Article

new media & society
1–22

DOI: 10.1177/14614448221146174
journals.sagepub.com/home/nms
SAGE

How online advertising targets consumers: The uses of categories and algorithmic tools by audience planners

Thomas Beauvisage
Jean-Samuel Beuscart
Samuel Coavoux
Orange Labs/SENSE, France

Kevin Mellet
CSO, Sciences Po, France

Abstract

Recent innovations in online advertising facilitate the use of a wide variety of data sources to build micro-segments of consumers, and delegate the manufacture of audience segments to machine learning algorithms. Both techniques promise to replace demographic targeting, as part of a post-demographic turn driven by big data technologies. This article empirically investigates this transformation in online advertising. We show that targeting categories are assessed along three criteria: efficiency, communicability, and explainability. The relative importance of these objectives helps explain the lasting role of demographic categories, the development of audience segments specific to each advertiser, and the difficulty in generalizing interest categories associated with big data. These results underline the importance of studying the impact of advanced big data and AI technologies in their organizational and professional contexts of appropriation, and of paying attention to the permanence of the categorizations that make the social world intelligible.

Keywords

Audience segments, big data, categorization, media planning, online advertising, social categories

Corresponding author:
Kevin Mellet, CSO, Sciences Po, 19 Rue Amélie, 75007 Paris, France.
Email: kevin.mellet@sciencespo.fr

The generalization of big data and machine learning from the mid-2010s has changed the way individuals are described (Cheney-Lippold, 2017). This behavioral and algorithmic turn in digitized areas of activities has brought a new set of data (likes, comments, browsing data, interests, etc.) that describe the social world as a heterogeneous assemblage without resorting to traditional demographic criteria. In a radical analysis of this evolution, big data algorithms are viewed as "post-demographic machines" (Rogers, 2009), which produce categorization operations from the unsupervised examination of large datasets. These meaningless descriptions would mark the end of theory, language, and social expertise in many professional activities (Anderson, 2008). Used as decision support systems, these technologies supposedly produce prediction at the cost of understanding, or knowledge without knowing (Andrejevic, 2013). More recent works, however, have pointed out the hybridization of behavioral, algorithmic categorizations, and classical, meaningful, demographic-based understanding of the social (Kotliar, 2020). The articulation of these two description regimes, however, has been barely explored empirically. How are behavioral data, big data, and machine learning articulated with demographic categories in data-intensive activities? This article investigates this question with a focus on the online advertising industry. Our study addresses this research gap in three ways. First, it complements the description of the online advertising industry targeting practices. Second, it contributes to the understanding of the socialization of big data categories (Bolin and Andersson Schwarz, 2015; Kotliar, 2020; Kotras, 2020), by describing the hybridization of traditional and algorithmic forms of audience description and targeting. Third, it brings empirical insight to the debate on surveillance capitalism and consumer segmentation.

Advertising is an ideal case to analyze the articulation of behavioral and demographic data in practice since it is historically structured around a demographic description of audiences and consumers. At its core, advertising is an audience marketplace (Napoli, 2003), where audiences are traded in the form of ad inventory, traditionally bundled into segments before being commercialized. In traditional media, the description of audiences is based on demographic data (Beuscart and Mellet, 2013) built out of panels by audience measurement companies: "Women 18–29," "household income > $75 K," and so on, and secondarily on audience interests inferred from the editorial environment: sports, entertainment, financial products, and so on. These forms of targeting are probabilistic, that is, based on aggregate characteristics of the audience. In online advertising, big data and artificial intelligence tools hold two promises: first, to produce a more precise and effective targeting; second, to automate the definition of the relevant targeting criteria for a given ad. In practice, these two dimensions are bundled: the main actors of the advertising industry, Google, and Facebook, have developed self-service ad-buying services, connected with automated systems able to purchase inventory, display ads, and optimize campaign efficiency. This ecosystem has been coined as "surveillance advertising" (Crain, 2021) since it relies on the accumulation of multiple layers of information about individuals' characteristics and behaviors.

In order to empirically assess this evolution, we study the uses of the Facebook advertising console by media planners. Media planners are the workers in charge of building ad campaigns for advertisers, by setting up targeting plans, buying corresponding ad inventories, and combining them into ad campaigns. Their activity has been highly affected by the digital turn of ad targeting, so much that they are now mainly named "audience planners." How do these practitioners choose targeting categories, arrange and combine them, and assess their relevance and effectiveness? What does the algorithmic machinery do to the way they manipulate audience segments? How can one characterize the advertisers' gaze on consumers, usually coined as "surveillance" in the literature?

Our results suggest that big data categories and tools have not replaced older forms of targeting but are selectively integrated within the constraints of media planning routines. In this process, targeting categories are evaluated along the following three criteria: efficiency, the measured capacity to make consumers click or buy; communicability, the ability to circulate through organizations; and explainability, the ability to produce accounts of these categories. These objectives help understand the persistence of demographic categories (as a coordination tool), the rapid development of audience segments specific to each advertiser (known as custom audiences), and the difficulty in generalizing the big data-driven categories (interest categories).

This article is organized as follows: (1) we summarize recent works and analysis on big data and categories, and how they are used for marketing purposes; (2) we describe our fieldwork, including how the Facebook console works, and our methodological approach to assess its uses; (3) we detail our empirical results on the use of targeting categories by media planners; (4) we discuss these results and (5) we finally draw perspectives for future research.

Big data and advertising categories

Implications of big data marketing for citizens and consumers

Recent developments in media planning are the latest step of a long-term process of data intensification and automation of advertising and marketing started in the 1970s (Crain, 2019; Turow, 2006). Digitization has increased the degree of individual customer surveillance through the combination of ever-expanding sources of data about demographic attributes, browsing behavior, purchase patterns, media consumption, mobility traces, and so on (Christl, 2017; Christl and Spiekermann, 2016). These data have been progressively assembled, processed, and channeled by the advertising industry for targeting purposes, feeding the process of individualization of audience characterization (Mellet and Beauvisage, 2020). Following a long-term trend toward the automatization of data exchanges and media purchasing (McGuigan, 2019), these data circulate among actors of the programmatic advertising value chain (Venkatadri et al., 2019).

The amount and the diversity of consumer data favor their algorithmic processing (Bolin and Andersson Schwarz, 2015). This results in the proliferation of "post-demographic" consumer categorizations that are at the center of this article. These categories allow the targeting of more precise and smaller segments, following a long-term trend of advertising innovation (Mackenzie, 2018). As less human agency is incorporated in their building, these categories lack explainability (Bolin and Andersson Schwarz, 2015); they have been coined "post-narrative," since they do not offer direct justification of the assembly of consumer segments they produce (Andrejevic, 2013). Research suggests that when some of these categories come to be used on a regular basis inside organizations, they require an important work of translation (Kotliar, 2020). Algorithmic categorizations are also more unstable and fluid. The proliferating algorithmic identities they produce about consumers are likely to change with the evolution of the data collection and algorithmic calculation (Cheney-Lippold, 2017), resulting in ever-changing categorizations and sorting of consumers. In the case of the Facebook Ad Manager, these data-driven categories, referred to as "interest" categories, are built algorithmically from several inputs: they are "translated" from users' explicit behaviors (liking a page) and "imputed" from their browsing traces; the categories also evolve according to the uses and the feedbacks of the advertisers (Cotter et al., 2021; García Martínez, 2016).

