Cookies, Pixels and Fingerprints

Donald MacKenzie

An edited version of this article was published by the London Review of Books on 1 April 2021: https://www.lrb.co.uk/the-paper/v43/n07/donald-mackenzie/cookies-pixels-and-fingerprints

For years, I impatiently clicked 'accept cookies' without, I confess, really knowing what a cookie was. It turns out that it's a small sequence of letters and numbers that a website generates and deposits in your browser. The latter then keeps sending the cookie back to the website when it makes a new request of the site, or when you visit the site again at some later date. The cookie thus identifies you (or, at least, your browser) to the website.

Getting interested in cookies led me to 'pixels'. When used as a tool to gather data, a pixel is a tiny, transparent image, and you can't see it on your screen. When I realised that, without knowing it, I must have downloaded pixels of this kind many times, I was rather spooked. Why would someone want to show you an invisible image? What does the pixel do? It turns out that it itself does nothing. The action that matters is your browser requesting the pixel, which it is prompted to do by computer code that a webpage downloads into your browser as it accesses the page. As a programmer explained to me, 'the code runs in the browser to gather as much information as it can and encodes that into the address of the image it requests'. What you view, what you buy, what you add to a shopping cart but don't actually buy, information about your browser – all that and more can be transmitted via this process. The most widely used pixel seems to be Facebook's, which the advertising technology firm QueryClick reckons is present on 30 percent of the world's 1,000 most visited websites. Facebook’s machine-learning algorithms use the data that the company’s pixels generate to optimise the delivery of ads, and Ian Bogost and Alexis Madrigal reported in the *Atlantic* last April that these pixels were playing exactly that role for the Trump campaign.

Another programmer prompted me to start asking about 'fingerprinting'. Fingerprinting a phone or laptop means scanning it for characteristics that will help to identify it at a later point or in a different context. What exact model is it? Which language setting are you using? Which browser, and which version of it? Precisely how does your browser render fonts on your screen? Even your phone’s or laptop’s level of battery charge can help an automated fingerprinting system keep track of it, at least over the short term. The first advertising executive whom I asked about fingerprinting, last May, told me it had a dubious reputation. It is 'clunky, not a particularly elegant solution', and above all if you engage in fingerprinting 'you're capturing a lot of information that is unique to essentially an individual device or browser … There's a little bit of a grey area there on should you really be capturing all of that information'. But not everyone seems to have his qualms. I've noticed websites increasingly asking me to 'accept' having my laptop or mobile phone scanned, for what I can only assume is fingerprinting.

Finding out about cookies, pixels and fingerprinting has involved dipping my toes in what Shoshana Zuboff famously dubbed 'surveillance capitalism'.[1] Her starting point is the observation that digital systems of any complexity spew out massive volumes of data, much of it what she calls 'digital exhaust', not essential for the operations of the system in question or for improving those operations. Zuboff argues that Google, in her view the primary inventor of today's surveillance capitalism, came to realise that this exhaust data, when processed by sophisticated machine-learning systems, could be used to predict users' behaviour. What kind of advertisement is likely to interest them enough that they will click on it? Will they go on to 'convert', as advertisers put it, in other words to buy the product in question?

Predictions of this kind are obviously of commercial value, and digital advertising was growing fast even before Google was founded in 1998 (Zuboff's history is a little too Google-centric). Traditional advertising, using billboards, newspaper ads, TV commercials and the like, was a somewhat haphazard process that demanded from advertisers a leap of faith in its efficacy. In contrast, techniques such as cookies and pixels – again, Google was not the pioneer of these – yielded copious quantities of data, which seemed to permit both the careful targeting of digital adverts to the desired audience, and the objective measurement of the success of those adverts.

