

Oxford Research Encyclopedia of Communication

Algorithms and Journalism

Rodrigo Zamith

Subject: Journalism Studies Online Publication Date: Feb 2019

DOI: 10.1093/acrefore/9780190228613.013.779

Summary and Keywords

Algorithms today influence, to some extent, nearly every aspect of journalism, from the initial stages of news production to the latter stages of news consumption. While they may be seen as technical objects with certain material characteristics, algorithms are also social constructions that carry multiple meanings. Algorithms are neither valueless nor do they exist in isolation; they are part of *algorithmic assemblages* that include myriad actors, actants, activities, and audiences. As such, they are imbued with *logics* that are only sometimes reflective of journalism's.

Algorithms have played an active role in a broader quantitative turn within journalism that began in the 1970s but rapidly accelerated after the turn of the century. They are already used to produce hundreds of thousands of articles per year through *automated journalism* and are employed throughout the many stages of human-driven newswork. Additionally, algorithms enable *audience analytics*, which are used to quantify audiences into measures that are increasingly influencing news production through the abstractions they promote. Traditional theoretical models of newswork like gatekeeping are thus being challenged by the proliferation of algorithms.

A trend toward algorithmically enabled *personalization* is also leading to the development of *responsive distribution* and *curated flows*. This is resulting in a marked shift from journalism's traditional focus on shared importance and toward highly individualized experiences, which has implications for the formation of publics and media effects. In particular, the proliferation of algorithms has been linked to the development of *filter bubbles* and evolution of *algorithmic reality construction* that can be gamed to spread misinformation and disinformation.

Scholars have also observed important challenges associated with the study of algorithms and in particular the opaque nature of key algorithms that govern a range of news-related processes. The combination of a lack of *transparency* with the complexity and adaptability of algorithmic mechanisms and systems makes it difficult to promote *algorithmic accountability* and to evaluate them vis-à-vis ethical models. There is, currently, no widely accepted code of ethics for the use of algorithms in journalism.

Finally, while the body of literature at the intersection of algorithms and journalism has grown rapidly in recent years, it is still in its infancy. As such, there are still ample opportunities for typologizing algorithmic phenomena, tracing the lineage of algorithmic processes and the roles of digital intermediaries within systems, and empirically evaluating the prevalence of particular kinds of algorithms in journalistic spaces and the effects they exert on newswork.

Keywords: algorithmic journalism, algorithms, audience analytics, automated journalism, computational journalism, ethics, filter bubbles, logics, journalism studies, transparency

The Algorithmic Turn in Journalism

The use of algorithms in journalism is not itself novel. However, the ongoing digitization of journalism (Boczkowski, 2005)—and, more broadly, of everyday life (Latzer, Hollnbuchner, Just, & Saurwein, 2016)—has made algorithms increasingly central to nearly every journalistic activity, from gathering information (Wallace, 2018) to producing stories (Waddell, 2018) to distributing information to citizens (Weber & Kosterich, 2018). Furthermore, the proliferation of algorithms has altered important social arrangements (Braun, 2015) and the logics that govern newswork (Carlson, 2018) and further changed the manner in which audiences engage with news content (Nelson & Lei, 2018) while making possible new tactics in disinformation campaigns (Marwick & Lewis, 2017). In short, journalism finds itself amid an *algorithmic turn* (Napoli, 2014) as its key components are transformed (Ananny, 2016; Carlson, 2018; Just & Latzer, 2017).

A considerable amount of scholarly attention has been paid to these developments in recent years. This article highlights the current state of knowledge across the key areas of inquiry that have received the most attention. It begins by describing different conceptualizations of the term “algorithm.” It then highlights how algorithms have been situated within broader systems and reviews the development of machine-driven journalism. Following that, the relationships between algorithms and both audiences and publics are examined. Approaches and challenges involving accountability in this realm are then described. Finally, the state of knowledge is evaluated and future research directions are identified.

Definitions of “Algorithm”

There is no single, widely accepted definition of the term *algorithm* (Gillespie, 2016), though two key strains are dominant in the literature: technical and social. Technical definitions typically focus on the material aspects of algorithms, such as the code that powers them. Those definitions are often critiqued as being *technologically deterministic*, or implying that the materiality of technology is primarily responsible for the development of social structures and cultural values. Technical definitions are more common in the realms of computer science and engineering, though variants are sometimes found in journalism studies. Social definitions emphasize the social processes

Algorithms and Journalism

that develop and recalibrate the technological artifacts associated with algorithms, which are in turn impacted by organizational and supraorganizational environmental conditions (Napoli, 2014; Weber & Kosterich, 2018).

Latzer et al. (2016) offer a useful technical definition: an algorithm is “a finite series of precisely described rules or processes to solve a problem” or accomplish a task, generally through “a sequence of stages that transforms input through specified computational procedures (throughput) into output” (p. 397). In this model, the algorithm performs its work during the throughput stage. Algorithms vary in both complexity and transparency (Shorey & Howard, 2016). For example, a simple algorithm might be specified in a user-auditable JavaScript file that detects the size of a device’s screen and, if it is below a threshold, redirects the user to a mobile-friendly page. Alternatively, an opaque server-side algorithm may draw from multiple inputs—a browsing history spanning multiple websites, the language used in user comments, and the location of a device—to automatically curate a content feed based on a model that is constantly evolving. Once designed, an algorithm may run autonomously and with little oversight from its creator (Zarsky, 2016).

Most scholarship within journalism studies draws on social definitions. Gillespie (2016) identifies three such types. First is algorithm as synecdoche, in which the term encompasses the “ill-defined network” (Gillespie, 2016, p. 23) or system that is comprised of a range of actors, actants, activities, and audiences. Second is algorithm as a talisman, where the term is used to highlight a set of value associations, such as unparalleled objectivity and rationality. Third is algorithm as a commitment to procedure, in which the term is used to reflect social phenomena driving the formalization of social facts into measurable data and the development of step-by-step interventions that can be enacted by a machine.

In adopting a social definition, scholars may be sensitized to the ways in which algorithms not only shape but are shaped by social contexts (Ananny, 2016). However, while referring to large and complex systems through the term “algorithm” can offer a more holistic lens, it also risks promoting “mystified abstractions” due to the difficulty of interrogating so many moving parts at once (Gillespie, 2016, p. 23). Ultimately, the intellectual aim of a work should guide the adoption of a particular definition. However, scholars should be aware that the term can mean different things and should therefore clarify their own use of it to avoid confusion.

