Draft. Published version in London Review of Books, Vol. 44 No. 9 · 12 May 2022

Blink, Bid, Buy

Donald MacKenzie on digital advertising and online news

Click on a link to an article on a news website. If you have a fast Internet connection, you'll start to see the article in a second or two, but the slots for adverts will usually remain empty for another couple of seconds. The ads that you'll see aren't generally chosen in advance. Instead, a near-instantaneous, automated auction of each slot goes on behind the scenes, determining the ad that you will be shown. Even your phone or laptop often joins in, itself gathering bids for the auction. Sometimes, though, an enigmatic pattern or grey rectangle appears instead of an ad.

What goes on in those few seconds is normally invisible, but it shapes the economics of journalism in profound ways. It is the site of sharp, subterranean conflict among news publishers, the big platforms (especially Google), and independent providers of AdTech, the technology of digital advertising. Sales of print newspapers have been falling for years, and revenues from online subscriptions alone are often insufficient to support large-scale, probing news reporting. So advertising income matters greatly.

In Britain, newspapers earned £4.6 billion a year from print advertising as recently as 2007, but by 2017 that had fallen to £1.4 billion, according to figures compiled by the consultancy Mediatique. Online advertising, in contrast, has boomed, but the big money from it has largely bypassed news publishers. Less than 5 percent of online advertising expenditure in Britain is on ads on newspapers' websites, and the resultant revenue (£487 million in total in 2017, reports Mediatique) falls well short of making up for the decline in print advertising.

An advertiser or advertising agency can show ads on a news website by striking a direct deal with the site's publisher, at a fixed price per thousand 'impressions'. (Each time an ad is shown to a reader counts as one impression.) But an advertiser can also buy ad slots, often far more cheaply, in what advertisers call 'open auctions' or the 'open marketplace'.

There, demand from multiple advertisers is brought together electronically with supply from anyone whose online content is accompanied by ad slots (news publishers are distinctly in the minority in the open marketplace). Each ad slot is auctioned separately, and its price is influenced not just by the quality of the website – which often matters less than one would hope – or by the nature of the content surrounding the ad slot, but by what, if anything, is known about you, the user to whom the ad will be shown in the coming couple of seconds.

Prices in these auctions are usually very low: the winning bid for an ad slot will not normally be more than a fraction of a penny. But fractions of pennies add up if earned many, many times over. I'm told, for example, that the *Guardian* alone displays about a billion online ad impressions every month, which brings in around £50-£70 million a year, between a fifth and a third of its total revenues. Without that, its newsroom would be much depleted.

The biggest-volume host of instantaneous auctions is Google's ad exchange, but news publishers and other big providers of online content also bring their supply of ad slots to market via AdTech systems known as Supply-Side Platforms or SSPs. Each SSP pumps out to potential buyers a 'bid stream' of each individual, momentary opportunity to show an ad, in around a second's time, in a specific slot on a particular site. These bid streams aggregate to a Niagara of ad slots, bought and sold via what participants call the 'programmatic pipes': a globe-spanning digital infrastructure of computer datacentres, AdTech platforms and fibre-optic cables.

To process the torrent flowing through the pipes and calculate whether each and every ad slot is worth bidding for – and if so how much – is very demanding. Rather than build their own system to do this, most advertisers and advertising agencies contract out the job to a Demand-Side Platform. One DSP tells me that it sometimes has to process tens of millions of ad opportunities a second, with only around a fifth of a second to respond to each one. If its system takes longer than that, it is simply timed out of the auction for that slot.

Though technologically very impressive, online advertising's programmatic pipes have often been economically opaque, leaking money in ways that are hard to trace. In 2016, the *Guardian* experimented by bidding for ad slots on its own website, and discovered that sometimes only 30 percent of what it paid found its way back to the newspaper. A couple of years later, a team at auditors and advisers PwC was commissioned by the Incorporated Society of British Advertisers to follow the money through the pipework more systematically. The team worked with fifteen huge advertisers, such as Nestlé, Unilever, Vodafone and British Airways, which spend in aggregate around £100 million annually in the UK via the programmatic pipes.

At the pipework's other end, twelve newspaper or magazine groups that are members of the UK's Association of Online Publishers joined the study. The PwC team tried to match ad impressions the advertisers bought with the ad slots the publishers sold. Negotiating access to data held by intermediaries such as SSPs and DSPs proved to be complicated – the PwC study ended up taking nearly two years – and difficulties such as differences in data formats often made matching impossible.

