JOURNAL OF COMMUNICATION

Media as Data Extraction: Towards a new map of a transformed Communications field

Journal:	*Journal of Communication*
Manuscript ID	JCOM-03-OA-2017-0143.R1
Wiley - Manuscript type:	Original Article
Keywords:	media, advertising, marketing, internet, personalization, smartphone, superstructure, data, digital, surveillance, capitalism

Media *as* Data Extraction: Towards a New Map of a Transformed Communications Field

Abstract

The Communications field must challenge traditional understandings of media in the face of a transformation in the dynamics of capitalism which prioritizes the generation of value from data based on continuous surveillance. New advertising and data-processing developments mean that while the term *media* may continue to attach to the distribution of narratives, researchers must now conceive it as the convergence of message-circulation technologies with data-extraction-and-analysis technologies that are linked to everyday objects increasingly typical of our new *mobile personalization* era. In fact, nothing less than a radical revision of the boundaries of the Communications field is required to adequately address the fundamentally altered social and economic order emerging from this ferment in the field of everyday life itself.

The Communications field is on the cusp of a major shift that it only barely recognizes. The recent convergence of message-circulating technologies (what we have called "media") with data-extraction-and-analysis technologies increasingly relegates the traditional content genres to a supporting role in a larger arena of knowledge production that we must also learn to associate with "media." The shift reflects a profound transformation of the social order. The Communications field, to retain its relevance, must give primary attention to media industries' surveillance and population constructions as key infrastructural aspects of economic life.

The developments are overturning three key verities held by many in generation of academics active around the time this journal published the first *Ferment in the Field* issue. One deeply held proposition was that media exercise social power through storytelling. A second, espoused by orthodox Marxist scholars, saw the media's storytelling role as creating society's "superstructure" that reflected its economic "base." Such media stories included widely shared games and dramas as well as nonfiction news and documentaries that provided ideological reinforcement of society's inequalities. A third principle shared by academics of all stripes was that these stories were forms of "*mass* communication," and so entirely divorced from whatever proximate responses receivers of those stories might provide. A person could read, hear or view a media-delivered message, but the sender couldn't immediately know its audiences responses, nor could it reply to those responses.

The generalizability of these propositions was already challenged in 1983. Placing media developments in the larger context of capitalism and control, Raymond Williams (1980) and Dallas Smythe (1981) argued that the orthodox divide between base and superstructure had never helped us understand media's role in capitalism. Both suggested that advertisers did not just

promote values through their stories but drove economic production itself through relations with commercial sponsors that encouraged consumption. Williams (1980, pp. 177-178) insisted that what drove the emergence of modern advertising was not the growth of national economies or markets as such, but the emergence of monopoly capitalism and its "growing desire . . . to organise and where possible control the market." Smythe went further, pushing a point presaged by Theodor Adorno and Max Horkeimer (1944) that individuals were performing economically important labor for media firms and advertisers through their activities as audiences. Smythe's insight contradicted the then-prevalent truism that audience members couldn't immediately interact with mass media, yet it proved deeply prescient. Even in the 1970s, "talk radio" and the spread of interactive cable television showed that immediate responses to mass-media —and their auditing of audience commercial labor—were possible. The notion that popular media had no ability to garner immediate audience feedback was demolished two decades later with the advertiser-sponsored internet. Google, Facebook, and many other digital publishers came to depend in part on audience labor to create the content of their websites and apps.

The deepest implication of these developments, though, does not concern the labor in which audience members engage. It concerns audiences' role as *resources for value extraction*—that is, as entities whose information can be used or sold by companies for a profit. The mere fact that a ratings firm says population segments watched a particular program already makes the segments potential resources for media producers and advertisers. What advertisers wanted, however, was not audience feedback but surveillance: the opportunity to record individual audiences' behavior with media. The internet fundamentally changed how this could be achieved. Virtually from its start in the 1990s, the commercial net allowed companies an unprecedented level of interactivity with and reconnaissance of the people visiting their sites.

