



ALEXANDER VON HUMBOLDT
INSTITUT FÜR INTERNET
UND GESELLSCHAFT

DRAFT PAPER prepared for the 1st Berlin Symposium
on Internet and Society
October 26th – 28th 2011

**Understanding the Role Played by
Algorithms and Computational
Practices in the Collection, Evaluation,
Presentation, and Dissemination of
Journalistic Evidence**

Conference Draft

Christopher W. Anderson

christopher.anderson@csi.cuny.edu

Paper prepared for the 1st Berlin Symposium on Internet and Society,

Oct. 25-27, 2011

Understanding the Role Played By Algorithms and Computational Practices in the Collection, Evaluation, Presentation, and Dissemination of Journalistic Evidence

C.W. Anderson

Christopher.Anderson@csi.cuny.edu; The College of Staten Island (City University of New York),

Department of Media Culture. 2800 Victory Boulevard. Staten Island, NY 10315.

ABSTRACT: This article advances a critical research approach to computational journalism. By “computational journalism” the article refers to the increasingly ubiquitous forms of algorithmic, social scientific, and mathematical forms of newswork adopted by many 21st century newsrooms and touted by many educational institutions as “the future of news.” The bulk of the article outlines a series of six lenses through which such a critical approach to computational journalism might be carried out. Four of these lenses are drawn from Schudson’s classic typology of the sociology of news—economic, political, cultural, and organizational approaches. In addition, the author adds Bordieuean field approaches and technological lenses to the mix. In each instance, the author discusses how particular approaches might need to be modified in order to study computational journalism in the digital age.

Introduction

In 2007, the news and financial market data provider Reuters (now merged with information and technology firm Thompson Company) purchased the text search firm ClearForest, “a maker of software used to search vast archives of news, Web pages and documents for relevant facts” (Auchard 2007) In the spring of 2008, Thompson Reuters launched the “Calais web service,” a web service that:

automatically attaches rich semantic metadata to the content you submit. Using natural language processing, machine learning and other methods, Calais categorizes and links your document with entities (people, places, organizations, etc.), facts (person “x” works for company “y”), and events (person “z” was appointed chairman of company “y” on date “x”) (“Calais Viewer”).

Calais, in short, automatically extracts “facts” from news (and other kinds of) narratives.

In the summer of 2009, a related service, called MemeTracker, made its debut. “MemeTracker builds maps of the daily news cycle by *track[ing] the quotes and phrases that appear most frequently over time* across this entire online news spectrum. This makes it possible to see how different stories compete for news and blog coverage each day, and how certain stories persist while others fade quickly.” Like Open Calais, the power of MemeTracker lies in its ability to extract nuggets of news (“memes”) from the swirl of narrative text and, over time, build connections and dynamic maps of these news nuggets.

In August of 2011, the venerable Google News supplemented its entirely algorithmically controlled front page with human-guided “editor’s picks,” featuring the news judgments of real, live human editors. As Megan Garber with the Nieman Journalism Lab noted:

When Google News launched in 2002, it did so with some declarations: “This page was generated entirely by computer algorithms without human editors.” And: “No humans were harmed or even used in the creation of this page. But the thing about humans is that, occasionally, they’re helpful to have around. Especially when it comes to the increasingly difficult task that is keeping track of the world as it twists and turns. Which is why, starting today, Google News is introducing a new section to its U.S. edition: Editors’ Picks, a display of original content that journalists (human ones!) have selected as editorial highlights from their publications. (Garber 2011)

The resulting debate over the role of human-editor hybrids in the new media ecosystem sheds light on the processes by which news is selected, filtered, and edited in the 21st century.

Perhaps the most interesting thing about all these examples-- Open Calais, MemeTracker, and Google News-- is that they begin turn our focus to the **objects of information, including the objects of news**. By objects of news I mean the fragments of information, data, and technology that are pieced together into various news stories. By objects of news I also mean news these news stories themselves, and the manner in which they circulate inside larger news ecosystems. The purpose of this working paper is to consider the **algorithm** as a new object of news that intersects with both journalistic practices and products, and ultimately affects the very definition of journalism itself. The second section of the paper discusses some possible ways in which communication scholars might want to approach the study of algorithms as journalistic objects. Embedded within this second section are a series of possible lenses for future research, which are addressed specifically as future research questions in the third and final section.

Objects of Evidence in Journalism: Thinking About Networked Things

Objects of news have always been part and parcel of the practice and study of journalism though they have rarely been discussed in precisely these terms. For instance, in sources as diverse as an introductory lecture in basic “news reporting and writing,” (fieldwork, November 2007), an online syllabus for budding citizen journalist (McGill n d), a guide for elementary school students about the nature of news (Niles n d), and a faculty handbook for the assessment of journalism and mass communication programs, writers make it clear that some of the most important objects of news include “observations, interviews, and documents” (McGill, n d). “There are three main ways to gather information for a news story or opinion piece,” writes the guide to journalism for children: “Interviews: talking with people who know something about the story you are reporting. Observation: watching and listening where news is taking place. documents: Reading stories, reports, public records and other printed material” (Niles n d). For budding

journalists on college campuses, the instructions are similar. A rubric for assessing the “core competencies and learning objectives” for journalism students lists the very first basic skill as “gathering information through interviewing, observation, and the use of documents, printed and computerized, to write fair and balanced news stories.” In these examples, documents, interviews, and observations each act as objects of news: lower-level fragments of news stories, forms of valid journalistic evidence (i.e., carriers of particular truth claims), and aspects of the reporting processes that have their own histories and their own ways of being integrated into larger processes of newswork.

