



Unregulated and Segmented Dark Ads on Social Media

Consumer Education and Regulatory Options

Mark Andrejevic, Robbie Fordyce, Nina Li, and Verity Trott
In collaboration with Daniel Angus and Jane Tan

December, 2020



Unregulated and segmented dark ads on social media: consumer education and regulatory options

Authored by Mark Andrejevic, Robbie Fordyce, Nina Li, and Verity Trott in collaboration with Daniel Angus and Jane Tan

Published in 2021

This project was funded by a grant from the Australian Communications Consumer Action Network (ACCAN).

The operation of the Australian Communications Consumer Action Network is made possible by funding provided by the Commonwealth of Australia under section 593 of the Telecommunications Act 1997. This funding is recovered from charges on telecommunications carriers.

Automated Society Working Group
School of Media, Film, and Journalism
Monash University

Website: www.monash.edu/arts/media-film-journalism/automated-society

Email: automated.society@monash.edu

Australian Communications Consumer Action Network

Website: www.accan.org.au

Email: grants@accan.org.au

Telephone: 02 9288 4000

If you are deaf, or have a hearing or speech impairment, contact us through the National Relay

Service: www.communications.gov.au/what-we-do/phone/services-people-disability/accesshub/national-relay-service/service-features/national-relay-service-call-numbers

ISBN: 978-1-921974-67-0

Cover image: Priscilla du Preez / Unsplash



This work is copyright, licensed under the Creative Commons Attribution 4.0 International Licence. You are free to cite, copy, communicate and adapt this work, so long as you attribute the authors and “Monash University, supported by a grant from the Australian Communications Consumer Action Network”. To view a copy of this licence, visit creativecommons.org/licenses/by/4.0

This work can be cited as: Andrejevic, M., Fordyce, R., Li, N. & Trott, V. 2020. Unregulated and segmented dark ads on social media. Australian Communications Consumer Action Network, Sydney.

Table of Contents

Figures and Tables	3
Acknowledgements	4
Executive Summary	5
Introduction	7
Literature Review	13
Ad Targeting	13
The Social Role of Advertising	14
The Rise of Dark Ads	16
The Changing Advertising Economy	18
Current Measures to Address ‘Dark Ads’	20
Methodology	24
Research Design and Participant Recruitment	24
The Ad Collection Tool	25
Automated Data Analysis	26
Application Method	30
Limitations	34
<i>Privacy</i>	34
<i>Mobile</i>	35
<i>Duplicates</i>	35
Methodological Complications	35
Technical Approach and Challenges	36
<i>In-line character-opacity obfuscation</i>	37
<i>‘Aria-labelledby’ obfuscation</i>	38
<i>Zero-size object obfuscation and margin obfuscation</i>	39
<i>Qualitative and image-classification tagging</i>	39
Interviews	40
Focus Groups	41

Results42

 Dark Ads Facebook Tool, Website and Mobile App 42

 Personas and Ad Patterns. 42

 Participant and User Responses and Literacy. 45

 Demographic Distribution by Ad Category 46

Conclusions50

Recommendations 51

 Recommendation 151

 Recommendation 251

 Recommendation 351

 Recommendation 4.51

 Recommendation 5 52

 Recommendation 6. 52

 Recommendation 7 52

Authors 53

References54

Figures and Tables

Figure 1	7
Figure 2	8
Figure 3	9
Figure 4	9
Figure 5	20
Figure 6	21
Figure 7	22
Figure 8	26
Figure 9	27
Figure 10	28
Figure 11	29
Figure 12	30
Figure 13	30
Figure 14	31
Figure 15	32
Figure 16	32
Figure 17	33
Figure 18	34
Figure 19	36
Figure 20	38
Figure 21	43
Figure 22	44
Table 1	37
Table 2	38
Table 3	49

Acknowledgements

The authors would like to thank all our research participants who donated their Facebook ads to our project. This was invaluable for the project's success and has demonstrated the effectiveness of our tool.

We would also like to thank the support of the Culture Media Economy Research Group in the School of Media, Film, and Journalism at Monash University.

We would also like to thank Dan Mattingly and Nik Matovinovic for their research assistance on this project. We are grateful to the assistance of the research assistants who assisted with tagging the advertisements: Rijul Baath, Dan Hunt, and Vaasanthi Palepu.

Finally, we would like to thank the students and staff in the School of Media, Film, and Journalism who helped us trial and refine our app dashboard.

Executive Summary

- **The online platforms that we access via our telecommunications infrastructure are fundamentally changing the way advertising works, with significant social consequences.** Advertising has an important role to play not just as a form of market signalling but also as a form of social messaging and, at times, as a means of discrimination and social sorting. We know that advertising has been used for progressive as well as pernicious purposes, providing public service messages, but also reinforcing harmful ethnic and gender stereotypes, and in some cases for spreading false or misleading information. For this reason, it is important to hold advertising systems and advertisers accountable for the forms of messaging they promulgate.
- **Accountability is much harder to achieve in online contexts than in the mass media era, thanks to the combination of data-driven micro-targeting with personalized devices and lagging regulatory frameworks.** Commercial messaging on social media takes the form of so-called “dark ads” that are visible only to those to whom they are delivered. This form of targeting means that the content is not generally available for public inspection and also that the pattern of distribution of ads is non-transparent, even to those who receive the ads. Moreover, online ads are often very short-lived, undergoing constant transformation, which renders them ephemeral and thus difficult to track and hold accountable for their content and their pattern of distribution.
- **The lack of accountability enables the spread of disinformation, problematic forms of targeting, and new forms of discrimination and stereotyping.** There are two issues at stake here: problematic content and problematic forms of distribution. For example, ads for housing or employment could be distributed in discriminatory ways: appearing only to men or only to white people. Forms of content that promulgate harmful stereotypes or false/misleading information could be distributed only to those unlikely to object to or to fact check the messages they receive.
- **Without any insight into the content of dark ads and their pattern of distribution, it is difficult to get a sense of the extent to which harmful practices are widespread and in need of public response and regulation.** For all practical purposes, without some form of supervision, online advertising exists in a de-facto unregulated Wild West, enabling advertisers and platforms to see what they can get away with. Laws against discrimination and false advertising exist, but they are difficult to enforce when the ads are fleeting, ephemeral, and not publicly available.

- **This project develops one model for providing accountability for dark ads on Australia’s most popular social media platform, Facebook.** By showing what accountability looks like, the research reminds us just how dramatically the advertising environment has shifted, and of the kinds of questions we need to be asking to ensure that dark ads are not abused in ways that are detrimental to society.
- **The research tool we developed represents the first attempt to make public not just what ads are being served to people online, but how advertising is distributed across demographic groups.** The tool enlists volunteers to install a browser extension that captures the ads appearing in their Facebook news feeds. When they install the extension, they are asked to provide some demographic information about themselves. The ads are collated by the tool and displayed in a form that can be filtered by demographic category. This allows us to query the database by, for example, sex, or income level, or party preference – or some combination thereof – to see which groups receive what types of ads.

Introduction

During the Australian Federal Election in late 2019, a series of customised ads appeared in the news feeds of Australians as part of a Liberal Party campaign accusing the Labor Party candidate, Bill Shorten, of seeking to raise taxes on popular car brands. There were two notable features of these ads that raised public concern about how they fit within a broader pattern of targeted advertising that received attention during the 2016 US Presidential campaign and the Brexit campaign in the UK the same year: the ads were both false and micro-targeted. As *The Guardian* reported, the ad campaign used, “Facebook ad functionality to target users with an interest in particular vehicle types to make the false claim about Labor’s policy” (Karp & Evershed, 2019). The ads drew on the ability of advertisers to target users with a demonstrated interest in a particular car brand – but likely also drew on the ability of Facebook to track user’s online activity. Based on that activity it is possible to infer what type of car someone is likely to own – or be interested in buying.

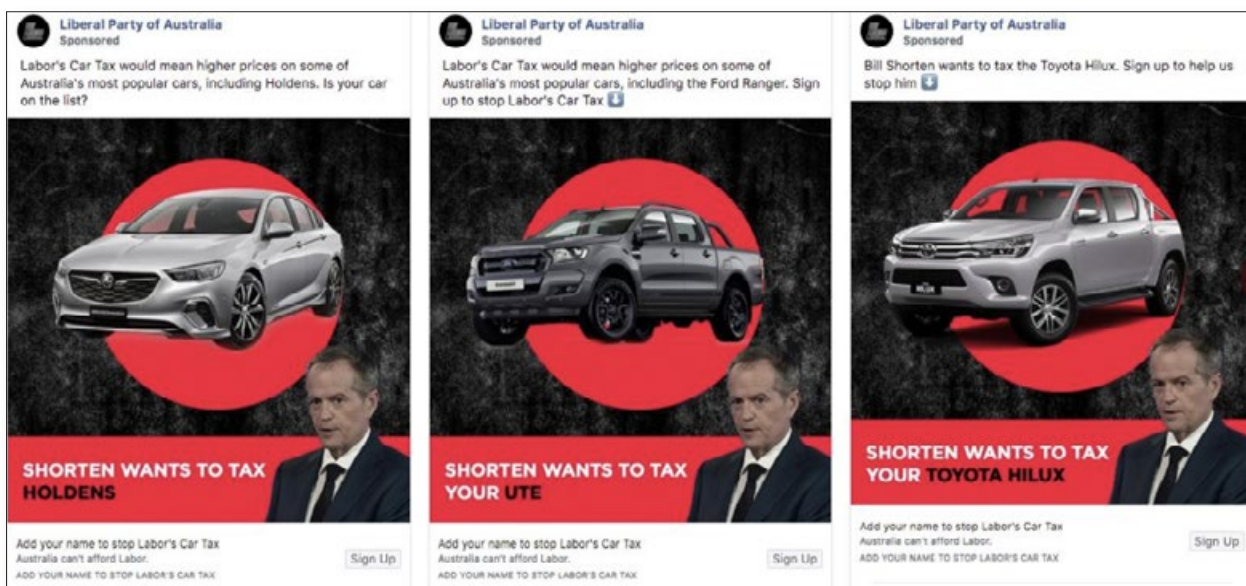


Figure 1: Targeted ads from the 2019 Australian Federal election campaign

The Australian campaign ads fit within a broader messaging trend that will likely play an increasingly important role in political advertising: the use of data collected from online platforms combined with the malleable character of digital messaging to create micro-targeted influence campaigns. The goal of such campaigns is to discern the interests and characteristics

of individual users or groups of users so as to be able to influence them more effectively with tailored content and imagery. This strategy was enlisted by the Brexit campaign, which invested almost \$4.5 million in a targeted ad campaign that developed ads built around wedge issues and micro-targeted at Facebook users based on profile data. The campaign also launched a data harvesting ad scheme that offered a prize of \$50 million Euros to anyone who could guess the results of all 51 games in the European Football Championship (an all but impossible challenge that no one won). The goal of the campaign was to get participants to provide personal data that could be used for relevant ad targeting (see Fig. 2).

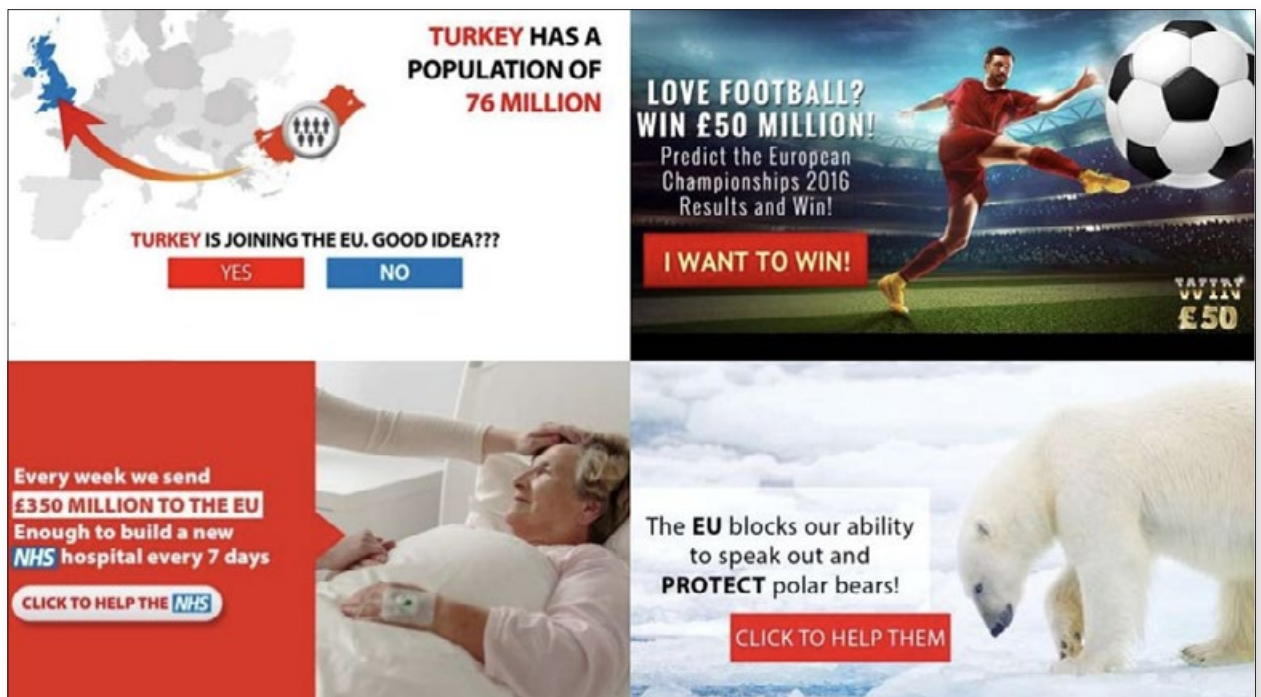


Figure 2: Examples of targeted leave ads on Facebook, including the data harvesting football contest.

Similarly, in the US 2016 Presidential election, the Trump campaign had an embedded Facebook team to assist in their online messaging strategy, which included the attempt to suppress voter turnout among African American voters by targeting them in key districts in battleground states with ads that falsely accused Hillary Clinton of calling them “super predators.” Because of the way targeted advertising works, those who received the ads would have been unaware that they were only visible to African Americans. Indeed, such a strategy might have worked to increase Clinton’s favourable rating among some pro-Trump groups. In other words, the targeted approach allows for advertising strategies that might significantly differ from approaches undertaken in a setting where the ads are widely available for public inspection by a broad audience of the general public.

The use of dark ads is not limited to political contexts. The investigative news outlet *ProPublica*, for example, demonstrated that it was possible to target job ads and employment ads to specific demographic groups based on race and gender – in violation of Federal anti-discrimination law in the United States (see Fig. 3).

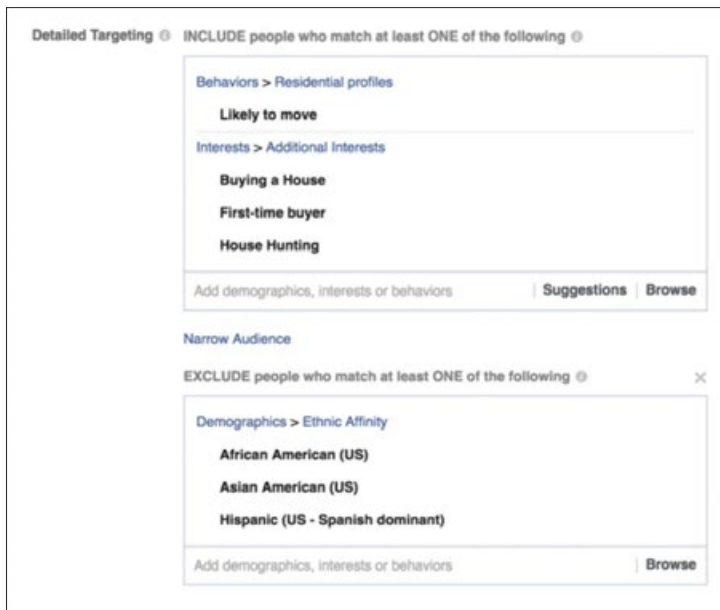


Figure 3: ProPublica demonstrates that it can purchase a housing ad on Facebook that makes it possible to exclude ethnic minorities – in violation of anti-discrimination laws. Facebook says it is no longer possible to purchase such ads.

Complicating matters is the fact that discriminatory patterns are not always or necessarily intentional. Because digital targeting happens on mass scale, much of the tailoring is automated. Data driven advertising systems learn how past ads perform and use this data to target new ads that they classify as similar in form or content. An advertiser may not



Figure 4: Harvard Professor Latanya Sweeney found that Google searches using names historically associated with African Americans were more likely to yield ads related to criminal activity: here she shows ads returned using her own name.