Initially, what was sold was relatively undifferentiated, such as – in the case of Google – ads linked to specific search terms or combinations of them: ‘cheap flights

[1] *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. Profile, January 2019, 691pp., £12.99. 9781781256855

to San Francisco', ‘erectile dysfunction', 'mesothelioma'. (That last term is notoriously expensive. You are unlikely to Google this dreadful form of cancer out of idle curiosity. There is a high chance that you have already been diagnosed with it and might be persuaded to join a law firm's justifiable but also potentially lucrative asbestos-exposure class action lawsuit.) Increasingly, though, it became possible to buy and sell the instantaneous opportunity to show advertisements to a *particular* individual (identified, for example, by a cookie) who had, a fraction of a second ago, entered a term into a search engine, clicked on a link taking them to a website, or was looking at Facebook or another social media platform. 'Real-time bidding', as this came to be called, was quite different from buying and selling advertising over cocktails on Madison Avenue. Automated 'advertising exchanges', in which algorithms could bid competitively for advertising opportunities offered by other algorithms, were set up. 'Markets in future behavior', as Zuboff calls them, had come into being.

The developments that made those markets possible are double-edged in their effects. Large-scale traditional advising is an expensive business. Automated advertising, targeted on what could be quite a limited number of likely consumers, opened up new, cheaper possibilities for start-up companies such as a small clothing brand or craft brewery. 'You can get your product in front of a global niche audience,' the co-founder of one start-up told me, and experiment quickly and cheaply with different wording and images on advertisements and different targeting strategies. On the other hand, targeting requires that data on individuals, their interests, their likely purchases, and so on need to be collected. For example, an advertiser’s algorithm will want to bid quite a bit more if it finds out that a potential customer previously had the product in question in a shopping cart but hasn't actually bought it, while a user with an Android phone is usually a less valuable target than the typically bigger-spending owner of an iPhone.

It is easy to see why Zuboff considers online advertising to be the prototype of surveillance capitalism, although she argues that it was only the starting point. Her *Age of Surveillance Capitalism* ranges widely, from uses of harvested data that (even if their adoption remains spotty) are all too plausible – such as in helping dictate access to or the price of life, health or motor insurance – to the downright bizarre, such as automated vacuum cleaners that try to make money on the side selling floor plans of their owners' apartments. The techniques of surveillance capitalism combine big data, machine learning, commercially available predictions of user behaviour, markets in future behaviour, behavioural 'nudges', and careful structuring of the ‘choice architecture’ (in other words, of the actions available to users and how they are presented). Zuboff's fear is that the ensemble already works all too well in shaping human action, and may become even more effective in determining how we behave, how we think and potentially even our personalities.

Surveillance capitalism, Zuboff writes, desires knowledge of more than the histories of our searches on Google or the contents of our abandoned shopping carts, 'more than my body's coordinates in time and space'. It is 'determined to march through my self'. It is 'imposing a totalizing collectivist vision of life in the hive, with surveillance capitalists and their data priesthood in charge of oversight and control'. She doesn't agree with the familiar remark that if a digital service is free, you're the product. Like an elephant slaughtered for the digital ivory of its data, '[y]ou are not the product; you are the abandoned carcass' (Zuboff seldom understates things). Surveillance capitalism, she believes, is increasingly stripping people of their autonomy, free will and genuine individuality.

Zuboff, clearly, is a fierce opponent of surveillance capitalism. But is she enough of a sceptic in relation to it? Does she overstate its powers? I've spent part of the past year or so finding practitioners of online advertising who are prepared to tell me about what they do and how successful – or unsuccessful – it is, and hanging out (first in person, then online) in the sector meetings in which those practitioners talk to each other rather than to the public. Although, as I've said, digital advertising is in Zuboff's view only the original form of surveillance capitalism, it remains crucial because it is still how much of Big Tech makes most of its money. Google's parent company, Alphabet, runs other businesses, but advertising currently makes up around 80 percent of Alphabet's revenues. For Facebook, that figure is even higher, at over 98 percent. Clustered around those giants of digital advertising is a big, broader ecosystem of mostly smaller firms (by no means all of them advertising agencies of the traditional kind) that sell advertising services, technologies and data.

Are all of the huge sums of money devoted to digital advertising well spent from the advertiser's viewpoint? It can be harder than you might think to know that with any certainty. The 19th-century Philadelphia department-store pioneer John Wanamaker is widely (although probably wrongly) believed to have said: 'I know that half the money I spend on advertising is wasted; the trouble is, I don't know which half.' Today's huge volumes of data don't necessarily change that as much as one might imagine. Everyone in the business would agree that the effects of what they call 'upper-funnel' advertisements are almost always small, and any effects that exist may not manifest themselves for weeks. (An upper-funnel ad, such as most of the 'display ads' you see on Facebook or when you visit a news website, is more akin to traditional advertising, in that it may be shown to you even if you've done nothing to indicate an interest in a purchase.) Online experiments to determine the effects of advertising of this kind can therefore involve looking for a very small needle in a very large haystack, as Randall Lewis and David Reiley – two of the increasing number of economists working for tech firms – put it. Even a well-designed experiment on over a million users may struggle to determine whether advertising of this kind is cost-effective.