When we start thinking about documents, sources, and direct interviews as *news objects*-- objects that present particular evidentiary claims and are embedded in specific historical trajectories-- we can see that there might be additional objects that *also* constitute fragments of journalistic evidence, but which are less universally discussed by either journalism educators or journalism scholars. The spectrum of possible news objects, in other words, is far more complex than the usual rundown of “documents, interviews, and observations” might suggest, and might include public forums, links, databases, web metric reports, T-1 internet lines, and tweets as potential fragments of a larger journalistic network. It might also include algorithms. I now want to turn to an overview of the relationship between journalistic practices, journalistic products, and one particular journalistic object— **the algorithm**-- along with some preliminary discussions of ways that we might want to study the computational journalistic practices that cluster around this object. What, exactly, is an algorithm? What could it possibly have to do with the way 21st century journalism is practiced and understood? And how does it relate to practices of computational journalism more broadly?

Use of the term “algorithm” dates back to at least the Middle Ages, and was originally used to distinguish scholars who utilized written, columnar calculation procedures from those performed calculations on an abacus (Stone 1972). In general, while there is no standard definition of algorithm, we can say that it constitutes a series of *rules* for accomplishing a particular *task* in a certain number of discrete *steps*. It is easy to see that, although algorithms were invented long before computers and other electronic calculation devices, the focus on

particular rules and repeatable steps would lead algorithms to a position of great prominence in the information age. In their simplest form, computers are devices that follow discrete rules and procedures in a rapidly repeatable fashion; they thus stand quite close to the algorithm in many respects.

Perhaps the most famous algorithm in recent times is the Google PageRank algorithm, which dominates the 21st century world of information retrieval. In the most basic sense, PageRank measures the “relevance” of a website by calculating the number of links to that site, and following this citation trail backwards in a series of increasingly complex and data-intensive steps. Google then uses this ranking to order its search results, thus bequeathing sites at the top of the results a tremendous amount of power, authority (and, it must be said, money. While it can be said that “Google” ranks these sites, it is more accurate to say that the socio-technical network that is Google, as mediated through its proprietary, socio-technical PageRank algorithm, ranks these sites. Journalists are only beginning to think about how algorithms might be used to manage their own informational workflows in a manner similar to that by which Google stores, retrieves, and ranks digital information on the entire web.

Understanding the relationship between algorithmic objects and journalistic practices, then, is ultimately a subset of a larger research question: understanding the interaction between large-scale data collection, algorithmic analysis, computational practices, and the production of public knowledge. I would argue that relationship between “big data” and communicative processes (such as computational journalism) has quickly emerged one of the central scholarly and methodological challenges of our time (Manovich 2011). If this is correct, then we need to find ways to bring old mechanisms of analyzing journalism to bear on these new computational practices and problems, including the problems of algorithmic journalism.

The kind of analysis I advocate for the study of algorithms and computational journalism here would pay careful attention to the means by which public policies and struggles over the transparency of open-government data impinge on newswork. It would examine, in detail, the unequal distribution of computational resources in 21st century journalism and critically dissect how this inequality is either impeding or facilitating journalism’s professional mission. It would look

at the dynamics of the journalistic field and the manner in which struggles by entities outside that field are impacting the diffusion of computational practices. It would pay careful attention to the organizational dynamics and work processes that either facilitate, or hinder the adoption of computational techniques inside the newsroom. It would engage in a careful, historically grounded analysis of journalistic culture and the role that culture plays in our understanding of what counts as “news”. It would, finally, bring a nuanced, socio-material perspective on technology to bear on questions regarding the relationship between computational technology and journalistic “sense-making.”

Astute readers will have noted that four of the lenses I advocate here— *political*, *economic*, *organizational*, and *cultural* frames— have long been considered to be the “classic” perspectives within the sociology of news (Schudson 2005) while the other two — *technological* and *institutional*, or “field,” perspectives— though not classic, are increasingly utilized in 21st century journalism research. I would also argue that two new additional frames may also be helpful in understanding journalism in an increasingly trans-institutional, technologically mediated age. These are not the only scholarly research perspectives on journalism, of course, but they are helpfully compact, and schematic enough to use as a starting point for analysis. The purpose of this categorical overview, in short, is to both examine the possible lenses through which to analyze a major new example of digital knowledge production in the 21st century, and to highlight some of the early work that has already been done in this vein. In order to accomplish this, let’s turn to a more in-depth discussion of these six categories and the way they might be applied to news algorithms and other computational practices.

Six Approaches to the Sociology of Algorithms and Computational Journalism

Michael Schudson’s “Sociology of News Production,” first published in *Media, Culture, and Society* in 1989 and revised several times since then, remains one of the key citations in any sociological analysis of journalism and news. The article has actually undergone several important mutations over the course of its 20-plus year history, with the key changes being

outlined rather frankly in the latest incarnation of the piece (“Four Approaches to the Sociology of News,” published in the equally oft-revised *Mass Media and Society* [5th edition]). The different descriptions of each of the categories amount to more than simply revisions for the sake of revision, particularly in the latest piece, which signals its intentions by announcing the “approaches” to studying news have now expanded from three to four. The relevant nuances of these category changes will be explored below. An equally important change in Schudson’s thinking—the argument that journalistic products are as real as they are “socially constructed”—lies outside the scope of this current paper, but would be valuable to analyze in its own right.