Facebook says it has changed its system to disallow such forms of discrimination – but micro-targeting makes it difficult to hold platforms accountable, especially when a range of other categories can be used as proxies for race, gender, and other protected categories. Moreover, it allows for new forms of discrimination and stereotyping that fly below the radar of anti-discrimination laws for housing or unemployment. It enables brands, for example, to target the groups they want associated with their products and to exclude other groups. Likewise, it can reproduce harmful stereotypes or reinforce historical patterns of discrimination. Consider, for example, the findings of Harvard Professor Latanya Sweeney: names historically associated with African Americans were more likely to trigger ads on Google associated with criminal activity and arrest records (see Fig. 4).

specifically request that an ad be delivered only to a specific demographic, but the automated system may optimise potential clicks by allocating the ad to particular groups or individuals in a way that looks biased by gender, ethnicity, or other variables. Even if the intention to discriminate is not present, the automated output may demonstrate patterns of discrimination as a result of the automated optimisation system. In this respect historical patterns of discrimination or stereotyping may be



“ Will we lose the ability to form a shared understanding of a political candidate, because each of us receive a completely different set of messages from that candidate? ”

reproduced in automated systems, as happens frequently online in realms ranging from predictive policing (which predicts crime based on past patterns of over-policing particular groups or geographic areas) to automated natural language processing, which reproduces historical forms of sex discrimination and stereotyping by, for example associating particular job categories with men or women.

The increasing significance of online advertising, then, raises a host of social issues that are likely to become increasingly pressing as people's media consumption and participation shifts increasingly toward digital media available in the home, at work, and via mobile devices wherever one happens to be. The trajectory is clear: online advertising is poised to become an increasingly dominant component of our media ecosystem, and this means we can expect increasingly granular forms of targeted, individualised messaging delivered to personal devices. We will not know who else is seeing these ads or what ads others are seeing. Ads may change and adapt from minute to minute according to the responses they receive in real time, which means the ad we saw a half hour ago may be very different from the one we are seeing at the moment – even if it's for the same product, service, issue, or candidate. The online ad ecology will become even more fleeting, ephemeral, and volatile than it already is, and equally will become harder to understand, predict, and regulate - even for the ad platforms themselves.

This raises some daunting questions: will forms of ethnic and gender stereotyping that, as a society, we have rejected, become resurgent once they can be targeted exclusively to those who are likely to respond to them and not be offended by them? Will we lose the ability to form a shared understanding of a political candidate, because each of us receive a completely different set of messages from that candidate? Will historical forms of stereotyping be reproduced in patterns that shape who sees what kinds of ads? Will new forms of discrimination emerge based on details unavailable to advertisers in the mass media context? Will advertisers increase their use of deceptive and misleading strategies on the assumption that is harder to detect these practices in a fast-changing online environment? If any of these developments are profitable – and many of them likely are – we can anticipate that there will be an incentive on the part of advertisers to pursue them, with socially detrimental consequences. The only way to ensure that this important part of our media environment – the customised commercial messaging that supports many of our digital media platforms – is held accountable to civil rights commitments and protections is to develop a system of accountability that can keep up with the changes in the advertising environment.

The challenges of providing accountability for online, targeted, 'dark' ads are multiple. An ad campaign may run hundreds or even thousands of variations of ads in response to ongoing forms of A/B testing – systematically varying elements including text, image, and design. This means ads can disappear in a matter of minutes or hours. The metrics that shape the selection and distribution of ads are non-public and proprietary, which means potentially discriminatory patterns are non-transparent and difficult to obtain. A comprehensive overview would require either total disclosure on the part of the platforms – which seems like a very long-shot proposition – or the ability to collect everything that all end users see – a daunting proposition that would require a data collection infrastructure on a par with that of the largest tech companies. Another possibility would be to crowdsource accountability – to counter the centralized power of the major platforms with tools that allows users or groups of users to explore which ads are being served to what groups in order to discern patterns of discrimination and the dissemination of false or misleading information.

The approach we have taken draws on the crowd sourcing model by inviting end users to install a web browser extension that collects the ads served to them in real time, combining them into a database that can then be sorted according to demographic filters. The tool also allows individual users to track all the ads they have received and to compare these to the overall patterns revealed by the database. There are a number of limitations to such an approach – covered in more detail in the Methodology section – but its primary purpose is to draw attention to the issues raised by dark ads through the development of a prototype for possible accountability strategies. The need for accountability will become increasingly pressing as advertisers continue to migrate from less malleable and targeted modes of commercial messaging in print and broadcast to online media – and as digital customisation comes to pervade all forms of commercially supported media.

The shift marked by digital interactivity is away from content-linked advertising to so-called programmatic advertising that follows users, whatever content they may be engaging with. In the mass media era, advertisers relied on proxy indicators of the composition of the audience to target ads: time slot and content served as prominent proxies (for example, daytime TV targeted advertising to those working at home – typically doing domestic labour – hence the term ‘soap opera’). The promise of interactive media is to bypass proxy measures by learning about particular users and targeting them regardless of time and content. If an online ad service knows that I’ve recently been searching for good deals on a bike, for example, bike store ads will appear in my news feed regardless of what time I’m online and what content I am engaging with. This approach will carry over into emerging forms of customisation associated with augmented reality: the information overlaid onto physical space (including ads) will be dynamic, following individual users, rather than the static forms of publicly visible messaging we associate with physical space. The future of digital advertising anticipates a world in which the model of online customisation comes to characterise the physical world. Just as two people may see entirely different messaging streams online, augmented reality and “smart” billboards create customised spaces: two people can look in the same direction and be exposed to entirely different information. If this information is displayed on a device, such as a smartphone or smart glasses, it will, once again, be visible only to its intended target. For all practical purposes not just the informational spaces we inhabit will be different from one another, but even the physical spaces – or the way we experience them. When that happens we will need tools as a society to understand the systems that customise our worlds for us – this project anticipates that time and the need to address it in advance.

Literature Review

Ad Targeting

The goal of targeting advertising is as old as promotional culture itself – as rehearsed in the familiar repetition of marketing pioneer John Wanamaker’s 19th Century lament: “Half the money I spend on advertising is wasted; the trouble is, I don’t know which half.” (Bradt 2016) With the advent of the mass media and the explosive growth of advertising over the course of the 20th century, the challenge for marketers became how to cut through “the clutter.” This led to the capture of new spaces and strategies for advertising, including the goal of making audiences “work” more efficiently by ensuring that the ads they received were likely to be in some way relevant to them. The concern was that, in the case of mass advertising, ads were being wasted insofar as they were reaching audiences to which they had no relevance or purchase. The rise of niche media was one result of this strategy: if you create a magazine specifically for ferret owners, companies advertising ferret products can be sure that the audience they are paying for is likely, as a group, to have a higher average level of interest in their products than a more general audience for a mass circulation newspaper. Niche media strategies rely on content as a proxy for audiences. The underlying goal for targeting, would be to bypass the proxy to get directly to the expressed and inferred interests of particular individuals. Shared devices and aggregated audiences pose a challenge for this level of customisation. However, the increasing penetration of personal devices, including smartphones and tablets, in conjunction with addressable accounts on platforms like Facebook, YouTube, and Twitter, provide an infrastructure for individualised data collection and content targeting. What looks to individual users like an increasingly personalised media infrastructure simultaneously appears to advertisers as a means of attaining advertising’s Holy Grail: individual messaging that has the capability to measure “return on investment” directly, since the behaviour of individual users (whether they click or not; purchase or not) can be captured by the platform. Thus, although there is a longer history of targeted advertising that can be traced back through the use of “zoned” editions of newspapers, direct-mail marketing, and geo-sorted magazine inserts, the rise of Internet platforms such as Facebook and Google ushers in an unprecedented combination of granular targeting, international scale, and opacity (see, for example, Turow, 1997).



“ ... advertising enabled the shift from a production oriented industrial society to a consumer oriented one.

The Social Role of Advertising

Advertising is typically overlooked when we think about the content of the media: primarily in the form of the news and entertainment that we focus on. This is because, for the most part, we do not seek out the ads; rather they are a by-product of the economic system that supports our access to news, information, and entertainment. Ads are the content that, for the most part, we skip over, try to ignore, and dismiss as largely irrelevant to our intended media experience, whether that be watching a video or a tv show, reading a news article, or listening to music.

From an economic perspective, of course, ads are a crucial component of our media economy – they support the production and distribution of the news and entertainment content. At the same time, they are an integral component of the creation and reproduction of a consumer

focused society. As the historian Roland Marchand (1985) argues, advertising enabled the shift from a production oriented industrial society to a consumer oriented one. The advent of first steam-powered machinery and then electricity combined with the standardisation of production and labour processes made manufactured goods cheap and plentiful in industrialised societies. However, people's way of life had to be transformed to encourage the levels of consumption necessary to keep pace with production. Advertising provided messaging that encouraged a shift from home-based production of staple items (clothing and food products) to the purchase of mass produced goods. It provided instruction in patterns and practices of consumption for workers disembedded from agricultural life and re-installed in urban manufacturing centres. In combination with branding, it helped producers gain control over pricing by highlighting the alleged benefits of particular products, and, over time, it built brand associations that have become one of the key determinants of brand value (Room, 1998). At the same time, it tapped into social hopes and fears, helping to frame fantasies of the good life and how this might be obtained via the acquisition of consumer goods that could improve one's love life, social and professional standing, and so on. Marchand (1985) traces the shift of advertising away from informational content (detailed product description) and toward associational marketing (that demonstrates the role products can play in achieving a desired and desirable lifestyle). As the historian TJ Jackson Lears put it, advertising has collaborated with other social institutions to promote, "dominant aspirations, anxieties, even notions of personal identity" (1995: 2).

In addition to its role in mobilizing consumption to keep pace with the productivity of industrialized mass production, advertising thus has a broader cultural significance. As Michael Schudson puts it advertising, "may shape our sense of values even under conditions where it does not greatly corrupt our buying habits" (1984: 23). For Schudson, the symbolic power of advertising comes from the way ads "make us mind, make us focus – and on some things rather than on others. Ads do not 'merely' reinforce existing social trends: they re-enforce social trends, and some trends and not others" (1984: 24). In this respect, he notes, "Advertising, whether or not it sells cars or chocolate, surrounds us and enters into us, so that when we speak we may speak in or with reference to the language of advertising and when we see we may see through schemata that advertising has made salient for us" (Schudson, 1984: 210).

Given its cultural role in shaping attention and reinforcing social trends, much work has been done on the role played by advertising in reproducing stereotypes, preconceptions, and dominant meanings and associations. Scholars and researchers have explored the role played by advertising in shaping attitudes toward female body image and beauty (Kilbourne, 1990); racial preconceptions and prejudices (Wilson and Gutierrez, 1995); and class (Marchand, 1985), among other areas of social life. For this reason, advertising has been, historically a key focus of attempts to challenge derogatory social stereotypes that legitimised forms of historical exploitation, abuse and violence. Challenging racism and sexism means critiquing the messaging systems that reinforce them – including racist and sexist ads.

As ads come to permeate contemporary life, the values and attitudes they select and reinforce become a core component of the informational atmosphere through which we move, in combination with the influence of the family, schools, and other arenas of meaning making and cultural production. The strategies and systems that shape this atmosphere, then, bear close scrutiny – especially during a time when advertising is undergoing profound shifts in its mode of production and distribution.

The Rise of Dark Ads

In the era of mass media, advertising strategies might be described as relatively unconcealed. For example, ads in newspapers were available to all who read them – and were preserved in the archives, thanks to media like microfilm and microfiche: they could be retrieved and studied. With the advent of recording technology, the same became true of radio and television advertising – which were also widely distributed, often to unintended audiences. We know, from the complaints in the United States to the Federal Communication Commission that controversial ads, or those deemed objectionable by some audiences, could generate response and, in some cases, government action. In the US, for example, viewers regularly file complaints about the “edgier” ads aired during the Super Bowl, the most watched media event of the year. These range from concerns about indecent content to claims of false or misleading content – claims that routinely receive media coverage and occasionally generate public discussion. In the United Kingdom, new regulations ban ads that reinforce gender stereotyping. The rules have already resulted in two ads being pulled: one that features, “dads bungling comically while looking after their babies” (Sweney, 2019). Like the US and the UK, many countries place restrictions on false, misleading, or socially harmful advertising strategies (including, for example, advertising targeting children). These regulations allow societies to make the cultural impact of advertising an object of public scrutiny, deliberation and collective response. They subject commerce to overarching social and democratic values – but they are only effective to the extent that advertising generates some kind of public impression and public record. One of the under-appreciated and under-studied drawbacks of commercial targeting from a social and cultural perspective is the fact that it renders a significant and influential part of culture socially opaque. This opacity is a stubborn one because it remains, from a commercial perspective, seemingly advantageous to marketers, advertisers, and campaign strategists.

The ability to avoid public scrutiny is, at least for some marketers, an added bonus of targeting – it frees up advertisers to develop and implement strategies that would have been objectionable and caused public and legal backlash where they subjected to shared public scrutiny. As the independent media outlet *ProPublica* discovered, for example, Facebook’s automated ad targeting system made it possible, in theory, for advertisers to discriminate by race in the delivery of ads for jobs and housing – potentially violating Federal law. *ProPublica* was able to prove this by setting up their own ad buy, however, they could not determine whether this feature had been put to use in the past by advertisers, precisely because there is no public record of the range of ads served to Facebook users over the years (Angwin et al, 2017).

With respect to gender, it is worth noting that recent research reveals Facebook’s own algorithms may reinforce stereotyping by introducing gender bias into the distribution of online ads (Ali et al, 2019). This research revealed that even when advertisers do not specifically target a particular gender, the algorithm might “decide” based on the success of past campaigns, that a particular ad is more likely to receive clicks from one or another gender. It will then distribute those ads accordingly, resulting in, “potentially discriminatory ad delivery, even when advertisers set their targeting parameters to be highly inclusive” (Ali et al, 2019). The combined result of the strategic use of “dark ads” by marketers and the perpetuation of gendered, racialized, or classed stereotypes by automated systems is the prospect of new configurations of stereotyping, discrimination, and the promulgation of

unaccountable forms of potentially socially harmful cultural messaging. The social constraints once relied upon to contain the pathological impulses of advertising fall away when advertising goes dark. Even when restrictions are in place in principle – to prevent alcohol or cigarette advertising to minors, for example, or to eliminate racial discrimination in housing ads – these become increasingly hard to enforce if the ads are not available for public scrutiny.

By every indication, advertising dollars are migrating toward the platforms that enable non-transparent customization strategies. Online ad spending globally outstripped television in 2016 and continues to grow rapidly (Molla, 2018). This spending is overwhelmingly captured by the large technology companies: “Google, Facebook, Amazon, Verizon, Microsoft, Twitter and Snap combined have 80 per cent of the market,” according to a 2019 report (B & T Magazine, 2019). The result is that the commercial media ecosystem is becoming increasingly reliant upon an opaque, automated advertising system.

This opacity cuts two ways: programmatic advertising, which relies on data-driven systems to seek out the most likely targets, can also be non-transparent to the companies who pay for ad placement; that is, they may not know which media outlets are hosting their ads. The activist organization Sleeping Giants has sought to shed light on this process by notifying large ad buyers when their ads crop up on outlets with which they might not want to be associated. Ad buyers can specify which outlets to include and exclude from their ad buy. Such accountability processes force advertisers to balance the expected payoff of algorithmic targeting against the potential cost of negative publicity.

An additional perceived benefit of dark ads is the ability to engage in ongoing forms of so-called A/B testing to attempt to approve ad effectiveness and response (Pasquale, 2018). The digital medium makes it easy to vary ad details across different groups, to see which changes result in higher click-through rates. This information can be coupled with research by the platform – such as Facebook’s notorious claim to advertisers that it, “has the capacity to identify when teenagers feel ‘insecure’, ‘worthless’ and ‘need a confidence boost’” (Levin 2017). Such information could facilitate the development of ad strategies targeted to reach young people at moments of heightened vulnerability. If these strategies were publicly visible, they might trigger a public backlash, but if they are limited to a target audience, they can evade public scrutiny and embolden advertisers to develop increasingly manipulative ads based on invasive forms of personal monitoring. It is worth noting that there are potential pro-social uses of dark ads, such as, for example, engaging in a preventative measure to target individuals who may engage in illegal activity (see for example forthcoming research by Henry [2019] on providing potential paedophiles with options about seeking help and avoiding criminal activity).

However, our research indicates that the vast majority of dark ads service commerce and politics, raising issues about public accountability and transparency. The following research takes as a starting point the need to develop a system for addressing the epochal shift associated with the transformation of advertising from a shared set of reference points to an increasingly individualised and opaque form of automatically curated culture.

