Improving privacy protection in the area of behavioural targeting

Zuiderveen Borgesius, F.J.

Publication date
2014

Document Version
Final published version

Link to publication

Citation for published version (APA):
Zuiderveen Borgesius, F. J. (2014). *Improving privacy protection in the area of behavioural targeting*.

2 Behavioural targeting

What is behavioural targeting, and how does it work? Behavioural targeting, also referred to as behavioural advertising or online profiling, involves monitoring people's online behaviour, and using the collected information to show people individually targeted advertisements.[75] The Interactive Advertising Bureau of the United States, a trade association for online and mobile advertising, describes behavioural targeting as follows:

> A technique used by online publishers and advertisers to increase the effectiveness of their campaigns. Behavioral targeting uses information collected on an individual's web browsing behavior such as the pages they have visited or the searches they have made to select which advertisements to be displayed to that individual. Practitioners believe this helps them deliver their online advertisements to the users who are most likely to be influenced by them.[76]

In a simplified example, an ad is shown on a website based on the inferred interests of that specific visitor: these interests can be inferred by an advertising network. An ad network is a firm that acts as an intermediary between websites and advertisers. The ad network might profile somebody who frequently visits websites about recipes as a food enthusiast. If that person visits a news website, the ad network displays

[75] See e.g. Federal Trade Commission 2000 ("online profiling") and McStay 2011 ("behavioural advertising").
[76] Interactive Advertising Bureau United States, Glossary.

advertising for restaurants or cookbooks. When visiting that same news website, somebody who reads a lot of legal blogs might see advertising for law books.

The chapter is structured as follows. Below is a glossary of some key terms. Section 2.1 and 2.2 introduce online advertising and the technology used for behavioural targeting. Section 2.3 to 2.7 sketch the process of behavioural targeting, divided in five phases: (1) data collection, (2) data storage, (3) data analysis, (4) data disclosure, and (5) targeting.[77] Section 2.8 concludes.

[77] Other authors also distinguish different phases of data mining, profiling, and data processing. See for instance Hildebrandt et al. 2008 (3 phases); Solove 2006 (4 phases); Solove 2008 (4 phases); Zarsky 2004 (3 phases), and Cabena et al. 1998, p. 43-44 (5 phases).

Glossary

Advertising network company

Advertising network companies, ad networks for short, connect advertisers to website publishers, and serve ads on websites. Using cookies or other technologies, an ad network can recognise a user when she visit websites on which the ad network shows ads.[78]

Advertising exchange company

Ad exchanges are automated market places where advertisers can trade with multiple ad networks in one place. The Interactive Advertising Bureau US provides the following description. "Ad exchanges provide a sales channel to publishers and ad networks, as well as aggregated inventory to advertisers. They bring a technology platform that facilitates automated auction based pricing and buying in real-time. Ad exchanges' business models and practices may include features that are similar to those offered by ad networks."[79]

Behavioural targeting

Behavioural targeting is the monitoring of people's online behaviour, to use the collected information to show people individually targeted advertisements.

Click-through rate

"The number of click-throughs per ad impression, expressed as a percentage."[80] For instance, if 3 out of a 1000 people click on an ad, the click-through rate is 0.3 %.[81]

[78] See for a more detailed description Interactive Advertising Bureau United States 2010.

[79] Interactive Advertising Bureau United States 2010. See also section 6 of this chapter.

[80] American Marketing Association dictionary.

Cookie

HTTP cookies, cookies for short, are small text files that a server can send to a browser. First party cookies are set by the website publisher, and third party cookies are set by others, such as ad networks. Third party cookies enable ad networks to follow people around the web. Tracking technologies that rely on storing information on a user's device that are used for purposes similar to HTTP cookies are sometimes called super cookies.[82]

Interactive Advertising Bureau (IAB)

The Interactive Advertising Bureau (IAB) is a trade association of online marketers, with branches in many countries. According to the IAB Europe website, "IAB Europe is the voice of digital business. Its mission is to protect, prove, promote and professionalise Europe's online advertising, media, research and analytics industries. Together with its members – companies and national trade associations – IAB Europe represents over 5,500 organisations."[83] The IAB also "promotes self-regulation for online behavioral advertising."[84] The IAB of the United States says on its website that one of its "core objectives" is to "[f]end off adverse legislation and regulation."[85]

Real time bidding

Real time bidding is a process where advertisers (or their intermediaries) bid on an automated auction for the right to reach a specific user, who is identified with a cookie. Real time bidding "creates a data market where users' browsing data are sold at auctions to advertisers."[86]

[81] See section 1 of this chapter.
[82] See section 2 of this chapter.
[83] Interactive Advertising Bureau Europe, website.
[84] Interactive Advertising Bureau Europe, website.
[85] Interactive Advertising Bureau United States, website.
[86] Castelluccia et a. 2013, p. 14. See section 6 of this chapter.

Website publisher

Website owners are often called website publishers.[87]

2.1 Online advertising

Behavioural targeting can be seen as the latest development in a decades-old trend of increasingly targeted advertising at smaller audience segments. Because media audiences became more fragmented in the 1970s, marketers started to pay more attention to targeting audience segments.[88] In the 1980s and 1990s direct marketing progressed to database marketing, "the use of customer databases to enhance marketing productivity through more effective acquisition, retention, and development of customers."[89] Marketers started to compile increasing amounts of consumer data.

In the early 1990s, marketers gave little attention to segmentation on the internet. Users were mainly well-educated, had relatively high incomes, and were based in a small number of Western countries. When more people started to use the internet, and more websites were published, segmenting and targeting became more important for advertisers.[90]

The trend towards targeted and personalised advertising is summarised well by the Association of National Advertisers, a trade association in the United States. At its hundredth anniversary in 2010, it adopted a Marketers' Constitution. "Marketing must become increasingly targeted, focused and personal," says the first article. The Marketers' Constitution adds that the "exciting, controversial, but extraordinarily

[87] The IAB describes a publisher as "[a]n individual or organization that prepares, issues, and disseminates content for public distribution or sale via one or more media" (Interactive Advertising Bureau United States Glossary).
[88] Turow 2011; McStay 2011.
[89] Blattberg et al. 2008, p. 4.
[90] See Cannon et al. 2007; McStay 2011, p 128-132; Newell 1997, p. 191; Turow 2006; Turow 2011.

important world of behavioral advertising offers enormous efficiencies to marketers and immense value to consumers."[91]

Behavioural targeting also aims to fulfill another desire of advertisers, who seek information on the audiences they reach.[92] A a famous phrase in marketing literature is: "I know half my advertising is wasted. The trouble is, I don't know which half."[93] Since the beginning of the twentieth century, measuring how many people are reached with advertising has been a continuous quest. Commercial mass media, such as newspapers and television, could be seen as providing audiences to advertisers.[94] Bermejo explains: "since the audience becomes a commodity, those who purchase it, advertisers, need to be certain that they are getting what they pay for."[95] Firms adapt the way they measure audiences if a new communication channel emerges. Different methods are applied to print, radio, television, or the web.[96]

For example, in 1914 American newspaper publishers established the Audit Bureau of Circulation. This organisation provided advertisers with figures about circulation, in order to dispel doubts that publishers were giving advertisers inflated figures.[97] Radio complicated matters. Counting the number of people who listen to a radio show is harder than counting how many newspapers are sold.[98] An audience measurement industry developed to provide advertisers with statistics about listeners.[99] Early research methods involved calling people at home to ask what they were listening to.[100] Later, firms such as Nielsen used recording devices called "audimeters" that were installed in households. Similar recording devices are still used for television

[91] Association of National Advertisers 2009.

[92] See Aaltonen 2011 (chapter 2); Bermejo 2007; Bermejo 2011.

[93] McStay 2010, p. 187; Turow 2006, p. 21. The quotation is attributed to different people.

[94] Turow 2006, p. 6.

[95] Bermejo 2007, p. 25. See also McStay 2011, p. 130-132.

[96] Bermejo 2007, p. 38-39.

[97] Andrejevic 2009, p. 82. The Audit Bureau of Circulation organisation still exists, now under the name Alliance for Audited Media. <www.auditedmedia.com> accessed 14 February 2014.

[98] Andrejevic 2009, p. 84. Bermejo 2007, p. 38-39.

[99] Bermejo 2007, p. 38-41.