What, then, are the different lenses through which we can analyze the sociology of news production, and how can each of them be applied to the study of computational journalism?

Politics and Public Policy

In its simplest form, the “political context of news-making” can be used to study the production of news on a macro or state-level, particularly insofar as different political systems co-produce different systems and styles of journalism. Political approaches to the study of news usually emphasize one or two unique, meta-theoretical points. First, the approach emphasizes that journalism is not co-terminous with democracy; contra Carey (Carey 2000), undemocratic states may possess their own, unique journalistic forms. Second and more importantly, the political approach to the analysis of journalism has been especially useful in the comparative study of news practices, particularly the different practices that exist within different democratic systems. Beginning with Hallin and Mancini’s 2004 *Comparing Media Systems* (Hallin and Mancini 2004) a growing and persuasive body of research has pointed to variations within capitalist democracies (between France, Great Britain, and Germany, for instance) as being particularly important causal factors accounting for differences in news (Benson 2010; Hallin 2004; Hanitzsch 2008; Waisbord 2000) In part, this is a reaction to earlier generations of news research that tended to posit “democratic, capitalistic” journalistic systems as structurally

homogenous when compared to the journalism of socialist or communist nation-states (Siebert et. al 1956).

At first glance, it is difficult to see how this political perspective might be at all useful in the analysis of algorithms and computational journalism. Is computational journalism not a micro-level or (at best) an organization-level phenomenon? How can it be studied in its political context without losing sight of what makes it unique? There is some truth to this claim. On the other hand, it is precisely by trying to fit the development of computational journalism within a macro-political frame that our attention is drawn to aspects of it that we might otherwise neglect. When analyzing computational journalism through a political lens, scholarship should initially focus less on comparative differences in political structure and more on the bureaucratic, policy-level initiatives that either allow computational journalism to thrive, or retard its growth. Developments in the world of so-called “open government initiatives” the role of large-scale databases in the crafting of public policy, the transparency and accessibility of government data and its use in democratic decision-making, government surveillance of online data and political activities; all these state-level initiatives may ultimately influence the forms of computational, data-driven journalism that news-organizations adopt.

In general, work on internet-afforded public policy looks at one of two phenomena—either digitally enhanced citizen-participation (Noveck 2009) or the more passive activity of making government data easily accessible and searchable via digital archives. While “open” and “wiki” government initiatives have received much attention from policy-makers, activists, and reporters, the scholarly community has yet to devote much attention to these phenomena. Noveck (2011) has been an exception, devoting the last several years to both open government policy-making in Washington D.C as a member of the Obama administration as well as academic research. Her work, however, spends little time discussing the relationship between wiki-government and journalism, and has little to say about news. One of the few academic pieces to directly tie developments in digital transparency to journalistic practices comes (again) from Schudson, who argues that the growth of what he calls “political observatories ... and the new availability of databases for public-interest research”

represent just one feature of the future for news. Yet it is a vital feature, and so far it has received little general notice. Political observatories do not replace journalists, nor do databases shove narratives aside. But the observatories are increasingly valuable partners for journalists, and databases lay new foundations for narrative. Both offer promise for developing the kind of public information that makes democracy possible.” (Schudson 2010, 8)

Of course, the flipside to digitally afforded public transparency is digitally-enabled surveillance, as Kreiss and Howard remind us (Kreiss and Howard 2010). But it is this complex overlap, embedded within a deep and nuanced understanding of that journalism’s political context, which a critical analysis of computational journalism would attempt to understand. The public policy decisions that enable or retard the growth of computational journalism are surely a ripe area for research. Ultimately, given a lengthy enough amount of time, scholars might even be equipped to relate differences in computational practices to system-level differences between a variety of political contexts.

Economics

Inside in a large conference room on the upper-floor of a center city office building, several employees were grilling the Vice-President of Philly.com, the news website for Philadelphia’s two daily newspapers, on her hiring priorities. The year was 2008, and the staff of Philly.com and the newspapers had gathered for an afternoon to plot their future. “You’ve been hiring all these web producers,” one reporter complained, “these kids who just take newspaper content and move it around Philly.com. But what it really seems we need are developers, computer programmers, silicon-valley types. Why aren’t we hiring them?”¹

There was a pause and some uncomfortable throat clearing. “Well,” the Philly.com Vice-President finally ventured, “the problem, you see, is that the web producers I hire are cheap. But the tech team you and I both know we *need* to hire? They aren’t cheap. I can hire six or seven

¹ The following anecdote is taken from the Authors’s newsroom fieldwork, conducted between 2007 and 2011.

web producers for what it costs to hire a single tech person. And we really just can't afford to do that."