The Changing Advertising Economy

In 2017, digital ad revenues surpassed television for the first time (Slefo, 2018). Last year, Facebook reported \$55 billion in ad revenue while Google reported \$116.3 billion (Facebook, 2019; Wodinsky, 2019). Together, the digital duopoly accounted for 60.1% of the total digital ad spending in the US in 2018 and was forecast to continue reigning over the ad world for the foreseeable future (Spangler, 2018; Wodinsky, 2019). This shift is arguably disruptive partly because digital entities like Facebook and Google, unlike their traditional counterparts, do not produce content themselves, although there are signs that things are changing. Rather, these platforms rely on their participatory, interactive structure, to solicit content—above all, via contribution from ordinary users (that is, UGC) and curation. During this process, platforms trace people’s web use, extract user data, and ultimately analyze and use that data to sell ads in a highly customised way and meanwhile to improve the architectural traits of the websites so as to encourage more user participation and contribution (Wu & Taneja, 2019). The appeal to advertisers lies in the ability not just to target consumers, but to experiment on them (by constantly changing ads to see how this affects response rates), and to monitor ad effectiveness on an individual level – something that is much harder to do in non-interactive contexts. Moreover, changes in media consumption habits also push in the direction of online access: Australians are spending more time using interactive devices than they are legacy media. The ABC has reported that, on average, Australians spend almost seven hours a day using some combination of tablets, phones and computers – a number that outstrips average TV viewing hours (under three a day – see Tilley 2018 and Nielsen 2018). As people spend more time using digital media, digital advertising comes to play an increasingly important role in the media ecosystem. All these terms point to “new advertising and data-processing developments” (Turow & Coudry, 2018: 415) that come under the umbrella of “platform” logics (Gillespie, 2017). Ad networks such as Google AdSense and Facebook Audience Network take on an increasingly important role as an intermediary layer of service providers to broker online marketing exchanges between numerous publishers/websites and advertisers.

Perhaps the defining characteristic of this shift in the online advertising ecosystem is toward automation, which results in what might be described as an increasingly value-focused and, simultaneously, value-blind system. In the former sense of the term, “value” refers to the ability to maximise the “return-on-investment” (ROI) sought by advertisers, thanks to the ability to experiment and monitor responses. Variations can be introduced automatically (colour scheme, layout, images, and even text), and then tested in real time to measure their impact. Such systems are also value-blind in the sense that they are driven by programmed imperatives that can be oblivious to broader social issues – such as discrimination or stereotyping. If, for example, the automated system “learns” that a particular ad is more likely to get the desired response from people of a particular ethnicity or gender, it can, autonomously, begin to discriminate in the ad delivery process. If particular images offend some viewers but appeal to others, the system can be adjusted to ensure that only those who are unlikely to be offended receive the images. This can certainly be beneficial in some contexts – to ensure, for example, that younger viewers aren’t exposed to ads with adult content. But, as the researchers at Algorithm Watch have demonstrated, it can also result in discrimination and stereotyping. The researchers placed ads on Google and Facebook for a range of different job opportunities, without stipulating any form of targeting, but found that the platforms’ automated system reproduced forms of social stereotyping. For example, ads

for truck driving positions were shown to almost 5,000 men but only 386 women, whereas ads for child care workers, running simultaneously, were shown to more than 6,400 women but only 258 men. This raises the important question regarding the role played by automated profit-maximising systems in the reproduction of historical forms of discrimination. The system does not “know” when it is engaging in regressive, racist or sexist activity – it is driven blindly by the goal of maximising clicks and responses. It may be driven by the patterns of activity related to past ads that were specifically targeted toward particular groups. This is one way in which social stereotyping can reproduce itself. By the same token, we, as a society, do not know the extent to which automated discrimination takes place because we cannot see into the system. This is the world we have created for ourselves: a customised digital virtual reality in which everyone gets their own “secret” messages invisible to others. Perhaps it is not surprising that other aspects of our information world are also becoming increasingly fragmented – advertising does not exist entirely separately from the news and entertainment it supports. Fragmentation cuts across our different messaging systems and becomes embedded even in the “background” messaging of the advertising world. We contribute to the creation of this world through submission to the comprehensive forms of data capture that support our online platforms: the more time we spend on our phones and laptops, the more we use these platforms, the more data we send to the advertising oligopolies that control this customised world in ways that remain opaque: we each see our own tiny shard of messaging, but the overall pattern within it remains inaccessible and illegible to us. If it results in new forms of discrimination, we are unlikely to find out and those who have access to the means of discerning it have little motivation to do so.

If we understand advertising to be more than simply an information signalling system (that is, a way of designating which products are available when and where at what price), but also a form of social messaging that enacts and reproduces social values, the shift in the way advertising operates takes on significant social consequences. Advertisers have always played a role in selectively reinforcing and amplifying social values in accordance with a narrow range of profit-maximising imperatives, but they have done so, by necessity if not by choice, under conditions of public transparency. In recognition of this fact, we have imposed social constraints on advertisers, limiting, for example the ability to target appeals for alcohol and cigarettes to young people, or the ability of pharmaceutical companies to make false or misleading claims about their products. Beyond legal regulation, advertising also needs to conform to generally acceptable social norms, or face the risk of brand-damaging negative publicity and boycotts. We witness these social controls at work – both progressively and regressively – when advertisers push the boundaries of what is considered to be socially acceptable at any given point in time. One of the promises of granular targeting is to be able to bypass “common denominator” social norms to target those who might be comfortable with messages that might, if publicised, result in public pushback. It also makes it possible for false or misleading messaging to be targeted to those unlikely to report it because it confirms their hopes, suspicions or prejudices. Additionally, it enables advertising practices that exacerbate forms of political polarization and social pathologies.

In the wake of the violent attack on the US Capitol building during the certification of the winning vote for President Joe Biden, a watchdog organization revealed that Facebook was serving targeted ads for military gear alongside posts about the planned insurrection (Mac and Silverman, 2021). It seems unlikely that this pairing was deliberate – most likely it was the result of an automated curation system that matched ads to user interests and topics. However,

the only reason the pairing was discovered was because it appeared on an account owned by a media watchdog organization that had created a Facebook account to monitor right wing violence online. It is unlikely that actual participants on online discussions would have complained about such an ad. The notion of ad “relevance” takes on a new and somewhat sinister meaning in such contexts: that those who receive the ad are precisely those who will not blow the whistle. The same is true of other forms of anti-social targeted advertising: forms of racist, sexist, and homophobic stereotyping can be targeted specifically to those who are unlikely to draw attention to them.

Current Measures to Address ‘Dark Ads’

The forms of targeting and segmentation enabled by the datafied intermediation of advertising is taking place at unprecedented “speed, scale, and intensity” (Turow & Couldry, 2018). Researchers have demonstrated the ability of advertisers to use platforms like Facebook to discriminate against protected categories of consumers, including age, race, and gender (Angwin et al. 2019) and the tendency of Facebook’s automated ad optimization to discriminate (Ali, et al, 2019). Empirically, there is also a nascent literature suggesting

data-based social discrimination in online ad delivery. For example, Sweeney (2013) shows that a greater percentage of background check service ads appeared for black-identifying first names than for white-identifying first names in searches on Google and Reuters which hosts Google AdSense ads. Similarly, Datta et al. (2015)’s study of Google Ad Settings found that setting the gender to female resulted in getting fewer ads related to high paying jobs than setting it to male. Important as this research is in showing the empirical patterns consistent with social discrimination, many of these initiatives are based on hypothetical scenarios in ad targeting and discrimination. Such approaches are “audit studies” insofar as they take the shape of, “field experiments in which researchers or their confederates participate in a social process that they suspect to be corrupt in order to diagnose harmful discrimination” (Sandvig, Kevin Hamilton, Karahalios, & Langbort, 2014: 5), often in fictitious scenarios. There has been no systematic research however from the consumer’s perspective: no one has compiled the evidence to demonstrate how targeting works in practice by showing how particular individuals receive very different ads based on demographic and psychological characteristics.

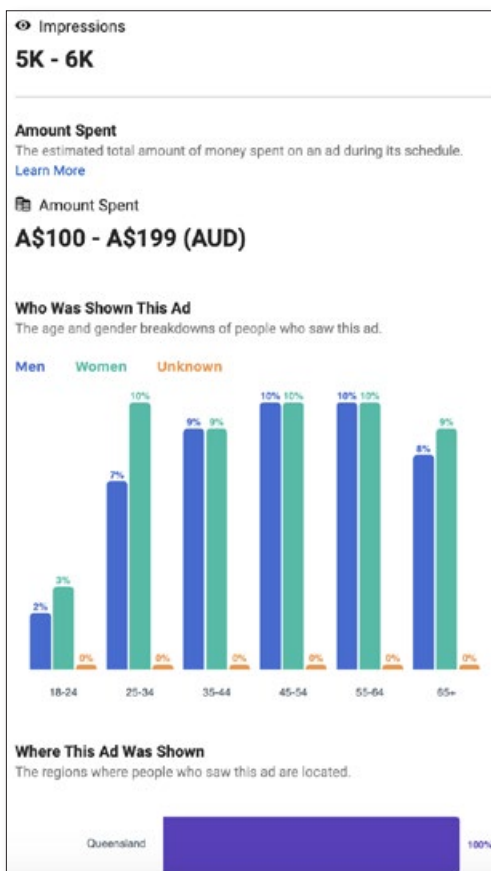


Figure 5: Facebook ad analytics for a “right to life” ad circulated during the 2020 Queensland elections.

The big challenge, of course, is to try to reverse engineer forms of opaque targeting that take place on a mass scale and are stubbornly difficult to study. Indeed, if we conceive programmatic advertising as a whole, it is rendered invisible and opaque to consumers and even industry professionals. Insufficient transparency has always been an issue plaguing the industry. Advertisers do not know where their ads are placed and who really see their ads, and publishers do not know who exactly buys their ad space or inventories. Our purpose here is not to probe into this blackbox as researchers in computer science and machine learning have done. Rather, we tackle the issue of opacity from consumers' perspectives, focusing on their interactions with targeting systems.

In the wake of public concern about targeted political messaging, Facebook made a symbolic attempt to create a political ad library in some countries, including the United States, but researchers have noted that this approach has been plagued with problems. As *The New York Times* reported in 2019, "The social network's new ad library is so flawed, researchers say, that it is effectively useless as a way to track political messaging" (Rosenberg, 2019). The library has undergone some functional improvements since then, but it still provides very little insight into the overall patterns of targeting and customization. For example, the ad library for Australia breaks down the reach of political ads by gender and location, and does not provide any information for commercial (non-political or non-issue-oriented) ads (see Fig. 5 and Fig. 6).

The image shows a screenshot of a Facebook advertisement for 'The Killer Coffee Co.' in Australia. The ad is titled 'About the Ad' and features a video thumbnail with the text 'SLAY YOUR DAY.' and a coffee bag with a skull logo. The ad text reads: 'It's always a good time for Killer Coffee. Rich, Smooth, Strong yet Bitter-Free. Are you game?' and 'Free Shipping across Australia & NZ'. Below the video, it says 'KILLERCOFFEECO.COM.AU' and 'Australia's Favourite Morning Wake-Up'. To the right, the 'About the Page' section shows the company's profile: '@killercoffeeco', 64,511 likes, Retail Company, and 20,358 followers. The 'More info' section describes them as 'Renegade roasters from Sydney crafting Industrial Strength coffee. Rich, Smooth, Strong yet Bitter-Free. #KillerCoffeeCo'.

Figure 6: Facebook does not supply any targeting details in the library for commercial ads in Australia.

The fact that no details are shown for non-issue-oriented ads means that it is difficult to trace whether commercial advertisers are discriminating by race, gender, and other protected categories. It also makes legal but potentially objectionable forms of targeting possible: for example, the ability to advertise different prices and rates to different individuals based on their perceived ability or willingness to pay.

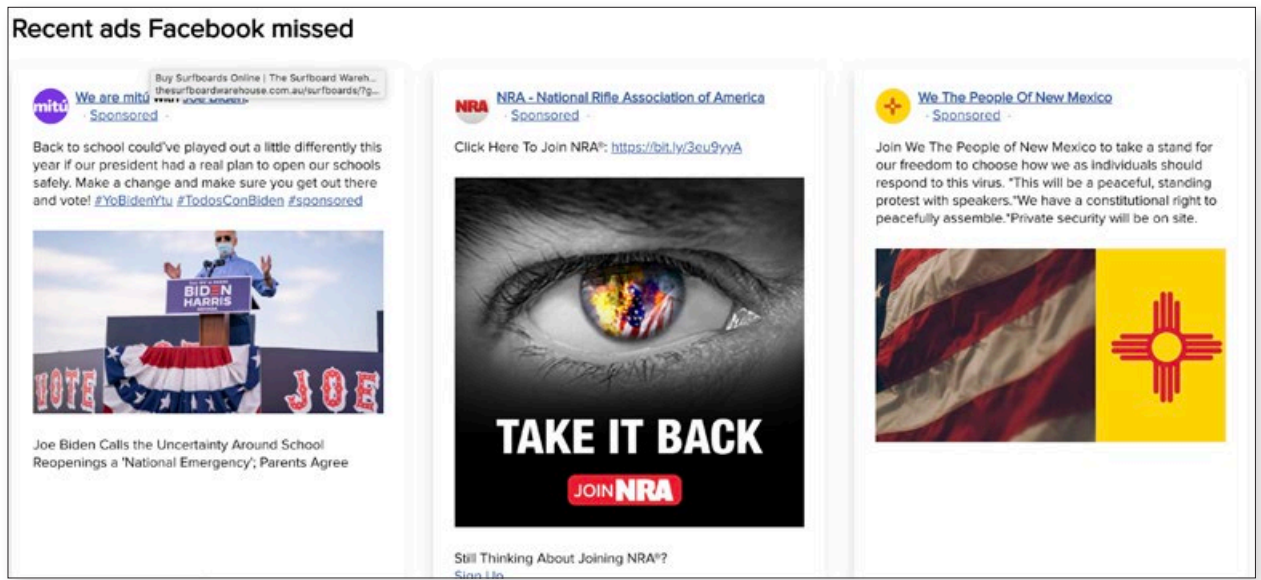


Figure 7: Political ads missed by Facebook but captured by NYU's Ad Observatory

The non-functionality of the Facebook ad library is unlikely to be an accident, given the demand on the part of advertisers engaging in strategies that rely on “dark ads.” From the perspective of political campaigns and a range of other advertisers, the ability to reach particular individuals or groups without being subject to public accountability is a feature, not a flaw, of the Facebook advertising proposition. As one public interest advocate put it, ‘you can go as narrow as you want, as false as you want and there is no accountability.’ (Angwin & Larson 2017)

To address the shortcomings of Facebook’s ad library, researchers at New York University created the Facebook Ad Observatory, which recruits volunteers to install a browser extension that shares the ads served to them on Facebook with researchers. The project is part of a larger ‘Online Transparency Project,’ which is devoted to improving the transparency of online political advertising. The Observatory researchers discovered that Facebook’s system sometimes misses political ads, which never make it into its ad library (see Fig. 7). It also pulls together data to monitor online ad spending, and overall patterns in the ads it is able to collect. The Observatory does important work in the political realm – but it does not address the broader issue of messaging beyond political advertising, and of the patterns of demographic targeting, because it does not collect demographic information about those who share the ads they see.

Perhaps not surprisingly, Facebook has issued a cease and desist letter to the Ad Observatory, claiming it is violating the platform's terms of service by 'scraping' ads en masse. The researchers have refused to back down, arguing individual users choose to share the ads with them – which is the core function of Facebook. The reluctance of Facebook to allow public scrutiny highlights a fundamental tension in the attempt to address the potential pathologies of ad targeting: if advertisers benefit from opaque forms of targeting, providing transparency can undermine the value of the service that Facebook and other platforms are selling. That is, if non-transparency is the point, calls for transparency threaten the online business model that supports many of the “free” online services we have come to rely upon. However, consumers have a right to know how the data they provide is being put to work online and how it shapes not just the news and entertainment they see, but also the advertising and commercial messaging they rely upon for both consumption and for functions like searching for jobs, housing, and educational opportunities.

We can also question whether Facebook or its staff properly understands its own advertising context; so much of its moderation of content of all forms has been outsourced to private companies with little in the way of clear chains of responsibility. Whether Facebook is interested in knowing more about its own content, or if it would rather remain agnostic of its own role in civic life is unclear. Across the online ad industry in general, the knowledge gap in this area is extensive with reports suggesting that other equivalent ad-based digital platforms, such as Google, do not have a strong awareness of the ads that are being served on their own platforms (Taylor, 2020); whether this applies to Facebook as well remains unclear, as Facebook has only limited cooperation with researchers and journalists.

Methodology

Research Design and Participant Recruitment

Our research design made use of three research methods: a digital ad collection process, a series of interviews with research participants who had provided us with ads, and a set of focus groups in the form of workshops with individuals who browsed and reflected on our dashboard. Each phase of the research informed decisions that we made about the subsequent phases, while also being important in influencing the design of our app and website dashboards.

The research was built on a method developed by the independent investigative news outlet *ProPublica*, which created a tool to capture political ads from Facebook in the aftermath of publicised news accounts of false and misleading political ads during the 2016 presidential election in the US (Larson, Angwin, and Valentino-DeVries 2017). The *ProPublica* tool is a browser extension that volunteers can install on their Chrome or Firefox Web browsers. The extension allows users to share the ads that appear in their feeds with *ProPublica*, automatically sending them to a database where a classifier targets the political ads for inspection. The goal is to gain some insight into messaging that might otherwise be seen only by those to whom it is targeted. However, the original *ProPublica* tool does not provide any user details, which makes it difficult to determine how ads are being targeted. Moreover, *ProPublica* drew upon its progressive readership to recruit volunteers, which means it only collected those ads served to a particular political demographic: they were unlikely to capture, for example, ads targeted toward the extreme right.