[100] Andrejevic 2009, p. 86.

ratings.[101] They are installed in the homes of a sample group of viewers, and record what television programmes are watched. Firms arrange panels to answer questions, in order to obtain demographic information about viewers of certain programs.[102]

Internet marketing

Formerly, "for-profit activities" were not allowed on the internet, but this prohibition was lifted in the early 1990s.[103] In 1994 the first banner advertisement was shown on the web, on the website HotWired.[104] Banner ads, or display ads, are rectangular ads on websites. The first ads on the web were bought in a manner comparable to advertising on television or in newspapers. On television, an advertiser pays a fixed fee, based on the expected number of viewers during a certain period. In print, the advertiser pays for the expected number of readers, based on circulation figures. On the web, it was possible to count the number of "impressions": the number of times an ad was displayed. In a "cost per mille" model, an ad network counts how often it shows an ad, and the advertiser pays for a thousand impressions.[105]

In the mid 1990s, many larger advertisers were still hesitant to spend money on web advertising. In particular, advertisers complained about the lack of information about internet audiences. For instance, before cookies, a website publisher couldn't tell the difference between visitors. A 1996 paper which was presented at an advertising conference complained: "twenty hits could mean 20 different people visited the site, or just one person clicked a computer mouse on the site 20 different times."[106]

[101] Andrejevic 2009 p. 87; Bermejo 2007, p. 41-42.

[102] Bermejo 2007, p. 108.

[103] See Murray 2007, p. 72-73. In 1992 the National Science Foundation Network still listed "for-profit activities" as "unacceptable uses", subject to some exceptions (NSFNET Backbone Services Acceptable Use Policy 1992).

[104] Turow 2011, p. 43. McStay 2010, p. 18. A banner ad on HotWired is usually referred to as the first banner ad, but McStay mentions that a Wikipedia entry speaks of a banner ad in 1993.

[105] Turow 2011, p. 43-44.

[106] Hong & Leckenby 1996, p. 7.

Advertisers successfully pushed for a different way of paying for internet advertising: a "cost per click" model.[107] In this model, an advertiser only pays the website if somebody clicks on the ad. Advertisers often buy ads through advertising networks. These ad networks typically use a cost per click model as well. There are more payment models for online advertising. For instance, in a cost per conversion model, the advertiser pays for every person that takes a certain action, such as buying a product. According to a report by the Interactive Advertising Bureau (IAB), around two thirds of all online advertising income is paid for per click, or per conversion.[108] The IAB is a trade organisation of online marketers, with branches in many countries.[109]

Few internet users click on ads. When an ad is shown to 1,000 people, on average between one and five people click on the ad. Hence, the click-through rate is in the order of 0.1 % to 0.5 %. To increase the click-through rate, ad networks aim to target advertising precisely. This gives firms an incentive to collect increasing amounts of data about individual internet users.[110] Since the 1990s, click-through rates have been falling dramatically. Prices for advertising are decreasing as well.[111] The number of websites however, keeps growing, so advertising space on the web is also growing. As the supply of advertising space grows, the prices go down.[112] Prices depend on many factors, and it's difficult to find exact numbers. Generally, an advertiser pays

[107] Turow 2011, chapter 2 and 3.

[108] Interactive Advertising Bureau 2013, p. 11. The report summarises such payment models as "performance-based pricing." Around 65% of the 2013 revenues in the US were priced on a performance basis. Around 33% of the revenues were priced on a cost per mille model.

[109] The website of the European branch says: "IAB Europe is the voice of digital business. Its mission is to protect, prove, promote and professionalise Europe's online advertising, media, research and analytics industries. Together with its members – companies and national trade associations – IAB Europe represents over 5,500 organisations" (Interactive Advertising Bureau Europe, website). The IAB of the US says on its website that one of its "core objectives" is to "[f]end off adverse legislation and regulation" (Interactive Advertising Bureau United States, website).

[110] Turow 2011; Strandburg 2013, p. 127.

[111] See e.g. Glaser 2014.

[112] Launder 2014.

between one and four euro for 1,000 ads (a cost per mille).[113] Website publishers receive about half of that amount; the other half goes to the ad network.[114]

There's "surprisingly scant research" on how effective or how expensive behaviourally targeted ads are, when compared to contextual ads.[115] A few papers suggest that behavioural targeting leads to an increase of advertising income for website publishers, but each of these papers is criticised for its methods.[116] For instance, a paper by Beales, sponsored by the Interactive Advertising Bureau US, says that behaviourally targeted ads are about 2.7 times as expensive for advertisers than ads sold in a "run of network" model. A "run of network" means that ads are presented completely randomly, usually on websites with the cheapest advertising rates. However, Beales doesn't compare behavioural targeting with contextual advertising. Contextual advertising concerns, for instance, ads for cars on websites about cars. Contextual ads are probably more expensive than completely random ads.[117]

Power relations in online media

On the internet it's possible to present a different ad to each individual. Therefore, the ads on a webpage aren't necessarily related to the content of that page. In print media, by contrast, groups of readers see the same ad.[118] By way of illustration, a printed newspaper with many golf players among its readers could be a good place for a golf club manufacturer to advertise. The newspaper assembles an audience, and provides the advertiser access to this audience.[119] The price of an ad is based, among other

[113] Turow 2011, p. 78. Mitchell reports on an average price for thousand viewers of $2.66 for an online banner ad (Mitchell 2012). Beales mentions a price for thousand viewers of $4 for a behaviourally targeted ad (Beales 2010, p. 3).

[114] Turow 2011, p. 78.

[115] Mayer & Mitchell 2012, p. 8.

[116] See e.g. Strandburg 2013, p. 100-105; Mayer & Mitchell, 2012 p. 8.

[117] See Mayer & Mitchell 2012, p. 8; Strandburg 2013, p. 100-105.

[118] Not all readers see the same ad in print media. Some print magazines and newspapers adapt advertising to regions. In one case, the cover of an US magazine showed a map on which the subscriber's address was circled (Carr 2004).

[119] See Bermejo 2007.

things, on the number of readers. The newspaper tells advertisers that it sells 100,000 copies, and shows research that says that 70 % of its readers play golf. With behavioural targeting, an ad network can show a golf ad anywhere on the web to a person whose profile suggests that he or she likes golf. An ad network doesn't have to buy expensive ad space on a large professional news website to advertise to an individual. The ad network can reach that individual when he or she visits an unknown website, where advertising space is cheaper.

Turow explains that publishers have less power in the online media environment than they had in the print environment. He quotes a digital marketing firm that says: "advertisers want to pay to reach the target audiences. They don't want to pay for the creation of content."[120] Advertising intermediaries and advertisers have more power than two decades ago. Hence, in the long run behavioural targeting may decrease ad revenues for some website publishers. Publishers that produce expensive content, such as online newspapers, might be better off with ads that aren't targeted at individuals. The editor of The Atlantic complains about the effects of behavioural targeting on the media: "[t]hen the digital transition came. The ad market, on which we all depend, started going haywire. Advertisers didn't have to buy The Atlantic. They could buy ads on networks that had dropped a cookie on people visiting The Atlantic. They could snatch our audience right out from underneath us."[121]

In addition to the advertisers' wish to segment audiences and to obtain information about the audience they reach, a third factor can help to understand the rise of behavioural targeting: the development of technologies that make behavioural targeting possible. Online advertising technology is discussed in the next section.

[120] Turow, 2011, p. 117.

[121] Madrigal 2013.

2.2 Advertising technology

In 1990, Berners-Lee invented the world wide web, an application that runs on the internet.[122] We use the web when we visit a website with our browser. The web consists of millions of web pages that are connected through hypertext.[123] Hypertext transfer protocol, or HTTP, is the network protocol that was developed for the web. The protocol enables communication between web browsers and web servers.[124] A web browser is software for users to browse the web, such as Chrome, Firefox, Internet Explorer, or Safari. A web server is a computer that holds the data of a web page. The hypertext transfer protocol includes the kinds of requests that a browser can ask to a server, and the different kinds of responses a server can send back to the browser. If somebody enters the webpage address (a URL, or uniform resource locator) in the browser, the browser sends that request to a server. The server sends back the requested documents, such as text or images. The server can record information about the computer that makes a request. Such "web logs" can include the time and date of the request, the IP address of the computer that makes the request, and information about that computer, such as the browser type and the operating system.[125]

The hypertext transfer protocol is stateless. This means that a web server sees each visit to a webpage as the web browser's first visit. After the browser has received the documents it requested, it breaks off the connection. When the user clicks a link, the browser contacts a server again. In short, a stateless system has "amnesia."[126] Statelessness wasn't a problem the first years after the web was invented, but in the early 1990s firms started thinking about online commerce. However, it was difficult

[122] The internet is "an electronic network that parcels application information into packets and ships them among computers over wires and wireless media, according to simple protocols (rules) known by various acronyms." Berners-Lee 2010, p. 83.