Traditionally the analysis of the "economic organization of news," like the analysis of the political context of news making, has been considered on a macro-level. With its origins in both Cold War dynamics (the division of the world into socialist and capitalist economic systems) and the political-economy critique of concentrated media systems, the economic analysis of news production (Baker 2002; McChesney 1995; Siebert 1956) might seem a poor way of understanding computational journalism. But just as in the political lens discussed above, drawing on economic categories to understand the development "big data" journalism might point our analysis in directions we might not immediately expect. Rather than simply focusing on the manner in which large economic systems intersect with various forms of journalistic production, scholars might critically interrogate the way that different institutionally specific resources constrain the options available to various news outlets and industry segments. Such scholarship could attempt to correlate forms of computationally enhanced news production with levels of institutional economic capital. Researchers in this tradition would take seriously the lament of the local newspaper executive, above, the fact that certain technologically-focused innovations appeared out of reach for less wealthy news organizations, and they would interrogate that lament by examining the role played by those seeking to level the playing field-- open-source software makers and volunteers-- in the more evenly spread diffusion of technology.

To date, there has been next to no work done on the relationship between economic resources and computational journalism. Nevertheless, white-papers like the recently released FCC report on the "Information Needs of Communities," with its emphasis on capacity gaps between local and nationally focused news organizations, can serve as a starting point for a less macro-level, more institutionally grounded analysis (Waldman 2011). Of course, a retreat from systemic economic perspectives does not obviate the feasibility of a systemic analytical lens over the long term; like the political context lens, drawing stronger and more generalizable correlations between economic system and journalistic form might be possible over time. In the interim,

however, such correlations would require extensive, critically focused empirical work on the relationship between economics and newsroom innovation.

Institutions and Fields

The shifting focus in the political and economic perspectives discussed above, away from macro-level generalizations and towards institutionally-grounded analysis, points us toward a third lens that might be useful in understanding computational journalism: the so-called “field” or “institutional” perspective. Schudson does not discuss field analysis in his classic series of papers on the sociology of news; indeed, Rodney Benson, who has led the way in importing field perspectives into journalism research, advances mezzo-institutional field research as part of a critique of classic sociological lenses on the production of news, including Schudson’s (Benson 1999; Benson 2004). I do not attempt to take sides in this dispute here, except insofar as I argue that *any one of the three scholarly lenses discussed so far — political, economic, and field — may be useful at different times and insofar as they attempt to answer different questions*. As a new object of study, scholarship on computational journalism would be best served by openness to a variety of methods and theories. And understanding the development of computational journalism from the perspective of the journalistic field has its own advantages, as well as its own drawbacks.

Drawing on Bourdieu’s highly influential work on the sociology of institutional power and applying it to the study of news, Benson argues that

The starting point for understanding the media field paradigm is Pierre Bourdieu’s general theory of fields (champs). Drawing on and modifying Weber’s sociology of religion, Bourdieu sees society as differentiated into a number of semi-autonomous fields (e.g., fields of politics, economics, religion, cultural production, etc.) governed by their own “rules of the game” and offering their own particular economy of exchange and reward, yet whose basic oppositions and general structures parallel each other. (Benson 1999, 466)

“Journalism,” according to this model, would best be understood as its own “field of power” whose general social dynamics are conceptualized as positional and relational. These dynamics

are based on the possession of a particular “habitus” as well as a stock of capital. Institutions (and individuals) would struggle to possess certain levels of socially-defined capital, as well as to alter the definition of what counts as legitimate capital in the first place. Understanding computational journalism in this fashion has at least two benefits. First, it adds a vector of power dynamics to an area of socio-technical life (technological innovation) too often understood from within an “all boats will rise” mentality. Second, and perhaps more importantly, it understands journalism relationally, placing it in the context of other institutional clusters and fields.

This analysis of computational journalism from a Bourdieuean or new institutionalist (Ryfe 2006) perspective would begin from the premise that a new field, the field of data journalism, is emerging alongside traditional fields such as computer science and news reporting. It would study the social, economic, and cultural power struggles that occur inside this field, and the way these struggles shape practices of newswork and journalistic products. Importantly, however, the dynamics of this field would never be seen as operating in isolation from other nearby fields. Developments in traditional journalism, or in computer science, would obviously play key roles in the development of data journalism. And clusters of seemingly unrelated institutions, like non-for-profit foundations, would also play a key role in the shape of the computational journalistic field. Lewis’ work analyzing the relationship between the Knight Foundation, news institutions, and cultural concepts of journalistic professionalism (Lewis 2011) is a particularly impressive example of this kind of research; similar methodologies and perspectives might be applied to the study of the role of the Knight Foundation, open source technology groups like the Mozilla Foundation, and other funders in the development of the computational journalistic field.

Organization-Level Dynamics

Despite the utility of political, economic, and field approaches to the study of computational journalism, each of these perspectives abstracts (to a greater or lesser degree) from the day-to-day organizational processes by which computational practices embed themselves in journalistic work. A fourth perspective, “the social organization of newswork,”

(Schudson 2005) takes us back down to earth, to ground-level newsroom dynamics. Research operating via this lens would examine how workflow routines, levels of technological adaptation, individual rivalries, bureaucratic divisions, and daily process imperatives affect the manner in which newsrooms integrate potentially far-reaching new technologies into their workflow. The manner by which news organizations incorporate powerful algorithms into daily news practices is only one example of this type of approach. While earlier research of the impact of technology on journalistic practice often posited a dynamic in which new technologies generate distinct editorial effects, Boczkowski (2004) has provided researchers with a far more nuanced model by which to understand the relationship between technology and newswork. For Boczkowski, production factors (organizational factors, work routines, and representations of users) shape adoption processes, which are themselves afforded by technological changes. Only at the end of this far more complex process do distinct editorial products begin to emerge.