For the 31 million impressions that the PwC team could match end-to-end, on average only 51 percent of what the advertisers spent reached the publishers. Two-thirds of the remaining 49 percent was absorbed by intermediaries' fees. (On average, 7 percent of what advertisers spent went to advertising agencies, and a further 8 percent to Demand-Side Platforms. Supply-Side Platforms charge their publisher clients a similar level of fee, and most transactions involve both a DSP and a SSP, and very likely also other vendors of data and technical services.) But almost

a third of the 49 percent simply couldn't be traced at all: PwC could not discover exactly where around 15 percent of advertisers' spending ended up.

It isn't surprising, therefore, that publishers in the UK and elsewhere have been making concerted efforts to re-engineer online advertising's pipework. Their first main initiative was to make the reader's device (phone, tablet, laptop or desktop) more central to the process of ad selection. Previously, as soon as a reader accessed a publisher's webpage, its ad server would immediately be informed that ad slots were about to become available. The ad server is the system that finally decides which advertisements to show you (it weighs up, for example, the value of the highest bid from the auction against the demands of the publisher's direct deals). These days, it isn't a machine that sits in a publisher's computer room, but a cloud service, in English-language news publishing almost always provided by Google.

UK news publishers' relations to Google are a fascinating, almost a psychoanalytic, topic. They often worry that advertising revenue that should be flowing to them is instead being earned by Google. Distrust of Google and/or resentment of its dominant position seem common. But much of the traffic to publishers' websites comes via Google, and Google's ad exchange and other services for advertisers are the single highest-volume channel of demand for ad slots. Google's ad server is integrated deeply into publishers' systems, making moving to a different server difficult. One publisher I spoke to told me that even though his parent company part-owned one of the handful of other ad server suppliers, shifting to it would have been very disruptive and too risky.

Turning your phone into an auctioneer enables news publishers and owners of other websites to reduce their dependence on Google without turning their back on either Google's ad server or the demand for ad slots that flows through Google's systems. What a publisher does is to add to its webpage 'headers' a snippet of computer code that instructs your browser, when it accesses a page, to send electronic messages to around half a dozen Supply-Side Platforms, inviting them to submit bids for the right to show adverts to you. (The header is the invisible initial part of a webpage that, among other things, tells your browser how to format the content that is about to follow. The header's role in triggering the auction led this approach, which originated in the US, to become known as 'header bidding'.)

Only once your browser has received the SSPs' bids does it alert the ad server, meaning that if Google's systems are to win the auction they have to improve on the highest bid that your browser has been able to obtain. Another attraction for publishers of header bidding is that the SSPs are placed in direct competition with each other, and one publisher told me he hoped that this would minimise what he called the 'fat' in SSPs' fees.

Google, in contrast, was not enthusiastic about header bidding, although it has gradually adjusted its systems to accommodate it. The issue is central to an anti-trust lawsuit against the company launched by a consortium of US states with Republican Attorneys General, who may even hope to break up Google. They accuse it, among other things, of 'schem[ing] to quash header bidding'. It's not a

charge Google accepts. '[W]e do not participate in header bidding … for good reason', the company told the UK Competition and Markets Authority in 2020. Header bidding, Google argued, 'is characterised by increased latency [electronic delays], reduced transparency and significant user trust and privacy concerns'.

Publishers themselves discovered that the original form of header bidding, in which your phone or other device is central to the auction, can have some of the disadvantages that Google cites. The time it takes your browser to establish its electronic links to Supply-Side Platforms can cause ads to load only slowly, and in the process data flows from your phone. In particular, if an SSP has previously deposited a cookie (a small string of letters and numbers that identifies your browser) on your device, it can retrieve that cookie, and so learn something about you. The availability of that information increases the bids your browser receives, but also means that the publisher loses a degree of control over data on its readership.

Greater control is indeed what news publishers often say they want when you ask them about the changes that they would like to see in how their ad slots are auctioned. Many publishers now see 'server-side' header bidding, a different form of the technique, as offering them more control. In this, bids are gathered from SSPs not by your phone or laptop, but by a computer server designated by the publisher. There's no longer a direct link between the SSPs and your device, which returns to a more passive role. Server-side header bidding is, however, more complex for a publisher to set up than the original version of the technique. A manager who is familiar with its challenges tells me that around ten person-years of highly specialist technical development work may be needed, which, as he says, is 'pretty heavy lifting for an individual publisher'.