[123] A website is a collection of web pages.

[124] See generally Gillies & Cailliau 2000.

[125] Kaushik 2007, p. 26-27.

[126] Schwartz uses the phrase "amnesia" in this context (Schwartz 2001).

to build virtual shopping carts for a web shop. In the web's stateless system, the web shop would see each browser request as coming from a new visitor.

Cookies

Cookies were invented to solve the problem of statelessness on the web.[127] One of the first popular web browsers was Netscape Navigator. In 1994, a 24-year old programmer at Netscape called Lou Montulli aimed to solve the problem of statelessness, to enable firms to build shopping carts for their websites. He invented cookies to give the web a memory.[128] Netscape quickly implemented cookie technology in its browser in 1994. Netscape didn't inform the browser users, and the browser didn't enable users to manage or refuse cookies.[129]

Cookies are small text files that a server can send to a browser. The browser saves that cookie. If the browser contacts that same server again, it sends back the cookie with its request. Like this, the server can recognise the browser. This is useful to remember the contents of a virtual shopping cart, language preferences chosen by a user, or to remember that a user is logged in. Session cookies are deleted when the browser is closed. Persistent cookies remain stored if the browser is closed and when the computer is turned off.

If a server places a cookie on a computer, in principle only servers from that same domain can read that cookie.[130] In brief, website X cannot read the cookies that website Y placed. If a user visits www.bookstore.com, that website may place a cookie on his or her computer.[131] Only servers from the same domain, such as bookstore.com or accounting.bookstore.com, can read that cookie. If the user later visits www.email.com, the servers from email.com can't read the cookies of bookstore.com.

[127] See generally on cookies St. Laurent 1998; Elmer 2004, chapter 6; Kesan & Shah 2003; Kristol 2001.
[128] The phrase "giving the web a memory" is borrowed from Schwartz 2001.
[129] Kesan & Shah 2003, p. 300; Turow 2011, p. 47-48.
[130] However, as explained below firms have found ways to work around this.
[131] The websites in the text are examples and aren't meant to refer to real websites.

Ad networks, however, have found a way to use cookies to track people around the web. “Third party cookies” are cookies that aren’t placed by the website publisher, but by a third party. If a user visits a website, say www.news.com, it seems that all elements on the screen are presented by news.com. But different parts of a website often come from different servers. For instance, a website might have a section, or “widget”, with weather information. The widget could be sent to the visitor’s browser from widgets.com. Social network site buttons on websites, such as the Facebook Like Button, are usually loaded from third party servers as well. Likewise, ads are usually sent to the visitor’s computer by a third party, for example from the domain advertising.com. This process is invisible for the visitor, who directed his or her browser to www.news.com. (When speaking of “third party cookies”, this study refers to cookies which aren’t operated by the website publisher, but by a third party, such as an ad network.[132])

To recognise internet users, ad networks also drop and read cookies on computers. In principle, such third party cookies are the same kind of cookies as the first party cookies that are used for digital shopping carts. But if a user first visits www.news.com, and then visits www.sports.com, an ad network that serves advertising on both sites can read its own cookies. By reading its cookies, an ad network can track internet users over all websites where it serves advertising, and can compile a list of websites somebody visits. “Cookies are used in behavioural advertising to identify users who share a particular interest so that they can be served more relevant adverts,” explains the Interactive Advertising Bureau UK.[133] Tracking people over various websites is sometimes called cross-domain tracking. Tracking within one website is also possible. An online store such as Amazon can recommend

[132] This study doesn’t use “third party” in the sense of the Data Protection Directive, which defines “third party” in article 2(f).

[133] Interactive Advertising Bureau of the United Kingdom 2009, p. 4.

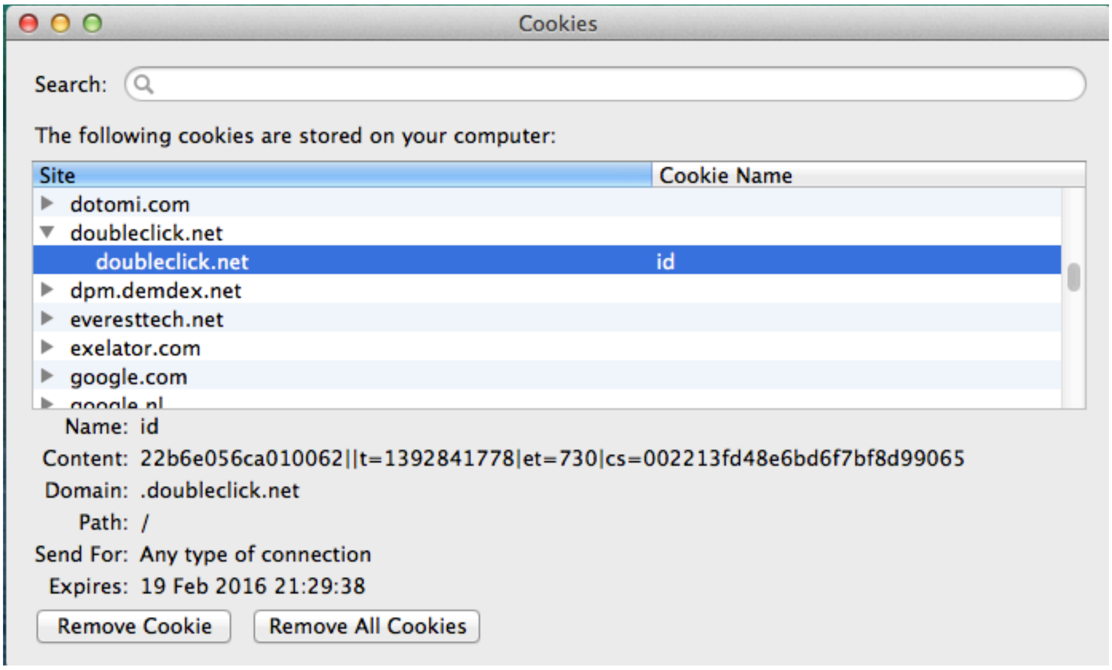
books based on a visitor's earlier browsing behaviour within the site.[134] This can be called "on site" or "first party" behavioural targeting.

The distinction between first party cookies and third party cookies is somewhat fuzzy. Firms can also use first party cookies for cross-domain tracking. For instance, firms can synchronise their own cookies with those of other firms. This way, a cookie that was installed as a first party cookie, can be used for cross-domain tracking.[135]

[134] Amazon has an ad network (Amazon 2014). If Amazon used data gathered through its ad network for its recommendations, this wouldn't be first party behavioural targeting.

[135] See section 6 of this chapter. See also Krux 2010; Tene Polonetsy 2012, p. 7; Castelluccia et al. 2013; Hoepman 2013.

Illustration 1. An example of a cookie



The following cookies are stored on your computer:

Site	Cookie Name
▶ dotomi.com	
▼ doubleclick.net	
doubleclick.net	id
▶ dpm.demdex.net	
▶ everesttech.net	
▶ exelator.com	
▶ google.com	
▶ google.nl	

Name: id
Content: 22b6e056ca010062||t=1392841778|et=730|cs=002213fd48e6bd6f7bf8d99065
Domain: .doubleclick.net
Path: /
Send For: Any type of connection
Expires: 19 Feb 2016 21:29:38

Name

In this case, the name of the cookie is “ID.”

Content

The content is the unique number of the cookie. A firm doesn’t have to allocate a unique number to a cookie. For instance, a cookie that is used to remember a website’s laguage setting could say “en” or “nl.”

Domain

The cookie is sent to the computer from the domain .doubleclick.net. In principle, only servers from the domain .doubleclick.net can access the cookie. In practice, firms have found ways to work around this.

