a networked sphere, marked by information abundance and hyper-reality. Changes in technology and the subsequent development of new types of news organizations have also created space for a different type of news worker, one with different routines and ambivalent attitudes toward traditional reporting practices (Agarwal & Barthel, 2013; Anderson, 2013). These news workers (sometimes called web producers or aggregators) often rely on the internet itself for stories. As Anderson (2013b) notes, these workers operate in a de-institutionalized setting. This allows them to distance themselves, and their work, from the traditions and habits of the institutionalized journalist. It remains an open question whether information gathered from outside the newsroom mirrors the voices of institutional actors. Therefore, the following research questions and hypotheses are drawn from the conversation above:

RQ3 To what extent does news content differ, in terms of news actors, when stories are drawn from a third-party compared to original reporting?

H5 Original reporting will draw on institutional actors for stories.

Case Study: Content Analysis of 12 US News Websites

METHOD

This study is based on a secondary analysis of the Pew Research Center News Coverage Index (NCI) data collected over a five-month period between January 2 and May 31, 2012.[9] The Index is a content analysis of the top news stories from a range of media outlets over the period, sampled at regular intervals each week. The data set contains 20,447 total stories. Since this study seeks to understand news content and journalism professional practice in online news media, only online news sites are analyzed (N=3,102).

Pew selects the websites for the NCI sample based on traffic rankings averaged over a period of seven months in 2009, drawing on data from Nielson Media Research and Hitwise, a marketing company that monitors web traffic. The websites in the sample therefore represent the most visited general interest news websites in terms of unique visitors. According to a separate Pew Research Center (2013) report, 10 of the 12 websites in the sample for the NCI remain the most visited on the web, based on data collected by comScore and Nielson in 2012. Only the Wall Street Journal Online and the Los Angeles Times fail to make more current lists. Pew coders (human employees) gathered news stories from each website twice a day in alternating patterns throughout the week, resulting in 30 stories per week. The NCI attempts to capture the most prominent news at any given time, so only the top five stories are captured and coded. Similar to capturing print newspaper stories "above the fold," the Pew coders were instructed to capture based on prominence. The relative size of the headline text and the

[9] The interpretations and conclusions drawn from this data, as well as any errors, are entirely the work of the author.

prominence of pictures associated with major stories are guidelines for determining which stories to capture (Pew Research Center, 2012).

Print	Television	News Aggregators
New York Times	CNN	Yahoo News
Washington Post	MSNBC	Google News
USA Today	Fox News	Huffington Post
Wall Street Journal	ABC News	
LA Times		

Table 1 Sample Breakdown of National News Websites

A custom codebook, created by the author, is used to prepare the data for analysis. In order to answer the research questions, the NCI codebook is used as a base, and additional categories are collapsed or created by the author. Statistical analysis of the data is performed using IMB's SPSS software. The analysis relies on cross-tabulation for categorical data and proportions. Finally, because the data contains both categorical and binary-coded variables, a logistic regression is used to test the likelihood of the independent variables to predict outcome variables (Tabachnick & Fidell, 2013).

Variables used in the Study

Organization Type is drawn from the NCI *sources* category. Websites are coded according the dominant, and mutually exclusive, media product each organization produces. For example, *The New York Times* and the *Washington Post* are coded as print oriented companies, while *Google News* is coded as a news aggregation company (see Table 1). The only exception to this rule is *Yahoo News,* which has a business partnership with a television company, *ABC News*. However, *Yahoo News* is coded as an aggregator, since its partnership with ABC did not commence until shortly before the NCI collected

the sample.[10] At least two websites in the news aggregator category, *Google News* and *Yahoo News,* filter content using algorithms (Hindman, 2012).

Professional Practice is measured in terms of the type of news work employed in each story. These variables capture how a story is reported. Types of news work under analysis in this study are *original reporting* by resident journalists, reliance on *wire service* and use of *third-party material*.

Original reporting is any story created by the internal staff of that news organization. Stories in this category may or may not have bylines. Original reporting coincides with what Anderson (2013b) refers to as "first-level" news processes. In theory, these stories would be created by reporters employed by the news organization under analysis. The following categories represent what Anderson (2013b) refers to as "second-level" news processes: *Wire* stories and *Wire/Staff* are those produced either by a news wire service or some combination of wire service and internal staff. *Third-party material* is based on the NCI code that instructs coders to record stories by "other news outlets." This variable stands for news material featured on a website that is *not* a product of original content created the website or a wire service.