We adapted the *ProPublica* tool by adding the capacity to collect some voluntarily provided demographic information and recruited a representative sample of paid participants to reflect the diversity of income, age, gender, and geography of Australia. This approach addressed two of the main issues associated with the *ProPublica* extension: the self-selection bias in their voluntary recruitment, and the inability to discern patterns of demographic targeting.

The cost of recruiting limited the sample size to 136 people. Participants were drawn from across Queensland, New South Wales, Victoria, South Australia, and Western Australia; within each state, our participants were drawn from different postcodes, including both urban and regional areas. Participants were asked to install our data-gathering tool, gathered ads for a prescribed period, and then were interviewed using a variation of the scrollbar qualitative research method to enable us to focus on the ads themselves. This was conducted in two phases, using an initial pilot sample of 10 people to explore how users were reflecting on their ads with the tool in April 2019, then with a further sample of 28 participants in May

2019. All participants who chose to complete the data-gathering and interview phases were compensated for their time.

Participants tended to reflect on the patterns of online activity that they felt played a role in shaping which ads they saw. These responses demonstrated a familiarity with the process of online ad targeting, with users, for example, being prompted to recall the specific set of travel searches that prompted a particular ad for a resort or a cruise trip. In many cases, the ad targeting pattern only emerged retrospectively. One woman, for example, had not noticed how many ads she had received for special education programs until she went over her ad feed. She then realised that the algorithm had been responding to searches she made for resources for her child on the autism spectrum. Our conversations prompted respondents to consider the level of detail of information about them that was used to target ads. These experiences were fed into the development of our focus groups to prompt new participants on their experiences and expectations. These experiences also aided us in the development of our workshops, as we designed these to respond to participants' concerns and anxieties about online tracking. Given the limitations of the recruiting process, we present the current project as a demonstration version of ad accountability: a "proof of concept" model for providing insight into patterns of ad targeting. We also believe it can serve as a useful tool for raising awareness about non-transparent changes in the advertising environment. For the first time in the history of the mass media, it is possible for dramatic changes in the form and content of advertising to take place without a shared recognition of the character and extent of these changes. Increasingly, we no longer experience advertising messaging as a shared experience, and the default assumption that we are all being treated the same, more or less, is likely to become less and less true.

The Ad Collection Tool

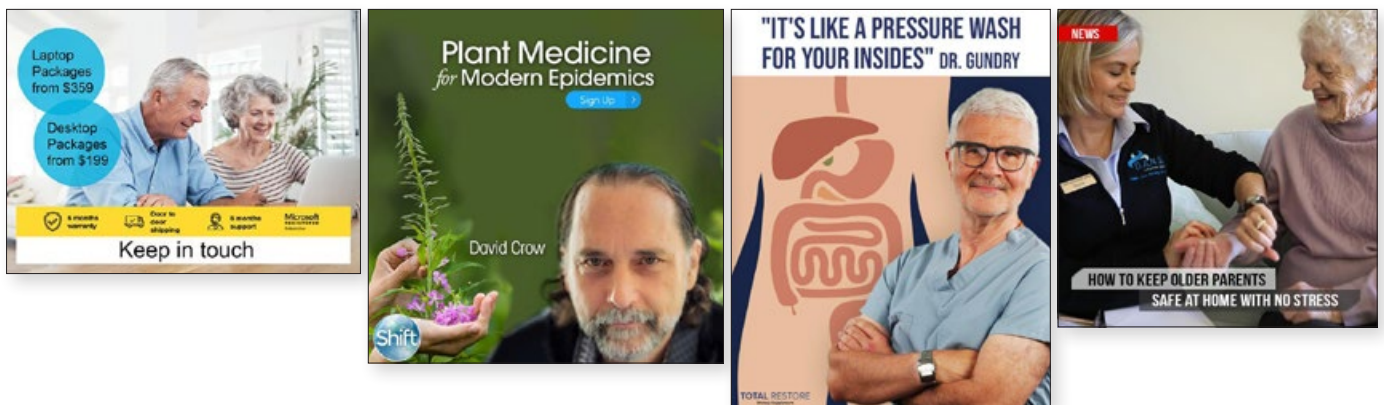
The ad collection process was designed to allow users to automatically send adverts to us that they encountered as they browsed Facebook on a desktop machine.

We employed a software designer to redevelop a version of the *ProPublica* Facebook data-gathering tool that would be deployable to our research context (see *ProPublica* (2017) for precise technical specifications). The tool was developed as a plugin for the Chrome browser on desktop computers, and would solely gather ads that appeared in the feed of the default main Facebook page. This allowed participants to have some control over our data gathering process insofar as they could elect to uninstall the tool or use alternative browsers if they wanted to control their degree of participation in the research project. The tool relies on the fact that sponsored content must be indicated as such on Facebook – this means that it can search the HTML code to determine which promotional content has been labeled as such. It is worth noting that Facebook does not encourage the use of the *ProPublica* tool and has repeatedly changed the way it codes the "sponsored" tag in order to hide it from automated ad collection. We had to update the tool several times over the course of the year in order to adjust changes by Facebook designed to avoid holding the company accountable for its ad targeting practices. The tool is not a mass data-scraping technology, rather it provides a way for individual users to decide to share information they receive online with researchers. In this respect it replicates one of the main functions of Facebook.

Automated Data Analysis

One of the defining challenges of providing accountability for online advertising is the ability to discern targeting patterns at scale. Companies like Facebook and Google have this data, but they are not required to share it publicly. Until this happens – and it should – we are left with only rare glimpses of how these giant corporations are shaping our media environment. Because of the possibility for illegal discrimination and the threat to public and democratic culture, these companies should be subject to advertising audits by independent organizations that have full access to their data on ad delivery and targeting. Any attempt to ‘reverse engineer’ targeting at scale confronts the issues raised by the sheer number of ads that flood online platforms. In the case of our pilot project, even a relatively short ad collection period for a small group yielded more than 10,000 ads – a number that is already unwieldy for analysing without some form of automated information processing. We anticipate then, that tools such as ours would require the use of automated systems for sorting and pattern detection. Text analysis tools, for example, might be able to reveal the types of appeals that are directed toward particular categories of consumers and whether these reinforce existing stereotypes or, indeed, create new ones. They might also help identify what types of products and services are being targeted to particular individuals. We are in the process of developing collaborations to automate the processing of the ads we have collected and have worked closely with Associate Professor Dan Angus at the Queensland University of Technology and his research assistant Jane Tan to use an automated image classification system to demonstrate one way of sorting the ads by using visual cues. The QUT image classification system inductively recognizes shared patterns across collections of images to create cluster of similar images. The level of similarity can be dialled up and down to create smaller and larger clusters of greater and lesser similarity. Taken to the limit, for example, the highest level of similarity would yield a cluster of one; whereas the widest tolerance for differences in the images would include the entire collection of photos.

Figure 8: Images from one of the similarity clusters detected by the automated image classifier.



We shared the data from our pilot project with the group at QUT to get a sense of how such a tool might be used to analyse data at scale for a more widespread implementation. It is important to note that, given the small sample size for the pilot project, we do not believe these results to be generalisable to Facebook users in Australia more generally. The goal is to demonstrate how such a tool might be used in conjunction with an automated image classification system.

The first step in the analysis was to create a database that included all of the images collected by the Dark Ad Tool. These images were limited to that seen by the Facebook users in their news feeds (that is, they did not include further images that the users might see if they were to click on the ad). We then used the ad classifier to identify clusters of images that were similar to one another but not identical. For a sense of the type of images clustered together by the classifier, see Fig. 8.

The classifier identified 66 clusters in our data set, although the size of these varied from just a few images to scores of them. We left in repeated images in order to get a sense of the overall volume of ads served and to address the fact that the same ad may have been served to different users. Once the tool created the clusters, we could get a demographic breakdown of each cluster based on the information provided by our participants. This breakdown allowed us to visually detect the demographic skew of ad cluster. The descriptions of the particular clusters, such as “sleek car” (Fig. 10) were done manually: we inspected the cluster to get a sense of what elements the photos it contained had in common with one another. We could see from each cluster, thanks to the data visualisation, which demographic characteristics were associated with it. Because the most robust demographic category in our data set was sex (in many other categories we had to small a range of participants for a meaningful analysis), we focus on this variable in the following discussion.

At this point the analysis takes on a qualitative component. The image classifier allows us to ask the question regarding how a particular image cluster might line up with the associations attached to particular demographic profiles. We note that this is a matter of interpretation, but that this is an important aspect of assessing how associations and stereotypes might be reproduced by targeted advertising. For example, the “sleek” looking car ads cluster illustrated in Figure 10 has a decidedly masculine skew in terms of those participants who were targeted by this cluster. By contrast, the “Sleeping Related” cluster had a very strong female skew (see Fig. 9).

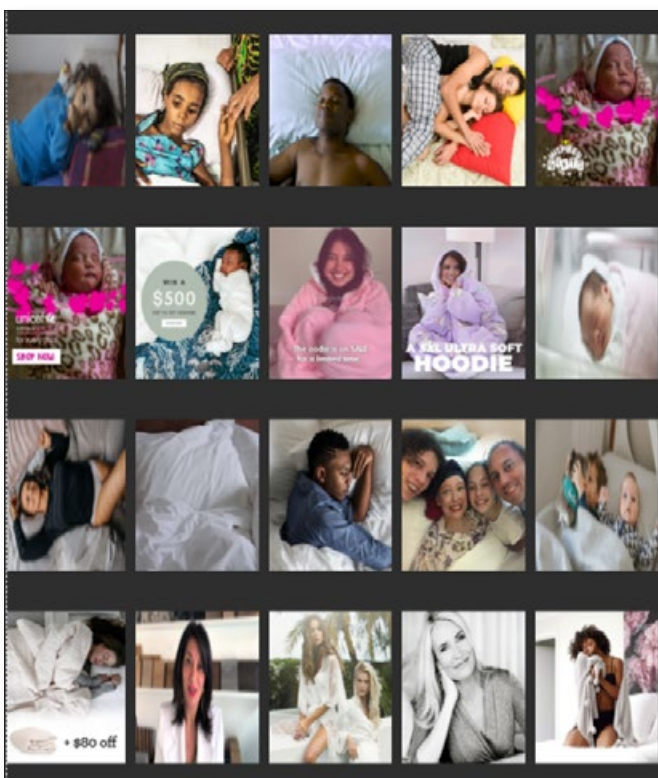


Figure 9: Sleeping Related Ad Cluster

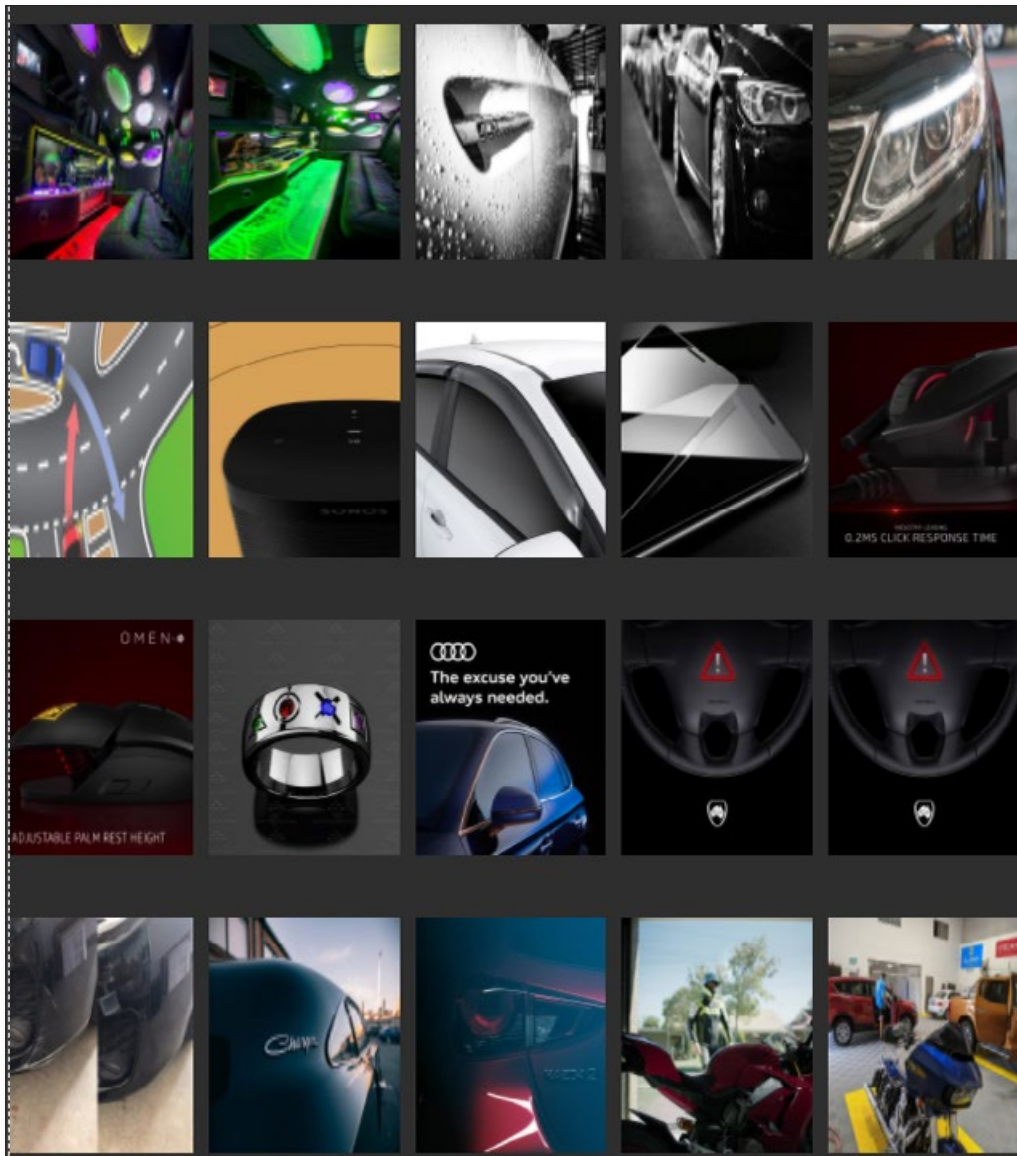


Figure 10: A screenshot of some of the images in the “Sleek Cars” cluster: this provides a sense of the commonality of the images grouped together by the image classifier.

It is not difficult to see some of coded differences between these two sets of ads. The car ads focused on abstracted car parts, feature dark colours, fetishise technology, and do not include images of people or sociality. By contrast, the sleeping cluster features softer colours, interiors, images of family life and domesticity. We might compare this with the “Dining” cluster (Fig. 11), which also features scenes of domestic sociality and was seen primarily by participants who identified as female.