Path

In this case, the path is set to "/". In short, this implies that the cookie can be accessed from, for instance, doubleclick.net/a, and from doubleclick.net/b.[136]

Send for

In this case, the cookie says "any type of connection." This means the cookie can be sent over an unsecured internet connection. Some cookies say "secure." That means they can only be sent over an encrypted connection, which would make it harder for other parties to intercept and read the cookies.

Expires

The cookie expires on 16 February 2016. (The screenshot was made on 19 February 2014.) However, the cookie and the expiry date can be renewed when the user encounters DoubleClick ads on the web. If no expiry date is set for a cookie, the cookie is deleted at the end of the session: a session cookie.[137]

Browsers

In 1995 the Internet Engineering Task Force (IETF), an international standards body, also started discussing ways to solve the problem of statelessness.[138] For a while the IETF considered more privacy-friendly standards than the Netscape cookie standard. But as the popular Netscape browser already supported cookies, the IETF rejected the

[136] Internet Engineering Task Force 2000, RFC 2965, article 3.2.2.

[137] Internet Engineering Task Force 2000, RFC 2965, article 3.2.2.

[138] The IETF's website says "the mission of the IETF is to make the Internet work better by producing high quality, relevant technical documents that influence the way people design, use, and manage the Internet" (Internet Engineering Task Force website).

idea of developing an alternative from scratch. Therefore the IETF set out to build on Montulli's work and to improve the Netscape standard.[139]

In 1996, the IETF started worrying about a "potential privacy threat in 'third party cookies'".[140] The IETF feared that cookies could be used to track people around the web. Therefore, the IETF wanted browsers to block third party cookies by default. "We added wording to the specification that either outright prohibits a browser from accepting third-party cookies ('cookies in unverifiable transactions'), or that permits a browser to accept them, provided they are controlled by a user-controlled option whose default value is to reject them."[141] Kristol, one of the authors of the IETF standard, says that while the IETF saw the theoretical possibility for cross-domain tracking, it didn't realise that ad networks were already doing this:[142]

> Strangely enough, when we added the words about "unverifiable transactions" [i.e. third party cookies] to the [the draft for the standard], our direct motivation was not advertising networks (which at best we were only dimly aware of at that time). Instead, [IETF member] Koen Holtman had independently discovered the theoretical potential to use third-party cookies for profiling and persuaded members of the subgroup that Europeans, at least, would be very troubled by the potential abuse of privacy they could promote.[143]

Meanwhile, the marketing industry had realised the potential of cookies. Trade publication AdAge discussed the usefulness of cookies in 1996. "Ever since the Web gained prominence as a commercial medium, marketers and publishers have

[139] Kesan & Shah 2003, p. 300-304.
[140] Kristol 2001, p. 159-160.
[141] Kristol 2001, p. 160.
[142] Kristol 2001, p. 166.
[143] Kristol 2001, p. 180.

demanded some way to understand how users move through their sites. Enter the cookie, technology developed by Netscape Communications Corp."[144]

The IETF released a standard in 1997. The standard "strongly encourages" browsers not to allow third parties to set cookies without the user's consent.[145] "A user agent should make every attempt to prevent the sharing of session information between hosts that are in different domains. Embedded or inlined objects [such as ads served by third parties] may cause particularly severe privacy problems if they can be used to share cookies between disparate hosts."[146] IETF's 1997 standard was met with hostility by the online marketing industry. Ad networks feared for their business model. One of the founders of DoubleClick, an early ad network, said that a default setting that doesn't allow third party cookies "is basically equivalent to not allowing them at all, because 99% of the population will see no reason to change the default." [147] Some firms said that large parts of the web are dependent on advertising.[148] Similar arguments are still used in current discussions about the regulation of cookies and behavioural targeting.

Kristol, who worked on the IETF standard, expected that browser vendors would implement privacy-friendly default settings. But the popular browser vendors, Microsoft and Netscape, basically ignored the 1997 standard and chose to allow third party cookies by default:[149]

[144] Carmichael 1996.

[145] Internet Engineering Task Force, RFC 2109, 1997, article 8.3. This document is a "request for comments", rather than a definitive standard.

[146] Internet Engineering Task Force, RFC 2109, 1997, article 8.3.

[147] Merriman 1997. DoubleClick was acquired by Google in 2007 (Google Investor Relations 2007). See on the power of default options chapter 7, section 4.

[148] Kristol 2001, p. 188. See on the economics of online advertising also chapter 7, section 2.

[149] In 2001 the IETF released a revised standard, RFC 2965. Again the standard emphasised that browser vendors should enable informed consent for third party cookies, and again browser vendors didn't follow the standard. Kristol notes that the paying customers for the major browser vendors weren't the browser users, but firms who profited, directly or indirectly, from third party tracking (Kristol 2001, p. 169-170). See regarding browsers also chapter 8, section 5 (on Do Not Track).

> We chose the default setting for third-party cookies because we felt it served the privacy expectations of users, especially European users, who, we inferred from European Union recommendations, might have high expectations. (…) Surely, we reasoned, [browser] vendors would choose to take such concerns into account for all users. Evidently we reasoned wrong. Vendors have steadfastly supported the advertising industry, leaving third-party cookies enabled by default.[150]

In 2014, most browsers still allow third party cookies by default. Perhaps this can partly be explained by the fact that most browser vendors have affiliated companies that carry out behavioural targeting and use third party cookies. In early 2013, Mozilla said it considered having its Firefox browser block third party cookies by default.[151] People would thereby have to change the browser settings to allow ad networks to track them.[152] The Interactive Advertising Bureau Europe was not amused, stating that "[t]he new Mozilla setting denies consumer choice, undermines industry efforts to responsibly manage control. Plus Mozilla threatens to completely undermine ad-funded content on the internet."[153] Later in 2013, Mozilla backtracked on its plans.[154]

Cookies don't offer a perfect tracking mechanism, because they identify a browser. A computer with two browsers installed (say Firefox and Safari) has two separate collections of cookies. If several people use the same browser on a computer, a cookie enables a website publisher to recognise the browser, rather than a person.[155]

[150] Kristol 2001, p. 166. The EU recommendation that Kristol refers to is: Article 29 Working Party 1999, WP 17, which said: "[c]ookies should, by default, not be sent or stored" (p. 3).

[151] Mozilla isn't in the behavioural targeting business. Mozilla does receive money from Google, which does behavioural targeting (see Mozilla blog 2011).

[152] Fowler 2013.

[153] Interactive Advertising Bureau Europe 2013. The remark that Mozilla undermines use control probably refers to the fact that the marketing industry offers people the possibility to opt out of receiving targeted advertising, but this system uses third party cookies. Hence, a browser that blocks third party cookies could be said to hinder the industry's opt-out system. See on the industry's opt-out system chapter 6, section 3, and chapter 8, section 5.

[154] Temple 2013b.

[155] This would be different if the users have separate accounts on the computer.

Furthermore, cookies rarely work in smart phone apps, and some browsers for smart phones block third party cookies by default. For instance, the Safari browser blocks third party cookies, which makes it harder for ad networks to track people's browsing behaviour on Apple devices.[156]

Beyond cookies

While many firms still use cookies for behavioural targeting, there are other ways to collect data for behavioural targeting. For instance, web beacons, or web bugs, are invisible elements on a web page or in an email message. Website publishers, or third parties such as ad networks, can operate a beacon. The firm that uses a beacon can see whether the web page has been visited or whether the email message has been opened or forwarded. Through a beacon, firms can set and read cookies as well. Beacons in emails can also be used to tie an email address to a cookie-based profile.[157]

People who want to avoid being tracked on the web can block or delete third party cookies. It has been estimated that 20-25 % of all internet users delete third party cookies.[158] This doesn't mean that people manually delete or block cookies. Anti-virus software sometimes deletes third party cookies. And as noted, Apple's Safari browser blocks third party cookies by default.

But some firms work around such browser settings.[159] For instance, Google bypassed the settings of the Safari browser.[160] There are many ways for firms to circumvent cookie deletion. In 2009 researchers found that firms used "flash cookies" for tracking.[161] Flash cookies are harder to delete than conventional cookies. Flash

[156] Felten 2012.

[157] Kaushik 2007, p. 28-30.

[158] Kaushik 2009, p. 129.

[159] Krishnamurthy & Wills 2009.