Algorithms and Systems

The proclivity toward social definitions has led to a focus on organizational- and systems-level theoretical frameworks that draw from sociology and science and technology studies. Those frameworks typically share an overarching understanding that journalism is comprised of a complex network (or networks) consisting of human and nonhuman constituents that are constantly interacting with one another and consequently influencing and being influenced by multiple domains of activity (e.g., journalism, law,

Algorithms and Journalism

and information technology). Key concepts from that literature, detailed in this article, have proven useful for examining the fluidity of and power differences within news-related networks, challenging notions of algorithmic neutrality by detailing how relationships between the social and the technical are reciprocated, and exploring how behaviors are structured when diffuse communities converge around shared interests in journalism.

Algorithmic Assemblage

Ananny (2016) calls attention to the notion of *algorithmic assemblage*, or the “institutionally situated computational code, human practices, and normative logics that creates, sustains, and signifies relationships among people and data through minimally observable, semiautonomous action” as a unit of analysis (p. 99). Such an approach moves away from treating “algorithms as fetishized objects” (Crawford, 2016, p. 89) and emphasizes the social contexts in which humans intersect with code and platforms. Moreover, the term calls attention to the reality of algorithms in practice: they rarely operate in isolation, and it is difficult to study an algorithm as a singular entity.

Braun’s (2015) work illustrates that reality by highlighting *transparent intermediaries*—companies and technical infrastructures whose operations generally go undetected by end users but that play a major role in shaping media products and the manner in which they are distributed. Braun argues that technical affordances in software designed by (often nonjournalistic) third parties can influence the flow and availability of content. For example, algorithms developed by YuMe, an advertising company, bridged the needs of MSNBC and its advertisers by pairing commercials with embedded videos only when the environment was deemed brand-safe. That, in turn, permitted MSNBC to offer users a video player with greater sharing functionality, which altered the way in which information was distributed. In short, YuMe was a crucial actor in the algorithmic assemblage oriented around MSNBC’s video player even though few users were aware of the company (Braun, 2015).

Similarly, algorithms are often built atop one another; they routinely reference existing code. For example, an algorithm that evaluates sentiment from user comments and generates a visualization of the discussion may rely on software libraries like the Natural Language Toolkit and jQuery, which in turn are developed under different software models that have distinct affordances and by teams of developers who have very different aims. Tracing the lineage of actants and deciding where an algorithm “ends” thus becomes an arduous task and may not be possible if a single link is closed-source.

Carlson (2018) adds that analyses of algorithmic assemblages not only point to institutional workings but also help highlight “the justificatory rhetoric that legitimate their knowledge structures vis-à-vis existing knowledge structures” (p. 1761). Indeed, algorithms are often described as being more neutral and objective (Ananny & Crawford, 2018), a discourse that is contrasted with the subjectivity of human newswriters. That discourse is especially important during a time of low trust in news media and leads to

Algorithms and Journalism

claims that algorithms are superior to their non-machine counterparts because of their ability to overcome inherent biases (Haim & Graefe, 2017). As Carlson (2018) puts it, “if the ideal of journalistic knowledge production is objectivity, then news algorithms are positioned as its apotheosis” (p. 1765).

Algorithms and Logics

Situating algorithms within an assemblage highlights that they are neither neutral nor objective (Gillespie, 2014; Klinger & Svensson, 2018). Algorithms are laden with values as they attempt to replicate logics within and outside journalism (Lindén, 2017). For example, Petre (2018) found that analytics companies discussed and engineered their products in ways that reflected some journalistic norms in order to gain legitimacy for their software and increase the likelihood that journalists would consent to analytics-driven labor discipline. In particular, there have been efforts to evaluate algorithms in relation to *media logic*, or “the process through which media present and transmit information” (Altheide & Snow, 1979, p. 10). At the root of that concept is the idea that newswork is guided by some set of formal and informal rules. Some scholars have argued that a single news media logic exists (e.g., Asp, 2014), though others believe journalism is simply too heterogeneous, complex, and divided to be treated singularly (e.g., Couldry, 2008).

The technical nature of an algorithm requires that rules and models be formalized and clearly specified through code. However, while newswork is guided by some rules that are easy to articulate and operationalize (e.g., word limits), many present difficult challenges (e.g., the inverted pyramid structure, which requires prioritizing information). It can also be difficult to identify the informal rules that guide news production. In particular, scholars have argued that much of newswork is based on *tacit knowledge*, or ideas and processes that are understood on a subconscious level but not easily articulated (Polanyi, 1966; Zamith, 2018A).

Imbuing human (media) logic(s) into an algorithm is thus a challenging task, and many scholars do not believe it is currently possible to completely do so (see Klinger & Svensson, 2018). For example, Lindén (2017) argues that although “journalism logic on its lowest level can easily be built into algorithm,” the “more complex forms of journalism are harder to turn into news applications due to the high degree of uncertainty around the application of rules” (p. 71).

Scholars have also argued that running an algorithm is not an abdication of human judgment; rather, it is an act of summoning the human labor, knowledge practices, and decision-making that went into the creation of that algorithm (Braun, 2015). This is well captured by the concept of *delegation*, or the reciprocal relationship between the social and the technical wherein a major effort is turned into a minor one through the use of technology, which in turn delegates behavior back onto the social (Latour, 1988). Put differently, when enacted, the logic employed by a media-related algorithm leads to the refinement of the formal and informal rules that govern media logic(s). As the use of

Algorithms and Journalism

algorithms proliferate, scholars must remain attune to changes in the logic(s) that govern newswork.

Algorithms and Boundaries

As algorithms become more salient within journalism, some actors once outside or on the margins of journalistic spaces have moved closer to the center (Usher, 2016). In particular, scholars have highlighted the growing role of technologists who often have no journalistic background (e.g., software developers, user interface designers, and data scientists) and the ways in which they interact with and alter interactions by existing actors (Karlsen & Stavelin, 2014; Lewis & Usher, 2016). That work has explored algorithms within three conceptual lenses that help make sense of the interactions between heterogeneous actors: trading zones, boundary objects, and social worlds.