Since algorithmic categories operate as "technical sieves" (Kockelman, 2013) for advertisers who target audiences, their use may have important consequences for consumers. First, many scholars have pointed out that the amount of data collected and the power of the data processing are very likely to give marketing professionals a strong advantage on consumers when shaping information and choice (Andrejevic, 2014), leading to a decline in consumer sovereignty and autonomy (Manzerolle and Smeltzer, 2011; Zuboff, 2019). Turow (2012) describes how online advertising is used to separate "targets" (who are offered interesting opportunities and caring customer service) from "wastes" (who are treated as second-class consumers). These categories may also be "weaponized" to target weaknesses in consumers (Nadler et al., 2018; Nadler and McGuigan, 2018). Many categories available for display targeting could be used for discrimination or harassment purposes (Speicher et al., 2018), and journalists and activists have uncovered illegal and discriminating uses of advertising targeting on Facebook (Angwin et al., 2017; Angwin and Parris, 2016). These concerns are particularly acute in political advertising, considering the potential consequences of targeting on voter registration and turnout. A research field has developed, mainly in the United States, in order to assess the consequences of the development of microtargeting techniques, defined in this literature in a very generic way as the use of individual databases in campaign targeting, on democracy (Barocas, 2012; Hersh, 2015; Hillygus and Shields, 2008). The Cambridge Analytica scandal has in particular brought public attention to psychodemographic targeting (Kosinski et al., 2013), a specific form of microtargeting based upon user data collected through Facebook's Partner Program.

Categories at work

Categories also shape the world of professionals who operate them (Bowker and Star, 2000). The deep change in advertising practices has long been a subject of matter for the professionals themselves. An abundant commercial and professional literature produced by the online advertising industry describes the evolution of targeting capabilities with online tools. Professionals envision a shift from media planning to "audience planning" (Monohan, 2019) and forecast the gains in efficiency as well as changes in professional practices. At the same time, research has unveiled the extent and the recklessness of the data collection by marketing professionals (Christl, 2017; Christl and Spiekermann, 2016; Turow et al., 2015), and exposed the behaviorist ideology shared by many actors as well as some deliberately manipulative practices (Nadler and McGuigan, 2018; Zuboff, 2019).

Professional discourses, however, are embedded in a strong data imaginary, that is partly disconnected from advertisers' daily practices (Beer, 2018; Simon, 2019). Advertising research stresses the difficulty to measure the actual lift created through targeted advertising; that is, beyond clicks and leads, the gains that would not have occurred without advertising targeting. Large-scale studies suggest that most of the time, due to methodological issues, the efficiency of targeting is largely overstated (Gordon et al., 2019; Lewis and Reiley, 2014), and that, overall, "digital ads don't work nearly as well as they're advertised" (Aral, 2020: 148). Studies in political advertising underscore the limitations of social media data, such as their poor quality or the difficulties to integrate them in the campaign's goal (Hersh, 2015).

In practice, audience planners must deal with a high variety of targeting choices, and cope with complex and sometimes opaque algorithmic tools. For instance, inside the Facebook Ad Manager and its partners, the North American and European population is characterized by hundreds of attributes, many of which are frequently renewed, and made available to any advertiser on Facebook (Venkatadri et al., 2019). As shown by research led in the field of "algorithms in practice," based on ethnographic approaches and comprehensive interviews with professionals, there is a gap between intended and actual effects of algorithms, which leads professionals to develop "buffering strategies" in order to control these effects (Christin, 2017). Recent ethnographic works on data analytic organizations describe a re-socialization process of algorithmic categories: to become usable in the organization's routines, audience categories are reprocessed, blended with existing organizational knowledge, renamed, translated into explicable concepts, and so on (Kotliar, 2020; Kotras, 2020). On top of making sense of algorithmic categories, advertisers must also decipher the opacity of the ad delivery algorithms. Research has shown that these algorithms are not neutral and favor, within a defined audience segment, the targets most likely to click on the ad, producing filter bubble effects (Ali et al., 2019b). For advertisers, algorithmic ad delivery creates uncertainty about the targets they reach and more generally reduces their control over the output of their actions (Kellogg et al., 2020). Overall, empirical studies of advertising practices tend to relativize the discourses of industry players and some academics about the power, accuracy, and efficiency of big data surveillance advertising, thus qualifying its effects on society. Our article contributes to this stream of research.

Research object, data, and method

Our research relies on interviews with media planners and their professional entourage. In this section, we first describe the Facebook Ad Manager and then present our data.

How targeting works in Facebook Ad Manager

At the time of our survey, in 2018–2019, the Facebook Ad Manager is the main entry point to purchase inventory on three platforms of the Facebook company, renamed "Meta" in 2021: Facebook, Instagram, and Messenger. It also provides access to affiliate inventories, particularly in the mobile world. The console is a very popular tool. In France, where this research was conducted, Facebook received more than 20% of digital

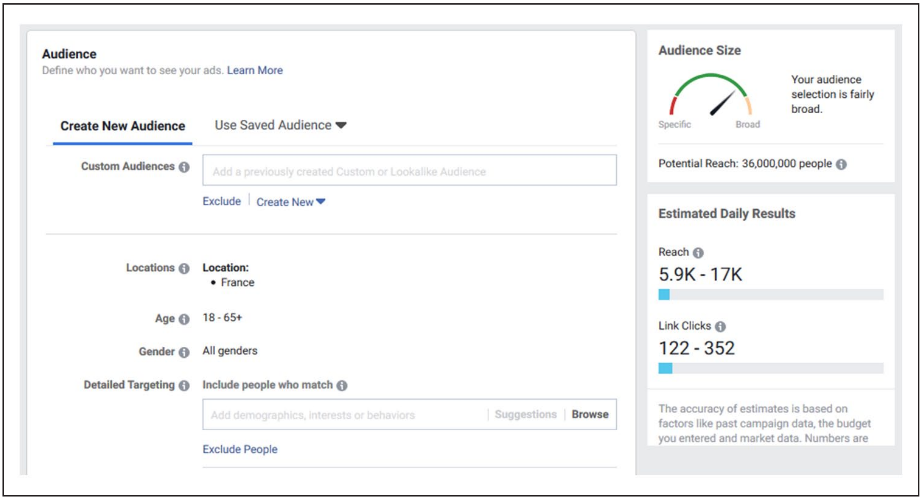

Figure 1. Audience setting in the Facebook Ad Manager (https://www.facebook.com/adsmanager/, retrieved 10th January 2018).

advertising spending, and served nearly 30,000 different advertisers in 2019 (source: France Pub, IREP, Kantar).