[160] Felten 2012; Mayer 2012. In the US, Google paid 22.5 million dollar to the Federal Trade Commission to settle charges it misrepresented privacy assurances to Safari users (FTC 2012, with further references). Also in the US, Google entered a 17 million dollar settlement agreement with multiple states in 2013 (Schneiderman 2013). At the time of writing, there's an on-going court case in the United Kingdom regarding the same matter (See High Court 16 January 2014, Vidal-Hall & Ors v Google Inc [2014] EWHC 13 (QB))

[161] Gomez et al 2009. See also Cranor & McDonald 2011; Ayenson et al. 2011; Hoofnagle et al. 2012.

cookies were placed through more than half of the 100 most popular websites in the United States.[162] European firms have used them as well.[163] Some firms use flash cookies to reinstall, or "re-spawn", regular cookies that were deleted by the user.[164] Trade publication Mediapost wrote in 2009 about a firm: "[w]hen Tatto began to develop its core behavioral frameworks and algorithms, it believed Flash cookies would remain the best way to slow the ability of consumers to delete cookies from their computers."[165] In sum, people deleted third party cookies to protect their privacy, and many firms re-installed those tracking cookies, on purpose, to circumvent people's privacy preferences.

There are more ways to re-install third party cookies that users have deleted. Computer researcher Kamkar shows that an identifier can be placed in fourteen different locations on a computer. He invented the "evercookie" that is stored in all these locations. It's therefore difficult to delete. The evercookie makes it possible to track an internet user when he or she uses different browsers on one device.[166] Identifiers that are used for purposes similar to third party cookies are sometimes called super cookies or zombie cookies.[167] "The entire point of new tracking methods," conclude Hoofnagle et al., "seems to be to ensure that users are ignorant of them."[168]

Another way to track people is by passive device fingerprinting. This technique involves recognising a device by looking at information it transmits, without first placing a cookie or similar identifier. A computer's browser can be recognised by looking at characteristics such as its settings, plug-ins and installed fonts. A device fingerprint is "a set of system attributes that, for each device, take a combination of

[162] Soltani et al. 2009.

[163] Helberger et al. 2011. See also Helberger et al. 2012.

[164] Soltani et al. 2009.

[165] Sullivan 2009.

[166] Kamkar 2010. See also Ayenson et al. 2011.

[167] Olsen 2011.

[168] Hoofnagle et al. 2012, p. 291. Hoofnagle adds about tracking: "in recent years, the methods have started to look more like computer hacking" (quoted in Temple 2011).

values that is, with high likelihood, unique, and can thus function as a device identifier.”[169] Researchers have fingerprinted smart phones by looking at the accelerometer, the sensor that measures vibration or acceleration.[170] Some firms use device fingerprinting for behavioural targeting. One firm claims to have fingerprinted 1.5 billion devices.[171] While some savvy users may know how to delete flash cookies and other identifiers, it’s very difficult to prevent one’s device being recognised by its fingerprint.[172]

People’s behaviour can also be tracked by installing software on their devices. Such software is called adware if it displays advertising. If people don’t like adware, they tend to call it spyware.[173] Adware is usually bundled with software installed by a user, such as file sharing software,[174] a music player[175] or a browser toolbar.[176] A firm called Flurry offers analytics software that app developers can include in their apps. Flurry’s analytics software is installed on over 1.4 billion mobile devices. Flurry also enables advertisers to target mobile users. In 2014, Yahoo announced that it would acquire Flurry.[177]

Deep packet inspection takes a different approach than the above-mentioned technologies. Deep packet inspection entails opening the digital packets that are sent over the internet, to look at the contents.[178] To illustrate, a firm called Phorm contracted with internet access providers to inspect their customers’ internet traffic. In

[169] Acar et al. 2013, p. 1. See generally on device fingerprinting Eckersley 2010; Joosen et al. 2013.
[170] Temple 2013a; Dey at al. 2014.
[171] Iovation 2013.
[172] Acar et al. 2013. See on device fingerprinting and EU law chapter 8, section 4.
[173] See on spyware and adware Federal Trade Commission 2005, p. 3-4.
[174] The popular file sharing software Kazaa was bundled with adware by 121 Media, which would later change its name to Phorm. McStay 2011, p. 20.
[175] Realplayer included spyware: Smith 1999. SONY included spyware on music CDs (Russinovich 2005; Federal Trade Commission 2007).
[176] For example, the firm Dollarrevenue enticed people to install a toolbar that also collected information. The Dutch telecommunications regulator fined the company one million euro based on the Dutch implementation of article 5(3) of the 2002 e-Privacy Directive, but the fine was overturned in appeal (College van Beroep voor het bedrijfsleven (Trade and Industry Appeals Tribunal), 20 June 2013, ECLI:NL:CBB:2013:CA3716 (Dollarrevenue/Autoriteit Consument en Markt). See in English: Libbenga 2007).
[177] Yahoo 2014 (Flurry).
[178] See generally Asghari et al. 2012; Kuehn & Mueller 2012; Kuehn 2013; Parsons 2013.

2006 a large access provider in the United Kingdom did tests with Phorm, without informing its subscribers. After media attention and parliamentary hearings, English access providers severed their business ties with Phorm. Later Phorm focused on other regions, such as South America and Asia.[179] Mobile operators can use deep packet inspection for behavioural targeting as well.[180] Deep packet inspection enables firms to access more data than web browsing behaviour. For instance, a firm that uses deep packet inspection can read the contents of email messages.[181]

Behavioural targeting isn't limited to the world wide web. For instance, providers of smart phone apps often enable ad networks to do behavioural targeting. Apps make use of the internet, but not necessarily of the web.[182] Many types of firms are interested in behavioural targeting income. For example, Akamai, an internet infrastructure provider that can see up to 30% of all internet traffic, is reported to inspect traffic for behavioural targeting.[183] In 2013, a Dutch firm in the smart TV business was found to track people's viewing behaviour. The firm had plans to use the data for behavioural targeting.[184]

Recent developments

In around 2007 the online marketing industry had recovered from the Dotcom crash of 2000. Since then, the online marketing industry is becoming increasingly centralised. Scale is important for behavioural targeting.[185] An ad network that can follow people over only a dozen websites may not be able to compile profiles that are detailed enough to improve the click-through rate on ads. Research shows that an increasingly small number of parties collects increasing amounts of data.[186] Large players such as Google, Yahoo, Microsoft and Facebook often buy smaller marketing

[179] See on Phorm McStay 2011, p. 15-42; Bernal 2011; European Commission 2009. See also chapter 6, section 3.
[180] Center for Democracy & Technology 2013, p. 6; Cisco 2014. See also Verizon 2014.
[181] This wouldn't work if the emails were encrypted.
[182] Berners-Lee 2010.
[183] Angwin 2010.
[184] College bescherming persoonsgegevens (Dutch DPA) 2013 (TP Vision).
[185] Brown et al. 2010, p. 74; Evans 2008; Evans 2009.
[186] Krishnamurthy & Wills 2009.

firms.[187] In 2012, 70% of all online advertising revenue in the United States went to the top 10 marketing firms, according to a report by the Interactive Advertising Bureau.[188] 89% of the revenue went to the top 50.[189] Another report says five firms, Facebook, Google, Yahoo, Microsoft and AOL, collected 51% of all income from display advertising in the US in 2013.[190] In 2009, that share was 38%.[191] By one estimate, Facebook and Google accounted for two thirds of all mobile advertising income worldwide in 2013.[192] In sum, there's increasing consolidation in the online marketing industry.

In autumn 2013, Microsoft and Google presented plans for their own proprietary tracking identifiers.[193] Apple already had a similar technology in place.[194] Such developments could lead to less competition in the behavioural targeting business. Rotenberg warns that people would have less control than with cookies: "the problem is about to get much worse – tracking techniques will become more deeply embedded and a much smaller number of companies will control advertising data."[195] For example, a smart phone manufacturer could decide to block tracking technologies of competitors on its phones. If an advertiser wanted to reach users of such phones, it couldn't choose any ad network, but would have to work with the phone developer.[196]

Currently behavioural targeting happens mostly when people use a computer or a smart phone. But the borders between offline and online are melting away.[197] Phrases such as ubiquitous computing, the Internet of Things, and ambient intelligence have

[187] Evans 2009; Angwin 2014, p. 31.

[188] Interactive Advertising Bureau 2013, p. 11.

[189] Interactive Advertising Bureau 2013, p. 11.

[190] Pew Research Center 2014. See also Pew Research Center's Project for Excellence in Journalism 2013.