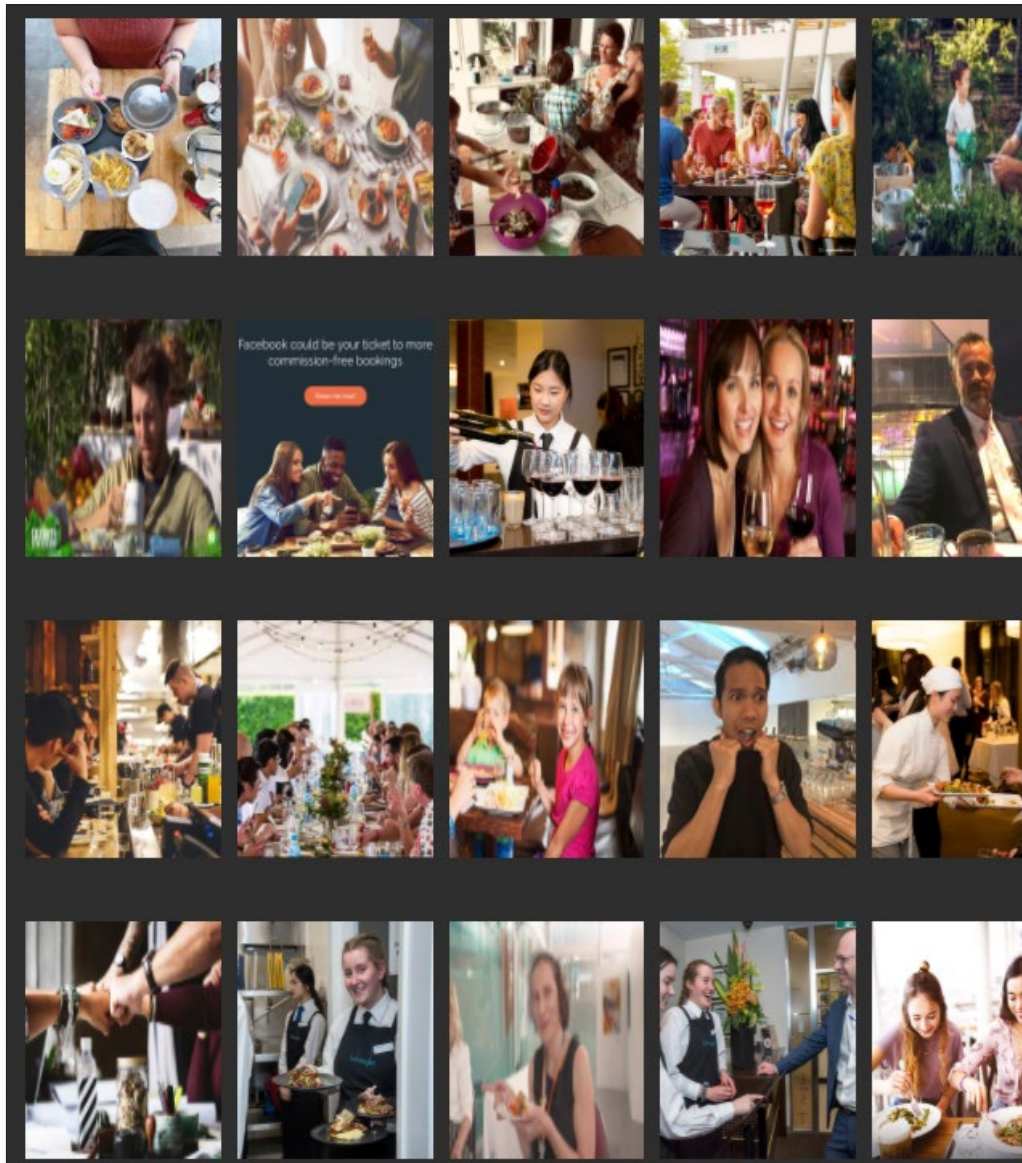


Figure 11: Dining Cluster

The screen shots provide only a random selection of the images included in each cluster, but it is possible to access all of the images to gain a sense of the overall character of the cluster. It is also possible to isolate particular images and see the demographic characteristics of those who received them. This sample of images provides an example of how the data collected by the Dark Ad Tool might reveal how design elements and their social associations are distributed demographically. We focus on automated image analysis because of the role these might play in raising awareness about ad targeting, and in advancing the discussion about forms of stereotyping reinforced by commercial messaging strategies. A range of tools can be enlisted for making sense of the data collected by the tool, and we will continue to invite collaborations as we find ways to recruit more participants to contribute to our ad database.

Application Method

When signing up to the tool, participants complete a short demographic questionnaire that is used during data capture to link ads to demographic characteristics. Users are also assigned a unique key that links their particular experience of Facebook to a dedicated marker. This key allows them to use the project Web site to see their individual ad streams (which are available uniquely to them).

Once the tool is installed, linked, and activated, the tool reads page information associated with the facebook.com domain when users are logged in. The core function of the tool is to allow the participant to automatically forward on ads to us that Facebook sends to them. For each ad, the collected data includes the associated image, image ‘carousels’ (rotating sets of ad images) or, in the case of adverts with videos, the first frame of the video. We also collect the name of the account that produced the ad, as well as the associated copy-written portion of the ad and the URLs for the links incorporated into the ad (see Fig. 12).

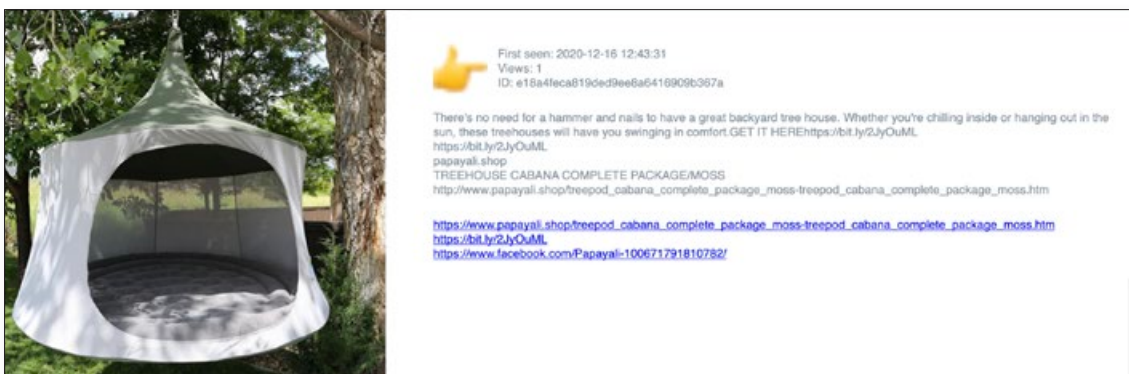


Figure 12: A sample ad collected by Monash’s Facebook ad collecting tool. The tool collects the ad image as well as the ad copy and the embedded links. These links can be followed to see what the user would encounter if the ad is clicked on.

The ads are displayed in the database in the order they appear, with the most recent at the top of the list of collected ads (see Fig. 13).

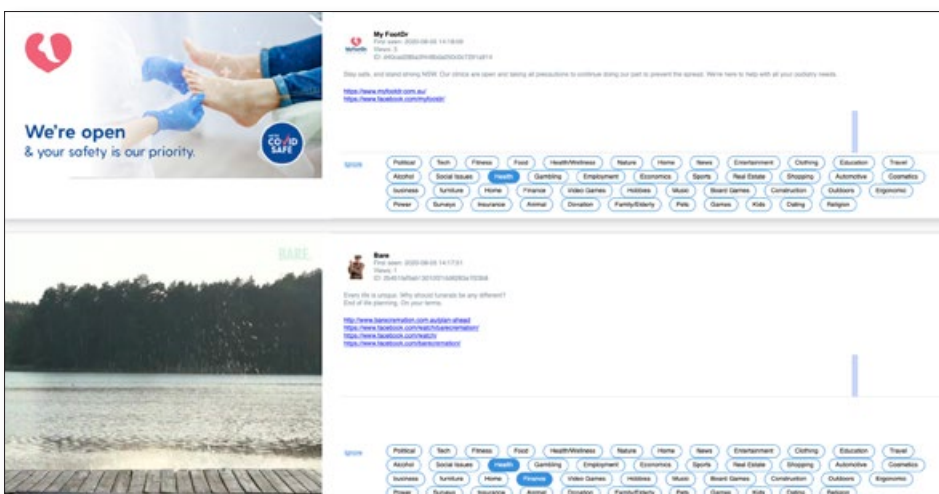


Figure 13: A screenshot of the ad display page for the Facebook ad collector. The ads appear in the sequence they were collected (most recent first). The tags at the side are added by the research tool, so that we can classify the ad according to content. The tags are editable and new tags can be added as different categories of ads appear. Ads can be tagged in multiple categories where appropriate. It would be possible to develop an ad classifier using machine learning, but we tabulated the ads manually with the assistance of a team of undergraduate research assistants.

The tags next to the ads allow us to code the ads by type – so that we can start to match the type of ad shown with the demographic background of the viewer (see Fig. 13).

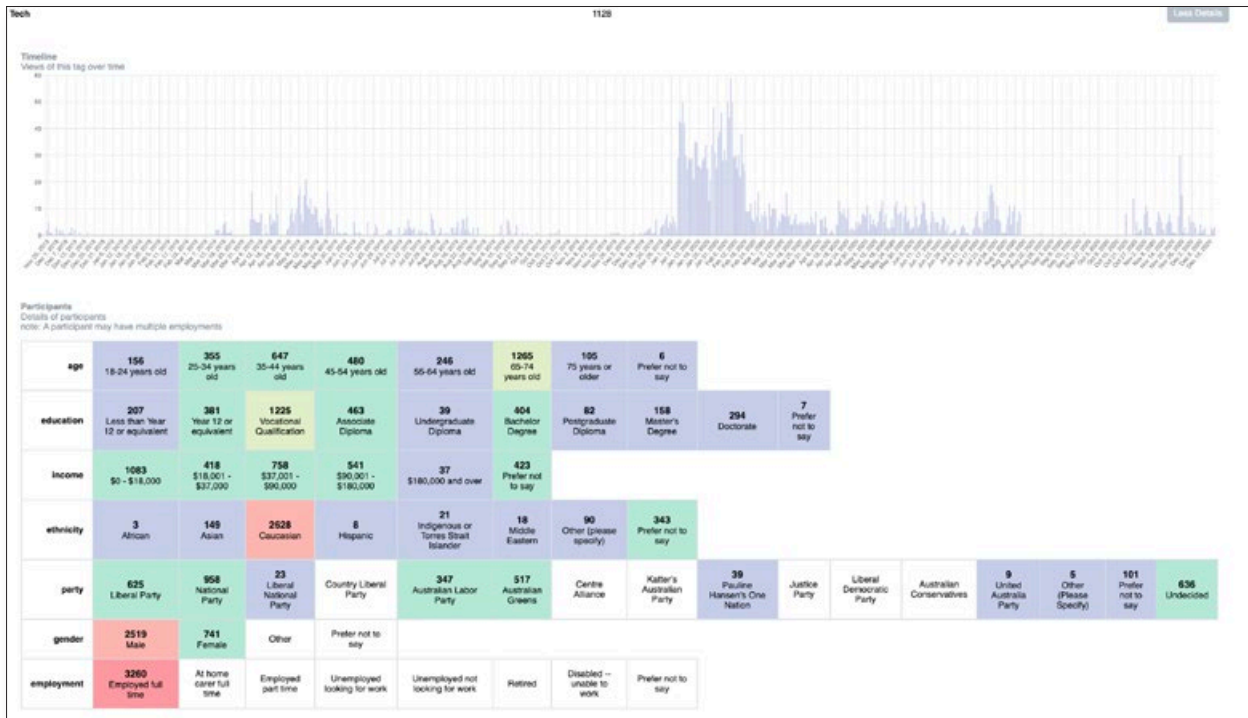


Figure 14: The ad tagging tool reveals a breakdown of who received which ads. For example, in this case a demographic analysis reveals that, in our pilot study, tech ads were overwhelmingly received by men.

One of the defining features of this iteration of the tool is the ability to filter the ads by demographic category. One or more filters can be chosen for ad sorting in order to obtain the results from increasingly specified groups of users. For example, a gender filter can be applied to see the differences between the ads received by women and those received by men (see Fig. 14).

The timeline feature allows a breakdown of when ads were seen, while the demographic table visualises ad views based on demographic features. This second table calculates ad views and not the number of participants or users who viewed an ad. The two examples below are for ads that have been tagged as 'alcohol' related.

What do these visualisations tell us? The timeline feature enables us to see when ads were viewed, which offers the potential to understand the advertising patterns over time, how advertising might be seasonal, and whether particular types of ads arise around key events (such as an election). However, the example above is simply a prototype due to our sampling size and makes clear when we tested our data collection tool – the peak of our data collection occurred over the summer of 2019-2020. With more consistent and longitudinal data collection, we have the potential to see temporal patterns in targeted data. For example, theoretically we could see at which times of the year ads related to alcohol are more frequently seen by users. This could be useful to know as concerns were raised in early 2020 about the frequency and volume of alcohol advertisements on social media during the pandemic.

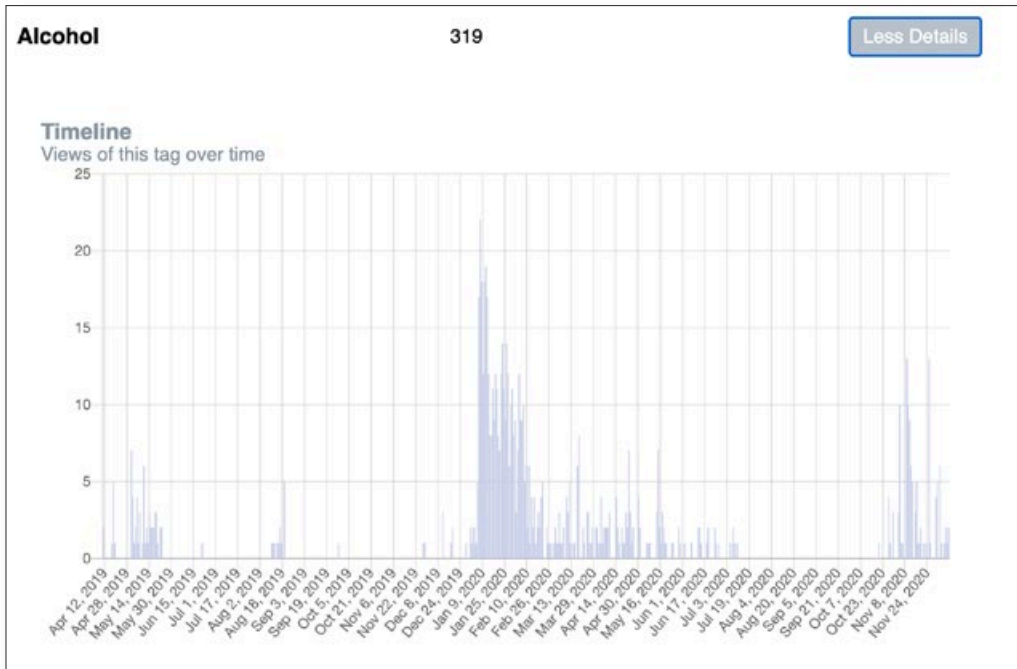


Figure 15

Participants
Details of participants
note: A participant may have multiple employments

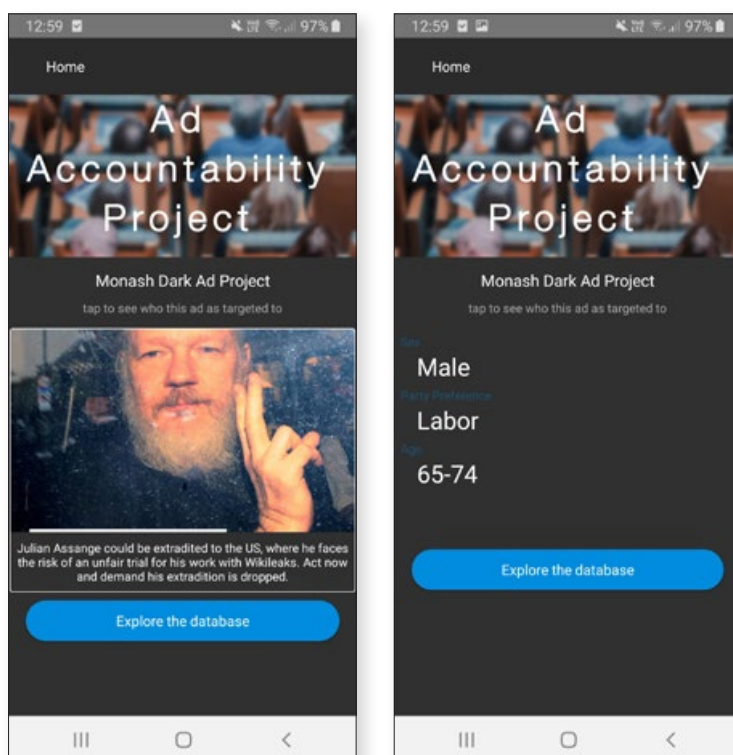
age	20 18-24 years old	100 25-34 years old	248 35-44 years old	135 45-54 years old	50 55-64 years old	227 65-74 years old
education	105 Less than Year 12 or equivalent	116 Year 12 or equivalent	243 Vocational Qualification	76 Associate Diploma	1 Undergraduate Diploma	80 Bachelor Degree
income	147 \$0 - \$18,000	147 \$18,001 - \$37,000	264 \$37,001 - \$90,000	145 \$90,001 - \$180,000	6 \$180,000 and over	75 Prefer not to say
ethnicity	1 African	15 Asian	670 Caucasian	2 Hispanic	24 Indigenous or Torres Strait Islander	1 Middle Eastern
party	139 Liberal Party	130 National Party	15 Liberal National Party	1 Country Liberal Party	55 Australian Labor Party	213 Australian Greens
gender	485 Male	299 Female	1 Other	1 Prefer not to say	0	0
employment	784 Employed full time	1 At home carer full time	1 Employed part time	1 Unemployed looking for work	1 Unemployed not looking for work	1 Retired

Figure 16

The second visualisation indicates the frequency of views by user demographic characteristics with the ‘hotter’ colours (e.g. dark red, orange) indicating a higher rate of views with ‘cool’ colours (blue, and dark green) illustrating lower view rates. This visualisation is interesting in that it shows views or impressions as opposed to the quantity of ads or user count. As a result, it highlights not just how many ads are in circulation but also the frequency of how often those ads are seen. This is an important dimension to consider in advertising, when common marketing tactics involve repetition (even of the same ad and delivered to the same user). Such repetition was also reflected in user experiences with several participants complaining about recurring ads, which often felt like spam. One participant in our focus groups voiced concerns over receiving the same type of ads ‘thousands’ of times and how it might work to create a ‘filter bubble’ and ‘limit’ one’s mind.