Marketers' needs got early attention. In fact, it was for tracking the same person's multiple purchases on a retailing website using a desktop browser that Netscape created the digital cookie in 1993. Advertising firms realized they could introduce cookies to sites on which they bought commercial messages and then use the cookies to recognize individuals (or at least their devices) as they travelled across other sites. The goal was to track what people were doing on the web so that new advertising messages could be *targeted at them specifically* in ways impossible before digital media. Over time such tracking widened to bring together *anything* that people did and even said on their online devices and non-digital ("offline") contexts.

The industrialized construction of "audiences" as targets of economic activity was not new: the process has been core to advertising and other media businesses (Ang 1991, Turow 1981). But these latest construction technologies connected advertising practices with broader long-term trends in the social construction of knowledge. Throughout the twentieth century actuaries, statisticians, accountants, engineers and computer scientists began building disciplines aimed at forecasting individuals' actions based on demographics and previous behaviors. Marketers increasingly paralleled practitioners in fields such as health care, political analysis, and insurance in their view that investigating as much as possible about an individual's background, contemporary activities, stated interests, psychological profiles, and social networks will lead to accurate predictions of their behaviors. All this occurred within a still wider shift in the definition of 'social knowledge' itself, driven by the predictive discourse of 'Big Data'. Advertisers' ability to exploit these approaches for more precise audience-targeting options was boosted through the stunning rise in advertising choices that websites brought about, boosted again with the rise of smartphone browsers and apps beginning with the iPhone in 2007.

What is distinctive about this latest approach to the audience is that it involves not simply finding customers for goods or services, but creating underlying *data profiles* of individuals which enables sorting them as more, or less, valuable to sellers. The desire to target based on profiles has given rise to a complex media-planning-and-buying world involving various forms of technological expertise to define, track, and trade audiences. At heart, this process of marketing through surveillance is, as Oscar Gandy (1993) suggested, *social discrimination*. Firms gather data about people, use the data to create labels about them or the groups to which they belong, and then act toward some as winners and others as losers based on whether executives of particular media sectors might (or might not) consider them present or future valuable users of their materials.

This industrialized social discrimination is taking place at unprecedented speed, scale and intensity. Media genres—for example, news, comedies, dramas, festivals—are often just the "bait" through which data extraction and discriminatory profiling are made possible. Digital media channels (websites, apps, internet connected TV networks) claim to learn far more than in prior decades about the demographics, lifestyles, personalities, and even relationships of their visitors. They collect data from login registrations, cookies, tracking pixels and other bits of code placed on their sites: so-called *first party data*. *Second party data* is when a site or app provides its information about individual visitors directly to an advertiser. *Third-party data* denotes information from data collection firms such as Experian, Acxiom, and Datalogix that websites, apps, or their advertisers can purchase about individuals they would like to target. Gathering these data, combining them, and determining what the output means for any particular advertiser vis a vis its audience involves computer-driven statistical analysis. The ultimate goal of such predictive analytics is for corporate entities to reach population segments, and even

certain individuals, those who would most likely be persuaded to purchase a particular product or service, with tailored messages wrapped in (and around) particular types of stories.

Opportunities to target such messages now take place largely in so-called "programmatic marketplaces" where advertisers bid to reach individuals with specific characteristics as they enter websites or apps. These domains offer up anonymous shoppers as well as individuals whose names, email addresses or other personally identifiable information are known. Some firms in the marketplaces tout the ability to find large numbers of "lookalikes" who have the characteristics of known individuals marketers aim to reach. In one facet of the programmatic arena, a *digital ad exchange* (run by Google, Rubicon, or another firm), publishers present the number and kinds of individuals they are selling. For example, an advertising network may be selling two million 18-25 year old software engineers looking for a car. (An advertising network is a firm that serves as a broker for many websites and apps that want to sell ads.) Increasingly, such sales of audience members take place in real time: That is, the ad network's selling of the right to reach those software engineers takes place as they are entering an app or website. In the split second before they bid, the computers of potential advertisers can discover from the network whether cookied customers of theirs are among the cookied customers the network is selling. The result of that matching might determine whether they bid and how much they bid to have the network's computers (or the exchange's computers) coordinate almost instantly the serving of specific messages to those individuals (Turow, 2011).