A *trading zone* refers to the “intermediate domain in which procedures could be coordinated locally even where broader meanings clashed” (Galison, 1997, p. 48). Put differently, trading zones are spaces in which diverse communities—sets of actors that have distinct logics and epistemological bases—come together to hammer out solutions. That coordination requires the development of sufficient mutual understanding to permit interdisciplinary productivity. Through those interactions, meanings and logics are adjusted across communities, though they need not fully converge.

Boundary objects refer to “objects which are both plastic enough to adapt to local needs and constraints of the several parties employing them, yet robust enough to maintain a common identity across sites” (Star & Griesemer, 1989, p. 393). They may be real or imaginary and are interpreted and used in distinct ways by different communities but must be rigid enough to be recognizable across more than one domain. Boundary objects thus serve as core entities that can link communities together and permit heterogeneous sets of actors to collaborate on a common task.

Social worlds refer to the networks of actors whose cooperative activities permit the production of objects for which that “world” (e.g., data journalism) is known (Becker, 2008). Social worlds require heterogeneous actors to develop sufficient agreement on meanings and conventions to organize labor. Within a given world, particular actors gain and lose status as members of those worlds come to see certain laborers and forms of labor as increasingly or decreasingly “good” or “valuable.” Actors may belong to multiple social worlds simultaneously and perform different functions within each, though the level appreciation for them and their labor may vary across worlds, which are themselves fluid.

These conceptual lenses have proven useful for examining how journalistic challenges are defined by different communities and how those communities come together in an attempt to solve those challenges through the use of algorithms and technology more broadly. For example, Lewis and Usher (2014) found that the lack of sustained and enduring exchanges made it difficult for chapters of Hacks/Hackers—a transnational grassroots organization that aims to bridge journalists and technologists—to generate

momentum for more in-depth sharing, especially when there was limited support from local institutions. Similarly, scholars have found a growing desire among journalists to collaborate with technologically oriented actors in order to develop algorithmic actants that advance journalistic values (Lewis & Usher, 2016; Lindén, 2017). Scholars have also argued that the mastery of algorithmic actants is being increasingly viewed as valuable work within journalism, and in particular its emerging forms (e.g., computational journalism and data journalism), thereby shifting power relations as the actors possessing such skills are accorded elevated status (Lewis & Zamith, 2017). In short, this stream of work has emphasized that algorithms are not only becoming more common in journalism but also gaining symbolic value, which has material implications for newswriters and consequences for the structures that govern newswriting.

Machine-Driven Journalism

Journalism is amidst a quantitative turn, with data and computation becoming increasingly central to the gathering, analysis, and publishing of news content (Coddington, 2015). The beginning of this turn is often attributed to the idea of “precision journalism,” which was introduced by Phil Meyer in the early 1970s and sought to make journalism more scientific by promoting the systematic collection and analysis of quantitative data (Coddington, 2015). Precision journalism spawned “computer-assisted reporting” (CAR), which emphasized the use of computation to realize the aims of precision journalism. CAR gained some popularity in the 1980s and 1990s but remained closely tied to investigative journalism (Parasie & Dagiral, 2012). In the early 2000s, the notion of “data journalism” began to gain traction. Data journalism is often seen as an outgrowth of CAR (Coddington, 2015; Splendore, 2016), though Hammond (2015) contends that the narrative of continuity obscures a broader epistemological shift shaped by the parallel emergence of the Big Data phenomenon. While data journalism also promotes the use of algorithms to gather and analyze information (e.g., scraping documents), it differs from the CAR tradition in its emphasis on using algorithms to help convey information, such as through responsive data visualizations, and is less coupled to investigative work (Zamith, 2019).

While CAR and data journalism both routinely involve the use of algorithms, they view algorithms as aides in a human-centric enterprise (Coddington, 2015). In the mid-2000s, machine-driven forms of journalism started to gain cultural capital within journalistic spaces. These forms view algorithms as co-producers of information, or even autonomous creators, and highlight distinctive value systems and routines (Splendore, 2016). Various terms are used to describe these forms, including “algorithmic journalism” (Diakopoulos, 2015; Dörr & Hollnbuchner, 2017), “automated journalism” (Carlson, 2015; Caswell & Dörr, 2018; Napoli, 2014), “computational journalism” (Karlsen & Stavelin, 2014), “machine-written news” (van Dalen, 2012), and “robot journalism” (Clerwall, 2014). This wide array of terms is often used interchangeably in the scholarship, which makes it difficult to quickly convey the centrality of algorithms within the object of study. Splendore (2016) helpfully delineates between *computational journalism*, which

Algorithms and Journalism

emphasizes the use of computation within the production process, and *algorithmic journalism*, which relies primarily on computation and automation throughout the process of news production (see also Coddington, 2015). However, in referencing the latter, it is perhaps best to avoid ambiguity and use the term *automated journalism*, or journalism driven by “algorithmic processes that convert data into narrative news texts with limited to no human intervention beyond the initial programming” (Carlson, 2015, p. 416).

Automated journalism already impacts day-to-day news production at mainstream news organizations. Narrative Science and Automated Insights, two commonly referenced organizations, already automatically generate hundreds of thousands of stories each year for media customers like The Associated Press, The Big Ten Network, and Forbes (Lindén, 2017). They typically draw on either structured information (e.g., play-by-play summaries) or semistructured information (e.g., annual corporate filings), identify the most significant information based on a set of rules, and generate an article by inserting machine-generated text based on that information into one of many predefined templates. However, scholars have identified a number of limitations with automated journalism systems, such as their inability to fully interpret or contextualize factual information (Dörr, 2016). Additionally, many journalists have lamented the creative limitations of current offerings, as well as their reliance on isolated and one-dimensional data feeds (Thurman, Dörr, & Kunert, 2017), making automated journalism currently suitable for only a few journalistic domains. Caswell and Dörr (2018) describe an approach for automatically generating complex event-driven narratives that is accessible to social scientists, and their study not only outlines the steps involved in natural language generation but also highlights the technical and editorial challenges of extending automation beyond simple descriptions.