[191] Pew Research Center 2014. See also Pew Research Center's Project for Excellence in Journalism 2013.

[192] Emarketer 2014. Pew Research Center 2014 says "nearly three quarters (73%) of (…) mobile display dollars [in the US] are collected by five companies – Facebook, Google, Pandora, Twitter and Apple."

[193] Soltani 2013 (about Google); Peterson 2013 (about Microsoft).

[194] Arnott 2013.

[195] Quoted in Tate 2013.

[196] The Interactive Advertising Bureau is actively discussing the future of behavioural targeting. For instance, it has a working group examining "privacy and tracking in a post-cookie world" (Interactive Advertising Bureau United States 2014).

[197] Hildebrandt 2011, p. 11. See also Greenfield 2006.

been used to describe – or promote – such developments.[198] When the new version of IP addresses is implemented (IPV6), there will be so many IP addresses that every object could have its own IP address.[199] If objects are connected to the internet, firms could use the data processed through those objects for behavioural targeting.[200] For example, a fridge that's connected to the internet could order groceries. Firms could analyse consumption patterns for marketing purposes.[201]

Some recent developments remind one of behavioural targeting in the physical space, like in the film the Minority Report.[202] For instance, an Italian firm sells mannequins with built-in cameras. The firm's website says that the mannequins "would make it possible to 'observe' who is attracted by your windows and reveal important details about your customers: age range; gender; race; number of people and time spent."[203] A drinks machine in Japan uses a camera to estimate age and gender of the user, to recommend drinks.[204] There are billboards with facial recognition technology that adapt their images to the people looking at the billboard.[205] One firm summarises: "a few years from now, we and other companies could be serving ads and other content on refrigerators, car dashboards, thermostats, glasses, and watches, to name just a few possibilities."[206]

[198] "Ubiquitous computing has as its goal the nonintrusive availability of computers throughout the physical environment, virtually, if not effectively, invisible for the user" (Weiser 1993, p. 71). The internet of things can be described as "a dynamic global network infrastructure with self configuring capabilities based on standard and interoperable communication protocols where physical and virtual "things" have identities, physical attributes, and virtual personalities and use intelligent interfaces, and are seamlessly integrated into the information network." (Bassi et al. 2011). Ambient intelligence refers to "digital environments in which the electronics are sensitive to people's needs, personalized to their requirements, anticipatory of their behavior and responsive to their presence" (Philips Research 2014; see also Van Den Berg 2009).

[199] See Internet Engineering Task Force 1995, RFC 1883.

[200] An article in the Pervasive Computing Journal describes possibilities for targeted advertising in a ubiquitous computing environment, and calls ubiquitous advertising "the killer application for the 21st century" (Krumm 2011).

[201] See Calo 2013a.

[202] Spielberg 2002.

[203] Almax 2012.

[204] Lies 2010.

[205] Chen 2012.

[206] Google 2013 (letter to United States Securities and Exchange Commission).

2.3 Phase 1, data collection

As previously noted, this study analyses behavioural targeting by distinguishing five phases: (1) data collection, (2) data storage, (3) data analysis, (4) data disclosure, and (5) targeted advertising. The distinction in five phases is a tool to analyse the behavioural targeting process. The distinction helps when analysing privacy problems and when applying data protection law in later chapters. The phases don't suggest a chronological description of the behavioural targeting process. Different phases overlap. For instance, selling data to another firm falls within phase 4, data disclosure. But the buyer that obtains data is in phase 1, data collection.

During the first phase of behavioural targeting, firms collect information about people's online behaviour. People's behaviour is monitored, or, as it is often called: "tracked."[207] Slightly adapting a description by the International Working Group on Data Protection in Telecommunications,[208] tracking could be described as collecting data on user activity from a computer or other device while using the internet in order to combine and analyse the data for commercial and other purposes.[209] This study uses the word "track" in a common, non-technical sense.[210]

Data collection for behavioural targeting happens on a large scale, and ad networks have a wide reach. For instance, major news websites such as the New York Times,

[207] In this study, the use of the word "monitoring" isn't meant to have a particular legal meaning. Article 3(2)(b) of the European Commission proposal for a Data Protection Regulation aims to make the Regulation applicable to non-EU firms that "monitor [the] behaviour" of data subjects residing in the Union.

[208] This "Berlin Group" was founded in 1983 and consists of representatives from Data Protection Authorities and other bodies of national public administrations, international organisations and scientists from all over the world.

[209] International Working Group on Data Protection in Telecommunications (Berlin Group) 2013, p. 1. The original definition is as follows: "the collection, analysis and application of data on user activity from a computer or device while using various services of the Information Society (hereinafter: the Web) in order to combine and analyze it for different purposes, from charitable and philanthropic to commercial. We consider various forms of market research to fall within this definition of Web Tracking, for example outreach measurement (the degree to which users are served with ads across the Web), engagement measurement (the degree to which users interact with services across the Web) and audience measurement (the degree to which micro profiles can be derived from users interacting with services across the Web)" (internal footnotes omitted). See for a similar definition Van Eijk 2012.

[210] See chapter 8, section 5 about the meaning of "tracking" in the context of discussions on the Do Not Track standard.

The Guardian, and BBC news allow ad networks to track their visitors.[211] In 2009 Gomez et al. analysed 400,000 websites and found that Google would be able to track people's browsing behaviour on 88% of the tested websites.[212] In 2010, 49 out of the 50 most popular American websites used tracking technologies.[213] Hoofnagle & Good found that in October 2012, a visit to the most 100 popular websites led to receiving 5493 third party cookies, from 457 different third parties. 21 of the most popular 100 sites placed more than 100 cookies. Various kinds of "super cookies" were placed through the top 100 websites as well. Moreover, the researchers found a trend towards more tracking when compared with an earlier test.[214]

Firms can collect detailed information about people's online activities, based on, for instance, what people read, what videos they watch, what they search for, and what they post on social network sites. Firms can collect up-to-date location data of users' mobile devices, data that people submit to websites themselves, and many other types of data. A 2010 industry report discusses some of the data that are collected for advertising:

> Every Web page's individual views, every word typed in a search query box (also known as the "database of consumer intentions"), every video download, and even every word in an e-mail may create one more data point that a marketer can

[211] On 23 February 2014, I found multiple third parties on all three websites, using Ghostery (Ghostery 2014). Ghostery is a browser plug-in which enables the user to detect and block third party tracking on websites.

[212] Gomez et al. 2009, p. 27. However, the researchers noted in 2009 that they "are not claiming that Google aggregates information from each of these trackers into a central database, though it does possess the capability to do so. It appears that they [Google] strive to keep data in silos" (p. 27). The Dutch Data Protection Authority found Google DoubleClick ads on more than 20%, and Google Analytics on more than 65% of the 8000 most popular websites in the Netherlands (College bescherming persoonsgegevens 2013 (Google), p. 12-13).

[213] Angwin 2010. The tracking-free website was Wikipedia.

[214] Hoofnagle & Good 2012.

> leverage and use to more precisely target the audience with customized media placement and messaging.[215]

Schedule 2.3 below gives an overview of the kinds of data that can be collected for behavioural targeting. Many categories in the schedule are adapted from a report on the future of advertising by Brown et al.[216] The categories serve as illustrations and sometimes overlap. Some categories concern the content of data; other categories consider the way in which data are captured.

In 2010, Brown et al. suggested that in the future, information about people's psychological and physical state might be used for targeted advertising as well.[217] Some game computers measure the player's heart rate (an example of physical state data), but currently this information isn't used for advertising.[218] In 2013, at least one firm enables advertisers to target people who play computer games with ads during times such as "congratulatory moments", or "moments of rescue."[219]

[215] Landry et al. 2010, p. 1. The report borrows the phrase "database of intentions" from Battelle 2003.
[216] Brown et al. 2010, p. 30-33.
[217] Brown et al. 2010, p. 39.
[218] Brown et al. 2010, p. 32.
[219] MediaBrix.

Schedule 2.3. Types of data processed for behavioural targeting

Web browsing data

A simple version of behavioural targeting concerns the collection of browsing behaviour, by an ad network for example. The data can reveal a lot about a person's interests. Information on a person's surfing behaviour can be seen as a category of media consumption data.