Content is measured according to the nature of news topics covered and the social role of major *news actors* driving each story. Variables related to content are the main dependent variables in the study. For topics, the Pew Research Center NCI codebook records 26 broad topical categories. Those categories are further separated by the author into *hard news* and *soft news* topics.

Hard/Soft News topics are measured following Patterson (2000) and Baum (2002). *Hard news* topics are those directly related to political institutions, public affairs,

[10] See http://abcnews.go.com/US/abc-news-yahoo-news-announce-online-alliance/story?id=14650998

education, environment and health policy. Davies (1996, p.108) defines *soft news* as "all news that isn't *hard news*." Taking a slightly more descriptive approach, this study defines *soft news* as all that which "eschews discussion of politics and public policy in favor…of celebrity gossip, crime dramas, disasters or other dramatic human interest stories" (Baum, 2002, p. 91). Where the news topic descriptions in the NCI codebook contain ambiguous or conflicting language, the topic is left out of the analysis. The resulting sample size for *hard* and *soft news* is 2,439 stories.

News actor is created by the author after coding all open-ended questions in the NCI's "lead newsmaker" category. This variable is similar to those employed in other news analyses that measure the "initiator" or "news driver" of a story (Livingston & Bennett, 2003; Segal, 1973; Wolfsfeld & Sheafer, 2006). In other words, who is the impetus for the story? In contrast to *hard* and *soft* distinctions that capture news outputs, the *news actor* category captures inputs. The NCI codebook calls on the Pew coders to write-in the name of a news actor when they are the focus of at least 50% of the story. Four major categories of actors are used in this study: *institutional news actors, news event actors, celebrities/sports* and *journalist actors*.

Up to two "lead newsmakers" are recorded in the NCI. These two categories contain multiple thousands of open-ended entries. For the purpose of this study, only the first lead newsmaker is coded, yielding 1,706 news articles with at least one lead newsmaker. Coders (the author and one other graduate student at the University of Texas, Austin) were instructed to separate newsmakers according their social position as media actors. In order to ensure consistency in this category, a reliability test was performed using two coders on a random sample of 10% of the valid cases. Following Riffe, Lacy and Fico (2008), performance testing yielded an overall agreement for lead newsmaker of 96% and Krippendorff's alpha was a high 0.94.

Institutional actors are those elected and appointed officials, their spokesman, staff or administration. Candidates seeking office are also considered institutional, as well as former office holders and political pundits, their direct targets and powerful, public corporate actors. These actors would typically be relied upon for what Cook (1998) refers to as the government information subsidy.

News event actors are all other news actors tied to timely events or stories in the news, all participants in a crime case (not including judges), terrorism suspects, characters in human interest stories or experts, but not elected officials or celebrities. This category draws on scholarship delineating event-driven stories from institutionally-driven stories (Bennett, Lawrence & Livingston, 2006; Lawrence, 2001). The key distinction in event-driven criteria, Lawrence (2001) notes, is that stories are more spontaneous, and not necessarily pegged to institutional actors.

Journalist news actors are those news makers that work for a news organization, or produce media content that is related to public affairs or the news industry in general. Following Bennett (2004), reporter-driven stories might be more investigative (as opposed to strictly tied to officials or events) and reflect, in theory, professional and personal news values (p. 296).

Celebrities/sports captures entertainment related news makers as they appear in top stories. This category of story driver is essential to accounting for what Postman (1985) called "infotainment," or the blending of information and entertainment. This category of news-driver has become a developing area of scholarship, particularly as many different types of news actors vie for relevance in public discourse (Williams & Carpini, 2011; Feldman, 2007; Feldman &Young, 2008). Celebrity or sports personalities are coded regardless of the story topic. This measure therefore reflects the degree to which entertainment drives content in top online news stories.

Other story variables in the NCI relevant for this analysis include one measure related to length: s*tory length*. *Story Length* is operationalized as the number of words in each story. Coders are instructed to cut and paste the entire story into a word processor and measure word count. As one pair of scholars note, story length can operate as a proxy for the intensity of reporting (Livingston & Bennett, 2003).