Scholars have found that readers are already unable to discern between automatically generated content and content produced by human journalists (Clerwall, 2014; Graefe, Haim, Haarmann, & Brosius, 2018). Additionally, cuing readers to an article’s algorithmic origin yields more favorable perceptions of some dimensions of credibility, quality, and readability (Haim & Graefe, 2017). These developments thus raise the question: “Why should news organizations allocate resources to human writers?” (Clerwall, 2014, p. 527). That question, in turn, may be viewed as a justification for valorizing technical labor within certain journalistic domains—though humans continue to be seen as offering more creativity and flexibility, as well as superior analytical skills (van Dalen, 2012).

Scholars have found surprisingly optimistic views of the impact of algorithms on the future of news production (given journalists’ historical techno-pessimism). For example, Lindén (2017) found that journalists viewed the use of algorithms, and automated journalism in particular, as a way to ensure the continuity of news operations by increasing efficiency and creating new forms of employment in journalism—though many also saw it as a threat that might drive further losses of editorial and noneditorial jobs (see also Carlson, 2015). These findings are consistent with research by Powers (2012), who focused more generally on technologically specific work. However, Lindén (2017) noted that future journalists may be most affected by automation as they would be

blocked from entry-level jobs that are easier to automate and through which they could begin to acquire important tacit knowledge.

Algorithms and Audiences

Over the past century, there has been a movement in media industries toward ever-greater rationalization of audience understanding, or the use of scientific methods to construct audiences based on data. Zamith (2018B) contends that journalism finds itself in the middle of a third wave toward the rationalization of audience understanding due largely to the proliferation of *audience analytics*, or “systems that enable the measurement, collection, analysis, and reporting of digital data pertaining to how content is consumed and interacted with” (p. 421). Those systems are driven by algorithms that log data requests and capture a range of user behaviors (e.g., how far someone scrolled down a page), aggregate data to identify patterns (e.g., trending stories) and translate data points into user-friendly metrics (e.g., “engagement”), and present them through intuitive interfaces.

Multiple audience analytics systems exist—Chartbeat, Google Analytics, and Parse.ly are among the most common as of 2018—and they are typically developed by organizations outside the journalism space (Belair-Gagnon & Holton, 2018; Petre, 2018). Those systems are often used in conjunction with one another and sometimes supplemented by a custom-built system specific to the news organization (or chain). It is clear from the literature that, by 2018, the majority of newsrooms employed at least one audience analytics system and monitored metrics at *some* level, though the extent of their use varies across sites (Zamith, 2018B). Christin (2017) points to three common resistance strategies: foot-dragging (refusing to engage with analytics), gaming (undermining analytics by aiming to inflate key metrics), and open critique (calling attention to shortcomings in analytics systems).

While different systems may focus on some of the same metrics, they employ very different algorithms to collect, synthesize, and present that information. Consequently, competing systems may provide radically different information about the same phenomenon (e.g., which stories are trending). Different actants may thus generate very dissimilar abstractions of audiences to newswriters, which in turn yield distinct constructions (Belair-Gagnon & Holton, 2018). That disconnect not only further confuses newswriters who already have trouble making sense of audience metrics (Graves & Kelly, 2010) but, more importantly, has serious consequences for how journalism is done (Zamith, 2018B). Indeed, constructed audiences inform decision-making at multiple levels, from calculations of newsworthiness (Zamith, 2018A) to organizational strategy (Petre, 2018). Despite their fetishized neutrality as byproducts of algorithms, Christin (2018) found that American and French journalists saw different things when looking at audience metrics and consequently reproduced cultural differences in their attitudes toward and uses of metrics.

Algorithms and Journalism

The algorithms that enable audience analytics also play an important role in editorial decision-making, which is becoming increasingly algorithmic in nature. *Gatekeeping theory*, in the context of journalism, is rooted in the notion that (human) editors serve as central nodes in the process of shaping news products and, through keeping their gates, decide not only what information gets through but also what it looks like once it has passed that gate (Shoemaker & Vos, 2009). Several scholars have found that algorithmically derived audience metrics like “popularity” are already exerting some influence on the placement and deselection of content, though most news organizations still use human editors to manage their website (Zamith, 2018B). However, by the mid-2010s, a great deal of news consumption was already occurring on the websites and apps of news aggregators, search engines, and social media companies, which are driven by algorithmic decision-making (Powers, 2017).

Gillespie (2014) has pointed to the rising importance of *public relevance algorithms* that select what is most relevant from a corpus of data and decide how to place that content in front of audiences. They include, for example, the algorithms behind the automated curation of social media feeds and sites like Google News. Gillespie observes that such algorithms blur important distinctions between editorial and distribution concerns more broadly. The proceduralized choices enacted by human-designed algorithms have in some ways supplanted traditional editorial logic, which depends on the subjective choice of experts (Gillespie, 2014). Crucially, those proceduralized choices depend on data that are structured and quantified. However, much of what newswriters value and find consequential cannot be easily measured, binned, or regressed (Braun, 2015). That, in turn, raises important questions about what is lost in deferring to algorithms.

In editing what users see, algorithms effectively administer a gatekeeping role (DeVito, 2017). From another perspective, journalism’s increasingly social, platform, and algorithmically driven orientation has rendered classic gatekeeping theory inadequate for explaining contemporary news production (Wallace, 2018). News organizations have become increasingly dependent on news aggregators, search engines, and social media sites for their traffic—and consequently become dependent on intermediaries that have different aims and their own, ever-changing algorithms (Nielsen & Ganter, 2018). To that end, Wallace (2018) proposes a *digital gatekeeping* model based on four types of gatekeepers (journalists, individual amateurs, strategic professionals, and algorithms) that employ distinct gatekeeping mechanisms that operate simultaneously and iteratively within a fluid network.

Algorithms and Publics

News media have long been viewed as a key set of agents in the formation of publics by brokering communication across myriad actors and offering a reservoir of common knowledge that serves as the basis for public deliberation and the formation of public opinion. The proliferation of new communication technologies in recent decades has made it easier for publics to fragment, raising concerns about algorithmically aided

polarization and the spread of disinformation through the manipulation of algorithmic logics.