In terms of the above example, we can see that a spread of demographics viewed alcohol related ads although there was a skew to male users (485 views) compared to female (299). This does not give us information about how many unique users viewed the ads and, in fact, this could be a small number of users who frequently saw alcohol related ads. However, the numbers prevalent within the other demographic categories reveal a fairly diverse range of users who saw alcohol ads, suggesting that in fact it was not just one or two users who saw many alcohol ads.

We also created a mobile app (illustrated below) for the project, allowing users to explore the database by using demographic filters. The home page of the mobile app begins with an example of an ad from the database and prompts users to click on the ad to reveal what persona or type of user saw the ad. In the below example, the ad about Julian Assange which asks the audience to ‘act now’ and join the cause in demanding for Assange’s expedition was seen by a male user aged between 65-74 and who voted for Labor.



Users then may choose to explore the database in which they are offered the demographic filters to help them search through and make sense of the ads presented. This operates in the same way as the public-facing website.

In addition to the mobile app, we also have a website that has both a public and private mode of accessing and exploring the dataset. In the public facing option, users are able to explore the entire database of ads and can filter using the demographic variables we have collected to see the kinds of ads that appear for particular types of users (this will be explored further in the next section). The private version of our website, accessible only by us, has some additional features that enable us to analyse and sort through the data.

Figure 17

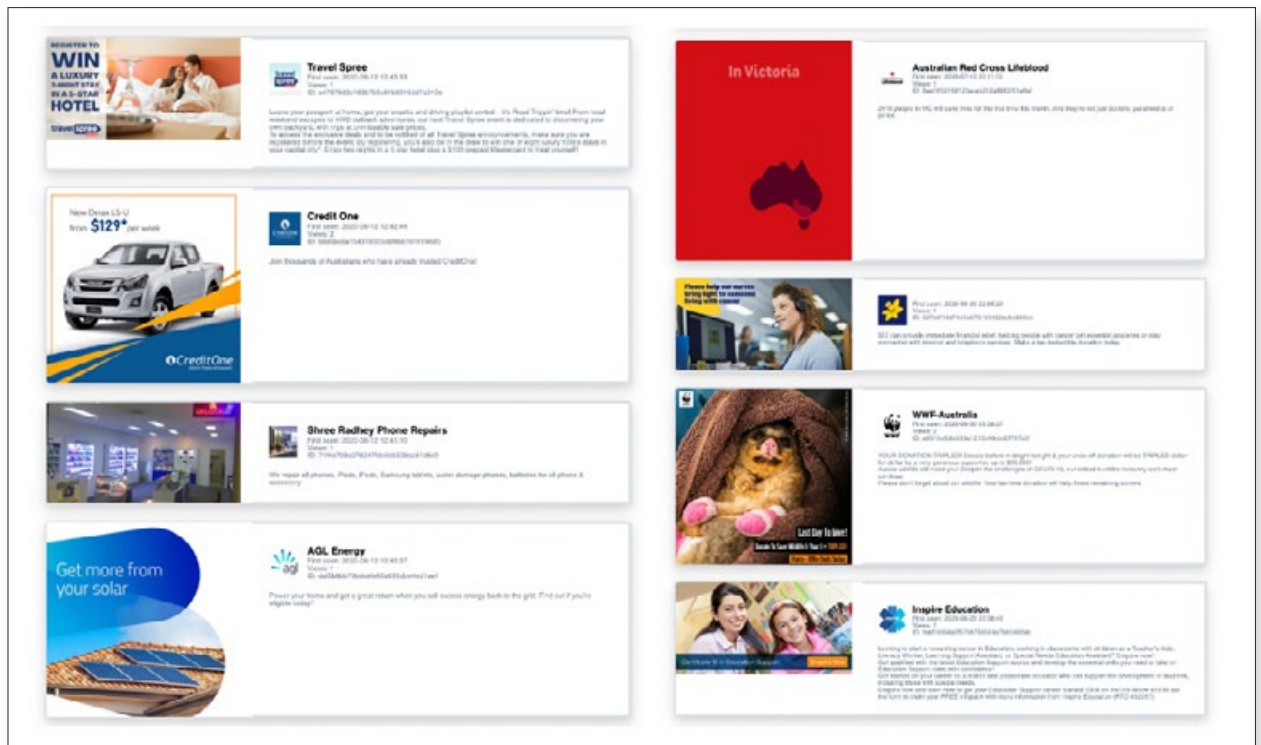


Figure 18: A short selection of ads varied by gender (self-designated ‘male’ on the left and ‘female’ on the right).

This functionality is supplemented by the tagging tool, which allows us to see the overall breakdown of who received which ads. However, for this tool to work, the ads need to be hand-tagged, which can be a costly and time-consuming process as the number of ads increases.

Limitations

The research tool demonstrates some of the challenges posed by the rapid growth of non-transparent ad systems. Whereas Facebook can gain insight into who receives which ads and how they respond, reverse engineering this process at scale would be a hugely expensive and time-consuming endeavour. We cannot hope to approach the granular level detail that is regularly recorded by Facebook about individual users – but we can begin to show emerging patterns in the distribution of dark ads. Managing a process like this at scale can become increasingly expensive and time consuming.

Our research design decisions have limitations in terms of scope and privacy; also due to technical aspects regarding the way that Facebook identifies unique ads, we have gathered data that superficially appear as duplicates, but represent unique ad buys.

Privacy

Our tool does not collect any user generated information (comments, likes/emotes, shares, or any interactions or clicks on the post by our participants), but is limited to sharing the

sponsored ads they receive while browsing online. We do not collect any information about our participants through Facebook; the only information about our participants comes from the demographic survey that they complete. Our tool does not gather any information about any other posts, whether by our participants, their networks, the pages or groups that they view, or other information. For technical reasons, our tool will read these posts as they are “seen” through the newsfeed, but this is purely in the process of identifying whether a post is an ad or not, and occurs solely on the client side, with no personal information being sent by our tool during this review. Our tool only gathered ads in the main Facebook feed that had been generated through Facebook’s ad manager tool. Ads in the Facebook sidebar, ads that appeared within other video content, or from other platforms (such as embedded youtube links) were not captured. Ads that are informal (i.e. ads between friends or within social groups, or forms of native advertising) were not captured by our tool. Our tool also does not capture video ads that are built into video playback. As such we only captured materials that were a product of specific targeting processes used by Facebook.

Mobile

Mobile development contexts are highly controlled, and this has restricted our ability to develop a tool that will gather mobile ad experiences. Mobile is a significant context of ad consumption. Some adverts are targeted specifically at mobile users, and some ads are affected by the location-based or geositional services that mobiles use. Both the iOS and Android operating systems have significant permission controls and built-in restrictions that prevent the easy development of apps for these contexts. Further to this, the Facebook mobile app is a closed platform with little capacity for our team to develop something that could reliably pull data from the Facebook app, developing a tool for mobile users was not feasible. This is an important obstacle for ad accountability, given the reliance of Facebook users – particularly younger ones – on the mobile app.

Duplicates

Some duplication of ads occurs because Facebook identifies some ads with identical images and copywriting as being different because they have different associated ad spending and targeting. Effectively each campaign produces a unique instance of an ad, even if the original post is the same. In one instance in our dataset, a small company has a high representation within our library simply because they engaged in dozens of small ad buys.

Methodological Complications

Our tool has some superficial similarities to webscraping, and thus our research design had to contend with Facebook’s use of anti-webscraping countermeasures. Our approach did not make use of scraping techniques: ‘spiders’ or other webcrawlers engage in a process of automatically identifying and following links within a page, and automatically sending data to a server. This form of webscraping is useful for gathering large amounts of data without a particular concern about a user who experiences the adverts; it also often involves accessing high volumes of data and in volumes that far exceed human browsing experience. Our research sought to represent the real-world experience of users, and thus this approach would not be appropriate. However, the measures that Facebook uses to reduce the impact of these approaches on their servers have impacts on our tool, specifically in terms of the way that we can identify and extract ads from the rest of a Facebook feed. We consider the privacy of our

participants to be extremely important, and we do not want to capture any of their personal posts or the posts of third party individuals. Many other projects that extract posts from Facebook are not seeking to investigate ads, so there are few established workarounds that we can draw upon. This leaves us with several methodological issues in our tool.

In discussions with researchers at New York University, we have identified that Facebook particularly seeks to engage in serial obfuscation of its adverts around politically sensitive periods, such as elections, apparently in order to prevent researchers from identifying patterns of advertising on Facebook around election time. This means that our project has been disrupted during the Australian election in May 2019, and during the US election in November 2020. At different points over these periods we had to update our tool to continue capturing ads.

A Facebook ad made through the ad manager is almost identical to a regular Facebook post, except with additional customisation, linking, predictive analytics, and data tracking for the person making the post. Indeed, we could think about a regular Facebook post as simply being an ad with reduced functionality. At an interface level, this is obvious insofar as the ads have a lot of similarities to a non-ad post within the Facebook interface. This creates an experiential similarity in the way that users relate both to their friends and to private companies, using the same commenting tools and affective reactions. This similarity, however, creates a technical problem in identifying and extracting ads for analysis.

Technical Approach and Challenges

Our approach relies on the way that Facebook identifies the status of its adverts to its users. On all ads, Facebook provides a visible tag as a part of the interface design that alerts users to the fact that the post is an ad. In English, this is the ‘Sponsored’ tag (see Fig. 19 below), and will be different in other languages.

While this is easy to read for a human, Facebook has invested considerable effort obscuring ads from scrutiny within its platform, making it difficult to identify an ad using technical

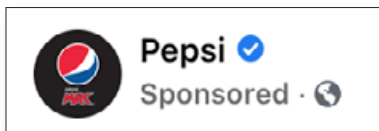


Figure 19

means that examine the page source. The page source is the document that describes how the website is supposed to appear, and is interpreted from a set of text and image files into a Facebook page, and is generally written in HTML and may include Javascript. Facebook seeks to restrict this behaviour through legalese in its Terms and Conditions, and through technical means by making it harder to identify and extract relevant data points in the page source.

The tension is between Facebook remaining transparent to its users, while seeking to be opaque to researchers, and it hinges on the use of the term ‘Sponsored’. The core identifying feature of any Facebook ad is the Sponsored label, and our tool works by finding this tag in the page source, and pulling information on the associated post. Facebook, however, has identified this practice as used by other researchers, and has sought to obscure this tag within the page source, so that it is visible to users, but not visible to our tool. It is possible to undo this obfuscation, but requires effort to identify the method of obfuscation and reverse the

process. While they will inevitably make more changes in the future, Facebook has used multiple processes of obfuscation throughout the life of our project: initially by changing character opacity, then through the use of a tag ‘aria-labelledby’, and more recently through zero-sized elements. Facebook does not remove these obfuscation techniques as it develops new ones, and as a result, the barrier to entry increases for those seeking to do legitimate research of Facebook. Individuals seeking to understand their own Facebook feeds, or those engaged in research without funding are thereby put at a significant disadvantage.

The first approach to obfuscating Facebook ads (in-line character-opacity obfuscation) was already in place when our project commenced. Facebook adjusted this approach to obfuscation during our project, and developed variations on this, including using random punctuation and hyphens. The obfuscation method also changes depending on the web browser used.

In-line character-opacity obfuscation

In this approach, Facebook inserts unused characters into the word ‘Sponsored’, and then sets them to be invisible so that the word still looks the same for users. Within the page source, the characters that comprise the word ‘Sponsored’ has been broken up into different ‘span’ or ‘class’ objects and salted with random characters; the visibility of these characters are set using an ‘opacity’ tag. The word would be rendered correctly for users depending on how the opacity tag is set.

At times these characters have been randomly chosen from the alphabet, other times they have been taken from the word ‘sponsored’ itself, and at other times they have solely been reproductions of the letter ‘s’. In effect, for our script, the word ‘Sponsored’ appears as variations on the theme of ‘SpSonSsoSredS’ or ‘SpSpsonssosredredsss’.

In order to obfuscate ads, non-ad posts also included the word ‘Sponsored’ using the same method within the human-readable timestamp data. This means that the fact that obfuscation is occurring is not sufficient for us to detect an ad, nor is it enough to see if the characters for the word appear in sequence. See Table 1 and Table 2 below for examples.

Non-ad - machine-readable layer	Non-ad - human-readable layer
YhoestetSrdlpoaeny athsaleahi c?s:to3mOr oAueIdMu	Yesterday at 7:30 AM
Yteihostosrerdsay piaoeTh alen2s:ci4ototOl PooreIdSuMm	Yesterday at 12:40 PM
t5Spolndshmooredi	51m
tSu2upohnsoreId	2h
Domercettgmboeffrsn l2cl sSpaaton c6osoi:r3ceO niAeahdarM	December 12 at 6:30 AM

Table 1: sample text from obfuscated non-ad page source data on Facebook

Ad - machine-readable layer	Ad - human-readable layer
tSSmfponostpoorensdotrerddscr	Sponsored
tSSptonpconsgsnoorrsdnhedeeud	Sponsored
Smfponostpoorensdotrerddscr	Sponsored

Table 2: sample text from obfuscated ad page source data on Facebook

It is possible to reproduce this obfuscated text on Facebook without any complex coding. This is best achieved using Chrome on a desktop or laptop computer. First, select the portion of text that refers to the date, timestamp, or time passed on any given post (see Fig. 20 below); copy the text; then paste the text into a text file such as Word. If done correctly, the outputted text will include the obfuscated characters.



Figure 20: selecting date text in Facebook.

As Merrill (2020) notes, this approach has the additional effect of impacting people using accessibility tools, with screen readers reading out the garbled text in full, effectively discriminating against users with vision impairments. These characters are kept in-line in sequence, and the task for our tool is to determine how the

sequence should be read by detecting the opacity of each character or set of characters. Our tool worked by parsing the opacity characteristics for each character, then reconstructing the output and evaluating it. If our evaluation determined that it was rendered as ‘sponsored’, then the ad would be captured.

The weakness in Facebook’s in-line character-opacity obfuscation method is that the data that our tool analyses is still ‘in-line’ with the information we are seeking to gather, in sequence with each post, and visible in the page source. In other words, the valuable portion of information that tells us whether a post is an ad or not is adjacent to the post itself. Facebook’s use of the ‘aria-labelledby’ tag creates a situation where this is no longer the case. These other methods are effectively variations on the method of obfuscation, so the general principles and problems remain the same, even if the technical workarounds are somewhat different.

‘Aria-labelledby’ obfuscation

The ‘aria-labelledby’ tag is designed for use with screen readers to allow the screen reader to properly format data for navigation purposes, and involves the creation of styles and other information that are located in different parts of the page source. The ‘aria-labelledby’ tag is used for helping to make webpages more accessible, but the functionality of this tag allows the Facebook page to load without inline-tags that identify the opacity within the source, and instead are defined elsewhere in the page source. This means that the in-line analysis we conduct is affected by this method.

The ‘aria-labelledby’ tag can be understood by reference to the way we might use the bold feature in an email. In normal convention, we bold words as we go, and this is easily read by humans. However, with the ‘aria-labelledby’ tag, we can instead imagine a long list of instructions and conditions that we read at the start, and then refer back to while we’re reading the email so that we know what sections are bold, and which are not. This would be difficult for humans to read easily, but is relatively trivial for a computer.