Results

The first overarching research question this paper asks is *how does professional practice differ across various news organization types*? Table 2 shows the breakdown of reporting methods and story length by organization type. Stories analyzed here are based on the total sample web news sites (N=3102). New aggregators are the least likely to employ traditional reporting (33%) and use wire service more than other organization types (24%). Print-based websites overwhelmingly (84%) favor original reporting. Television news websites featured original reporting about half of the time (47%) and used staff/wire combination in stories more than other organization types (9.7%). The Chi Square tests confirmed difference in proportions among organizations were not due to chance and statistically significant ($X^2(6, N=2672) = 982.05, p < .001$).

H1 predicts print-based websites will rely less on third-party content. As Table 2 shows, print-based news websites rarely appropriate third-party content (1%). They are also the least likely to do so compared to other organizational types. Therefore, **H1** is confirmed.

H2 hypothesizes that news aggregator websites will favor third-party content. News aggregator websites feature third-party content more than any other organization type, about a third of the time (29%). However, News aggregator websites, as a category, also rely on original reporting a little over third of time (33%). **H2** is partially confirmed, in terms of aggregator websites employing third-party sources as featured content. One should qualify the findings, since there is still some original reporting featured on these sites.

	Original Reporting (%)	Wire (%)	Wire/Staff (%)	Third-Party Stories (%)	Story Length (Mean/SD)
Print	83.6	6.8	0	1.0	911/610***
Television	47.3	17	9.7	3.3	682/424
News Aggregators	33.1	23.6	1.0	29.3	705/611

*p<.05, **p<.01, ***p<.001

Table 2 Professional Practice and Story Length by Organization Type[11]

H3 hypothesizes that news aggregator websites will feature shorter stories. As Table 2 shows, print-based news organization's stories tend to be longer. A significant difference was found in the means between print and the television and news aggregator groups ($F(2, 1617) = 30.8, p < .001$). In addition, a post-hoc test using Tukey HSD reveal the print group has longer stories than the others, but no difference is found between the television and aggregator groups. News aggregator websites feature stories that are shorter than print-based organizations, but are not statistically different in length compared to television-based websites. Therefore **H3** is upheld with qualification.

RQ2 asks how aggregator websites differ in terms of professional practice, reliance on institutional actors and story length. There are differences across the board within the news aggregator websites (see Table 3). *Google News* has almost no original reporting (1.1%), while *Huffington Post* has nearly half (54%) and *Yahoo News* slightly less than half (43%). *Google News*' reliance on third-party content (68%) coincides with the lowest proportion of stories that rely on institutional actors (48%). *Huffington Post* uses third-party content less than the others in the category (7%). Both *Google* and *Yahoo*

[11] Proportions do not add to 100% due to the way the NCI measures story format. The NCI includes another variable not used in this study, multimedia content. That variable was left out because it is not clear weather multimedia is created by a third-party source or produced by in-house reporters.

feature shorter stories than other news organizations, and the difference is statistically significant at the $p<.01$ level. *Huffington Post* stories are similar to the mean for the entire sample. *Huffington Post* also features institutional actors more than other web aggregators (76%), followed closely by *Yahoo News* (73%).

	Original Reporting (%)	Third-Party Stories (%)	Institutional News Actor (%)	Story Length[12] (Mean/SD)
Yahoo News	43.3	14.2	72.6	664/363**
Huffington Post	53.8	7.0	76.4	750/275
Google News	1.1	67.8	48	644/384**

*$p<.05$, **$p<.01$, ***$p<.001$

Table 3

News Aggregator Websites: Professional Practice, Institutional Actors and Story Length

H4 hypothesizes increased reliance on third-party stories will favor soft news topics. Reliance on third-party content tends to favor soft news topics (Figure 1) compared to original reporting (40% in third-party compared to 20% of original reporting) and features slightly less hard news topics (60%). Original reporting overwhelmingly features hard news topics (78%). Chi-square tests also show a difference in hard and soft news topic preference at a significant level ($X^2(3, N=2061) = 56.72, p < .001$).

[12] Asterisks for story length indicate p values for the means compared to the entire sample.