For reasons such as this, the UK's news publishers have set up their own collective AdTech development effort, the Ozone Project. It brings together publishers with quite different political orientations, such as the *Guardian*,*Telegraph* and News UK, publisher of the *Times* and *Sun*. The biggest obstacle to collaboration, says an AdTech specialist who works for right-of-centre titles, was long-standing 'emotional attachments … Once you let go of that, and actually understand that your true competitors are the Facebooks, the Googles, the Amazons … that's when it becomes easier, I suppose.' The Ozone Project does the heavy lifting of setting up server-side header bidding, and also provides advertisers with a more direct, shorter, less leaky programmatic pipe through which to bid for ad slots.

Leaks from the pipes are not the only way in which news sites lose potential advertising revenue. Another is 'ad-blocking', in which a news story or other online material is judged, usually by an automated system, to be unsuitable for advertising, either in general – such as because the story reports on distressing events or controversial issues – or for reasons that are specific to the product or service being advertised: no tanning salon wants its ads to appear alongside an article on skin

cancer. (This form of ad-blocking is entirely different from the software that users themselves sometimes install, which tries to stop any ads being displayed.)

Newspapers usually classify their content into broad categories such as ‘news’, ‘politics’, ‘sport’ and ‘travel’. Some advertisers simply don’t allow their adverts to be shown alongside ‘news’. In November, for example, the editorial staff of one major UK newspaper quite reasonably chose to classify its reporting of COP26 not under ‘environment’ but as ‘news’ or ‘politics’. Their doing that caused advertising alongside that reporting to dry up almost completely.

The ad-blocking of ‘news’ is particularly widespread at the moment, I’m told, because of the extensive coverage of the war in Ukraine. While it’s easy to see why advertisers don’t want their ads appearing alongside reports on atrocities, they seem to be underestimating responsible news publishers’ efforts to avoid this. One publisher, for example, tells me that in the early weeks of the Ukraine invasion it entirely removed its homepage masthead ad slot, because of the news content likely to appear beneath it.

Big advertisers, though, don’t simply rely on how publishers classify articles. Because they place ads in a wide range of online contexts, they have learned that it’s wise to employ an AdTech firm that specialises in ‘verification’. These firms check that ads are displayed for long enough to count as viewable, and that the viewer is actually a human being, rather than a computerised ‘bot’ that simulates clicks on ads on a website designed to earn money from unwary advertisers. These verification services now also include trying to prevent ads appearing alongside content that the advertiser considers unsuitable.

The giant flood of requests for bids for ad slots that flows through the programmatic pipes is scrutinised by verification systems, not just by Demand-Side Platforms. If the system judges a slot to be unsuitable and therefore stops the DSP that acts on the advertiser's behalf from bidding for the slot, it leaves no visible trace. That form of ad-blocking, though, doesn’t always work perfectly. So there is often a second line of defence: the verification firm’s system does another check just before an ad is displayed. (The time that this check takes can be another reason why ads are often slow to load.)

If the final check suggests a probable violation of brand safety or suitability, and if it’s too late for the publisher’s ad server to find a different advert, the verification firm’s system provides a stock neutral image to fill the ad slot. One such image, a pattern evoking white clouds against a blue sky (then employed by verification provider DoubleVerify), became briefly famous at the start of the coronavirus pandemic, when it was spotted in the masthead ad slots on the *New York Times*’s and *Wall Street Journal*’s homepages, normally prime real estate for online advertising.

A rough but plausible estimate is that UK news publishers lost £50 million in the early months of the coronavirus pandemic to ad-blocking of their coverage of it, causing Culture Secretary Oliver Dowden to appeal both publicly and privately to advertisers in April 2020 to relax their blocking. A year later, though, one could still find articles

being ad-blocked even when connected to the pandemic only indirectly, such as via a discussion of working from home. Another of 2020's news stories that seems to have been the subject of widespread ad-blocking was the Black Lives Matter protests that followed the killing of George Floyd. *Vice* reports that its coverage in this area attracted less than half the advertising revenue it would otherwise have expected.