Despite hype from companies involved, none of these activities is straightforward. The process of tracking, profiling, and targeting is fraught with challenges for the entire digital ecosystem. Advertisers pay to reach people with messages on websites and apps plagued by click fraud and ad blocking. Click fraud takes place in pay-per-click advertising when the owners of

digital locales that post the ads are paid an amount of money based on the number of visitors to those domains who click on the ads. An unscrupulous site or app owner will pay people or create automated scripts to click on ads simply to accumulate money. Ad blocking is the act of using a type of software (less commonly, computer hardware) to remove advertising content from a webpage or app. While industry players debate the specific reasons for these activities, they agree they cause substantial economic losses. The developments worry advertising and media firms greatly, and they have mobilized to address them through technologies that detect click fraud, refusals to allow individuals to enter sites with ad blockers on, payments to ad blocker firms to allow ads through, and more generally through a rise in native advertising as well as technologies to circumvent ad blockers.

Through all this, websites and apps have developed their industrial constructions of audiences with the goal of attracting people to their stories or "content." The streaming video firm Netflix, for example, uses data it has about its members to decide what kinds of video series and movies to produce; one result was the U.S. version of the series *House of Cards*. For many in the media realm, the next logical step is to attract audience members to personalized ads by sending them personalized menus of news and entertainment —and even differently produced entertainment and news narratives--based on the data profiles constructed about them (Turow, 201, pp. 125-28).

Culturally significant though they are, though, technologies involved in the targeted presentation of media stories represent today only one facet of how audience construction and social power are today converging within a larger transformation of Western society toward what Zuboff (2015) calls surveillance capitalism. We might talk more simply of a transformation of capitalism that draws on today's massively-enhanced capacities for

surveillance. Ever-growing endeavors by capitalist enterprises, supported by governments, attempt to profit by generating information about everyday life. The result is a dynamic, iterative system in which data collection through continuous automated monitoring provides a core means to organize, indeed govern, people's actions for economic value through statistical aggregation, profiling and predictive analytics. Much of this information gets collected, analyzed, and acted upon through devices through which narratives can be distributed. But for the purposes of capitalism what matters is not the link to narrative transmission, but rather *the underlying transformation of the internet into a space of continuous data extraction through surveillance*. Or as Bruce Schneier put it: 'the business model of the internet *is* mass surveillance' (Schneier 2013, added emphasis). In this world, the term 'media content" cannot be limited to narratives that the Communications field has traditionally studied—the news, entertainment, advertising and other storytelling genres. It must also include the broad gamut of data that are critical for a different kind of storytelling: the discriminatory profiling of individuals and groups that companies carry out. Often unrelated to the production of tales *for* audiences, the media involved in these new narrative activities have hardly been studied at all by communications researchers. Ignore these wider infrastructures of communication, and we miss the transformation of social and economic order—and control—under way today.

Consider the use of media in shopping, an early sector where surveillance capitalism is flexing its muscles. The purchase of products through the internet has become a taken-for-granted activity for many in the twenty-first century. Clicking to items on the web and via apps often puts into play a cavalcade of activities that reflect profiles of the shopper as created from previous behaviors in the digital store, from its brick-and-mortar incarnation (if one exists), from data the merchant has gathered about the shopper from third parties, and from information

inferred about the individual based on their technology uses. Based on the profile, the store may display a personalized agenda of products. Prices, too, may be geared to what the store knows about individual customers and how much it values them. A top executive from a major retail chain told one of us that if a store's tracking of a shopper's purchases shows they have been slowing in previous months, the store might try to regain loyalty by offering a lower price for a product than it would offer a solidly loyal customer. "Loyalty" in this world of surveillance capitalism has taken on a meaning starkly different from previous decades. Frequent shopper schemes, for example, are set up as much to encourage data retrieval as they are to promote recurring purchases. Loyalty, above all, means adherence to the social system of data extraction that sustains the latest stage of capitalism: Subjects' economic contribution takes the form not of labor but of submission—and often resigned submission—to being a natural resource for data extraction (Cohen 2014, Turow, 2017:254-56).