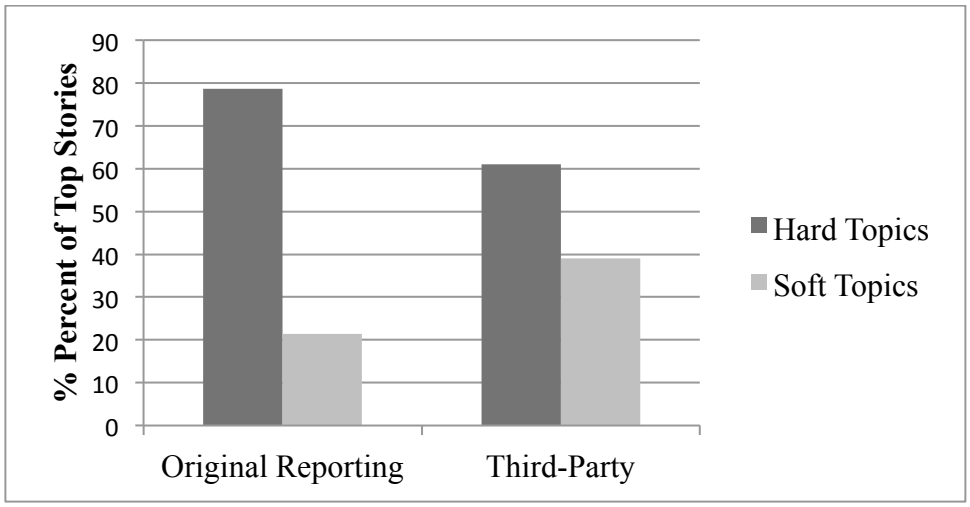

Figure 1 Hard and Soft News Topics by type of Professional Practice

	Institutional	Event	Celebrity/Sport	Journalist
Original Reporting	71(2.4)***	22(.46)***	7(.54)**	1(0)
Third-Party	56(.65)*	23(.8)	20(3.4)***	1(.76)

Table shows proportions and odds ratios in parenthesis, asterisks represent significance levels. *p<.05, **p<.01, ***p<.001

Table 4 Type of News Actor in Original and Third-Party Reporting

RQ3 asks to what extent news content differs, in terms of news actors, when stories are drawn from a third-party compared to original reporting. **H5** suggests original reporting will draw on institutional actors for stories. Stories created with original reporting (see Table 4) favored institutional actors (71%) over event-related (22%) and celebrity actors (7%). Original reporting was one-and-half times (140%) more likely to favor institutional actors, but 54% less likely to feature event-related actors and 46% less likely to feature celebrities, therefore, **H5** is supported. In contrast, stories gathered from a third-party source are nearly two-and-a-half times more likely to feature celebrity

content. They are also less likely (by 35%) to favor institutional sources. Though third-party content sourcing is slightly less likely to feature event-related actors, the relationship was not statistically significant. The overall lack of journalists as featured news actors is evident, as both news gathering practices mentioned journalists as news makers about in about 1% of stories.

Discussion and Conclusion

Aggregation has become a key feature of the online news landscape. As a professional practice, it represents a developing sense of journalistic values, one that, perhaps, relies less on institutional actors. As a prominent destination for news on the web, aggregation websites represent a fluid space for shifting practices in news work. In either case, the emphasis on third-party content tends to favor entertainment and celebrity news-drivers over public officials. Third-party sourcing also favors soft news topics, in this case, news topics not tied to public affairs or government institutions. News aggregation websites also seem to display less intensity in their reporting, at least in terms of story length. These findings are in line with Hamilton (2004) and Patterson (2000), who suggested a trend toward entertainment and audience demand in news media, even before the explosion of online content.

The results also point toward a complex news environment online, one where professional practice varies according to the influence of parent companies and individual newsroom cultures. Print-based newsrooms, in this analysis, seem reluctant to "shovel" content from other sources, at least not for featured stories placed prominently on their websites. Print-based websites also feature original reporting more than other organizations and display the most intensity of reporting in terms of story length. One might speculate that print organizations are mainly transferring content from the printed product to the website (Quandt, 2008b).

The differences in content patterns between original reporting and third-party content collecting routines suggest that different conceptions of newsworthiness are at work. Future research might explore the impact of sourcing material in the hyper-real environment of the web. For example, as Cook (1998) notes, news making represents a

negotiation between reporters and officials. But what are the contours of that relationship when the reporter is more concerned with audience metrics and checking content filtering apps to determine newsworthiness? One possible answer is a slide toward a more fragmented news sphere, where general interest news sites are relegated to competing for audiences with more celebrity news and courtroom drama.