In 2018–2019, purchasing ad space on Facebook Ad Manager is a five-step process: the advertiser or her contractor sets a goal (promoting a Facebook page, generating visits on a website, etc.), defines one or more target audiences, determines a daily or weekly budget, uploads the ad design, and finally launches the campaign. To define an audience, three main targeting categorizations are available (see Figure 1):

- "Custom audiences," that is lists of individuals that are already known by advertisers. The latter may upload a list of identifiers (email addresses, phone numbers, HTTP cookies, etc.);
- demographic categories based on declarative (age, gender, location) or technical (location, language) data;
- "Detailed targeting," also referred to as "interest categories," includes a large array of attributes, most of which are categories inferred from users' actions (likes, comments, interactions . . .). Andreou et al. (2018) and Speicher et al. (2018) have identified 614 "curated attributes" that combine demographic characteristics and stabilized interests, and more than 240,000 "free form" interest categories.

Moreover, Facebook provides two main algorithmic tools to automate audience choice: lookalike audiences and automated ads. Lookalike audiences are meant to extend "custom audiences" by computing their statistical twins among the Facebook users. Campaign optimization lets an algorithm find the most responsive audience through trial and error, without the need to specify a precise audience.

Meta’s advertising services are constantly evolving. Since 2019, Meta’s ad manager has undergone a series of changes related to targeting options and campaign optimization, especially to exclude sensitive targeting categories, and enhance campaign automation. However, the three means to define the audience (custom audience, demographic categories, and interests) remain the main entry points in Meta’s 2022 ads manager; hence, Meta’s Ad Manager changes since 2019 do not substantially affect audience planners’ work.

Data and method

We followed a grounded theory research design (Glaser and Strauss, 1967). Our main empirical source in this article is a series of semi-structured interviews with French advertising professionals. Table 1 displays an anonymized list of respondents and their official job title. Although none of the interviewed professionals had “audience planner” as a job title, all of them either could be described as such, as their main activity is to define audience segments on advertising consoles (8), or they manage workers who do (7), advise customers to do so (3), or order such work from agencies (2). In all cases but that of two customers ordering audience planning from agencies, our respondents have firsthand experience with advertising consoles even though they may have moved to a management or consultant position. Consultants, especially, often do audience planning for their customers. We thus covered positions at almost every level of the advertising food chain, while focusing on one operation, the production of audience segments.

Since this research was one in a series of studies of the advertising world, most participants were recruited through snowball sampling, starting from informants from previous studies. We conducted 16 interviews between January 2018 and January 2019. Three interviews were with multiple respondents, so that we interviewed 20 people in total. Most interviews were face-to-face; two interviews were conducted with a video meeting software for distance reasons. Interviews lasted between 45 and 120 minutes, and 70 minutes on average.

We performed semi-structured interviews. We first asked respondents to describe their jobs and their usual customers. We then went over the various kinds of targeting, and systematically asked whether they used them or used to, in which conditions, and in which proportions, collecting specific, recent examples. We then asked respondents about automation tools and how they fit into their work. This guide was adapted to respondents who were not directly involved in audience segmentation.

Each interview was fully transcribed. The interviewers wrote a two-page summary based on this transcription, which was discussed among the four authors. During this first analysis phase, salient themes emerged. We adjusted the interview guide and planned further interviews based on those preliminary results. We then coded all interviews with a qualitative data analysis software (QDA Miner), using a thematic coding scheme inductively built. The main level 1 codes used in this essay were “setting targets” (respondents describing the process of audience selection), “targeting” (with subcodes for interest categories, demographic, custom audience), and “quality assessment.”

Table 1. List of interviews conducted.

Type	Code	Company size	Occupation
Buyels: advertisers and advertising firms	advertiser 1	medium	1. R&D manager 2. data and development manager 3. data scientist
	advertiser 2	medium	Data manager
	advertiser 3	large	Digital tools project manager
	advertiser 4	large	Media data manager
	media agency 1	large	Programmatic advertising buying team manager
	media agency 2	large	Strategy and digital innovation consultant
	programmatic agency	medium	1. Head of programmatic advertising 2. Data analyst
	social media agency 1	medium	CEO
	social media agency 2	medium	1. Web Marketing director 2. Facebook projects manager
	social media agency 3	medium	CEO
	social media consultant 1	small	Independent consultant
	social media consultant 2	small	Independent consultant
	social media consultant 3	small	Independent instructor
Sellers and intermediaries	ad exchange	large	Vice-president in charge of development
	ad network 1	medium	Activation and data manager
	ad network 2	large	Head of advertising

CEO: chief executive officer; R&D: research and development.

In the following section, we present the results in the form of interview excerpts (verbatims). We indicate, when possible and relevant, the recurrence of the phenomenon evoked in the verbatims by mentioning the coded theme, the number of verbatims dealing with the theme (v = X) and the number of interviews concerned (i = Y), such that they appear in our coding scheme.

Results: building audiences out of big data categories

How do new technologies and categories fit within media planners' daily activity? How do they articulate with existing targeting practices based on demographic categories? Our study shows that advertisers constantly arbitrate and combine these different targeting modes. There are three main processes through which categories are integrated into real-world professional practices: their intrinsic efficiency (the cost-benefit and quality assessment of these categories); their communicability, in order to help actors (along the advertising value chain) to coordinate; and their explainability, in order to circulate among actors, and to reinforce impact on the agency/autonomy of these professionals (their agency and their ability to carry on their skills and expertise).

Efficient categories: reach, cost, and volume

The first important and maybe unsurprising result of our inquiry is that targeting categories and algorithms must prove their worth in the light of efficiency metrics at work in the industry, in order to get integrated into the daily practices of advertising professionals (*Quality assessment: Metrics and KPIs*; v = 36; i = 15). These metrics may vary from one campaign to the other, but most of them are structured by measurable actions from consumers (click, buy, fill a form) related to costs of campaigns. These costs come in three components: the somewhat fixed human and technical cost of running a targeting campaign, the price of a given audience (related to the demand for it), and the proportion of targeted users who react to the ad of the campaign. In practice, economic calculation drives advertisers' choice of targeting categories: many such categories may be relevant and accurate, that is, have high rates of response from targeted users, yet not be viable because these audiences are too expensive, or too few to cover overall costs (*Targeting: Audience size*; v = 28; i = 10).