Most newsroom ethnographies either explicitly or implicitly adopt an organizational routines perspective, with detail-rich, on the ground observations often bringing to light just how slow an halting organizational change actually is. Alongside classic work in this vein (Epstein 2000; Gans 1979; Tuchman 1978), the last decade has witnessed a veritable outpouring of ethnographic newsroom research (Anderson 2010a; Boczkowski, 2004a; Cottle, 2007; Eliasoph, 1997; Klinenberg, 2005; Paterson and Domingo, 2011) research often prompted by macro-level technological change but whose conclusions have often been that ground-level journalistic changes are far less dramatic than might be assumed. And while much of this research has looked at technology in general—the use of crowdsourcing, blogs, Twitter, etc— little of it has explored computational journalism, if by computational journalism we mean (as I argued above) “the combination of algorithms, data, and knowledge from the social sciences to supplement the accountability function of journalism” (Turner 2009). Royal’s provisional research, which analyzed the organizational integration of the *New York Times* Interactive New Department into larger work routines at the *Times*, is one of the few exceptions (Royal 2010). There is much room for additional scholarship in this area. Not only would it be easy to apply recently reinvigorated ethnographic methods to the study of computational practices, but it would be intriguing to see

whether differences existed in the findings of earlier and later digital newsroom ethnographies, and what these differences might say about journalism in a digital age.

Cultural History

The intersection of technological change and newsroom practice is most often analyzed on the level of organizational routines; nevertheless, the growth of big datasets and computational practices can be expected to affect journalistic culture as much as journalistic work. Schudson draws on Sahlins to argue that cultural aspects of society, “while they may be uncovered by detailed historical analysis, cannot be extrapolated from features of social organization at the moment of study. They are part of culture— a given symbolic system within which and in relation to which reporters and officials go about their duties” (Schudson 2005, 187). There is much debate about the role played by culture in the operation of journalism. Benson, referring to forms of culture grounded in national differences, argues that the term should be abandoned. It must be admitted that there is a slight lack of clarity in Schudson’s discussion of the journalistic culture; he does occasionally refer to national cultures and the role they play in the construction of news— national differences that, as Benson argues, may be more usefully be seen as the outcome of field dynamics rather than symbolic systems per se. Nevertheless, Schudson’s arguments about culture and journalism are far subtler and far ranging than Benson allows. From the perspective of praxis, culture might be seen as the accrued, long-term detritus of daily newswork, a detritus that creates the symbolic background against which journalists guide their actions. Or we may see it as the background symbolic system that guides many of the most important and obvious but least understood aspects of journalistic practice. The culture of journalism is invoked by Schudson to guide researchers toward what I would call “interesting but tricky problems”: “journalist’s vague renderings of how they know news when they see it ... [why] news stories are so often personified... [why] reporters write of persons and not structures, of individuals not social forces ... and [journalistic] assumptions about narrative, storytelling, human interest, and the conventions of photographic and linguistic presentation.” (188-89). Obviously, all

these questions can be reoriented to shed light on the nuances of data journalism. Does using big data create a difference sense of what news is? Does data encourage a greater emphasis on structures and social forces as opposed to personalities and incidents? The analysis of organizational routines might begin to answer these questions, but newsroom sociology cannot definitively answer them. For Schudson, Zelizer (Zelizer 1992) Carey (Carey 2008), and others, only a cultural analysis informed by nuanced historical research can do the heavy lifting required to answer these questions.

To date, most of the detailed historical research on the relationship between journalism's symbolic systems, computer programming, and big data remains to be done. While he might not call his research cultural history, Powers work on the evolution of journalistic "computational talk" from 1975 to the present day (Powers, forthcoming) is one possible exception; for Powers, discussions of computers and journalism usually frame these devices as providing continuity of work routines, threats to journalism, and opportunities for professional reform. Powers lengthy analytical time frame and his focus on value, technology, and work thus marks one model to be emulated by computational journalism scholarship. Research that embraces an even longer time-frame, such as Barnhurst and Nerone's analysis of the "forms of news" (Barnhurst and Nerone 2001) or my own work on "news objects" (Anderson, forthcoming) can also be seen as providing openings for the future analysis of the impact of big data on journalistic sense-making.

Technology and the News

In this paper, up until this point, "technology" has been discussed as a form of embedded materiality that intersects with journalism via the mediation of cultural, political economic, organizational, or institutional factors. To conclude, I want to advance the notion of technology as an independent lens of analysis without advocating, I hope, a pernicious form of technological determinism. For most sociology, an over-emphasis on the role played technology in the construction of news constitutes the primary sin, one to be assiduously avoided. But is there a way to talk about technology and the news on its own terms, without reducing said technology to

either a political, economic, cultural, or social construction? I would argue that, to a limited degree, there is.

Some key aspects of computational journalism are themselves partly technological in nature; they are in part determined by the vast increase in computer processing speed and a growing data ubiquity, as seen over the last thirty years. These technological developments are not, in and of themselves the primary determinants of the shape computational journalism has taken in the past or may take in the future. I hope that each of the five lenses discussed so far has properly inoculated readers against visions of technologically guided progress; the meaning and impact of computational journalism is as much affected by politics, economics, institutions, culture, and organizational routines as it is by microchips and Moore's Law. Nevertheless, while not determinative, a scholarly concern with actual role played by materiality and technology in the processes of journalism might emphasize at least three areas: the oft-hidden intersection between *imagined values* and *engineering design* during the construction of journalistic artifacts, the increasingly *hybrid* nature of newsroom sorting and filtering technologies, and the changing status of *journalistic evidence* fostered by the exponential increase in available evidentiary forms in the digital age. To date, little serious work has been done in any of these areas, though extensive examples from field of science and technology studies provide examples of what such a research agenda might look like if applied to journalism (Collins and Evans 2002; Dunbar-Hester 2009; Latour and Woolgar 1976; Latour 2010; Law 1990; Pickering 2003)..