We now are able to return to Bennett's (2004) question: Does the rise of a new order represent a sea change in gate-keeping practices, or are we witnessing a hybrid model where traditional news making is simply supplemented, but not replaced? Ten years after his model was proposed it seems as though the "entertainment-technology" order is in open competition with the "reporter-organization" order. News organizations still rely on public officials and staged news events. Print-based organizations seem to be rolling their routines into the digital product. One might suggest, based on these results, that news aggregation simply represents the acceleration of the economic, audience-driven order that has characterized the competitive news landscape since the 1980's (Baum, 2007; Hamilton, 2004).

However, if we are to take the work of some recent scholars into account, the demands of the web are signaling the introduction of a new type of news worker, one not necessarily bound by traditional practice. These news workers are less enthusiastic about the objectivity norm (Agarwal & Barthel, 2013), employ divergent logics of news production (Boczcowski, 2010) and appeal to alternate epistemologies in terms of evidence for stories (Anderson, 2013). These insights, in combination with the evidence of a complex mix in news inputs and outputs outlined in this study, call for more careful reflection. It may not be so simple, as Bennett puts it, to assume a "multimodal system of the sort that exists in most eras in which a dominant gatekeeping pattern….is supplemented- as opposed to supplanted" (p. 302). We may be entering an era where it is

difficult to ascertain what the dominant order is. We may also be hard-pressed to determine where one order ends and another begins. We may be better off giving up on the attempt to place the complexities of the modern media environment into theoretical boxes. We should, instead, look at the interaction of human actors within their unique environments. By doing so, we will be better equipped to address the growing complexity of news work as a pluralistic enterprise, sometimes governed by organizational pressures, sometimes economic, or sometimes by entirely different incentive structures.

For example, in the case of aggregation as news work, individual web producers might not adopt uniform principles and values. In Bennett's' model, we can make certain assumptions about the intentions of individual journalists to act as watchdogs. We can also assume there will be certain relationships with sources. These conclusions become less certain if we consider the web as a source (instead of personal relationships) and individual content curators who may not fit the mold of traditional journalists. Future work should look more closely at those aggregators (as organizations and as news professionals), especially those who orbit actors in the political sphere.

This study, like any other, also includes limitations. For one, it re-purposes data designed and collected with aims outside those of the author. Second, it is not clear what "original reporting" and "third-party" content means in the real world. These are proxy measures, based on the Pew NCI codebook. It may be that web producers are putting bylines on the content they collect from other sources when they summarize a story. These stories could, by nature of Pew's coding instructions, be counted as original reporting. A study more reflective of news production online might combine ethnographic or other deep qualitative work with content analysis to determine the exact degree to which content is produced by "shoe-leather" routines or digital content curating. Second, one should always be cautious of dualistic categories that stand for

content as a whole. For example, the distinction between hard and soft news used here does not take an account of the presence of politically relevant news frames or themes related to policy or other matters potentially useful in a democratic society. These limitations do not detract from the findings. It remains clear that third-party sourcing, as a routine, favors certain topics and news actors over others. These findings are concurrent with both previous literature in political communication and the ethnographies cited herein.

Finally, this study is one of the first to take up Agarwal & Barthel's (2013) call to consider the centrality of news aggregation to the development of online routines, norms and the professional development of online journalism. This discussion puts forward a nuanced definition of aggregation, both as news work and as a news organization. It also provides evidence for the elevation of audience in terms of a shift to celebrity news actors and shorter news stories, at least when compared to traditional reporting. It is also evident that individual newsroom cultures might be a factor in why one professional practice, routine, or subsequent set of values are chosen over others.

Cautiously, this report refrains from making normative judgments on the role of information in democratic societies. Though the online news media seems to be sliding toward tabloid journalism, this study only considers the top 12 US news sites. Future work should explore the quality of news in other production contexts, like collaborative filtering or niche news environments.

References

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. (2013b). Rebuilding the news. Temple University Press

Agarwal, S. D., & Barthel, M. L. (2013). The friendly barbarians: Professional norms and work routines of online journalists in the United States. *Journalism*, 1464884913511565.

Arthur, C. (2006). What is the 1% Rule? *The Guardian*. Retrieved from: http://www.theguardian.com/technology/2006/jul/20/guardianweeklytechnologysection2

Baum, M. A. (2002). Sex, lies, and war: How soft news brings foreign policy to the inattentive public. *American Political Science Review*, *96*(01), 91-109.