Algorithmically generated interest categories ("detailed targeting") are a typical example: they have been lauded by the advertising industry because of their ability to reach fine-grained responsive audiences and are the keystone of discourses about the targeting capabilities of online advertisement. But our respondents are more careful. Admittedly, all of them put forward one or two success story displaying their dexterity in finding relevant audience categories (*Success stories: Targeting*; v = 21; i = 12). These stories are often about niche products: "a customer organized a charity run [. . .] we targeted people who have done similar races, 'Iron Man' races, 'Color me' Runs, Mud races [. . .], we only talked to about 12000 people, but they were highly qualified!" (Social Media Consultant 2). However, they also describe how, most of the time, interest targeting leads to poor efficiency:

> We had a real estate program in Montpellier, we targeted the people who were in Montpellier and X kilometers around. [. . .] We had a fairly broad, purely geographical targeting at the beginning, [then] we added data about real estate interests, from the Acxiom database [. . .] And from there, we not only reduced our volume but our cost per contact exploded. I think that is because the target was too narrow and in the end it worked much better when we [did not target]. Everyone may want to buy a flat and the wide targeting worked much more than the very precise targeting, so we quickly turned back. (Social Media Agency 2)

Most of the interviewees explain that interest targeting often produces low volumes of targets that generate high costs and require maintenance (*Setting targets: trade-offs*; v=25, i=9). Online advertising players very often have to make the trade-off, highlighted by several interviewees, between highly efficient but very small audiences, on the one hand, and larger audiences with less relevance, on the other hand. Most campaigns need to not only optimize the acquisition cost but also increase visibility and sales. While an array of small targets may be relevant, interest targeting entails coordination and management costs. First, since campaigns often rely on the combination of several tools, the coordination of small-sized audiences' campaigns is more complex, posing technical problems of synchronization between tools. Second, the management of campaigns made up of several audience segments is more expensive. Ideally, each target group should be addressed a dedicated visual creative and a customized message, which increases the costs of creation; in addition, the daily monitoring of a set of audiences requires more attention and, ultimately, increases labor costs:

> Activating all these data means managing as many campaigns with very precise levels of the audience. Today, quite frankly, we don't know how to do it. Every time we try, we face a limit in terms of people and resources we can to the task. It's impossible for us; we cannot run all our campaigns with very thin targeting levels and audience segments. There is a technical limit and a human limit. (Advertiser 4)

Finally, the sheer number of interest categories makes them difficult to know: ordinarily, advertisers would rather test a large amount of possibly relevant categories, keeping those that perform well. "We test everything, everything and anything, and in fact there are things that work and we don't really know why" (Social Media Agency 3).

As a result, interest-based targeting has not become the main practice in online advertising. Rather, it is considered useful in the absence of first-party data or of a developed customer base, especially for a young company with a niche product (*Targeting: Interest categories*; v=44; i=10). "At the beginning of a campaign, you must start by making assumptions, and you imagine the audience in terms of their interests. There, I think, interests are useful at this step, and they work very well" (Social Media Agency 3). For other interviewees, especially those working in large agencies, the balance sheet of interests is rather negative: "Facebook's promise of interest targeting, in fact, did not perform as well as advertisers had hoped" (Social Media Agency 1).

The search for efficiency at low cost has also favored another form of targeting, based on messages directed to already known customers or visitors. Such products are labeled "custom audiences" by Facebook; advertisers can upload their "own," "first-party" audiences, that is, targeting a list of known customers or prospects, as targeting criteria and address specific messages to them (*Targeting: Custom audience*; v=47; i=11). Advertisers can also rely on visits to product pages, as well as on a wide range of behavioral signals, including interactions with ads, interactions with the brand's content on social networks, the use of a search engine to reach the site, and so on (*Targeting: Behavioral*; v=9; i=7). All these "engagements" with the brand are tracked and can be used as filters to form advertising segments.

Such direct marketing practices are not new to marketing operations, and originate in the development of credit services in department stores in the 1920s (Lauer, 2012). But online advertising has made it possible to extend it to online stores' visitors, besides rather than actual customers, and to "recognize" them in the anonymous crowd of web users. These techniques are also referred to as retargeting, and they are almost as old as online advertising, and have always been considered efficient. In this sense, custom audiences enlarge the direct marketing paradigm: it primarily targets people who are already in the advertiser's customer database. Custom audiences are considered the most effective strategy by industry professionals because they say that it guarantees targeting people who are truly interested in a particular brand or a product. They are said to be efficient because the ground for the assumption that a customer might be interested in the product, previous interaction with the brand, is much more solid than it is for interest-based targeting. In other words, the specificity of these custom audiences makes them valuable advertising assets, even in the form of small-sized segments—that is, the lookalike algorithms, allowing these small segments to be extended to form large audiences of statistically similar profiles, are very commonly used to articulate idiosyncratic quality and audience volume:

> What works best is retargeting: retargeting people who have already been on your site, or who have already bought something, or retargeting through existing customer databases. These people have already expressed an interest for the brand, so they ought to be not far from converting. (Social Media Agency 2)

Custom audiences make it possible to target audience segments built on any advertiser's available data on consumers, including browsing data and past purchases. One of the most preferred uses of custom audience segments is reaching customers for a particular product or offer (reactivation, upgrading, complementary products, and services) or, in contrast, excluding users who are already equipped from an earlier campaign:

> We are able to target only people who have large cell phone plans, to offer them luxury phones, for example. We can imagine a lot of scenarios and set them up every quarter based on our specific goals and creatives. (Advertiser 3)

In other words, by evaluating their targeting options through the lens of efficiency, advertisers tend to favor more easily reachable audiences, consumers that have already shown a sign of interest in the brand or product and are more susceptible to performing a measurable positive reaction to the campaign.

Communicable targeting: coordination along the advertising chain

The second expectation toward targeting categories is communicability. Targeting categories need to have the ability to circulate beyond the advertising professionals who operate them. Digital advertising campaigns are part of larger communication and marketing efforts, which require orchestration. In this respect, audience categories are conventional representations of consumers that operate as coordination resources (Bowker and Star, 2000; Desrosières, 2002).

Depending on the scope and on the objectives of the campaign, advertisers may rely on generic standardized categories that can be easily appropriated by different actors, or on the contrary resort to idiosyncratic categories that favor coordination inside an organization. Some categories, such as demographic categories, can circulate easily along the overall advertising practitioners; others, such as custom audiences, can be more easily matched with internal data from the company, and facilitate coordination inside the firm. Our fieldwork shows that actors rely mostly on demographic categories to coordinate along the advertising industry, beyond the digital, and on custom audiences to coordinate at the company level (*Advertiser/agency relationship*; v=24; i=11).

Demographic targeting remains a keystone of audience planning precisely because it is simple and widespread and, thereby, useful in coordinating marketing work (*Targeting: Demographic*; v=34; i=13). Contrary to interest categories, which are only available for digital advertising, they are shared by all outlets and are used to coordinate advertising strategies in global campaigns. For large advertisers and their agencies, online advertising is a continuation of other campaigns, especially television campaigns, where demographic targeting is the norm. Facebook is especially useful to reach the youth, who watch less television:

> Overall, what I'm looking for, when we talk about audience planning, today, and what advertisers are looking for, is to recreate demographic targeting on digital media. (Social Media Agency 1)

Because of this need to coordinate campaigns, filtering by demographics is always the first step audience planner take to translate advertisers' brief into audience segments (*Setting targets: Advertiser's brief*; v=28; i=11), even when other methods such as interest categories or lookalike audiences are used (*Targeting: Lookalike*; v=20; i=11). "When we advertise, demographic data, such as age, gender, spoken language, are really the basics" (Social Media Consultant 2). In other words, demographic targeting has been reinforced rather than weakened by online advertising tools.