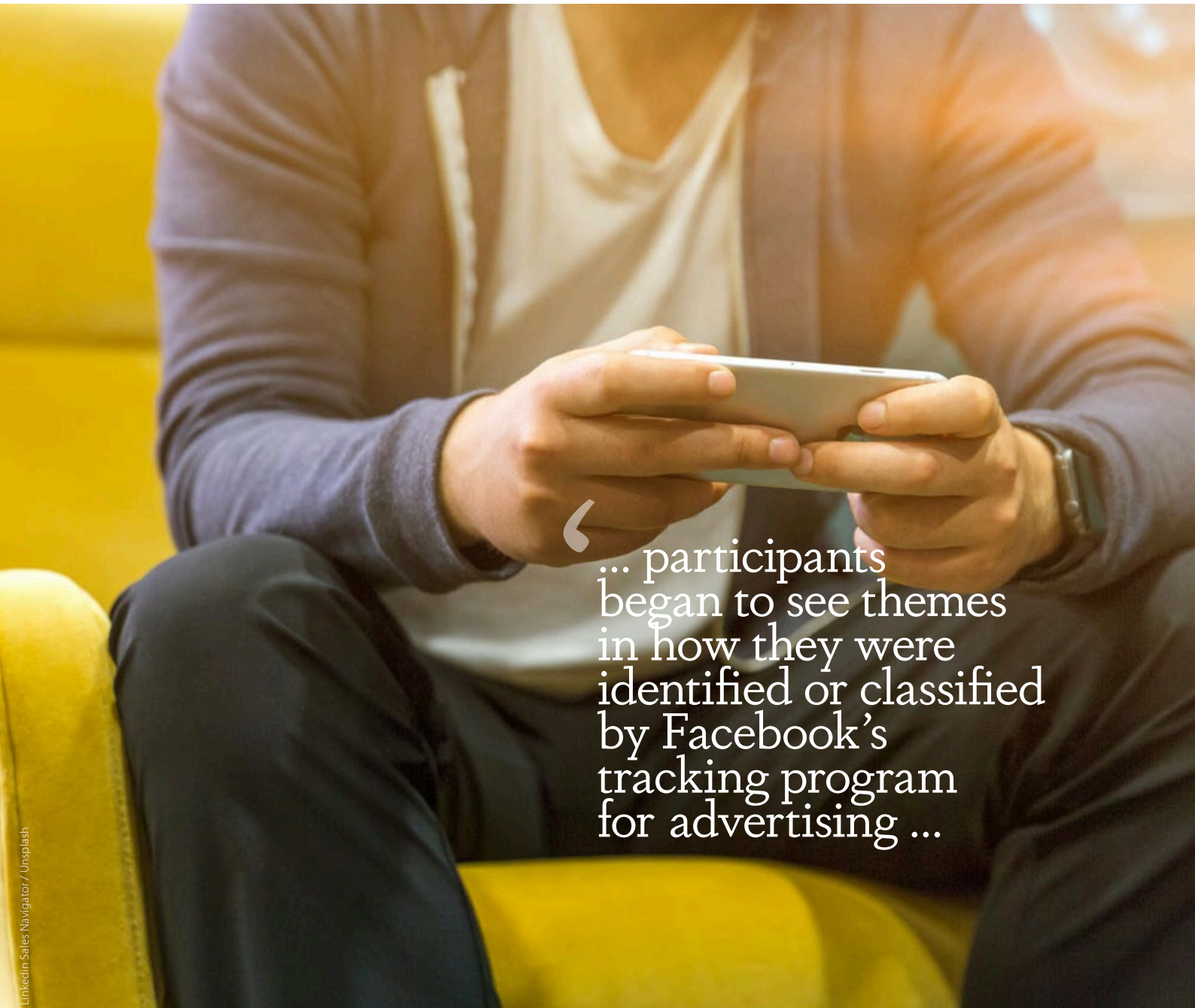
In the case of our tool, the page source loads with this set of instructions which have to be interpreted by our tool to determine which posts need to have their ‘Sponsored’ tag made visible for users, and which are to be left obfuscated. We cannot simply try and disentangle adjacent code; we now have to address the whole document.

Zero-size object obfuscation and margin obfuscation

At some point in late-September 2020, Facebook incorporated a new obfuscation technique that included posts with zero size, and also shifting label text to outside of the margins of their respective ‘span’ containers. Both these approaches take advantage of how different areas of the page are represented, and effectively generates sequences of text where some of the data appears in-line, but with other parts of the text being rendered invisibly in other areas of the page. As such, these approaches are variations on the earlier obfuscation technique. However, in some cases this includes generating entire posts that do not render on the page. These posts would be represented in the page source, but would not render for a regular user browsing Facebook. This is similar to having an image in a document that has been resized to 0 pixels in one dimension - it is still present for the computer to analyse, but not visible to users. Our tool now detects this and is capable of rendering the posts correctly. In all cases, a technical solution for these issues has been provided in open source form at the *ProPublica* (2017) or *ProPublica* (2020) github repositories.

Qualitative and image-classification tagging

The tool has helped us to generate a large database of Facebook ads. We have employed research assistants to manually tag our ads with a set of qualitative content tags. These tags describe the general theme or industry that the ads relate to, and allow us to review how different types of advert relate to different demographics and industries. Our research project is in the process of collaborating with researchers at the Queensland University of Technology to use an AI for image-classification, which will allow us to efficiently sort the large volume of images and evaluate them in terms of patterns, colours, or tendencies that might shape how ads are assigned to different Facebook users. As Kayser-Bril (2020) notes, images are a key part of how Facebook assigns ads. Facebook’s image classification system makes decisions about who to target based on image content – for instance, images containing truck drivers are much more likely to target men than women, with the opposite being the case for images containing child care workers. Because of this, we are seeking to tag and categorise the posts and images to see if there are further details we can extract about the ads beyond what we already know.



‘ ... participants began to see themes in how they were identified or classified by Facebook’s tracking program for advertising ...

Interviews

Interviews have a common methodological approach to the media scroll back methodology described by Robards and Lincoln (2017). The scroll back methodology seeks to understand user experiences of social media platforms by using the platform itself as a part of the interview process. Participants will ‘scroll’ through their own social media feed as relevant to the researcher’s scope and provide a prompted commentary that addresses different features of the tool. Through this, participants provide a detailed, reflexive, and personal response to the platform in a way that is capable of exposing thoughts or

patterns that were not previously conscious, or at least not considered significant enough to utter. While the process of scrolling through a social media feed may sound ineffective or unstructured, in practice it involves a great deal of methodologically useful cues and reminders that allow researchers to focus questions on gathered research data, rather than dealing with hypotheticals, anecdotes, and assumptions. This grounds the interview process in material that is relevant to the research questions.

Our variation of this method involved using solely the ads sourced during the ad collection process. By focusing our participants' attention on the ads alone, we were able to get them to focus on an aspect of their own social media that they tended to ignore or shift to the background. Through this, participants began to see themes in how they were identified or classified by Facebook's tracking program for advertising, and proposed their own theories about what logical processes were being used to establish targeting for their advertising. This process of highlighting the ads for our participants ended up being confronting for some, as they were unaware of how many ads they'd seen, or were perhaps perturbed by their sense of how accurate the ads might be.

We contacted 18 of the participants who had been most active on Facebook during the data collection period for follow up interviews discussing their experience of advertising on Facebook. The researcher and the participant would both read the participant's ad feed on our database. Participants were asked to review ads, describe them, think about whether they were familiar, and consider why they might have been targeted by the ads. During the interviews, participants would articulate their theories and ideas about Facebook, tracking, targeting, and their relationship to Australian civic life. Interviews were recorded, transcribed, and coded using an open coding approach.

Focus Groups

We ran several focus groups with participants who were separate from our originally recruited individuals. Focus group participants were recruited in various contexts from students, staff, and conference attendees. Participants were given access to our ad library, and encouraged to explore the data base, and experiment with filtering options. Participants were tasked with trying to construct demographic profiles based on themselves or people they knew, and reflect on the similarities or differences from their own advertising experiences. Participants were prompted to reflect on why these differences existed, and what decisions Facebook might have made about targeting these ads at specific individuals, or at different scales (local/global, gender, age, etc).

Results

Several focal points have emerged from our pilot project that begin to make transparent how ads are received by Australian social media users. While advertisers have the capability to select and target particular demographics and types of users, Facebook’s algorithmic sorting and recommendation functionality also influence who receives and sees particular ads. Our project focuses on what ads users actually see in practice as a result of the complex and non-transparent entanglement of processes that shape the delivery of ads. Making visible the end result is one of the operative accountability mechanisms in anti-discrimination regulations. Intent matters, but so do outcomes, and one of the achievements of anti-discrimination policy is to include a focus on outcomes in the form of “disparate impact” – whether intentional or not (Datta & Datta, 2018).

Dark Ads Facebook Tool, Website and Mobile App

The ad collection tool along with the public facing website and app are the primary outcomes of this study. They operate as prototypes that can be used for a number of purposes and future studies that can take a more focused approach to understanding the impact of dark ads on various populations. They also work to build consumer and user literacy and awareness by making visible the patterns and differences in the array of ads delivered on Facebook.

Personas and Ad Patterns

In the current iteration of the interface of our website, visitors are invited to select a combination of variables to create an online “persona” in order to see what types of ads people with the selected attributes have encountered on Facebook. This public facing interface allows members of the public to experiment with the data by selecting different combinations of demographic variables to see how this changes the array of ads collected.

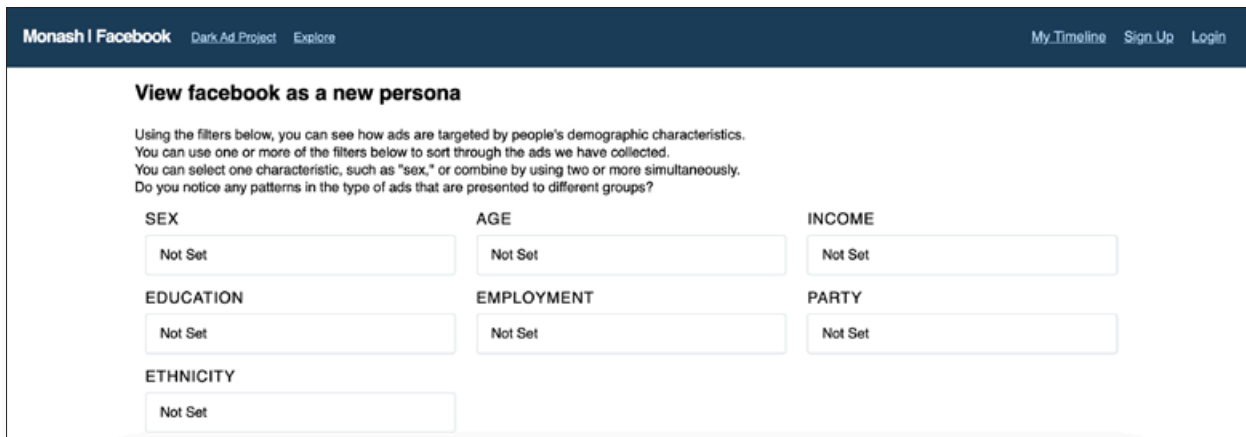


Figure 21

In the workshops and focus groups we have conducted, our participants are invited to engage with this site and explore the database. By navigating through the dataset by selecting different combinations of filters, first, our participants were able to explore their own personas and reflect on how accurate the vision of themselves is reflected in the ads. Some of our participants reported that when they entered their own demographic information, they encountered many of the same ads that they have seen on their own Facebook feeds. The result is that they tend to notice how they are associated with a particular cluster - and the type of messaging with which this cluster is associated. Of course, this finding depends a lot on a match with the database. Because of the limited initial sample, there is not broad coverage of demographic groups (for example, one Asian male in his 20s reported that he couldn't find any ads that matched his demographic profile in the pilot database). Much depends on the ability to recruit a broad-ranging group of participants to install and donate their ads.

Second, our participants were able to explore other personas to see how ad targeting and customization is being played, an experience which is otherwise unattainable in an individualised mode of digital consumption. Several advertising patterns emerged within our dataset when the demographic filters are varied to see how the content and form of the collected ads change. For example, by setting the filter from 'female' to 'male', our participants reported to see a very different set of ads. The example below compares a selection of ads side-by-side varied by gender. Ads filtered by "male" can be seen on the left while ads filtered by "female" are on the right. A brief selection of ads like this is only suggestive, since the ads are displayed in the order in which they are received, meaning they may cluster by users (it is likely, for example, that each set documented here for each demographic came from the same person during a particular browsing session). Nevertheless, even a small sample, varying the filter can yield some speculative patterns, both in terms of content and form.

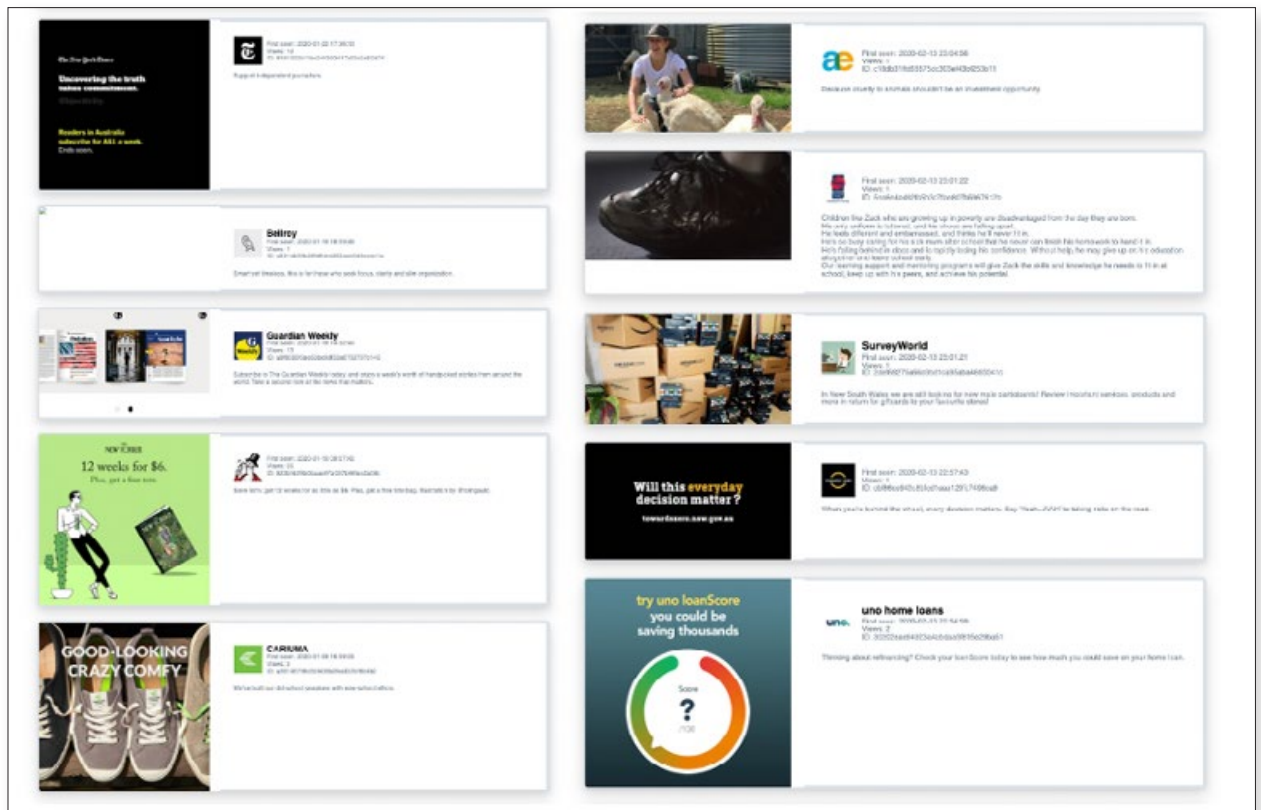


Figure 22: A short selection of results from the ad database comparing different values for the education filter. The figure on the left shows results from entering the variable ‘doctorate’ in the filter; that on the right from entering “less than year 12 or equivalent.”

We noticed, in our pilot group, for example, a strong skew of gambling, tech, and military recruitment ads (by gender: primarily male), as well as of craft, clothing, cooking, and birth control ads (primarily female). Such associations are not necessarily surprising given the gendered sets of expectations and biases that are reproduced by advertising logics. But they point to the ways in which ad targeting reproduces commercial stereotyping. Some other patterns are perhaps less obvious, such as the gender skew in certain forms of business and entrepreneurial self-help programs (depending on what might be read into these ads).

We also include, in the above figure, a short selection of ads varied by education level. There are some ready-to-hand associations between reading material and education level - with publications like *The New York Times* and the *New Yorker* targeted toward the group with a higher stated level of educational attainment, and loan savings services to those with lower levels of education attainment.

Our pilot study also presented results varying based on party preference that align with familiar lifestyle associations. For example, the Greens party preference was coupled with yoga and fitness ads (also highly skewed toward female participants), and there was a connection between right-wing politics and the military.

Participant and User Responses and Literacy

We recorded participant responses and reactions to both exploring the dataset as a whole (in interviews and focus groups) and for those who donated some of their data, we asked them to reflect on their personal stream of ads (in follow up interviews with participants). Our workshop became a discursive locus where our participants collectively shared their own stories of being targeted by ads and reflected on that experience of dark ads. Most of our participants had the experience of ‘being followed around by an ad’ online, and they reported that it usually happened when they searched an item on X website and later on constantly encountered associated ads in Y website and others. Some participants noted a more pervasive form of online surveillance. For example, one participant recounted her perception that when she spoke about a particular item or topic in person when her Facebook was active or “on”, she would shortly begin to see related ads. She reported an increase in awareness about the possibility of Facebook listening to her conversations since the first time she noticed this perceived form of targeting. In fact, this observation prompted her to conduct her own research into why it might happen. She attributed the correlation between the ads she was seeing and her conversations to her allowing Facebook to access her microphone – a frequent but debunked explanation for hyper-granulated targeting. The fact that so many people think Facebook must be listening suggests that, at least in some cases, the data-driven profiling can be very accurate.

Our workshop participants are also generally aware that the ads they receive online are different from what other people might receive. This awareness is prompted by accidental experiences of glancing at someone else’s (oftentimes family members, partners, friends etc) screens when/because they are in sight. For example, one participant shared that she and her partner would receive different ads given their different search histories and even if they search the same product, her partner (male) would receive ads of another colour or model of the product that is tailored to men.

In general, our participants believe that the online ads they receive accurately reflect their online behaviours. And they are surprised by the level and forms of discrimination and social sorting that is being played out in this process. Some female participants felt particularly strongly about the gendered pattern of dark ads and how that would reinforce gender inequality. One participant commented that from users’ perspectives, receiving the same ad for numerous times would limit users’ minds, suggesting a kind of filter bubble in the advertising space in addition to that in news consumption. This comment suggests an important line of research: the ways in which advertising can create its own informational filter bubble – one distinct and perhaps even more pronounced than the forms of customization associated with news and entertainment content.