A trend toward algorithmically enabled *personalization*, or the individualization of content and experiences, is often at the center of these critiques (Just & Latzer, 2017). For example, Braun (2015) argues that the proliferation of algorithms, and public relevance algorithms (Gillespie, 2014) in particular, has helped drive a shift toward *responsive distribution* whereby messages are selectively distributed based on algorithmic surveillance and the interpretation of individuals' past behaviors. That, in turn, leads to fluid group memberships within *networked publics* (boyd, 2007) or *calculated publics* (Gillespie, 2014) based on algorithmic decisions, thereby shaping key conditions under which individuals congregate, socialize, and exchange information. Thorson and Wells (2016) similarly refer to *curated flows* in contending that each member of the public now sits at the epicenter of multiple, intertwined content flows. The incoming flows are highly individualized and may be affected by user choices (e.g., subscribing to a topic), user behaviors (e.g., topics they previously accessed), the choices and behaviors of those in their networks (e.g., the topics subscribed to and previously accessed by friends), and algorithmic biases based on strategic calculations (e.g., advertising considerations). Consequently, one's information experience—and media's effect on that person—may also be highly individualized.

This shift toward personalization is a marked change from journalism's traditional focus on shared importance (Thurman, 2011). While journalists have long personalized news to a certain extent (e.g., tweaking a national wire story to emphasize a local angle), their efforts were restricted to large audiences (Thurman & Schifferes, 2012). Algorithms, in contrast, can easily scale and thus shift the focus away from "what deserves attention for this community?" and toward "what does this individual want?" (Anderson, 2011). As Carlson (2018) argues, "this shift represents a core departure from how journalism has been understood and cannot be contained as an extension of journalism's professional logic" (p. 1765).

The shift toward personalization has also impacted the ways in which members of publics interact with one another and are exposed to phenomena. A phenomenon that has received considerable attention is the emergence of *filter bubbles*, where users sort themselves and are sorted into spaces where information perceived as agreeable is privileged (Pariser, 2012). Algorithms play a central role in that sorting process: they not only create the possibilities for self-sorting (e.g., providing filters that may be selected) but also work invisibly to sort users into particular spaces by highlighting certain content, users, and paths. That, in turn, has been linked to the increased fragmentation and polarization of publics as users congregate in *echo chambers* that strengthen preexisting beliefs (see Flaxman, Goel, & Rao, 2016). Scholars have found that individuals are often unaware of the extent to which news sources and intermediaries track user data and generally do not understand how algorithms are used to deliver personalized experiences (Bucher, 2017; Powers, 2017). However, recent evidence suggests that fears about algorithmic filter bubbles in the context of online news may be exaggerated (Flaxman et

Algorithms and Journalism

al., 2016; Haim, Graefe, & Brosius, 2018), and it has been posited that news algorithms could be purposely designed to bridge bubbles (Zamith & Lewis, 2014).

Scholars have also observed that *algorithmic reality construction* (Just & Latzer, 2017) can be gamed when actors take advantage of the rules and biases of the algorithms that govern selection and visibility. In particular, scholars have become increasingly concerned about the role of algorithms in enabling and facilitating the spread of misinformation and disinformation. Bots are now routinely used on Twitter to amplify disinformation and negative information about particular individuals (Marwick & Lewis, 2017). Similarly, “Google bombs” have been used to surface content in search results by creating large, artificial link networks to fool Google’s algorithms into classifying that content as more authoritative (Bar-Ilan, 2007). There are also concerns about the inverse: the threat of invisibility as algorithms make content or individuals disappear from view (Bucher, 2012). While deliberate exclusion may be viewed as censorship, “incidental” algorithmic invisibility emerging from algorithmic curation may be viewed as less insidious, despite having similar consequences (Shorey & Howard, 2016). However, contemporary concerns about epistemology—including the “fake news” phenomenon—should not be understood exclusively through the lens of algorithmic reality construction (Waisbord, 2018). Parallel sociocultural changes within and beyond the field of journalism over the past two decades must also be incorporated into examinations of epistemology and the diffusion of information, requiring scholars to revisit linkages between journalism and publics. As Waisbord argues, scholars must situate the study of journalism and truth within a broader communication ecology that is sensitive to the shifting conditions for public communication and truth-telling, be attune to journalism’s fragmented state and inability to serve as the dominant definer of news as truth, and be more critical of new technologies—including algorithmic artifacts—instead of focusing on their potential to “save” news and promote democratic engagement.

Algorithms and Accountability

As algorithms have become more salient in journalism, scholars have begun to emphasize the importance of *algorithmic transparency* at multiple levels, from algorithmic mechanisms to the software system’s logic (Diakopoulos, 2016). This is consistent with a broader trend toward adopting transparency as a norm within journalism, a phenomenon driven partly by declining trust in news media. Moreover, transparency is often seen as a valuable check on the power of algorithms and a way to reduce the authoritarianial confusion around them (Carlson, 2018). However, algorithms are regularly described as “black boxes” because the ones most interesting to scholars are complex and opaque, making them difficult to study as empirical objects. This opacity is largely driven by public relations concerns, competitive dynamics, and ulterior motives (Diakopoulos, 2015; Latzer et al., 2016). For news algorithms in particular, a lack of business incentives for disclosure and a concern about overwhelming users with information serve as key obstacles for greater transparency (Diakopoulos & Koliska, 2017).

Some scholars have argued that algorithmic transparency is insufficient. The transparency ideal is driven by a logic whereby “observation produces insights which create the knowledge required to govern and hold systems accountable” (Ananny & Crawford, 2018, p. 974). Put differently, it equates the ability to peer inside systems with the power to govern them, yielding a potentially dangerous “transparency illusion” (Heald, 2006). Moreover, the technical sophistication of a system may render it incomprehensible even to its creators (Burrell, 2016), especially if some aspects of an algorithmic system never take durable, observable forms (Diakopoulos, 2016). For example, adaptive systems are constantly learning from new input, and any effort to open up an algorithm’s source code and training and testing datasets is thus temporally bound. Consequently, that effort would yield only a snapshot of an algorithm’s functionality, missing prior and future iterations that complicate the notion of a “single” system (Ananny & Crawford, 2018).