Finally, demographic categories are also necessary for audience planners to account for their actions. "We must be able to explain [our targeting] to our customer" (Social media agency 2). Demographics are intelligible to all actors: they are part of the standard language in the field of advertising, easier to understand and easier to explain.

However, custom audiences, as they proceed from company's internal data, offer many affordances for a tight articulation of online advertising with the rest of the marketing and commercial activity (*Interviewees: Organization and coordination*; v=37, i=11). Through the exploitation of first-party data, advertisers and their agencies produce taxonomies that make sense only to them. The use of custom audience tools, therefore, is aligned with this advertiser-centric view of the world; it enables the translation of internal consumer categorizations into advertising targets. One of the most preferred uses of custom audience segments is reaching customers for a particular product or offer (reactivation, upgrading, complementary products and services) or, in contrast, excluding users who are already equipped from an earlier campaign. Within a sports-media advertiser, for example, the data science team has reconstructed a "cycling fan" target from its own website browsing logs and uses this as input to develop a campaign linked to a cycling event:

> Tomorrow, I'm going to email subscribers who are cycling fans to tell them that, say, we're broadcasting [a cycling tour]. To avoid sending an email to everyone, we're going to extract the people who have seen more than 15 articles and who have watched one live show about cycling over the last four months. (Advertiser 2)

The creation of such local segments might quickly be made profitable; a small set of segments is frequently used throughout the company, and the campaigns' feedback helps to refine and stabilize them. Segments are then enriched with the results of new campaigns, contributing to the overall knowledge of the company about its customers. These categories resemble interest categories, being the result of a similar process of data analysis, but interest categories are created in the abstract and not on an ad hoc basis. This idiosyncratic turn in targeting is fundamental: on top of traditional, generic, descriptive categories (age, gender, income, family situation, interests) used for media planning, advertisers rely increasingly on descriptions oriented toward intent and conversion as well as on local understandings of the consumption of the advertiser's products and services.

The transparency of algorithms and the autonomy of advertising professionals

The last expectation toward targeting categories is accountability. Besides interest-based targeting, the main alternative to demographic targeting comes from automation (*Targeting categories: Automation*; v = 16; i = 7). Contemporary advertising consoles, such as Facebook's, offer means to turn a very broad description of the targeted audience into precise audience segments, through a process of trial and error. Algorithms will try out dozens of segments, measure their return on investment, and narrow down the audience to the most cost-efficient one. "Basically, Facebook is saying: 'Don't bother with detailed targeting, put in all you've got, and we'll find the right people for you'" (Social Media Agency 3). However, it does not give feedback on the attributes of the reached audience. Ad buyers know how many people clicked their ad, but not who they were. Moreover, the algorithms that determine the most efficient audience are not made public, nor are they explained in detail (*Quality assessment: Access to data*; v = 13; i = 6). Not only are they opaque to mundane users with no statistical background, but even data scientists working in advertising agencies have no information on how they work, despite repeated requests. "It is totally a black box" (Media Agency 1). Among our respondents, the use of optimization is very popular among individual consultants, who see it as a means of satisfying their clients' demand for simplicity and efficiency, but it is controversial among those who work in agencies, for large customers, because audience planning falls under their professional jurisdiction and these tools diminish their role and importance in the advertising production chain.

The transparency of algorithms is often discussed in the literature on the ethics of artificial intelligence (Ananny and Crawford, 2018; Eubanks, 2018). The opacity of algorithms raises accountability issues that have dire consequences on inequalities in sectors like health or justice. Similarly, in advertising, opacity contributes to a distrust in the quality of data and automated targeting. Advertising platforms are judge and jury, providing the tools to create audiences and the tools to measure their efficiency, but several advertisers say they are "not interested in watching Facebook measure itself" (Media agency 2). Moreover, platforms keep the data to themselves and, thereby, do not allow media agencies and advertisers to make their own verification (*Trust issues with Facebook and Google*; v = 11; i = 8). Finally, they do not give information on who, precisely, was targeted: "we do not know what fits in exactly" (Social Media Agency 3).

Besides accountability, audience planners value intelligibility: they want to improve through feedback, and they would rather work with tools with results that can be described in natural language, and then be shared with all actors in the value chain. Interest-based targeting often makes intuitive sense, relying on common sense associations, which is an advantage over automated targeting. However, audience planners often have to try out several loosely related targets to find the best-performing ones. In the end, demographic categories appeared as the last reliable interpretive scheme. Why do people interested in "travel" and "cooking" click more often on real estate ads than those categorized as interested in real estate, as one respondent found out?

> When someone likes a real estate developer's [Facebook] page [and thus enters the "real estate" interest category], it is because they have already bought something and want to be kept up to date, but they won't buy a second one right away.

Moreover, every advertiser on the real estate segment uses this category, so the price of the eyeball rises. However, "travel and cooking" work because "when you think about it, being interested in travel means one has money to travel, and thus a bit of money to invest" (Social media agency 2). In short, demographics are the ultimate explanation advertisers rely on to interpret their results.

Second, intelligibility is necessary in feedback loops. The work of audience planners entails trials and errors. Whenever an audience has been defined based on the customer's brief, there are many possible implementations in the ad manager. Audience planners build several, monitor their results, and adjust them during the campaigns. They need to be able to make sense of these different audiences in order to provide such adjustments. "The value of data is also to provide insights, and we can understand them only by activating them, and looking for trends in performances to find out what made our campaign work" (Programmatic Agency). Feedback loops require intelligibility: it is only when the audience planner understands which targeting worked that they can focus on this or look for new ways to cover poorly targeted audiences. In the long run, intelligibility is key to the craft of audience planners. The centrality of intelligibility explains why automated targeting is still frowned upon by audience planners.

Discussion

This research analyses how the data-intensive advertising tools, and some of the categories they produce, are integrated in the working routines of advertising professionals. In early academic works on big data technologies, these tools have been described as post-demographic machines (Rogers, 2009), operating a switch from a "Gaussian" to a "Paretian" world (Bolin and Andersson Schwarz, 2015). We rather observe a process of "hybridization" (Kotras, 2020) of demographic categories with algorithmically processed ones such as interests or lookalike audiences. In this last section, we discuss these results with regard to the organizational dimension of data categorization, and the public debate on data surveillance.

The organizational socialization of big data categories

Demographic categories remain essential in all communications along the advertising value chain: they populate the briefs sent by advertisers to agencies, the accounts of the campaigns made by agencies to their advertising clients, as well as the internal discussions between professionals operating the dashboards and interfaces (Ariztia, 2015). Demographic categories are hybridized with new forms of consumer categorizations, rather than replaced by them, because they are essential to the representation of the activity and to the coordination of the advertising work along the value chain. Several scholars have noted that algorithmically produced categories often need to be "translated back" into more classical narrative categories (Bolin and Andersson Schwarz, 2015). Observing the negotiations around the naming of algorithmic clusters of consumers within marketing agencies, Kotliar (2020) analyses this operation as a "return of the social," by which unexplainable categories get labeled with approximate but workable social descriptors.