As competition ramps up between internet retailers (particularly Amazon) and their brick-and-mortar counterparts, the same dynamics of surveillance that retailers use online and on apps are making their way into brick and mortar locations, transforming the physical space of social life, too (Turow, 2017). The smartphone is the medium at the heart of this. Tracking an individual's smartphone can actually begin far from the store, as shoppers select store coupons at home and print them or download them to their smartphones. GPS and cell tower technologies can follow the phone as the shopper moves outdoors, and programmatic location marketplaces can send messages and discount offers tailored to individuals based on their profiles and distance from the retailer. "Geoconquesting" tactics can help encourage shoppers walking near a particular store to divert and go to a competitor. Once in the store, a panoply of devices—Bluetooth BLE, Wi-Fi, undetectable sounds, lighting systems, and more—

pinpoint the smartphone, interact through it with the shopper, and offer personalized prices (through discounts) as the individual stands in front of particular items. Many instore recognition systems depend on the shoppers' downloading an app (not necessarily the retailer's app) that can interact with the location technologies in the store. Apps are the hooks that lure users into specific data-extraction zones that correlate with domains of physical space. Near field communication (NFC) chips can allow interactions between retailers and manufacturers with some smartphones around specific packages. Enhancements in surveillance promise ever richer lodes of data extraction. Facial recognition and other biometric identification systems are among the potential veins. They are not yet widely used for marketing purposes, but some observers believe that is only a matter of time (see Turow, 2017, pp. 236-37).

The interactions of these various devices with individual shoppers continuously reflect and adjust profiles that lead the retailers to identify those they deem "high value" and to act better toward them compared to others they see as average or money losers. To those who get the best deals and service in such a system this probably sounds perfectly acceptable. But for every person who feels that way, there will be plenty who don't. People whose buying history shows they are mostly bargain shoppers yielding the retailer small or no profit margins will be shown few discounts, or maybe none at all. If fashion shoppers are cherished regulars, for example, special mirrors with cameras may remember their shape and help them match clothes without even trying them on.

Increasing in prominence, then, is a media-centered data-extraction regime that encourages and actualizes many dimensions of discrimination throughout the field of social action. Computers, smartphones, their webs and their apps are implicated in this overlay, but

so is the internet of things: lamps, thermostats, refrigerators, music equipment, cars and other devices that are connected to one another, to corporate cloud computers, and to smartphones, tablets, gaming consoles, laptops, and other computers of the individuals who own the devices. The term *media* must encompass these and the many other tools of everyday life that become vehicles for transmitting messages about their users to, and for receiving them from, organizational actors. In fact, lines between the internet of things and traditional media can be blurry. Amazon's Echo communication cylinder, for example, can follow a person's commands to raise the home temperature, move a shopping list between the Echo and the phone, play the latest National Public Radio and BBC news broadcasts, and deliver a new nonfiction book—all while sending records of these interactions to Amazon for the individual's profile.

These changes reflect that in today's new media environment the ability to use media in private is vanishing for all but the highly technically adept. In the 1970s Raymond Williams saw television as a person's private—unwatched--window on a quickly moving world, and he coined the term *mobile privatization* to describe that development. We now are entering a quite different age of *mobile personalization*. In it, individuals, *wherever they are*, are (a) continuous targets of personally-adapted messages and content; (b) potential producers of content and so, in some sense, personalizers of the media environment; and (c) the subjects of continuous tracking and profiling, with (a) and (b) generating the incentives that drive (c). It is (c), our subjection to continuous tracking and profiling via media platforms, which makes surveillance capitalism possible.

This transformation is so fundamental that it renders irrelevant older boundaries around the objects of communications research. It also signals a deeper transformation of the

economic and social order in which infrastructures of communication ('media') exist. Earlier generations of communications researchers believed that media are crucial mainly for creating detachable cultural *superstructures*. Even Williams and Smythe tied their analysis of the links between media and society's economic base to the creation of stories in various ways. But media as we should now understand them connect far more broadly and deeply than Williams and Smythe imagined to society's core economic structure. Media are key technologies for social control through data management.

The broad convergence of message production and distribution with data-extraction-and-analysis activities thrusts the Communications field into a world of knowledge production indissolubly linked to discriminatory decision-making. When a firm collects the circumstances under which families turn up or down their thermostats, it generates data the firm can use in its general business processes as well in its particular pricing strategies toward that family. Similarly, when doll or television set manufacturers track what people say to those items (two controversial recent examples), they may be able to profile their owners in ways affect the creation of TV sets and dolls in the aggregate as well as personalized sales messages the owners receive. Moreover, depending on regulatory policies of the societies involved, the data from all these constructions may be sold on the open market to other corporations that will integrate them into their general production, distribution, and data processing. Such data inform industrial activities, often personalized ones, even as they are used to construct ideas about customers and society at large. This work, in turn, fuels more industrial activities. If we still retain the language of base and superstructure, we can say that they are ever more tightly intertwined, informing each other and in that way influencing the social-political order. Perhaps it is better to drop the distinction between base and superstructure altogether.