Increased attention has been paid to the development of an *ethics of algorithms* that not only emphasizes accountability in lieu of transparency but also critically evaluates the broader algorithmic assemblage. For example, Ananny and Crawford (2018) argue that it is more beneficial to hold systems accountable by looking across systems rather than peering inside them, thus viewing them as broad sociotechnical systems that do not merely contain complexity but rather enact complexity by linking and intertwining with assemblages of humans and nonhumans. Indeed, certain news algorithms, such as those that suggest trending stories, are highly dependent on non-news algorithms that operationalize popularity or make available share counts, and news actors are being increasingly forced to reshape their practices to suit the non-news algorithms their organizations depend on (Belair-Gagnon & Holton, 2018; Nielsen & Ganter, 2018). Recent scholarship has argued for more expansive ethical models that focus on “satisfactory” relations through traceable, operationalized concepts that can be identified across the many elements of algorithmic assemblages (see Ananny, 2016). Scholars have explored questions about the ethical and legal responsibilities of news algorithms, with some arguing that such actants cannot themselves be held accountable for ethical lapses due to their lack of agency (Klinger & Svensson, 2018) and that legal standards in the United States make it difficult to hold algorithms (and their creators) liable for libel, though fault may be easier to show in some European jurisdictions and by private citizens (Lewis, Sanders, & Carmody, 2018). There is, as of 2018, no widely accepted code of ethics for the use of algorithms in journalism.

Discussion and Research Directions

It is apparent from the rapidly developing body of literature at the intersection of algorithms and journalism that algorithms have begun to influence, to some extent, nearly every aspect of journalism. Their impacts may be observed from the initial stages of news production (e.g., story selection) to the latter stages of news consumption (e.g., commenting on stories). However, this body of work is still in its infancy. While a number

Algorithms and Journalism

of valuable insights have already emerged, important theoretical and empirical questions have not been comprehensively addressed.

There is definitional contestation in the literature, with scholars using the term “algorithm” to refer to both technical and social phenomena. While this is not uncommon, especially during the early development of a research area, it can lead to conceptual confusion. There is no easy resolution to that issue: an “algorithm” can be so complex and dependent on other components of a system that it defies a simple definition. However, the development of a comprehensive typology of algorithms and/or algorithmic assemblages would be especially helpful for promoting conceptual clarity for future work.

Scholarly efforts to describe the material components of algorithms and situate them within broader systems are already paying dividends for understanding how they are impacting journalism. For example, sizable bodies of work have emerged assessing how particular algorithms (e.g., Facebook’s News Feed) and algorithmically driven platforms (e.g., Twitter) are impacting news distribution practices and citizens’ news consumption patterns. Similarly, scholars have begun to theorize about the growing valorization of technologically specific labor and empirically observed ways in which algorithmic surveillance is altering routines and notions of “success.” However, there are still many opportunities for empirically assessing the extent to which potentially disruptive algorithmic innovations are manifesting themselves in journalistic spaces like newsrooms and the impact they are having on different actors’ attitudes and behaviors. For example, it remains unclear how prevalent computational journalism and automated journalism are in different kinds of newsrooms, and few large-scale, systematic studies of the content produced through those forms have been performed. Such work will be essential for evaluating emerging theories as the body of literature matures.

However, as the scholarship has aptly noted, newswriters and news organizations are becoming increasingly dependent on algorithms developed and enacted by a range of digital intermediaries—some of which are more visible than others—that have disparate and sometimes conflicting aims and logics vis-à-vis those in journalism. Those intermediaries have begun to receive scholarly scrutiny, but scholars will need to examine them even more closely in the coming years. In particular, scholars will need to revisit key models for explaining journalism and retrace the lineage of news phenomena while examining the recalcitrant processes that result from shifting arrangements.

One issue that is likely to become particularly important in coming years is algorithmically enabled amplification and manipulation, especially in the contexts of misinformation and disinformation. The scholarship in this stream has thus far focused on distributional processes and the formation of homogeneous communities, such as curated flows and filter bubbles, and the ways in which misinformation can be quickly amplified as a result of the rules and biases that guide public relevance algorithms. Such phenomena will only become more important as publics further fragment and democratic institutions continue to be challenged. However, an emerging threat is that of algorithmically generated disinformation. For example, computer scientists have already

Algorithms and Journalism

been able to generate seemingly authentic videos in which a public figure speaks from a doctored transcript by applying computer-generated imaging and machine learning algorithms to archived footage. This will introduce new challenges not only to core journalistic practices like verification but also to social practices around news consumption as institutions and authority are further contested.

In examining these developments, scholars must find a way to deal with one particularly vexing problem: the inability to peer inside proprietary algorithms and make sense of ever-changing, complex mechanisms and systems. While transparency is an important condition for effective public governance and scholarly evaluation, scholars have offered compelling arguments for why it is insufficient for algorithmic accountability. Some scholars have already attempted to reverse engineer algorithms and audit open-source code, contributing to understandings of algorithmic logics and providing insight for the development and evaluation of ethical models. However, it is apparent that, in order to advance, this body of work will require scholars who have technical expertise and the development of partnerships with private companies. The former demands more interdisciplinary work and the latter a compelling business incentive to overcome the perceived business risks associated with disclosure.

Further Reading

Ananny, M., & Crawford, K. (2018). **Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability.** *New Media & Society*, 20(3), 973-989.

Carlson, M. (2018). **Automating judgment? Algorithmic judgment, news knowledge, and journalistic professionalism.** *New Media & Society*, 20(5), 1755-1772.

Caswell, D., & Dörr, K. (2018). **Automated journalism 2.0: Event-driven narratives.** *Journalism Practice*, 12(4), 477-496.

Diakopoulos, N., & Koliska, M. (2017). **Algorithmic transparency in the news media.** *Digital Journalism*, 5(7), 809-828.

Gillespie, T. (2014). The relevance of algorithms. In T. Gillespie, P. Boczkowski, & K. Foot (Eds.), *Media technologies: Essays on communication, materiality, and society* (pp. 167-194). Cambridge, MA: MIT Press.

Gillespie, T. (2016). Algorithm. In B. Peters (Ed.), *Digital keywords: A vocabulary of information society and culture* (pp. 18-30). Princeton, NJ: Princeton University Press.

Just, N., & Latzer, M. (2017). **Governance by algorithms: Reality construction by algorithmic selection on the Internet.** *Media, Culture & Society*, 39(2), 238-258.

Napoli, P. M. (2014). **Automated media: An institutional theory perspective on algorithmic media production and consumption.** *Communication Theory*, 24(3), 340–360.