Our work brings new evidence to this thesis. First, advertising professionals mostly use ordinary language and classical categories dealing mostly with demographic markers (gender, age, but also habits and tastes). These categories and their combinations are more abundant and volatile than they used to be, they nevertheless structure accounts of professional activity. Second, when confronted with tools producing untranslated algorithmic categories, such as the lookalike audiences or the Facebook automation of campaigns, advertisers underline the cost of the lack of explainability: it diminishes their ability to learn from their successes and failures, to capitalize on the dynamics of trial and error now characterizing the process of audience planning; and it makes it harder to explain and value their work and expertise in audience understanding and targeting. When they resort to such untranslated algorithmic categories, advertisers weigh gains in efficiency with losses in accountability.

The tension between stable and common knowledge demographic categories, and big data originated automated targeting, opens the way to a variety of "buffering strategies" (Christin, 2017) deployed by professionals to temper the impact of algorithms on their daily work. In their search for efficiency, advertisers explore many possible targeting options, combine various "ad sets" and easily move budgets according to their respective performances. However, the organizational need to interface Facebook targets with other audiences or customer categories, inside or outside the advertiser's organization, mitigates the use of algorithmic targeting. It explains the success of "custom audience" categories, which translate organizational categories into advertising targets, sometimes enriched with online behavioral traces. Custom audience tools offer a privileged and non-algorithmic way to build targets that will easily circulate and allow coordination at the company level because they integrate part of its existing knowledge. As noted by Kotras (2020), this articulation between big data and organizational categories can be seen as a condition of their epistemic success, rather than a temporary overlay.

A practical and idiosyncratic gaze

By focusing on the advertisers' gaze and practices, our study offers an additional layer to the description of the advertising ecosystem. Previous research has described the advertising ecosystem in terms of "surveillance marketing" (Pridmore and Lyon, 2011) or "surveillance capitalism" (Zuboff, 2019). The surveillance analysis rightly focuses on the sheer amount of personal data that is routinely collected by marketing professionals and put into circulation in the advertising ecosystem (Christl, 2017; Crain, 2019). It adequately describes the potential use of this marketing data by state surveillance actors (Lyon, 2014) and the increased risk of using Facebook advertising tools to discriminate against people and fabricate sensitive or illegal audience segments (Ali et al., 2019a; Angwin et al., 2017; Cabañas et al., 2018; Cotter et al., 2021).

Nevertheless, it also suggests that advertising professionals carefully construct well-defined representations of consumers and then finely steer their behavior. To what extent do advertisers daily "surveil" their targets? When accounting for their practices and skills, online advertising specialists put forward, rather than a precise knowledge of the audience attributes, their ability to test a large number of possible options, their pragmatism in selecting the audiences, their agility in multiplying trial-and-errors to reach the desired metric of efficiency. Professionals have a practical mind-set, build ad hoc, ephemeral audiences, and focus on success metrics. In doing so, they anticipate the fact that a lot a targeting tools perform poorly and that the success or failure of some audiences are not always easy to account for. Basically, targeting tools, in practice, are not tools for knowing populations but rather than means for finding people who will react to a message at a lower cost. Advertising professionals are faced with an ocean of targeting possibilities, they choose the most efficient ones for their short-term objectives, considering their coordination and accountability imperatives.

In this context, advertisers tend to follow the easy path and often rely on audiences they already know. The use of custom audience categories combined with lookalike audiences appears to be the most popular in the ordinary practice of the advertisers that we have studied (in different contexts and with widely different purposes). This idiosyncratic turn of targeting is a significant shift in the online advertising landscape. Our results on this point are in line with existing empirical accounts of online political targeting campaigns, in which small audiences built up from local political databases and fundraising campaigns serve as the starting point for advertising campaigns augmented by lookalike audiences (Hersh, 2015; Madrigal and Bogost, 2020). For marketing professionals, the use of custom audience fosters a company-centered view of the audience, where targets are described in concentric circles ranging from well-known customers to unknown prospects. Advertisers classify audiences according to their links to the company: registered consumers, consumers who have filled out a form, consumers who have visited the website, consumers who made a contact on Facebook, consumers with whom the company never made contact, and so on. Rather than separating audiences between "targets" and "wastes" (Turow, 2012), the advertisers we interviewed take special care of the consumers they know (whose profile is recorded in their database), who can be addressed with sophisticated strategies, while trying to catch new contacts with less predictable outcomes. To the surveillance metaphor, suggesting advertisers accessing every relevant characteristic of the population they scan, we may prefer to describe the daily work of online marketing as herd guarding, where shepherds jealously and carefully take care of their recorded customers, regularly address them, and, from time to time, go on a raid to gain new cattle and enrich the flock. In an ecosystem providing an almost infinite number of data points about consumers, companies are engaged in a process of capitalization of the targets they know. Consequently, these audiences are always at the risk of being over-solicited, known consumers being the object of extra-care by advertisers, possibly trapped in commercial filter bubbles, stuck in the sticky links of the custom audience.

These results help us understand, to a certain extent how commercial advertising acts on society. The advertising gaze is often described in the literature as a surveillance work, aiming precise and extended understanding of individuals based on digital selves (Cheney-Lippold, 2017) that aims to manipulate individuals with powerful algorithmic tools for the benefit of Big Tech companies (Birch and Cochrane, 2022; Zuboff, 2019). Audience planners' practices, however, do not exactly fit this description. Google and Meta are, of course, powerful companies, and they accumulate large datasets on consumers. However, when it comes to advertising power, the mechanisms of targeting are still a combination of relatively precise information, professional intuition, and a large part of trial-and-error. Hence, the power of algorithmic advertising seems more to rely on its ability to deploy commercial filter bubbles through a myriad of customized audiences, competing for degraded consumers' attention (Hwang, 2020).

Conclusion and further research

This research brings three contributions to the existing literature on the social consequences of big data. First, it complements the description of the online advertising industry targeting practices: computer generated categories and optimization are not just overstated (Aral, 2020; Gordon et al., 2019), they also neither match the power of demographics as standards, nor the idiosyncratic meaningfulness of custom audiences. We show that targeting categories undertake three trials: efficiency, communicability, and explainability. Second, this research contributes to the understanding of the socialization of big data categories (Bolin and Andersson Schwarz, 2015; Kotliar, 2020; Kotras, 2020), by describing the hybridization of traditional and algorithmic orders of audience segments' description. Third, it brings empirical insight to the debate on surveillance capitalism and consumer segmentation, by stressing on the importance of local custom audiences capitalized from advertisers.