What of the Communications field? It must surely give primary attention to surveillance as a key infrastructural dimension of economic and social life. In the process, what we now call media must encompass far more objects, devices and environments than one typically encounters at a Communication conference or in a Communication class. This, in turn, forces us as "communication scholars" to clarify what value we can add to understanding the new media system. Communications researchers will need to be in dialog, and have credibility with, colleagues in information schools, engineering schools, surveillance and privacy law departments, marketing and retailing departments. Anything less will mean ceding authority to speak for and about communications to other disciplines and interdisciplinary traditions, marginalizing the Communication field just when its continued insights are most needed.

The societal stakes in developing this expanded research agenda are also high. The massive increase in data production, processing, and profiling depends on activities of continuous surveillance which are fundamentally in tension with the basic freedoms on which democratic processes depend. Worse, they may undermine the basic levels of trust and legitimacy on which civic, economic and political institutions depend. But here also lies an opportunity. Through its specialized understanding of the interpersonal, organizational and industrial dynamics that fuel the changing media, the Communications field is in a privileged position to illuminate these tensions and so contribute fundamentally to society's understanding of an economically-driven transformation of the very nature of social order (Couldry and Hepp, 2017). Conducting critically inflected research on these new dynamics will not be easy. To forestall controversies and uncomfortable questions, companies typically make it difficult for members of the public and even government regulators to understand the

surveillance activities that they implement with respect to advertising, marketing, and internet-of-things. It is, nevertheless, quite possible to unpack the essential aims and workings of most of the activities through wide, careful reading of industry trade magazines, attendance of industry conferences, close analysis of the technologies used, and interviews of relevant executives. (For examples, see Kreiss, 2016, and Turow, 2011 and 2017.) For competitive reasons, these sorts of explorations are unlikely to expose closely held algorithms that guide the specific profiles of and actions toward audience targets in particular surveillance scenarios. They can, however, reveal enough about how commercial surveillance works in various institutional sectors to contribute to substantive and theoretical debates among academics; to spark concerns by citizens which, in turn, are registered in everyday life in ways we can study as researchers; and to suggest policy directions for activist groups, foundations, government agencies, and corporations.

Raymond Williams already grasped nearly four decades ago the drive of corporate power to "increasingly demand not a sector, but a world, not a reservation but a whole society . . . within which to operate" (1980, p. 194). But bold as he was, Williams did not predict the fusion of economic, social and data-extraction spaces that is being engineered today. Any ferment in the Communications field needs to be adequate to confront the societal ferment that is already under way.

References

Ang, I. (1991). Desperately seeking the audience. London, UK: Routledge.

Cohen, J. (2015). The biopolitical public domain. Available at SSRN: https://ssrn.com/abstract=2666570.

Couldry, N. & Hepp, A. (2017). The mediated construction of reality. Cambridge, UK: Polity.

Draper, N. (2014). Reputation, Inc. Philadelphia, PA: University of Pennsylvania doctoral dissertation.

Kreiss, D. (2016). Prototype politics. New York, NY: Oxford University Press.

Schneier, B. (2013). The Public-Private Surveillance Partnership, *Bloomberg BusinessWeek*, July 31.

Smythe, D. (1981). On the audience commodity and its work. In D. Smythe (Ed.), Dependency Road (pp. 22-51). New York, NY: Ablex.

Turow, J. (1981). Getting books to children. Chicago, IL: American Library Association Press.

Turow, J. (2011). The daily you. New Haven, CT: Yale University Press.

Turow, J. (2017). The aisles have eyes. New Haven, CT: Yale University Press.

Williams, R. (1980) Culture and materialism. London, UK: Verso.

Zuboff, S. (2015). Big other: Surveillance capitalism and the prospects of an information civilization. Journal of information technology, 30, 75-89.