Media consumption data

Behavioural targeting can also use other types of media consumption data. For instance, a firm that offers video content on the web or smart TV could register what a person watches.[220] In some cases, software to play music or video files sent information back to the vendor.

Search data

Major search engine providers, such as Bing and Google, store all search queries of their users. The providers personalise the search results based on earlier behaviour of the user. The search queries can be used for behavioural targeting.

[220] Brown et al. 2010, p. 31.

Other intentional data

Search data can be seen as a category of intentional data: information that shows people's intentions. Firms can also infer intentional data in other ways. For instance, a person who uses an online mortgage calculator might be interested in obtaining a new mortgage.[221] And users of price comparison sites are likely to be interested in buying the product of which they compare prices.

Transaction data and pre-transaction data

Transaction data relate to what people have bought or rented.[222] Online shops can use such data for behavioural targeting. Banks and credit card firms have access to transaction data as well, but in Europe they don't seem to share such data for behavioural targeting.[223] An example of pre-transaction data is information about which products a person views in an online shop.

Demographic data

Demographic data concern for instance a person's gender or age. A book on database marketing gives the following examples: "age, sex, family size, family life cycle, income, occupation, education, religion, race, nationality."[224]

Psychographic data

These are data about a person's character. Lifestyle, social class, and personality are examples from marketing literature.[225]

[221] Business Wire 2012.

[222] Brown et al. 2010, p. 31.

[223] In the US, credit card companies often share data about customer purchases with direct marketers. See e.g. Dwyer v. American Express Co. 625 N.E.2d. 1351 (Ill. App. 1995).

[224] Newell 1997, p. 150.

[225] Newell 1997, p. 150.

Communication contents

People's communications can also be analysed for behavioural targeting. Some email providers analyse the contents of email messages for marketing purposes. A well-known example is Google's Gmail service.[226] Social network site providers can also analyse the contents of messages.[227]

Social data

Social data concern relationships between people.[228] People with friends that drive a Toyota may be interested in a Toyota too. Social network sites such as Facebook and LinkedIn, email service providers, and mobile operators, have access to social data.[229] Some firms automatically scan the web, searching for information about people's relationships on social network sites, or to extract information from blog post, tweets, etc.[230] Marketing firm 33Across specialises in social data, and says that it reaches "over 1.25 billion users."[231]

Self-provided data

Website publishers can ask people to provide information. It's often reasonably clear when a firm requests data for marketing purposes, for instance when a website asks for information before a visitor can download something. But sometimes people might not realise that data will be used for marketing, for example when a firm uses a game or a quiz to entice people to disclose information.[232] Search data can also be seen as a category of self-provided data.

[226] Yahoo also scans the messages in its email service for advertising (Gallagher 2012).
[227] Soltani &Valentino-DeVries 2012.
[228] Brown et al. 2010, p. 31.
[229] See for a definition of social network sites boyd & Ellison 2007.
[230] McStay 2011, p. 5. See also Gürses 2010, p. 100-101.
[231] 33 Across 2012.
[232] See e.g. College bescherming persoonsgegevens (Dutch DPA) 2009 (Advance Concepts).

Subscription data

Firms can ask people to provide information when they sign up for a service. Subscription data are a category of self-provided data.

Location data, fixed

Examples of fixed location data are an address and a ZIP code. An IP address often gives a rough indication of a computer's location. A location could give information about a person's environment, for instance whether he or she lives in a suburban area, a city centre, or a rural area.[233]

Location data, mobile

Mobile location data refer to mobile devices, such as phones or tablets. Mobile location data can show where a person is in almost real-time.[234] Various parties have access to such location data. Smart phone apps sometimes send location data to the app provider, or to ad networks.[235] Some in the industry have high hopes for advertising on mobile devices.[236] If a person's profile suggests that he or she likes Italian food, a pizzeria might advertise a deal when he or she is in the area around lunchtime. Some firms track people's movements in shops by analysing signals emitted by people's phones, such as Bluetooth and Wi-Fi signals.[237]

[233] Center for Democracy & Technology 2009, p. 16.

[234] Center for Democracy & Technology 2009, p. 16.

[235] Thurm & Iwatani Kane 2010.

[236] Peterson 2012.

[237] See Future of Privacy forum 2013.

Contextual data

Contextual data refers to data about content.[238] For instance, contextual data can concern the language and the subject matter of a web page. If a car manufacturer buys advertising space on a website about cars, that would be called contextual advertising. Ads can be matched automatically to a site's content, by having software analyse the website's text.[239] Many behavioural targeting firms aim to take the website content into account as well. A cruise operator probably doesn't want its ads to be shown next to news about a ship disaster.

Environmental data

Environmental data concern for example local conditions such as the weather.[240]

Time-related data

Many firms adapt advertising to the time of day.[241] For example, advertising for restaurants might be shown around dinnertime.

Offline data

Offline data is a catch-all phrase for data that are collected from sources other than the internet. For instance, supermarkets use loyalty card programs to collect transaction data.[242] There are various ways of tying such data to online profiles.[243] The offline/online distinction is becoming less relevant, as more devices are being connected to the internet.

[238] Brown et al. 2010, p. 32.

[239] See for instance Google Adsense 2014.

[240] Brown et al. 2010, p. 32. This study uses the phrase contextual data for data about content. Brown et al. use the phrase contextual data differently as the overarching term for data that aren't about a person but about the environment.

[241] Brown et al. 2010, p. 32. See also McStay 2010, p. 44; McStay 2011, p. 5.

[242] See Pridmore 2008.

[243] In the United Kingdom, marketing firm Yahoo has enriched online profiles with data obtained from loyalty cards (Charlton 2010).

2.4 Phase 2, data storage

In the second behavioural targeting phase, firms store the data, tied to a unique identifier such as a cookie. For instance, a profile of a person might contain a list of websites that somebody visited. Or a profile might contain a person's interest categories, such as "cooking & recipes" or "mountain & ski resorts."[244] A profile is a "set of correlated data that identifies and represents a data subject."[245] An individual profile generally refers to a single person.[246] For ease of reading, this study also uses the word "profile", rather than "individual profile." Instead of individual profile, phrases such as "data double",[247] "data shadow",[248] or "digital dossier" are also used in literature.[249]

In computer science, nameless individual profiles are referred to as pseudonymous. "A pseudonym is an identifier of a subject other than one of the subject's real names." [250] Firms using behavioural targeting often call individual profiles "anonymous", when they don't tie a name to the profiles. Group profiles don't contain information about a specific person, but about a group or a category.[251] Unlike individual profiles, group profiles can be anonymous. "Anonymity of a subject means that the subject is not distinguishable from the other subjects within a set of subjects."[252] Chapter 5 returns to the topic of pseudonymous data, and shows that data protection law generally applies to such data.

Some firms have individual profiles on hundreds of millions of people. For instance, Facebook had over 1 billion monthly active users in 2014.[253] Google says its "Display

[244] The examples taken from Google Ad Settings 2014.
[245] Hildebrandt & Backhouse 2005, p. 106.
[246] Hildebrandt & Backhouse 2005, p. 106. In the context of this study, a pseudonymous profile may contain information about multiple users of one computer.
[247] Haggerty & Ericson 2000.
[248] Garfinkel 2000, p. 70. Garfinkel says the phrase "data shadow" was coined by Alan Westin in the 1960s.
[249] Solove 2004, chapter 2.
[250] Pfitzmann & Hansen 2010, par. 9.
[251] Hildebrandt & Backhouse 2005, p. 106.
[252] Definition taken, and slightly adapted, from par. 3 and footnote 18 of Pfitzmann & Hansen 2010.
[253] Facebook says it had "1.35 billion monthly active users as of September 30, 2014" (Facebook 2014).