There is greater confidence in 'search ads': ads of the kind you get shown when you Google something of commercial interest. You are then lower in the funnel: you may be only a couple of (eminently detectable) clicks away from buying. Here, though, the problem is disentangling the causal effect of an advertisement from mere correlation. For several years, I’ve worn a pair of hiking shoes made by an Italian firm, Scarpa, but they are no longer waterproof, so this autumn I needed to replace them. The shop I usually go to was shuttered, so I bought my new pair online. When I was searching for ‘Scarpa walking shoes’, I got shown ads. Did those ads cause me to buy these shoes? Even if I had been shown an ad for a particular retailer, and had gone on to buy my shoes from them, how sure could the firm be that I would not have done so without it having paid for the ad?

What has become a famous set of experiments demonstrates that apparent success in advertising may not always be real. The auction site eBay used to pay

Google and other search engines to display an ad even if someone's search was simply for ‘eBay’ or included the word 'eBay'. Such ads can seem very effective: they are often clicked on, and many are followed by ‘conversions’ (someone who searched for 'eBay shoes', for instance, often went on to buy a pair on the site). In 2012, however, three economists working for eBay found that levels of traffic on its site were scarcely affected when it stopped these advertisements: those who weren't shown them seemed to find their way to eBay anyway. The trio then started experimenting more systematically by turning off all eBay’s Google search advertising in randomly chosen metropolitan areas, while continuing it in others, and concluded that the $50 million a year that the firm was spending on this advertising was not earning a positive return.

If you have skin in the game, if your pay or career is influenced by whether the advertising for which you are responsible is seen as effective, you may well be wary of discovering that it isn't. Other well-known firms often do what eBay had been doing: paying to show an ad when the firm's name is entered into a search engine. Despite the eBay experiment, that *might* still make sense: firms often fear that if they don't, a competitor could get top slot on the results page, even though Google's rules on the use of trademarks make it difficult, for example, simply to bid for the search term 'British Airways' if you are actually a different airline. What is striking, though, is that other big firms simply don't seem to have tried to find out whether they too were wasting their money. Two researchers, Justin Rao and Andrey Siminov, who looked for traces of analogues of the eBay experiments in a large, detailed advertising dataset, couldn't find them. As they put it, perhaps people who had spent money on such advertising did not want to take the risk that 'past expenditure could be revealed as wasteful’ and that cuts in budgets and headcounts might ensue.

As far as I can tell, the majority of those who work in online advertising firmly believe that the data collection and experimentation possible with digital systems do make the measurement of success possible. I think that they would all agree, though, that what they call 'attribution' is a very difficult issue: Wanamaker’s problem in modern guise. A firm that advertises on any scale it is likely to buy search ads, standard display ads, video ads, ads in other apps, and so on, perhaps along with, for example, conventional TV ads (on what is now condescendingly referred to as 'linear TV'). When a 'conversion' occurs, the customer who buys a product or service has probably seen a variety of the firm’s advertisements. How, then, should credit for the purchase be 'attributed', in other words divided up among those advertisements? Because different people often have responsibility for different types of advertisement, they once again have skin in the game. 'Attribution is a very touchy subject in any company', another advertising executive told me. Marketing teams can be 'really attached' to a particular way of dividing up credit (for example between immediately pre-purchase search ads and earlier display ads), 'so they don't change the attribution model easily'.