In a technologically oriented study of computational journalism, researchers might first wish to probe the manner in which journalistic values refract through the material lens of digital design. Here, Nissenbaum's work on "values in design" (Nissenbaum 2004) might serve as a potential guide. In designing semantic tagging systems to assist Google and other search engines in the proper sorting of news content, how do engineers filter newsroom notions such as Authorship, the byline, and the fact though taxonomical systems originally designed for use in computer science? What aspects of the traditional story, in short, appear to be the most valued in the artifact building process, and what aspects are neglected? This focus on the manner by which human needs are embedded within and yet refracted through technological artifacts could draw

researchers attention to a second aspect of computational journalism: the hybrid nature of newsroom sense-making technologies. Data-crunching algorithms and other increasingly invisible information ordering devices are neither entirely material, nor are they entirely human— they are hybrid, composed of both human intentionality and material obduracy. In this way, the human becomes partially obdurate and the material partly intentional (Latour 1993). Third and finally, the analysis of computational journalism through a technological lens would focus on the manner in which the explosion of digital “objects” - Tweets, links, blog posts, databases, etc— changes journalists evidentiary calculations about what counts as proper story evidence. While none of these three angles amount to a methodological research prescription *per se*, a technological approach to computational journalism would be as much about asking particular types of research questions and emphasizing different areas of scholarly interest than the previously discussed lenses.

Future Research

Given the previously outlined argument that algorithms should be approached as potential news objects, as well as the claim that computational and “big data” journalism is one of the key research domains for journalism and communications studies in the future, what are some specific topics that media scholars might wish to investigate in this burgeoning area? Readers will note that a number of potential research topics have already been posed, in passing, in the sections above. I want to summarize four of these topics here, and conclude with some thoughts on the importance of cross-national, interdisciplinary research to the study of algorithms in the news, and to the study of computational journalism in general. I want to note that these research areas primarily serve as a starting point for future discussion and do not amount to a full-fledged research *per se*.

Potential items for future research include:

- **Transformations in Journalism Education and Newsroom Socialization:** Given that journalists are now confronted with a series of new digital objects, sources of evidence, and methods of quantitative analysis, will the training of journalism (either formal or informal) change accordingly? How are journalism schools seeking to adapt their curricula to face these new empirical and methodological challenges? And can we compare these shifts in education and socialization across time; for instance, to the rise of so-called “precision journalism” and “computer-assisted reporting” in the 1970s and 80s (Meyer 1991)?
- **Comparative Analysis of Political System Transparency Initiatives and their Impact on Journalism:** In the United States, there have been a number of recent government-led initiatives to push for greater transparency of public information and government data (Schudson 2010). Obviously, however, public-sector attitudes towards transparency differ cross-nationally (see differing attitudes towards the Google Streetview program in the United States and Europe, just to name one recent example). And to the degree that public-sector data transparency affects journalism, different attitudes towards that transparency will affect journalism differently. Analysis of computational journalism as embedded within political systems, then, should be approached from a cross-national perspective as well. This line of analysis would obviously consider legal and regulatory questions as well.
- **Computational News Routines:** Scholarship on journalism in the internet age has begun to make considerable progress in understanding the manner by which larger changes in communicative systems (the so-called change from a one-to-many to a many-to-many informational system) are impacting newswork (see Singer et. al. 2011). Little research has yet been done, however, on the way that algorithms and computational practices are affecting news routines, individual rivalries, bureaucratic divisions, and daily process imperatives in newsrooms. Ethnographic research of this sort would go a long way to advancing our understanding of computational journalism. It would also be interesting to compare newer and older research.

- **Technologies and Tools Adapted by Journalists:** How do journalists understand their own technological tools? What cultural meanings do they invest in algorithms, data-extraction analysis, and other “objects” of computational work? And do their understandings of what news is and the role it plays in society change as a result? This research would adopt the technological and culturalist perspectives on computational journalism outlined above. Such work would explore at least three areas:
 - the intersection between *imagined values* and *engineering design* during the construction of journalistic artifacts,
 - the *hybrid* nature of newsroom sorting and filtering technologies,
 - and the changing status of *journalistic evidence* fostered by the increase in evidentiary forms in the digital age.

— It should be obvious that research of the kind I outline here would require crossing borders in at least two important ways. First, this research would have to be cross-national. Indeed, given a research agenda that focuses so heavily on technology, cross-national research is the best guard against investing technology with deterministic properties. By watching how different organizational, political, regulatory, and economic systems refract technological processes, scholars can gain a more nuanced understanding of the impact of algorithmic and computational practices on journalism.

Second, the research I advocate here would have to cross disciplinary borders as well as national ones. Given the wide variety of systems at play in the emergence of computational journalism, researchers would need to be drawn from fields of computer science, law and regulation, science and technology studies, and journalism studies. The rise of the algorithm inside newsrooms is a powerful development, one that requires an “all hands on deck” approach from scholars hoping to understand these dramatic changes in the way we live, work, communicate, and understand public issues in the 21st century.