Baum, M. A. (2007). Soft news and foreign policy: How expanding the audience changes the policies. *Japanese Journal of Political Science*, *8*(1), 115.

Benkler, Y. (2006). The wealth of networks: How social production transforms markets and freedom. Yale University Press.

Bennett, W. L. (2004). Gatekeeping and press-government relations: A multigated model of news construction. *Handbook of political communication research*, 283-314.

Bennett, W. L., Lawrence, R. G., & Livingston, S. (2006). None dare call it torture: Indexing and the limits of press independence in the Abu Ghraib scandal. *Journal of Communication*, *56*(3), 467-485.

Bimber, B. (2003). *Information and American democracy: Technology in the evolution of political power*. Cambridge University Press

Boczkowski, P. (2010). *News at work: Imitation in an age of information abundance*. University of Chicago Press.

Brannon, J. (2008). Maximize the medium: Assessing obstacles to performing mulitimedi journalism in three US newsrooms. In C. Patterson & D. Domingo (Eds.) *Making the news online: The ethnography of new media production* (99-112). Peter Lang

Castells, M. (2009). Communication power. OUP Oxford.

Cawley, A. (2008). News production in an Irish online newsroom: Practice, process and culture. In C. Patterson & D. Domingo (Eds.) *Making the news online: The ethnography of new media production* (45-60). Peter Lang

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

Curran, J., Iyengar, S., Lund, A. B., & Salovaara-Moring, I. (2009). Media system, Public knowledge and democracy: A comparative study. *European Journal of Communication*, *24*(1), 5-26.

Deuze, M., & Marjoribanks, T. (2009). Newswork. *Journalism*, *10*(5), 555-561.

Domingo, D. (2008). When immediacy rules: Online journalism models in four catalan online newsrooms. In C. Patterson & D. Domingo (Eds.) *Making the news online: The ethnography of new media production* (113-126). Peter Lang.

Feldman, L. (2007). The News about Comedy Young Audiences, The Daily Show, and Evolving Notions Of Journalism. *Journalism*, *8*(4), 406-427.

Feldman, L., & Young, D. G. (2008). Late-night comedy as a gateway to traditional news: An analysis of time trends in news attention among late-night comedy viewers during the 2004 presidential primaries. *Political Communication, 25*, 401-422.

Fico, F. G., Lacy, S., & Riffe, D. (2008). A content analysis guide for media economics scholars. *Journal of Media Economics*, *21*(2), 114-130.

Fishman, M. (1980). *Manufacturing the news, Vol. 143*. Austin: University of Texas Press.

Gans, H. J. (2004). *Deciding what's news: A study of CBS evening news, NBC nightly news, Newsweek, and Time*. TriQuarterly Books.

Garcia, E. (2008). Print and online newsrooms in Argentinean media: Autonomy and professional identity. In C. Patterson & D. Domingo (Eds.) *Making the news online: The ethnography of new media production* (61-76). Peter Lang.

Grueskin, B., Seave, A., & Graves, L. (2011). Chapter Six: Aggregation. *Columbia Journalism Review*. Retrieved from http://www.cjr.org/the_business_of_digital_journalism/chapter_six_aggregation.php?page=all

Hamilton, J. (2004). *All the news that's fit to sell: How the market transforms information into news*. Princeton University Press.

Herman, E. S., & Chomsky, N. (2008). *Manufacturing consent: The political economy of the mass media*. Random House.

Hindman, M. (2012). Personalization and the future of news. EUI Working Papers.

Isbel, K. (2011) What's the law around aggregating news online? A Harvard Law report on the risks and the best practices. *Nieman Journalism Lab*. Retrieved from http://www.niemanlab.org/2010/09/whats-the-law-around-aggregating-news-online-a-harvard-law-report-on-the-risks-and-the-best-practices/

Keith, S. (2011). Shifting circles: Reconceptualizing Shoemaker and Reese's theory of a hierarchy of influences on media content for a new media era. *Web Journal of*

Mass Communication Research, 29. Retrieved from Fhttp://www.scripps.ohiou.edu/wjmcr/vol29/29.html#reconcept

Keller, B. (2011). All the Aggregation That's Fit to Aggregate, *New York Times*. Retrieved from http://nyti.ms/htb7Xk

Klinenberg, E. (2005) 'Convergence: News Production in a Digital Age', *Annals of the American Academy of Political and Social Science* 597(1): 48–64.