Shorey, S., & Howard, P. N. (2016). **Automation, big data and politics: A research review.** *International Journal of Communication*, 10, 5032–5055.

References

Altheide, D. L., & Snow, R. P. (1979). *Media logic*. Beverly Hills, CA: SAGE.

Ananny, M. (2016). **Toward an ethics of algorithms: Convening, observation, probability, and timeliness.** *Science, Technology & Human Values*, 41(1), 93–117.

Ananny, M., & Crawford, K. (2018). **Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability.** *New Media & Society*, 20(3), 973–989.

Anderson, C. W. (2011). **Deliberative, agonistic, and algorithmic audiences: Journalism's vision of its public in an age of audience transparency.** *International Journal of Communication*, 5, 529–547.

Asp, K. (2014). **News media logic in a new institutional perspective.** *Journalism Studies*, 15(3), 256–270.

Bar-Ilan, J. (2007). **Google bombing from a time perspective.** *Journal of Computer-Mediated Communication*, 12(3), 910–938.

Becker, H. S. (2008). *Art worlds*. Berkeley, CA: University of California Press.

Belair-Gagnon, V., & Holton, A. E. (2018). **Boundary work, interloper media, and analytics in newsrooms.** *Digital Journalism*, 6(4), 492–508.

Boczkowski, P. J. (2005). *Digitizing the news: Innovation in online newspapers*. Cambridge, MA: MIT Press.

boyd, D. (2007). Why youth (heart) social network sites: The role of networked publics in teenage social life. In D. Buckingham (Ed.), *Youth, identity, and digital media* (pp. 119–142). Cambridge, MA: MIT Press.

Braun, J. A. (2015). *This program is brought to you by . . .* New Haven, CT: Yale University Press.

Bucher, T. (2012). **Want to be on the top? Algorithmic power and the threat of invisibility on Facebook.** *New Media & Society*, 14(7), 1164–1180.

Bucher, T. (2017). **The algorithmic imaginary: Exploring the ordinary affects of Facebook algorithms.** *Information, Communication & Society*, 20(1), 30–44.

Algorithms and Journalism

- Burrell, J. (2016). **How the machine “thinks”: Understanding opacity in machine learning algorithms.** *Big Data & Society*, 3(1), 1-12.
- Carlson, M. (2015). **The robotic reporter.** *Digital Journalism*, 3(3), 416-431.
- Carlson, M. (2018). **Automating judgment? Algorithmic judgment, news knowledge, and journalistic professionalism.** *New Media & Society*, 20(5), 1755-1772.
- Caswell, D., & Dörr, K. (2018). **Automated journalism 2.0: Event-driven narratives.** *Journalism Practice*, 12(4), 477-496.
- Christin, A. (2017). **Algorithms in practice: Comparing web journalism and criminal justice.** *Big Data & Society*, 4(2), 1-17.
- Christin, A. (2018). **Counting clicks: Quantification and variation in web journalism in the United States and France.** *The American Journal of Sociology*, 123(5), 1382-1415.
- Clerwall, C. (2014). **Enter the robot journalist.** *Journalism Practice*, 8(5), 519-531.
- Coddington, M. (2015). **Clarifying journalism’s quantitative turn.** *Digital Journalism*, 3(3), 331-348.
- Couldry, N. (2008). **Mediatization or mediation? Alternative understandings of the emergent space of digital storytelling.** *New Media & Society*, 10(3), 373-391.
- Crawford, K. (2016). **Can an algorithm be agonistic? Ten scenes from life in calculated publics.** *Science, Technology & Human Values*, 41(1), 77-92.
- DeVito, M. A. (2017). **From editors to algorithms.** *Digital Journalism*, 5(6), 753-773.
- Diakopoulos, N. (2015). **Algorithmic accountability.** *Digital Journalism*, 3(3), 398-415.
- Diakopoulos, N. (2016). **Accountability in algorithmic decision making.** *Communications of the ACM*, 59(2), 56-62.
- Diakopoulos, N., & Koliska, M. (2017). **Algorithmic transparency in the news media.** *Digital Journalism*, 5(7), 809-828.
- Dörr, K. N. (2016). **Mapping the field of algorithmic journalism.** *Digital Journalism*, 4(6), 700-722.
- Dörr, K. N., & Hollnbuchner, K. (2017). **Ethical challenges of algorithmic journalism.** *Digital Journalism*, 5(4), 404-419.
- Flaxman, S., Goel, S., & Rao, J. M. (2016). **Filter bubbles, echo chambers, and online news consumption.** *Public Opinion Quarterly*, 80(Suppl. 1), 298-320.

Algorithms and Journalism

Galison, P. (1997). *Image and logic: A material culture of microphysics*. Chicago, IL: University of Chicago Press.

Gillespie, T. (2014). The relevance of algorithms. In T. Gillespie, P. Boczkowski, & K. Foot (Eds.), *Media technologies: Essays on communication, materiality, and society* (pp. 167–194). Cambridge, MA: MIT Press.

Gillespie, T. (2016). Algorithm. In B. Peters (Ed.), *Digital keywords: A vocabulary of information society and culture* (pp. 18–30). Princeton, NJ: Princeton University Press.

Graefe, A., Haim, M., Haarmann, B., & Brosius, H.-B. (2018). **Readers' perception of computer-generated news: Credibility, expertise, and readability**. *Journalism*, 19(5), 595–610.

Graves, L., & Kelly, J. (2010). **Confusion online: Faulty metrics and the future of digital journalism**. New York, NY: Tow Center for Digital Journalism.

Haim, M., & Graefe, A. (2017). **Automated news**. *Digital Journalism*, 5(8), 1044–1059.

Haim, M., Graefe, A., & Brosius, H.-B. (2018). **Burst of the filter bubble?** *Digital Journalism*, 6(3), 330–343.

Hammond, P. (2015). **From computer-assisted to data-driven: Journalism and big data**. *Journalism*, 18(4), 408–424.

Heald, D. (2006). Varieties of transparency. In C. Hood & D. Heald (Eds.), *Transparency: The key to better governance?* (pp. 25–46). Oxford, U.K.: Oxford University Press.