A recent book by Tim Hwang, *Subprime Attention Crisis: Advertising and the Time Bomb at the Heart of the Internet*,[2] draws a provocative parallel between the dependence of Big Tech on advertising revenues and the dodgy mortgage-backed

[2] Farrar, Straus and Giroux, 165 pp., October 2020. 978-0-374-53865-1

securities that triggered the near-implosion of the global financial system in the 2008 banking crisis. For example, more than a few of the mortgages packaged into those securities were fraudulent, while online advertising too is dogged by deception such as 'bot fraud', in which the clicks on adverts are generated by computer programs, but advertisers pay for what they think is a human audience. (A lot of effort is devoted to detecting fraud, but the bots, I'm told, are getting more sophisticated. They often used to run on computers in static, identifiable locations such as poorly-policed datacentres, and did little more than pretend to click on adverts. Now malware can install them on the phones or laptops of genuine human users. Today's bots can fill in forms and simulate a human being's hesitant mouse hovers.) I'm not confident, however, that Hwang is right in his overall argument that online advertising's effectiveness is already low and is declining: that seems true in some domains but not others. But if he is right, it would suggest that surveillance capitalism's power to shape human behaviour is less than Zuboff fears.

What is unequivocally the case, though, is that many advertising practitioners are under constant pressure to provide data that show that money is being well spent. There are 'chains of persuasion', so to speak: an advertising agency or other supplier of advertising services has to demonstrate the cost-effectiveness of those services to its client’s marketing department. That department, in its turn, has to justify its budget to senior managers, and sometimes the latter are answerable to owners, such as private equity firms, who take a close interest in how much money is spent and on what. One practitioner to whom I spoke told me that a well-known hotel chain had been one of her clients, and part of her job had been to produce data-laden PowerPoint slides bearing the chain's logo and in its preferred format, so that its marketing department could directly use them in presentations to their bosses.

Many of the divides in the world of advertising have to do with what data should be generated and how, and who should control access to it. At issue is not simply the targeting of advertisements, important though that is, but also ways of measuring advertising's success and attributing the credit for that success. In December, I opened my copy of the *Financial Times* and found a full-page ad by Facebook – a *print* newspaper ad, by *Facebook*! – denouncing Apple: 'We're standing up to Apple for small businesses everywhere'. The ad wasn't explicit on exactly what had riled Facebook, but the heart of the issue is that every iPhone or iPad has an IDFA, an Identifier for Advertisers, which unequivocally identifies the particular device.

If you are an advertiser, big or small, knowing a device's IDFA is pretty useful: it eliminates, for example, any need for clunky and perhaps legally questionable fingerprinting. Previously, a user who wanted to block access to their phone's or iPad’s IDFA had to know that it had one (I didn't), open up its settings, find the relevant setting, and make the change. Now, Apple is going to require every app – on pain of banishment from its App Store (a penalty that even Facebook is not prepared to pay) – to get each user’s explicit permission to access the IDFA or to track them in other ways. It's not going to be like accepting cookies. The interface Apple is using makes it just as quick to deny permission as to give it, and Apple (which treats the App Store as its estate, and sets the rules) won't allow apps to restrict the services they provide to those who don't. The assumption in the industry is that only a small minority of users will agree to be tracked.

Restrictions on tracking, and not being able to access iPhones’ or iPads’ IDFAs make it harder to target advertisements to potentially big-spending app users. (Much of the revenue from game apps, I'm told, comes from a small minority of users who, having downloaded them, go on to spend hundreds or even thousands of dollars on in-game purchases.) Apple's changes also potentially screw up 'attribution'. If you click on an ad for a game or other app (let's say it's an ad on Facebook), are taken to the App Store and install the app, Apple will still send a message to Facebook telling it about the installation. But the message will not now contain the IDFA of your device. So, as journalist John Koetsier explains, neither Facebook nor the app developer may be able reliably to attribute post-installation revenue to the ad. Such ads may no longer seem cost-effective, and the bidding algorithm acting for the app developer will not bid so much to have them shown. It sounds like a small thing, but it could trigger a substantial loss of revenue for Facebook and other companies that carry ads that seek to persuade people to install apps: that business is worth around $80 billion annually.