Kopper, G. G., A. Kolthoff and A. Czepek (2000) 'Research Review: Online Journalism

Lawrence, R. G. (2001). Defining events: Problem definition in the media arena. *Politics, discourse, and American society: New agendas*, 91-110.

Lewis, S. C. (2012). The tension between professional control and open participation: Journalism and its boundaries. *Information, Communication & Society*, *15*(6), 836-866.

Livingston, S., & Bennett, W. L. (2003). Gatekeeping, indexing, and live-event news: Is technology altering the construction of news? *Political Communication, 20*(4), 363-380.

McChesney, W. (2008). The political economy of media: Enduring issues, emerging dilemmas. Monthly Review Press.

Mitchelstein, E., & Boczkowski, P. J. (2009). Between tradition and change A review of recent research on online news production. *Journalism*, *10*(5), 562-586.

O'Sullivan, J. and A. Heinonen (2008). Old Values, New Media: Journalism Role Perceptions in a Changing World. *Journalism Practice, 2*(3): 357–71.

Patterson, T. E. (2000). *Doing well and doing good: How soft news and critical journalism are shrinking the news audience and weakening democracy-and what news outlets can do about it*. Joan Shorenstein Center on the Press, Politics and Public Policy, John F. Kennedy School of Government, Harvard University.

Pew Research Center (2014). State of the News Media. Retrieved from http://www.journalism.org/2014/03/26/how-big-is-the-digital-news-world/

Pew Research Center (2013). State of the Media. Retrieved from http://stateofthemedia.org/2013/digital-as-mobile-grows-rapidly-the-pressures-on-news-intensify/digital-by-the-numbers/

Pew Research Center (2012). News Coverage Index, 2012. Retrieved from http://www.journalism.org/by_the_numbers/datasets

Pew Research Center (2011). Google: Parasite of Provider? Retrieved from http://www.journalism.org/2011/05/09/google-parasite-or-provider

Pew Research Center (2006). The Aggregators. Retrieved from http://www.journalism.org/2006/11/27/the-aggregators/

Oreilly, T. (2005). What is Web 2.0? Design Patterns and Business Models for the Next Generation of Software. Retrieved from http://papers.ssrn.com/sol3/papers.cfm?abstract

Quandt, T. (2008) News tuning and Content Management: An observation study of old and new routines in German online newsrooms. In C. Patterson & D. Domingo (Eds.) *Making the news online: The ethnography of new media production* (77-98). Peter Lang.

Quandt, T. (2008b). News on the World Wide Web? A comparative content analysis of online news in Europe and the United States. *Journalism Studies*,*9*(5), 717-738.

Reese, S. (2001). Understanding the global journalist: A hierarchy-of-influences approach. *Journalism Studies*, *2*(2), 173-187.

Reinemann, C., Stanyer, J., Scherr, S., & Legnante, G. (2012). Hard and soft news: A review of concepts, operationalizations and key findings. *Journalism*, *13*(2), 221-239.

Ryfe, D. (2012). Can Journalism Survive? *An inside Look at American Newsrooms. Cambridge: Polity*.

Schudson, M. (1989). The sociology of news production. *Media, culture and society*, *11*(3), 263-282.

Shoemaker, P. J., & Reese, S. D. (1996). Mediating the message. White Plains, NY: Longman.

Sigal, L. (1973). *Reporters and officials: The organization and politics of newsgathering*. Lexington, MA: Heath.

Tabachnick, B. G., & Fidell, L. S. (2001). Using multivariate statistics. Pearson Education Inc.

Tuchman, G. (1978). The news net. *Social Research*, 253-276.

Tuchman, G. (1973). Making news by doing work: Routinizing the unexpected. *American Journal of Sociology*, *79*(1), 110.

Waldman, S. (2011). *Information Needs of Communities: The Changing Media Landscape in a Broadband Age*. DIANE Publishing.

Williams, B. A., & Carpini, M. X. D. (2011). *After broadcast news: Media regimes, democracy, and the new information environment*. Cambridge University Press.

Wolfsfeld, G., & Sheafer, T. (2006). Competing actors and the construction of political news: The contest over waves in Israel. *Political Communication*, *23*(3), 333-354.