How can an automated system judge the suitability for advertising of a news story or other web content? A crude but still surprisingly common way of doing it is for an advertiser or advertising agency to have a 'blocklist': a list of keywords – several hundred, I'm told, or perhaps even several thousand – that characterise content alongside which their ads must not appear. The exact contents of blocklists are kept confidential, but it seems as if those in place in 2020 must have contained terms such as 'coronavirus', 'pandemic', 'lockdown', et cetera.

In 2017, though, verification provider Integral Ad Science did publish a list of the twenty then most commonly blocked terms. First was 'explosion', then 'terror'. Others included 'dead', 'shooting', 'gun' and 'kill'. Domestic terrorism seems originally to have prompted the inclusion of these words in blocklists, but today the same words will most likely be ad-blocking reporting of the war in Ukraine. Publishers also say that a wide range of articles can be penalised accidentally. Many things explode, metaphorically at least; footballers shoot; one can dress to kill.

And quite surprising words appear on blocklists. Scott Gatz, of the LGBTQ-oriented US electronic publisher Q.Digital, tells me that one big advertiser entered into a direct deal with it to advertise alongside its Pride coverage, but no ads then appeared on Q.Digital's websites. He asked to see the blocklist that the advertiser had given to its verification provider. On it were 'lesbian', 'gay', 'bisexual', 'transgender', and 'queer' – quite sufficient to block all Pride-related advertising. Gatz says that he quite often finds these words on blocklists.

'Muslim' is another word that seems to be common on blocklists, as is 'Trump', and 'politics' is sometimes a blocked word as well as a category to be avoided. Exactly why such terms are present on blocklists is unclear (I haven't been able to talk to anyone who admits adding them), and inertia may be part of the reason they remain there: several people tell me that they suspect that blocklists are seldom pruned. Certainly, Gatz finds that the members of staff in an advertiser or advertising agency who design campaigns usually do not have responsibility for its blocklist, and, as in the Pride case, may not even be aware of its contents.

No-one I've heard speak about blocklists really likes them. The verification firms, in particular, urge advertisers to use their more sophisticated services. They deploy 'crawlers': programs that systematically digest the content of all the webpages on which their clients are likely to bid for ad slots, sometimes also analysing the links from those pages to other sites. Their systems use artificial-intelligence techniques to classify all this web content, 'reading … the adjectives as well as the nouns', as Richard Reeves of the Association of Online Publishers puts it.

‘It’s absolutely monumental,’ says one verification provider, ‘the number of crawls and categorisations that are going on every millisecond [thousandth of a second]’. Crawling ranges far beyond news websites, but the latter change continuously, so are crawled intensively: one publisher estimates that every new addition to his newspaper’s site will be crawled and categorised in terms of brand safety and suitability within fifteen minutes to an hour.

This giant-scale algorithmic reading of the news (and of other web content) already seems to have considerable influence on the flow of advertising money. While many welcome this as better than the mechanical use of blocklists, publishers are not always happy with its results. Gatz, for example, says that Q.Digital’s webpages are often classed as ‘adult’, a category that includes sexually explicit material, and is often avoided by advertisers.

Nor have AI-based techniques entirely displaced blocklists: fears of damaged careers promote caution. An executive in one advertising agency warns that you can ‘lose a client’ by inadvertently advertising alongside content the client regards as unsuitable. A manager in another agency says that she herself is prepared to advertise without a blocklist. But she understands those who won’t: ‘it’s so hard from having so many safety nets to going, we are just going to remove this … It’s like bungee jumping and going, we are going to remove the net below you’.

If, though, you are looking for a single cause of journalism’s economic difficulties, a publisher said to me, the third-party cookie might be the top candidate. That’s a cookie that is deposited in your browser, not by the website you are visiting, but by a different system, operated perhaps by a Supply-Side or Demand-Side Platform. As the sociologists Thomas Beauvisage and Kevin Mellet point out, those cookies form an infrastructure that permits the same individual – or at least their browser – to be identified in multiple online contexts and by multiple organisations. (There are giant matching tables that, e.g., translate the cookie by which you are known to an SSP to the different cookie set by a DSP.) Why, then, advertise expensively to you via a direct deal with a high-quality news website, you can be reached with the same ad more cheaply on a site that doesn’t change often and/or has content that can be produced without paying professional journalists?