There was a mix of reasons in participants’ responses as to why they thought they were receiving particular ads. Some participants connected the ads to previous searches they had conducted on Google or Facebook, especially in terms of searching items to purchase (like a vacuum cleaner or signing up to a new health insurer). Other participants reported connections between some of the ads and Facebook groups they were part of or pages that they follow. Other connections participants made included: associations with their children or other relatives’ schools; recent travel; an interest or actions by a friend or another contact; online shopping; use of services (such as Netflix); or because of a demographic characteristic

(e.g. one participant reflected they were seeing ads for Noni B likely because of their age and gender rather than because they would shop there).

During the interviews some of the participants reflected on how they felt after looking at their personal ad stream. Some participants felt they should be more vigilant about what they look for via Google Search as they felt many of the ads they saw were from that source. Others did not mind seeing ads as long as they felt they were relevant and in line with their personal interests. Still others were surprised to see how their own behaviour generated legible and consistent patterns for advertisers.

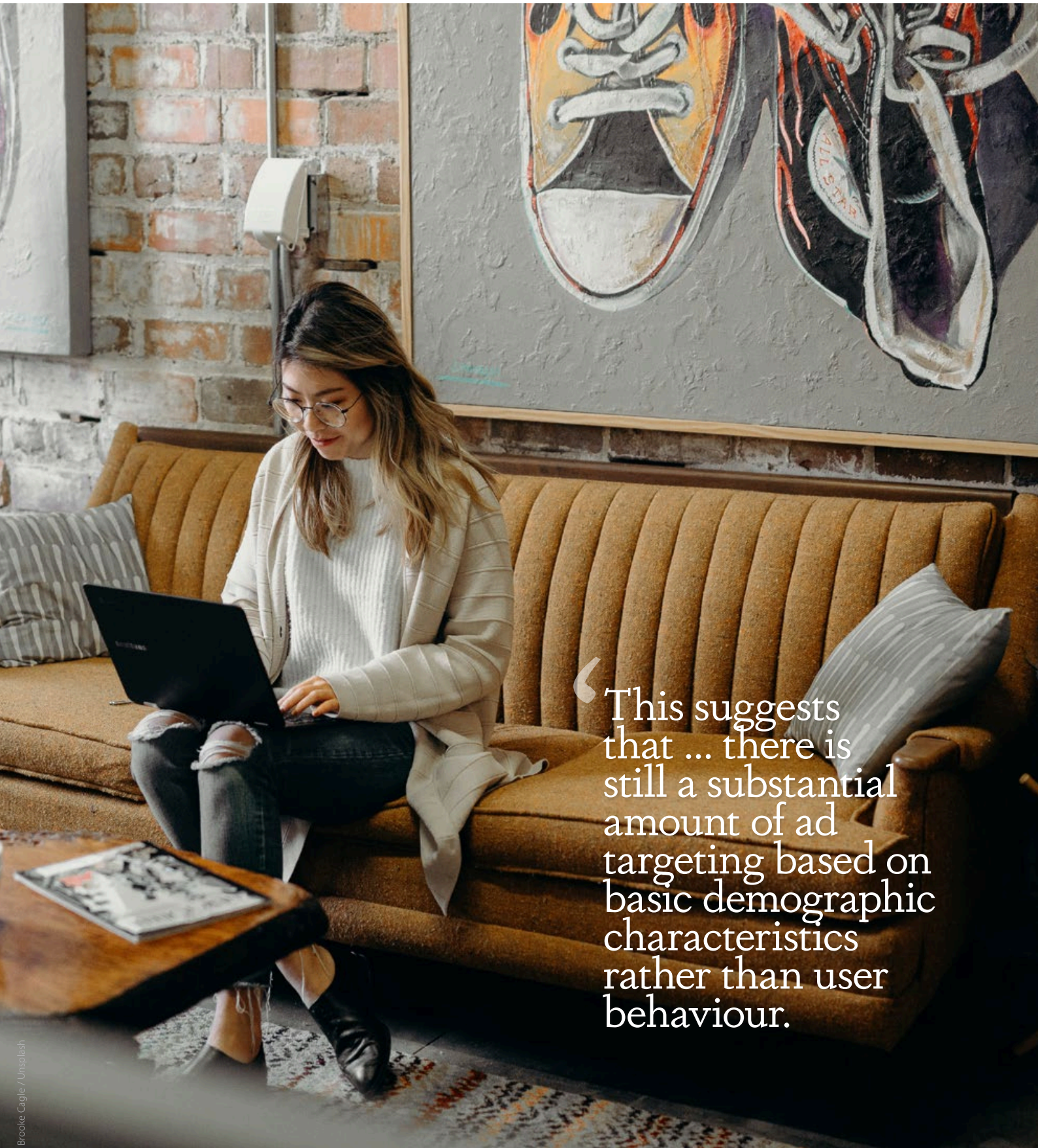
Some of the participants also felt the ads were only half accurate in terms of aligning with their interests and online behaviour, while there was a portion of ads they could not determine a reason as to why they might have been part of the target audience. This is perhaps less surprising than it might seem at first, given the experimental character of online advertising, which is always in the process of learning which groups are likely to respond to which ads. Participants also reflected that the ads often were ‘annoying’ and ‘too repetitive’ when the same ad would recur yet had no relevance to them or they could not identify what it was for. There were several ads that users noticed and that emerged in our dataset that appeared to be clickbait or scams in which there was no clear advertiser or item being advertised, opening up another avenue for possible research.

In our focus groups, some participants expressed surprise at the similarities in their newsfeed with the persona matches on our website and revealed that they had engaged in several tactics in an attempt to minimise the targeting of ads that they received (e.g. denied and blocked cookies associated with particular sites that track one’s behaviour across other websites, and cleared or changed the ad preferences in their own account settings). This suggests that, at least in some cases, there is still a substantial amount of ad targeting based on basic demographic characteristics rather than user behaviour. In saying that, in our follow up interviews with participants who donated their data by installing the browser extension, there were several cases in which the ads that were collected from them were very specific in their targeting.

In the follow up interviews, we asked the participants to reflect on the stream of ads that was collected from their specific accounts and if they could explain to us why they thought they had received some of these ads. There were a few instances in which the participants initially expressed hesitancy to reveal to us the personal reasons they believed contributed to this ad targeting. Some of these reasons ended up being related to frequency of gambling and betting, and in one case a user disclosed her son was disabled, which is why she believed she was receiving ads about children and disability services.

Demographic Distribution by Ad Category

While our database was small, we did recruit participants to be roughly representative of the country by gender and geographic distribution. We attempted representativeness by age, but this was more difficult to achieve in the sampling. With these caveats in mind, we were able to go through the distribution of ads by category to get a sense of the patterns that emerge from Facebook’s targeting system. We present these results more as a demonstration of proof-of-concept – that is, what can be done with an application like this if it were implemented on a broader scale – than as a fully representative sample of how ads are



“ This suggests that ... there is still a substantial amount of ad targeting based on basic demographic characteristics rather than user behaviour.

currently distributed in Australia. Because gender was the most representative category in our sample (although, in the end, we did have a male skew, as reflected in the findings), we focus on results by gender, but we also pull out other patterns where these look suggestive. For example, it was clear that our sample skewed toward the conservative side of the political spectrum, with self-identified Liberal Party or National Party supporters generally in the majority. But when it came to music ads – typically for concerts or new releases, there were more impressions for Labor supporters, which demonstrates a strong skew toward Labor, insofar as it overcame the bias of the sample. When it came to political advertising – either for issues or candidates – there was a strong skew toward the National party. Rather than positing these results as ones we could generalise, we invoke them as possible avenues for future research on the role of advertising in perpetuating forms of stereotyping.

Our paid sample wasn't broad enough in many categories to make national level generalizations. For example, one active user in a particular demographic group could heavily skew the results. When it came to ads for alcohol, for instance, the chart makes it look like the Labor party is disproportionately represented – but this could have been the result of one particular user with a taste for whiskey. However, the tool makes it possible, when implemented at a broader scale, to identify patterns of the distribution of ads across demographic categories. We are currently in the process of developing approaches for recruiting a broader ranging set of participants. One possible approach, for example, would be to open the tool up for use in classes on digital media at the university level. The tool could be used for class projects exploring the operation of online advertising. If we were able to recruit, say, six large undergraduate classes for the major part of a term, we would have a much larger pool – and one that might be more diverse in most categories (except for age). It might also be possible to promote the tool as an ongoing project through media articles that focus on the issues raised by dark ads. Information about the shift in the information environment tends to trickle out: an article about Google targeting ads to people with ethnic names, another about ads for military gear being served alongside social media posts advocating armed insurrection, and so on. There has not been sufficient coverage of the social and cultural consequences of the wholesale transformation of the advertising environment.

Table 3: Top 10 ad categories by frequency

Ad Topic	Number of Unique Ads	Gender Skew
Health/Wellness	1464	F
Shopping/retail	1266	M (slight)
Business	1159	M
Technology	1128	M (strong: 3:1)
Education	1110	F (slight)
Food	989	M (slight)
Travel	939	M (slight)
Entertainment	936	Even
Finance	838	M (strong, almost 3:1)
Clothing	700	F
Home	634	Even

Conclusions

Our research project examined dark ads in an Australian context using data drawn from our participants' Facebook feeds. We then sought to obtain perspectives and experiences from our participants to understand how they saw these ads operating in their lives. Experiences were diverse. Some were anxious about ads, others felt powerless, while many felt like the only thing to do was to block the ads themselves. Blocking ads does insulate users from their content, but it does not prevent their behaviour being monitored for the purposes of so-called collaborative filtering: finding out which audiences might be most receptive to particular types of content. Nor does it prevent them being profiled based on their activities – the data about their interests and behaviour is still captured by online platforms. Our participants' theories ranged from the apathetic (that most ads were for 'everyone') through to the individually micro-targeted through to the conspiratorial (with ideas that all their devices and all their behaviours were being fed into a digital advertising complex). Opinions on the ad targeting varied from seeing it as highly prescient through to a vague and inaccurate process. In many cases, the impacts of online ads were seen as an individualised self-replicating phenomenon: ads are bad because they are a product of tracking, tracking is bad because it leads to ads. Missing from these responses is a sense of the broader social consequences of unaccountable forms of advertising – the possibility of stereotyping, discrimination, and anti-social messaging.

From our research, we can see that the Australian public lacks a suitable awareness of how online ads work and lack a lot of nuance in terms of its effects. Regulatory interventions in other jurisdictions have been successful in mobilising changes within Facebook and other online companies. The Australian Federal Government's *News Media Bargaining Code* (ACCC 2020) has demonstrated that the government is not afraid of regulating social media giants, and we see an opportunity to take this further in making productive developments in a way that enhances democratic processes in Australia. Our recommendations detail the specifics that we see as important for these developments.

Recommendations

Recommendation 1

Social media platforms should be required to provide publicly accessible and searchable ad libraries that make it possible to sort ads by the demographic characteristics of those who received them.

Recommendation 2

Social media platforms should be required to design and implement socially responsible algorithms and prevent potentially discriminatory ad delivery and distribution. Increasing attention has been paid to discriminatory patterns in online content distribution, but little has been done with advertising.

Recommendation 3

Regulatory bodies and Internet Service Providers should make adherence to accountability provisions a condition of access to end users.

Recommendation 4

Increased public awareness and discussion of how dark ads work and the ramifications of dark ads to structural discrimination is urgently needed. The dark ads Facebook tool, website and mobile app derived from this study could serve as a starting point for public engagement. This would involve improving consumer and user literacy about dark ads as a safeguard to help mitigate against the effects of manipulative targeting. Although we should not rely on such measures when users cannot control what ads they receive; the increased literacy will help users consider their own options in terms of being aware of the ads, understanding why they are seeing certain ads and perhaps adopting strategies to take more control over what they see (e.g. using ad blockers, changing their 'interests' etc).

Recommendation 5

Social media platforms should be required to improve and streamline their mechanisms for the public to lodge complaints about inappropriate and offensive ads. These mechanisms should include a defined response time, so that users can see that complaints have been registered and actioned.

Recommendation 6

Social media platforms should be required to be specific about the sources of data that lead to them being targeted for advertising purposes. At present, users receive very general and non-specific information, if at all. By allowing users to have a better idea of the information that is being used to target them, then they would be better informed about the way their data is being used, and how their online and offline behaviour generates their advertising profiles.

Recommendation 7

Social media companies should support public, non-profit research efforts to provide accountability, providing access to data that makes it possible to assess the outcomes of their algorithmic patterns of targeting. They should agree to ban strategies of obfuscation that thwart accountability efforts.

Authors

The authors are members of the Automated Society Working Group, Monash University:



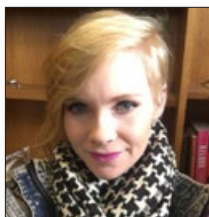
Professor Mark Andrejevic researches the social and cultural aspects of digital media technologies, with a particular emphasis on automated data collection and decision-making. His work explores the intersection of popular culture, surveillance, and media technology. He is the author of three monographs, more than 70 journal articles and book chapters, and co-edited two books.



Dr Robbie Fordyce's work is focused on the use of digital media technologies by the general public, examining what barriers, concerns, and risks people encounter with new technology, and how these technologies can empower or enhance our everyday lives.



Dr Nina Li's work focuses on digital culture, global media and cultural industries, media history and political economy, television studies, Chinese media, among others. Her recent project examines how online video emerged as an alternative cultural institution to television in China and its consequences for cultural politics and global communication.



Dr Verity Trott's research combines big data analytics with an ethnographic approach to examine the political, cultural and social dimensions of digital technologies from an intersectional and feminist standpoint. Her recent research explores digital feminist activism and expressions of masculinity online.

References

- ACCC. 2020. News Media Bargaining Code. ACCC.org.au. Accessed 16 December, 2020. Available at: www.accc.gov.au/focus-areas/digital-platforms/news-media-bargaining-code
- Angwin, J. & Larson, J. 2017. Help Us Monitor Political Ads Online. *ProPublica*. 7 September, 2017. Available from: www.propublica.org/article/help-us-monitor-political-ads-online
- Australia, House of Representatives 2000, *Debates*, vol. HR103, pp. 2-9
- Bradt, G. 2016. Wanamaker Was Wrong – The Vast Majority Of Advertising Is Wasted. *Forbes*. 14 September, 2016. Available at: www.forbes.com/sites/georgebradt/2016/09/14/wanamaker-was-wrong-the-vast-majority-of-advertising-is-wasted/?sh=37c10211483b
- Grant, A. & Howartz, D. (eds) 2011, *Australian Telecommunications Regulation*, 4th edn, CCH Australia, Sydney
- Kayser-Bril, N. 2020. Automated discrimination: Facebook uses gross stereotypes to optimise ad delivery. *Algorithm Watch*. 18 October, 2020. Accessed 15 December, 2020. Available at: algorithmwatch.org/en/story/automated-discrimination-facebook-google
- Larson, J., Angwin, J. & Valentino-DeVries, J. 2017. How We Are Monitoring Political Ads on Facebook. *ProPublica*. 5 December, 2017. Available from: www.propublica.org/article/how-we-are-monitoring-political-ads-on-facebook
- Merrill, J. B. 2020. Facebook didn't mark ads as ads for blind people for almost 2 years. *Quartz*. 12 February, 2020. Accessed 14 December, 2020. Available at: qz.com/1800064/for-blind-facebook-users-ads-havent-been-labeled-as-ads
- Murphy, T. & Whitty, N. 2007, 'Risks and human rights in UK prison governance', *British Journal of Criminology*, vol. 47, no. 5, pp. 798-816
- Neilsen. 2018. Screen Time Still an Australian Pastime. Nielsen.com. 3 April, 2018. Available at: www.nielsen.com/au/en/insights/report/2018/screen-time-still-an-australian-pastime. Accessed 23 April, 2021.
- ProPublica*. 2017. facebook-political-ads project repository. *Github.com*. Available at: github.com/propublica/facebook-political-ads Accessed 11 December, 2020.
- ProPublica*. 2020. social-media-collector. *Github.com*. Available at: github.com/OnlinePoliticalTransparency/social-media-collector Accessed 14 December, 2020.

Robards, B. and Lincoln, S. 2017. Uncovering longitudinal life narratives: scrolling back on Facebook. *Qualitative Research* 17 (6): 715-730.

Taylor, J. 2020. Automated discrimination: Facebook uses gross stereotypes to optimise ad delivery. *The Guardian*. 14 December, 2020. Accessed 14 December, 2020.

Available at: www.theguardian.com/technology/2020/dec/14/scam-bitcoin-ads-using-unauthorised-celebrity-images-fool-tens-of-thousands-of-australians

Tilley, C. 2018. How we are spending our internet minutes. *ABC News*. 2 July 2018.

Available from: www.abc.net.au/news/2018-07-02/how-we-spend-our-internet-minutes/9923822?nw=0