Our investigation faced various limitations. First, it is limited to the French advertising market. The rate of adoption of automated solutions might vary along national markets and further research may show that. Second, we focused on one technology, the Facebook console. Although it is one of the dominant devices on the market, and although we completed our empirical investigation with a thorough review of the literature on the display advertising market as a whole to ensure the consistency of our results, the balance of targeting techniques might be different on other platforms. Third, this research documents the uses of advertising for consumer goods and services. Our results may not generalize to other areas, such as political advertising, where the stakes are different. Finally, we recorded what audience planners say they do rather than their actual work. This said, our findings are internally consistent among our respondents. Moreover, we presented these results to industry professionals, who agreed that it depicted current practices adequately, and nothing contradicted our interpretation in the trade press. Nevertheless, further ethnographic research on audience planning in the making is needed. It would be interesting to initiate research on the very places where advertising and marketing categories are manufactured—places which typically remain closed to public or scientific gaze, and accessible by too rare insider accounts (Ariztia, 2018; García Martínez, 2016). The significant place taken by "custom audiences" in digital advertising campaigns also invites us to study the categorization and classification operations carried out by the advertisers themselves, from their databases of customers, prospects, or loyalty card holders, insofar as they are used more and more as a basis for the selection of advertising targets for online media advertising. The manufacture of customer segmentations, in particular, among retailers, traditionally relies on calculations based on transaction histories; but they are then often "dressed" and "translated" into more meaningful terms such as "executive manager mom" or "buoyant boomer," which contain implicit demographic attributes (Sunderland and Denny, 2011). These specific categories give structure to the organization and communication of large companies. Do they persist in economic environments increasingly structured around big data and machine learning? Second, we can question the fate of demographic categories in social worlds other than online advertising. The worlds of insurance (McFall et al., 2020), recruitment or social treatment policies (Eubanks, 2018), or public statistics in developing countries (Blumenstock et al., 2015) would probably be very fruitful fields of investigation.

Authors' note

Jean-Samuel Beuscart is now affiliated with Telecom Paris, i3, Institut Polytechnique de Paris, France. Samuel Coavoux is now affiliated with ENSAE, CREST, Institut Polytechnique de Paris, France.

Declaration of conflicting interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was carried out as part of the Algodiv research project (ANR-15-CE38-0001) funded by the French National Research Agency.

ORCID iDs

Samuel Coavoux https://orcid.org/0000-0001-7991-3555
Kevin Mellet https://orcid.org/0000-0002-2200-0170

References

Ali M, Sapiezynski P, Bogen M, et al. (2019a) Discrimination through optimization: how Facebook's ad delivery can lead to skewed outcomes. *arXiv:1904.02095 [cs]*. Available at: http://arxiv.org/abs/1904.02095 (accessed 4 April 2019).

Ali M, Sapiezynski P, Korolova A, et al. (2019b) Ad delivery algorithms: the hidden arbiters of political messaging. *arXiv:1912.04255*. Available at: http://arxiv.org/abs/1912.04255 (accessed 16 June 2022).

Ananny M and 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 (2008) The end of theory: the data deluge makes the scientific method obsolete. *Wired*, 23 June. Available at: https://www.wired.com/2008/06/pb-theory/ (accessed 9 April 2020).

Andrejevic M (2013) *Infoglut: How Too Much Information Is Changing the Way We Think and Know*. London: Routledge.

Andrejevic M (2014) Big data, big questions. The big data divide. *International Journal of Communication* 8: 1673–1689.

Andreou A, Venkatadri G, Goga O, et al. (2018) Investigating ad transparency mechanisms in social media: a case study of Facebook's explanations. In: *Proceedings of the network and distributed system security symposium (NDSS)*. Available at: https://mislove.org/publications/Explanations-NDSS.pdf (accessed 9 December 2022).

Angwin J and Parris T (2016) Facebook lets advertisers exclude users by race. *Propublica*, 28 October. Available at: https://www.propublica.org/article/facebook-lets-advertisers-exclude-users-by-race (accessed 9 April 2020).

Angwin J, Varner M and Tobin A (2017) Facebook enabled advertisers to reach "Jew haters." *ProPublica*, 14 September. Available at: https://www.propublica.org/article/facebook-enabled-advertisers-to-reach-jew-haters (accessed 7 April 2020).

Aral S (2020) *The Hype Machine: How Social Media Disrupts Our Elections, Our Economy and Our Health—and How We Must Adapt*. New York: HarperCollins.

Ariztia T (2015) Unpacking insight: how consumers are qualified by advertising agencies. *Journal of Consumer Culture* 15(2): 143–162.

Ariztia T (2018) Consumer databases as practical accomplishments: the making of digital objects in three movements. *Journal of Cultural Economy* 11: 209–224.

Barocas S (2012) The price of precision: voter microtargeting and its potential harms to the democratic process. In: *Proceedings of the first edition workshop on politics, elections and data*, November, Maui, Hawaii, pp. 31–36. New York, NY: USA - Association for Computing Machinery.

Beer D (2018) *The Data Gaze: Capitalism, Power and Perception*. Thousand Oaks, CA: SAGE.

Beuscart JS and Mellet K (2013) Competing quality conventions in the French online display advertising market. *Journal of Cultural Economy*, 6(4): 402–418.

Birch K and Cochrane DT (2022) Big tech: four emerging forms of digital rentiership. *Science as Culture* 31: 1–15.

Blumenstock J, Cadamuro G and On R (2015) Predicting poverty and wealth from mobile phone metadata. *Science* 350(6264): 1073–1076.

Bolin G and Andersson Schwarz J (2015) Heuristics of the algorithm: big data, user interpretation and institutional translation. *Big Data & Society* 2(2): 1–12.

Bowker GC and Star SL (2000) *Sorting Things Out: Classification and Its Consequences*. Cambridge, MA: MIT Press.

Cabañas JG, Cuevas Á and Cuevas R (2018) Facebook use of sensitive data for advertising in Europe. *arXiv:1802.05030*.

Cheney-Lippold J (2017) *We Are Data: Algorithms and the Making of Our Digital Selves*. New York: New York University Press.

Christin A (2017) Algorithms in practice: comparing web journalism and criminal justice. *Big Data & Society* 4(2): 1–14.

Christl W (2017) *How Companies Use Personal Data against People. Automated Disadvantage, Personalized Persuasion, and the Societal Ramifications of the Commercial Use of Personal Information*. Vienna: Cracked Labs. Available at: http://crackedlabs.org/en/data-against-people (accessed 12 April 2019).

Christl W and Spiekermann S (2016) *Networks of Control. A Report on Corporate Surveillance, Digital Tracking, Big Data & Privacy*. Vienna: Facultas. Available at: http://crackedlabs.org/en/networksofcontrol (accessed 12 April 2019).

Cotter K, Medeiros M, Pak C, et al. (2021) "Reach the right people": the politics of "interests" in Facebook's classification system for ad targeting. *Big Data & Society* 8(1): 1–16.

Crain M (2019) A critical political economy of web advertising history. In: Brügger N and Milligan I (eds) *The SAGE Handbook of Web History*. London: SAGE, pp. 330–343.

Crain M (2021) *Profit Over Privacy: How Surveillance Advertising Conquered the Internet*. Minneapolis, MN: University of Minnesota Press.