But third-party cookies are on the way out: by late 2023, they will most likely be rejected by all major browsers in their default configurations.[1] Users can still be identified in other ways, but no single framework for doing this has been adopted sufficiently widely to form a fully equivalent infrastructure, and regulators may resist the latter’s creation. In November, Elizabeth Denham, then the UK’s Information Commissioner, warned that her office, which has long-standing concerns about the sharing of data about users in ad auctions, ‘will not accept proposals based on underlying adtech concepts that replicate or seek to maintain the status quo’.

[1] Donald MacKenzie wrote about the coming death of the third-party cookie in the *LRB* of 1 April 2021.

There is therefore a revival of interest in ‘contextual targeting’: advertising focused not on the identity of the user, but on the exact content with which the user is interacting, right at this moment. News websites, with their rich, diverse and ever-changing content, fit well with that approach. That’s grounds for a little optimism.

For who has a deeper and more up-to-date understanding of the content of a news website than its publisher? A verification company, with its ever-active crawlers and its AI-based analytical tools, may nevertheless hope to come a close second, and several verification companies have added contextual targeting to the services they offer advertisers.

And that has become yet another source of conflict. The publishers I’ve spoken to believe that it is they, not AdTech firms, who should offer advertisers the information needed for contextual targeting on their websites. One publisher even speculated about denying AdTech’s crawlers access to those sites. ‘If you’re a *Times*, *Telegraph*, or an *FT* [*Financial Times*], you can probably do that’, he says, but ‘if you’re a *Mirror*, a *Mail*, or a *Sun*, then it’s going to be harder’. If a news site is, as he puts it, a ‘mass-market proposition’, then it will of necessity depend on open auctions, and when bidding in them advertisers need the reassurance of the automated assessment of brand-safety risks.

Another way, though, in which things are changing is that there is a growing sense among advertising practitioners that they need to take responsibility not just for the content of adverts, or even for brand safety as narrowly conceived, but for the wider impact of ad spending. As the influential UK-based pressure group, the Conscious Advertising Network, points out, this means not funding hate speech or misinformation, even inadvertently, but also not being reliant on the mechanical use of blocklists. (A current concern of the group’s co-founder, Harriet Kingaby, is blocklists that include words related to climate change.)

If you’re a big advertiser, to advertise ‘consciously’ almost certainly means being more selective. On average, the firms that took part in the PwC study each advertised, in just three months, on over 40,000 websites in the UK alone. That’s a scale likely to defeat genuine human oversight.

But how to be selective in your ad spending? Surely it must involve supporting activities that contribute to human betterment. None of us can have first-hand knowledge of more than a tiny sliver of what goes on in the world; much of what we know we have to learn from others, including journalists. Their profession – and, at its best, journalism is a profession in the fullest sense – requires the experience to distinguish unfounded, attention-seeking speculation from frequently mundane truth, and the courage to pursue truth even when the latter is awkward, discomfiting and dangerous. Intermediaries’ excessive fees damage the economic base of this crucially important activity. Indiscriminately ad-blocking ‘news’ incentivises the production instead of bland lifestyle articles.

It is encouraging that the *Guardian*'s subscribers and 'supporters' (who make regular donations) now number more than a million. The *Financial Times* also has more than a million digital subscribers and the *New York Times* more than ten million. But a handful of successful titles is not enough. Local and regional newspapers, for example, have suffered especially severely from declines in circulation and advertising revenues, but can still make a big contribution to their communities: local, not just national, decision-makers need held to account. And there are, of course, outlets that specialise in reaching the very audiences whose ethnic, religious or gender identity or sexual orientation involves one of the words that too often populate blocklists. Such outlets are worth support in and of themselves, and also offer a way for advertisers to broaden the reach of their advertising.

Advertising revenue alone is of course seldom going to provide adequate support for extensive in-depth journalism. That requires diverse sources of income, above all sales and subscriptions, but also initiatives such as the *Guardian*'s supporters scheme, revenue from events, perhaps philanthropy and (in some areas, such as local journalism) maybe even central government support. Nor is it going to be easy to achieve sustainably higher advertising revenue. But advertising that keeps in mind the need to support independent, professional journalism may start to deserve being called 'conscious'.

Donald MacKenzie