Just, N., & Latzer, M. (2017). **Governance by algorithms: Reality construction by algorithmic selection on the Internet**. *Media, Culture & Society*, 39(2), 238–258.

Karlsen, J., & Stavelin, E. (2014). **Computational journalism in Norwegian newsrooms**. *Journalism Practice*, 8(1), 34–48.

Klinger, U., & Svensson, J. (2018). **The end of media logics? On algorithms and agency**. *New Media & Society*. [Advance online publication]: 1–18.

Latour, B. (1988). **Mixing humans and nonhumans together: The sociology of a door-closer**. *Social Problems*, 35(3), 298–310.

Latzer, M., Hollnbuchner, K., Just, N., & Saurwein, F. (2016). The economics of algorithmic selection on the Internet. In J. M. Bauer & M. Latzer (Eds.), *Handbook on the economics of the Internet* (pp. 395–425). Northampton, MA: Edward Elgar.

Lewis, S. C., Sanders, A. K., & Carmody, C. (2018). **Libel by algorithm? Automated journalism and the threat of legal liability**. *Journalism & Mass Communication Quarterly*. [Advance online publication]: 1–22.

Algorithms and Journalism

Lewis, S. C., & Usher, N. (2014). **Code, collaboration, and the future of journalism.** *Digital Journalism*, 2(3), 383–393.

Lewis, S. C., & Usher, N. (2016). **Trading zones, boundary objects, and the pursuit of news innovation: A case study of journalists and programmers.** *Convergence*, 22(5), 543–560.

Lewis, S. C., & Zamith, R. (2017). On the worlds of journalism. In P. J. Boczkowski & C. W. Anderson (Eds.), *Remaking the news: Essays on the future of journalism scholarship in the digital age* (pp. 111–128). Cambridge, MA: MIT Press.

Lindén, C.-G. (2017). **Algorithms for journalism: The future of news work.** *The Journal of Media Innovations*, 4(1), 60–76.

Marwick, A., & Lewis, R. (2017). **Manipulation and disinformation online.** New York, NY: Data & Society Research Institute.

Napoli, P. M. (2014). **Automated media: An institutional theory perspective on algorithmic media production and consumption.** *Communication Theory*, 24(3), 340–360.

Nelson, J. L., & Lei, R. F. (2018). **The effect of digital platforms on news audience behavior.** *Digital Journalism*, 6(5), 619–633.

Nielsen, R. K., & Ganter, S. A. (2018). **Dealing with digital intermediaries: A case study of the relations between publishers and platforms.** *New Media & Society*, 20(4), 1600–1617.

Parasie, S., & Dagiral, E. (2012). **Data-driven journalism and the public good: “Computer-assisted-reporters” and “programmer-journalists” in Chicago.** *New Media & Society*, 15(6), 853–871.

Pariser, E. (2012). *The filter bubble: How the new personalized web is changing what we read and how we think.* New York, NY: Penguin.

Petre, C. (2018). **Engineering consent.** *Digital Journalism*, 6(4), 509–527.

Polanyi, M. (1966). *The tacit dimension.* Garden City, NY: Doubleday.

Powers, E. (2017). **My news feed is filtered?** *Digital Journalism*, 5(10), 1315–1335.

Powers, M. (2012). **“In forms that are familiar and yet-to-be invented”: American journalism and the discourse of technologically specific work.** *Journal of Communication Inquiry*, 36(1), 24–43.

Shoemaker, P. J., & Vos, T. P. (2009). *Gatekeeping theory.* New York, NY: Routledge.

Shorey, S., & Howard, P. N. (2016). **Automation, big data and politics: A research review.** *International Journal of Communication*, 10, 5032–5055.

Splendore, S. (2016). **Quantitatively oriented forms of journalism and their epistemology.** *Sociology Compass*, 10(5), 343–352.

Star, S. L., & Griesemer, J. R. (1989). **Institutional ecology, “translations” and boundary objects: Amateurs and professionals in Berkeley’s Museum of Vertebrate Zoology, 1907–39.** *Social Studies of Science*, 19(3), 387–420.

Thorson, K., & Wells, C. (2016). **Curated flows: A framework for mapping media exposure in the digital age.** *Communication Theory*, 26(3), 309–328.

Thurman, N. (2011). **Making “The Daily Me”: Technology, economics and habit in the mainstream assimilation of personalized news.** *Journalism*, 12(4), 395–415.

Thurman, N., Dörr, K., & Kunert, J. (2017). **When reporters get hands-on with robo-writing.** *Digital Journalism*, 5(10), 1240–1259.

Thurman, N., & Schifferes, S. (2012). **The future of personalization at news websites: Lessons from a longitudinal study.** *Journalism Studies*, 13(5–6), 775–790.

Usher, N. (2016). *Interactive journalism: Hackers, data, and code*. Champaign, IL: University of Illinois Press.

van Dalen, A. (2012). **The algorithms behind the headlines.** *Journalism Practice*, 6(5–6), 648–658.

Waddell, T. F. (2018). **A robot wrote this?** *Digital Journalism*, 6(2), 236–255.

Waisbord, S. (2018). **Truth is what happens to news.** *Journalism Studies*, 19(13), 1866–1878. Advance Online Publication: 1–13.

Wallace, J. (2018). **Modelling contemporary gatekeeping.** *Digital Journalism*, 6(3), 274–293.

Weber, M. S., & Kosterich, A. (2018). **Coding the news.** *Digital Journalism*, 6(3), 310–329.

Zamith, R. (2018a). **On metrics-driven homepages.** *Journalism Studies*, 19(8), 1116–1137.

Zamith, R. (2018b). **Quantified audiences in news production: Assessing the knowledge.** *Digital Journalism*, 6(4), 418–435.

Zamith, R. (2019). **Transparency, interactivity, diversity, and information provenance in everyday data journalism.** *Digital Journalism*. [Advance online publication.]

Zamith, R., & Lewis, S. C. (2014). **From public spaces to public sphere.** *Digital Journalism*, 2(4), 558–574.

Zarsky, T. (2016). **The trouble with algorithmic decisions: An analytic road map to examine efficiency and fairness in automated and opaque decision making.** *Science, Technology & Human Values*, 41(1), 118–132.

Rodrigo Zamith

Journalism Department, University of Massachusetts Amherst