Key Research Questions

- How are algorithms influencing journalistic work routines in terms of both gathering and presenting news?
- How do open data / transparency initiatives in different countries influencing computational / data journalism?
- What are some of the key algorithmic processes affecting journalism today (both from within the news industry and from outside it?). How are these algorithms constructed as man-machine hybrids? How do they intersect with editorial practices?
- Do levels of field-specific capital (financial, cultural, etc) affect the uptake of computational journalistic practices?
- Is there a “computational journalistic” culture? Is distinct from traditional journalistic culture?

Bibliography

- Anderson, CW (2011). Between creative and quantified audiences: Web metrics and changing patterns of newswork in local US newsrooms. *Journalism: Theory, Practice, Criticism*. 12(5): 550-566.
- Baker, CE (2002) *Media, Markets, and Democracy*, Cambridge, UK: Cambridge University Press
- Barnhurst, KG and Nerone JC (2001) *The Form of News: A History*. New York: The Guilford Press
- Benson R (1999) Field theory in comparative context: A new paradigm for media studies. *Theory and Society* 28(3): 463-498
- Benson R (2004) Bringing the sociology of media back in. *Political Communication* 21(3): 275-292.
- Benson R (2010) What makes for a critical press? A case study of French and US immigration news coverage. *The International Journal of Press/Politics* 15(1): 3
- Berry DM (2011) The computational turn: Thinking about the digital humanities. *Culture Machine* 12(0): np.
- Boczkowski PJ (2004) The processes of adopting multimedia and interactivity in three online newsrooms. *Journal of Communication* 54(2): 197-213
- Carey JW (2000) 'Journalism and Democracy Are Names for the Same Thing.' Available at: <http://www.nieman.harvard.edu/reports/article/101943/Journalism-and-Democracy-Are-Names-for-the-Same-Thing.aspx>.
- Carey, JW (2008) *Communication As Culture: Essays on Media and Society*. New York: Taylor & Francis
- Medill.northwestern.edu (200x) Coders wanted: Journalism scholarships available for programmers / developers. Available at <http://www.medill.northwestern.edu/admissions/page.aspx?id=58645>
- Cohen P (2010) In 500 billion words, new window on culture. *New York Times*, December 17, 2010
- Cohen S, C Li, J Yang, and C Yu (2011) Computational journalism: A call to arms to database researchers. In *Proceedings of the 5th Biennial Conference on Innovative Data Systems Research*. Sailorman, CA: ACM.
- Dabello M (2011) A genealogy of digital humanities. *Journal of Documentation* 67(3): 480-506
- Daniel A, T Flew, and C Spurgeon (2010) The promise of computational journalism. In *Proceedings of the 2010 Australian and New Zealand Communication Association*. Canberra: ANZCA
- Diakopoulos N, F Kivran-Swaine, and M Naaman (2011) Playable data: Characterizing the design space of game-y infographics. In *Conference on Human Factors in Computing Systems (CHI)*. May, 2011
- Diakopoulos N, M Naaman, and F Kivran-Swaine (2010) Diamonds in the rough: Social media visual analytics for journalistic inquiry. In *Conference on Visual Analytics Science and Technology (VAST)*. October, 2010
- Diakopoulos N and Naaman M (2011) Towards quality discourse in online news comments. In *Proceedings of the ACM 2011 Conference on Computer Supported Cooperative Work*.
- Domingo D and Patterson C (2011) *Making Online News (Volume 2): Newsroom Ethnography in the Second Decade of Internet Journalism*. New York: Peter Lang
- Dossick CS and Neff G (forthcoming) Messy talk and clean technology: Communication, problem-solving and collaboration using building information modeling *The Engineering Project Organization Journal*
- Epstein EJ (2000) *News From Nowhere: Television and the News*. Lanham, MD: Ivan R. Dee
- Gans HJ (1979) Deciding what's news: A study of CBS evening news, NBC nightly news, Newsweek, and Time. New York: Pantheon
- Hallin D and Mancini P (2004) *Comparing Media Systems: Three Models of Media and Politics*. Cambridge, UK: Cambridge University Press
- Hanitzsch T (2008) Comparing journalism across cultural boundaries.: State of the art, strategies, problems, and solutions. In: Weaver D and Loffelholz M (eds). *Global Journalism Research: Theories, Methods, Findings, Future*. London: Blackwell
- Hayles NK (2011) How we think: The transforming power of digital technologies. In Berry, DM (ed) *Understanding the Digital Humanities*. London: Palgrave
- Irfan E and Diakopoulos N (2007) Computational journalism: CS 4803CJ/8903CJ. Available at: http://www.cc.gatech.edu/classes/AY2007/cs4803cj_spring
- Journalism.Columbia.com (2010) Journalism-computer science joint degree. Available at: <http://www.journalism.columbia.edu/page/275-dual-degree-programs/278#computer>.
- Klinenberg E (2005) Convergence: news production in a digital age. *The ANNALS of the American Academy of Political and Social Science*, 597(1): 48