Desrosières A (2002) *The Politics of Large Numbers: A History of Statistical Reasoning*. Cambridge, MA: Harvard University Press.

Eubanks V (2018) *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*. London: St. Martin's Press.

García Martínez A (2016) *Chaos Monkeys: Obscene Fortune and Random Failure in Silicon Valley*. New York: HarperCollins.

Glaser B and Strauss A (1967) *The Discovery of Grounded Theory. Strategies for Qualitative Research*. Chicago, IL: Aldine.

Gordon BR, Zettelmeyer F, Bhargava N, et al. (2019) A comparison of approaches to advertising measurement: evidence from big field experiments at Facebook. *Marketing Science* 38(2): 193–225.

Hersh E (2015) *Hacking the Electorate: How Campaigns Perceive Voters*. New York: Cambridge University Press.

Hillygus DS and Shields TG (2008) *The Persuadable Voter: Wedge Issues in Presidential Campaigns*. Princeton, NJ: Princeton University Press.

Hwang T (2020) *Subprime Attention Crisis: Advertising and the Time Bomb at the Heart of the Internet*. New York: FSG Originals.

Kellogg KC, Valentine MA and Christin A (2020) Algorithms at work: the new contested terrain of control. *Academy of Management Annals* 14(1): 366–410.

Kockelman P (2013) The anthropology of an equation: sieves, spam filters, agentive algorithms, and ontologies of transformation. *HAU: Journal of Ethnographic Theory* 3(3): 33–61.

Kosinski M, Stillwell D and Graepel T (2013) Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences of the United States of America* 110(15): 5802–5805.

Kotliar DM (2020) The return of the social: algorithmic identity in an age of symbolic demise. *New Media & Society* 22(7): 1152–1167.

Kotras B (2020) Mass personalization: predictive marketing algorithms and the reshaping of consumer knowledge. *Big Data & Society* 7(2): 1–14.

Lauer J (2012) Making the ledgers talk: customer control and the origins of retail data mining, 1920-1940. In: Berghoff H, Scranton P, and Spiekermann U (eds) *The Rise of Marketing and Market Research*. Springer, pp. 153–169.

Lewis RA and Reiley DH (2014) Online ads and offline sales: measuring the effect of retail advertising via a controlled experiment on Yahoo! *Quantitative Marketing and Economics* 12(3): 235–266.

Lyon D (2014) Surveillance, Snowden, and Big Data: capacities, consequences, critique. *Big Data & Society* 1(2): 1–13.

McFall L, Meyers G and Hoyweghen IV (2020) The personalisation of insurance: data, behaviour and innovation. *Big Data & Society* 7(2).

McGuigan L (2019) Automating the audience commodity: the unacknowledged ancestry of programmatic advertising. *New Media & Society* 21(11–12): 2366–2385.

Mackenzie A (2018) Personalization and probabilities: impersonal propensities in online grocery shopping. *Big Data & Society* 5(1): 1–15.

Madrigal AC and Bogost I (2020) How Facebook works for Trump. *The Atlantic*, 17 April. Available at: https://www.theatlantic.com/technology/archive/2020/04/how-facebooks-ad-technology-helps-trump-win/606403/ (accessed 20 April 2021).

Manzerolle V and Smeltzer S (2011) Consumer databases, neoliberalism, and the commercial mediation of identity: a medium theory analysis. *Surveillance & Society* 8(3): 323–337.

Mellet K and Beauvisage T (2020) Cookie monsters. Anatomy of a digital market infrastructure. *Consumption Markets & Culture*, 23(2): 110–129.

Monohan B (2019) Audience planning ate media planning. Available at: https://www.adexchanger.com/data-driven-thinking/audience-planning-ate-media-planning/ (accessed 7 April 2020).

Nadler A and McGuigan L (2018) An impulse to exploit: the behavioral turn in data-driven marketing. *Critical Studies in Media Communication* 35(2): 151–165.

Nadler A, Crain M and Donovan J (2018) *Weaponizing the Digital Influence Machine. The Political Perils of Online Ad Tech*. New York: Data & Society Research Institute.

Napoli PM (2003) *Audience Economics: Media Institutions and the Audience Marketplace*. New York: Columbia University Press.

Pridmore J and Lyon D (2011) Marketing as surveillance: assembling consumers as brands. In: Zwick D and Cayla J (eds) *Inside Marketing: Practices, Ideologies, Devices*. Oxford: Oxford University Press, pp. 115–136.

Rogers R (2009) Post-demographic machines. In: Dekker A and Wolfsberger A (eds) *Walled Garden*. Amsterdam: Virtueel Platform Amsterdam, pp. 29–39. Available at: http://www.virtueelplatform.nl/downloads/2446_alledgarden_ch04_rogers.pdf (accessed 9 December 2022).

Simon FM (2019) "We power democracy": exploring the promises of the political data analytics industry. *The Information Society* 35(3): 158–169.

Speicher T, Ali M, Venkatadri G, et al. (2018) Potential for discrimination in online targeted advertising. In: *Conference on fairness, accountability and transparency*, 23 and 24 February 2018, pp. 5–19. New York, NY: New York University, NYC - PMLR.

Sunderland PL and Denny RM (2011) Consumer segmentation in practice: an ethnographic account of slippage. In: Zwick D and Cayla J (eds) *Inside Marketing. Practices, Ideologies, Devices*. Oxford: Oxford University Press, pp. 137–161.

Turow J (2006) *Niche Envy. Marketing Discrimination in the Digital Age*. Cambridge, MA: MIT Press.

Turow J (2012) *The Daily You: How the New Advertising Industry Is Defining Your Identity and Your Worth*. New Haven, CT: Yale University Press.

Turow J, McGuigan L and Maris ER (2015) Making data mining a natural part of life: physical retailing, customer surveillance and the 21st century social imaginary. *European Journal of Cultural Studies* 18(4–5): 464–478.

Venkatadri G, Sapiezynski P, Redmiles EM, et al. (2019) Auditing offline data brokers via Facebook's advertising platform. In: *The World Wide Web Conference*, New York, 13 May, pp. 1920–1930. Association for Computing Machinery.

Zuboff S (2019) *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. New York: PublicAffairs.

Author biographies

Thomas Beauvisage is a sociologist at the Social Sciences Department of Orange Labs (Sense). His research topics cover digital markets, online advertising, and personal data.

Jean-Samuel Beuscart is Assistant Professor in sociology at Telecom Paris and IPP in Palaiseau, France, affiliated with i3/CNRS research center. His research interests focus on digital consumption, online market devices and sustainable digital practices.

Samuel Coavoux is Assistant Professor in Sociology at ENSAE and IPP in Palaiseau, France, affiliated with Crest research center. His research interests include cultural consumption, digital practices, and social inequalities.

Kevin Mellet is Assistant Professor in Sociology at Sciences Po, affiliated with CSO (Sciences Po, CNRS). His work mainly draws on economic sociology and science and technology studies to study market techniques in the digital age.