Similar dispute swirls around that traditional tool of advertising, the cookie. A long-standing principle in the design of browsers is the 'same-origin policy', which keeps a browser's interactions with different websites separate, so limiting the damage that a malicious website can do. One consequence of the policy is that a cookie is readable only by the website that set it in the first place: a browser won't send the cookie to a different website. That could make the tracking of users across websites (and also attribution) horrendously difficult. From almost the very beginning of online advertising, the workaround has been what's called a third-party cookie: one set not by the website being visited but by a different firm's system. For example, the advertising technology firm DoubleClick, set up in 1996 in what was becoming New York's 'Silicon Alley', operated a system that generated advertisements on many different websites. When a user visited one of those websites, that system would, in the course of displaying an advert, deposit a cookie of its own in the user's browser, and then be able to read it when the user visited another of the sites, so making it possible to connect up a user's behaviour on different sites.

As with Apple's IDFA, the third-party cookie is far more consequential than at first appears. There is a network effect at the heart of digital advertising, initially discovered above all by DoubleClick. As more website publishers engaged DoubleClick to generate adverts, the more widely disseminated the firm's cookies became, so DoubleClick gained more information with which to target those adverts, and thus became ever more attractive as an advertising partner. Perhaps the single most pivotal moment in the field's history was in 2007, when Google bought DoubleClick, fusing the former's unparalleled technical expertise (and increasing dominance of search advertising) with the latter's network-effect prominence in display advertising. Although automated advertising may make life easier for new entrants in other industries, network effects of this kind help make the business itself something of an oligopoly, dominated by Google and Facebook (although with Amazon and Apple playing increasingly important roles).

I knew how important cookies are to online advertising from a fine article on the topic by the sociologists Kevin Mellet and Thomas Beauvisage, so I was not taken aback when the hot topic at the industry meetings I attended in 2020 turned out to be not the coronavirus crisis but the coming 'death' of the third-party cookie. Two mainstream browsers, Mozilla’s Firefox then Apple's Safari, had made the blocking of such cookies part of their default configurations. (If I'm right that there has been a shift towards 'fingerprinting', this above all is likely to have been the driver.) Given the centrality of third-party cookies to the field's history, I also wasn't surprised to discover that divides on the issue run through corporations as well as between them. One such divide became apparent in January 2020, when Justin Schuh, Director of Engineering for Google's Chrome (which accounts for over 60 percent of browser use globally), said that it too intended to start blocking third-party cookies by the end of 2021. That caused consternation among at least some of Google's advertising staff: 'People are like, wait, what is Chrome doing there? How are we supposed to do our ads? Why didn't they talk to us first?… Maybe some executive somewhere had seen it and agreed to this … It certainly wasn't communicated to us.'

Decisions such as Chrome’s and Apple’s are often criticised on the grounds that they reinforce the dominant position of Big Tech by denying crucial data to their smaller competitors. Apple’s IDFA decision, for example, is being challenged legally in France on competition law grounds, by a consortium of organisations whose members include not just advertising technology firms but also *Le Monde*. But I think it is a mistake always to be cynical. The engineers and programmers who work on digital advertising or in other roles for Big Tech do often have strong views, by no means always in alignment with the economic interests of their employers. One, for example, tells me that reading Zuboff's *Age of Surveillance Capitalism* 'made me feel bad', and in private life he uses ad-blocking software.

There is, nevertheless, something unsettling – especially in the midst of a pandemic that has forced so much of commerce and of everyday life to move online – about being brought face-to-face with the extent to which crucial decisions that shape what can and can't happen in that sphere are made by private companies and those who work for them. The two main tools of public policy that have been applied so far are data protection law and competition law. They leave crucial issues largely untouched, such as the way in which indiscriminate digital advertising can inadvertently fund hate speech or the dependence of so much of serious journalism on revenue from online advertising, and they have only limited purchase on the shockingly high proportions of that revenue that can get absorbed by intermediaries' fees and markups rather than reaching publishers.

And, crucially, data protection measures and policies designed to enhance competition are often implicitly at odds. It is not too difficult for a big corporation to implement a data protection regime that requires it to get its users' permission for what it does with their data, because its data transfers can be entirely internal, and such a corporation may even welcome rules that stop it sharing that data with other companies. Complying with a regime of that kind is potentially much more onerous for the ecosystem of independent firms, which often need to pass data to one another. Is it possible to preserve for small companies the advantages of digital advertising, while curbing invasive data gathering, and to do both while stopping the field becoming ever more an oligopoly? I hope so, but momentum has yet to build behind any clear blueprint for doing so.