- Kreiss D and Howard P (2010) New challenges to political privacy: Lessons from the first U.S. Presidential race in the web 2.0 era. *International Journal of Communication* 4 (2010)
- Latour B (1993) *We Have Never Been Modern*. Harvard: Harvard University Press
- Latour B (2004) Why has critique run out of steam? *Critical Inquiry* 30 (2): 225-248
- Lewis SC (2011) Journalism innovation and the ethic of participation: A case study of the knight foundation and its news challenge. Unpublished doctoral dissertation, University of Texas-Austin
- Manovich L (2011) Trending: The promises and the challenges of big social data. Available at <http://manovich.net/2011/04/23/new-article-trending-the-promises-and-the-challenges-of-big-social-data>
- McChesney RW (1995) *Telecommunications, Mass Media, and Democracy: The Battle for the Control of US Broadcasting, 1928-1935* Oxford: Oxford University Press.
- McPherson T (2009) Introduction: Media studies and the digital humanities. *Cinema Journal* 48(2): 119-123
- Meyer, P (1991) *The New Precision Journalism* Bloomington, Indiana: Indiana University Press
- Michel J , YK Shen, AP Aiden, A Veres, MK Gray, JP Pickett, and Hoiberg D (2011) Quantitative analysis of culture using millions of digitized books. *Science* 331(6014): 176-82
- Ngyuen D (2010) Scraping for journalism: A guide for collecting data. Available at: <http://www.propublica.org/nerds/item/doc-dollars-guides-collecting-the-data>.
- Nissenbaum H (2004) Privacy as contextual integrity. *Washington Law Review*. 79: 119-158
- Noveck B (2009) *Wiki Government: How Technology Can Make Government Better, Democracy Stronger, and Citizens More Powerful*. Washington DC: Brookings Institute Press
- Noveck B (2011) The single point of failure. *Innovating Government: Information Technology and Law Series 20 (2): 77-99*
- Nussbaum E (2009) The new journalism: goosing the gray lady. *New York Magazine*, January 11, 2009.
- Powers M (forthcoming)
- Rosenbloom PS (2010) Towards a conceptual framework for the digital humanities. Available at <http://pollux.usc.edu/~rosenblo/Pubs/TCFDH.pdf>
- Rothberg M (2010) Quantifying culture? A response to Eric Slauter. *American Literary History* 22 (2): 341
- Royal C (2010) The journalist as programmer: A case study of *The New York Times* interactive news technology department. Presented at *The International Symposium For Online Journalism*. Austin, TX
- Ryfe DM (2006) Guest editor's introduction: New institutionalism and the news. *Political Communication* 23(2): 135-144
- Schudson M (1997) The sociology of news production. In Berkowitz DA (ed) *Social Meanings of News: A Text Reader*. London: Sage, pg 7-20.
- Schudson M (2010) Political observatories, databases & news in the emerging ecology of public information. *Dædalus* 139(2): 100-109.
- Schudson, Michael. 2005. Four approaches to the sociology of news. In Curran J and Gurevitch M (eds.) *Mass Media and Society, (4th ed)* London: Hodder Arnold
- Segel E and Heer J (2010) Narrative visualization: Telling stories with data. *IEEE Trans. Visualization & Comp. Graphics (Proc. InfoVis)* 16 (6): 1139-48.
- Siebert, FS, T Peterson, and Schramm W (1956) *Four Theories of the Press*. Champaign, IL: University of Illinois Press.
- Singer, J.B., Hermida, A., Domingo, D., Heinonen, A., Paulussen, S., Quandt, T., Reich, Z., Vujnovic, M. (2011). *Participatory Journalism: Guarding Open Gates at Online Newspapers*. New York: Wiley-Blackwell.
- Slauter E (2010) Revolutions in the meaning and study of politics. *American Literary History* 22(2): 325-
- Stoeffel K (2011) Huffington Post hires Andrei Scheinkman to lead interactive news features team. Available at <http://www.observer.com/2011/media/huffington-post-hires-andrei-scheinkman-lead-interactive-news-features-team>
- Stone W (1972) Abacists versus algorists. *Journal of Accounting Research* 10(2): 345-350.
- Taylor, Megan (2009). Washington Post's 'web ninjas' build map-timeline combo. Available at <http://www.pbs.org/mediashift/2009/02/washington-posts-web-ninjas-build-map-timeline-combo047.html>.
- Tuchman, G. 1978. *Making News*. New York, NY: Free Press.
- Turner F and Hamilton JT (2009) *Accountability Through Algorithm: Developing the Field of Computational Journalism*. Available at http://dewitt.sanford.duke.edu/images/uploads/About_3_Research_B_cj_1_finalreport.pdf
- Waisbord SR (2000) *Watchdog Journalism in South America: News, Accountability, and Democracy*. New York, NY: Columbia University Press
- Waite M (2009) The key lesson I learned building PolitiFact: Demos, not memos available at <http://www.mattwaite.com/posts/2009/apr/27/key-lesson-i-learned-building-politifact-demos-not/>.
- Waldman S (2011) *The Information Needs of Communities: The Changing Media Landscape in a Broadband Age*. Washington DC: Federal Communications Commission

Yong E (2011) The Renaissance man: how to become a scientist over and over again. Available at: <http://blogs.discovermagazine.com/notrocketscience/2011/06/08/the-renaissance-man-how-to-become-a-scientist-over-and-over-again/>.

Zelizer B (1992) Covering the Body: The Kennedy Assassination, the Media, and the Shaping of Collective Memory. Chicago: University of Chicago Press.