Network reaches 83% of unique Internet users around the world."[254] But some lesser-known firms also have information about many people, such as the Rubicon Project ("600 million monthly unique users"),[255] and AddThis ("1.7 unique users worldwide").[256]

Firms can enrich individual profiles by tying data sets together. For instance, some firms can tie a name or an email address to a profile. Providers of social network sites such as Facebook know the names of many users. An email provider that uses behavioural targeting could tie an email address to many of its profiles. If a firm knows the name behind a profile, it could use the name to add more data to the profile.[257] Behavioural targeting profiles can be detailed. The ValueClick firm tells its advertising customers: "our database stores an average of 204 attributes for 97% of all online users."[258]

Some firms tie data collected on one device to data collected on another device: "cross device targeting." Somebody who searched for a car on his or her computer might be targeted with related ads on his or her phone.[259] If somebody uses the same email or social network account on both devices it's easy to link the devices to one person. Another way to link a person to multiple devices is looking at the IP address. If somebody uses his or her smart phone and laptop at home, both devices may use the same IP address every night. It's also possible to follow somebody while he or she uses various devices by analysing that person's browsing behaviour. "Users have very specific browsing patterns," explains Hoepman. "Everyone has his personal list of favourite websites (recall that your top five favourite movies are quite identifying). In

[254] Google AdWords 2014. "The Display Network is a collection of partner websites and specific Google websites – including Google Finance, Gmail, Blogger, and YouTube – that show AdWords ads. This network also includes mobile sites and apps."

[255] Rubicon Project.

[256] AddThis 2014.

[257] See e.g. Charlton 2010.

[258] Elsewhere on the website, ValueClick adds that the number concerns all online users "tracked by ComScore in the US" (ValueClick). In 2014 the firm merged with other firms to form Conversant Media, which claims to be able to target 263 million people (Conversant 2014).

[259] See e.g. Harper 2011.

fact, your browsing history becomes unique after a few visited websites. And people read their favourites in a fixed order."[260] A firm called Drawbridge uses this technique, and claims it has connected "over 1 billion customers across devices."[261] The firm claims it can also recognise different users of one device by analysing their behaviour.[262]

Some firms add data gathered offline to online profiles.[263] Even when the name of the person behind a profile isn't known, it may be possible to do this. For instance, a firm that knows in which neighbourhood a computer's IP address is located, could add information about the average housing price in that neighbourhood to a profile. One American firm uses the location of IP addresses to infer "120 demographic variables" about people, including information such as "life stage, affluence, home ownership, auto interests, political affiliation, and social connectivity."[264]

[260] Hoepman 2014. For a majority of users, the browsing history is unique (Castelluccia et al. 2013a). Regarding identifying people by their favourite movies, Hoepman refers to Narayanan & Shmatikov 2008. See also Sivaramakrishna 2012; Cain, Miller & Sengupta 2013.

[261] Drawbridge 2014.

[262] Cain Miller & Sengupta 2013.

[263] Combining offline and online data is sometimes called "onboarding" (Federal Trade Commission 2014, p. 27).

[264] Semcasting.

Schedule 2.4. Examples of individual profiles

- The person with ID xyz on his or her computer, that uses IP address 146.50.68.36, visited the following 2000 websites:

(1) hockey.com,

(2) basketball.com,

(3) soccer.com,

(…)

(1998) redrunningshoes.com,

(1999) blackrunningshoes.com,

(2000) bluerunningshoes.com."

- The person with ID xyz on his or her computer likes sports and running shoes.

2.5 Phase 3, data analysis

In phase 3, firms analyse the data. Somebody who reads a lot of legal blogs could be profiled as a person who is interested in the law. A firm may or may not delete the data it has collected about somebody’s online behaviour after deducing that person’s interests.[265]

Data can be analysed in various ways. For instance, data mining is the process of finding new information in data sets. Data mining can be described as “the nontrivial extraction of implicit, previously unknown, and potentially useful information from data.”[266] Data mining doesn’t have to begin with a hypothesis. Software is used to analyse the data in order to find correlations, and these correlations can be unexpected.[267] One firm found that customers who buy certain accessories for their cars often default on their credit. As the New York Times reports, “[a]nyone who purchased a chrome-skull car accessory or a ‘Mega Thruster Exhaust System’ was pretty likely to miss paying his bill eventually.”[268] Conversely, people who bought felt pads for under the feet of their furniture to prevent scratches on the floor, almost always repaid their credit without problems.[269]

Firms may also construct predictive models.[270] For example, a firm might find the following model. *If a person visits website A, B, C and D, there’s a 0.4 % chance that the person clicks on ads for product E.* Siegel defines a predictive model as follows.

> A mechanism that predicts a behavior of an individual, such as click, buy, lie, or die. It takes characteristics of the individual as input, and provides a *predictive score* as output. The higher the score, the more likely it is that the individual will exhibit the behavior.[271]

[265] Schunter & Swire 2013, p. 10-16.

[266] Frawley et al 1992, p. 58. See on data mining Custers 2004; Barocas 2010; Barocas 2014; Zarsky et al. 2013.

[267] Siegel 2013, p. 98. See in more detail: Barocas 2014, p. 54-56.

[268] Duhigg 2009. See also Brunton & Nissenbaum 2011.

[269] Duhigg 2009.

[270] Predictive models are roughly comparable with non-distributive group profiles (see Hildebrandt 2008).

[271] Siegel 2013, p. 26 (emphasis original). Predictive models can also be used to predict when somebody lies, or to predict how old somebody is likely to become: to predict behaviour “such as click, buy, lie, or die.”

Siegel gives an example of a predictive model that was used for online advertising. A publisher of a website where people could search for scholarships wanted to improve the click-through rate on the site's ads. The following predictive model was found.

```
IF the individual

    is still in high school

    AND

    expects to graduate college within three years

    AND

    indicates certain military interest

    AND

    has not been shown this ad yet
```

> THEN the probability of clicking on the ad for the Art Institute is 13.5 percent.[272]

In brief the model says: if a website visitor fits in four categories (the input), there's a 13.5 % chance that he or she clicks on an ad for the Art Institute (the output). 13.5% might not seem like a high number, but the probability that a random website visitor clicked the ad was only 2.7%.[273] Siegel says it's unclear why people who expressed an interested in the military are more likely to click on the ad. He adds that causation is irrelevant: "it's important not to assume there is a causal relationship."[274] Whether people who expressed interest in the military see the ad as relevant is of little interest

[272] Siegel 2013, p. 26.

[273] In general, click-through rates are much lower. See section 1 of this chapter.

[274] Siegel 2013, p. 27.

for the website; what is of interest is whether or not people are likely to click. Likewise, an ad network doesn't need causal relations. The goal is improving the chance that a person will click on an ad. As an aside, this implies that claims that behavioural targeting leads to "more relevant" ads should perhaps be taken with a grain of salt.[275]

By definition, predictive models aren't always accurate when applied to individuals. To illustrate, when a predictive model says that there's a 60% chance that people who visit sports websites also like running shoes, it's still possible that a person who visits sports websites doesn't like running shoes. And a person with an IP address from a neighbourhood with expensive real estate might be a poor student, renting a small room in an expensive villa. A book on data mining and marketing explains that predictions don't have to be accurate to increase profit.

> The fact is that, to take a typical application of data mining to direct marketing, *95 percent of the people picked by data mining to be likely responders to an offer will not respond*. In other words, at the level of individual consumers, *data mining predictions are nearly always wrong*. (…)
>
> The reason that data mining is valuable, despite being so very inaccurate, is that although only 5 percent of the people predicted to respond actually do so, that may be a significantly higher number than would have responded if no data mining model had been used. The ability of data mining to identify a population within which we can expect a 5 percent response rate, instead of the 2.5 percent response rate we could achieve

[275] Google says its behavioural targeting system makes ads "more relevant" and "more interesting" (Wojcicki 2009). See also Interactive Advertising Bureau Europe - Youronlinechoices (about).

> without data mining, makes it worthwhile from a business point of view.[276]

In short, accuracy isn't needed for behavioural targeting to be a good business decision. A firm doesn't have to predict accurately to improve profits. "Predicting better than pure guesswork, even if not accurately, delivers real value," notes Siegel.[277] Any improvement to the click-through rate is welcome. Say the chance that random internet users click on an ad for chairs is 0.1 %. An ad network could improve the click-through rate on the ad if it had the following predictive model: *If a person visits more than 10 websites about furniture every week, there's a 0.4 % chance that the person clicks on ads for chairs*. Hence, the predictive model, while not very accurate in predicting people's interests, can lead to a 400% improvement of the return on investment.

Behavioural targeting typically involves profiling. Hildebrandt offers the following definition.

> Profiling is the process of "discovering" correlations between data in databases that can be used to identify and represent a subject and/or the application of profiles (sets of correlated data) to individuate and represent a subject or to identify a subject as a member of a group or category. In the case of group profiling the subject is a group (which can be a category or a community of persons).[278]

[276] Berry & Linoff, p. 20 (emphasis original). See also Danna & Gandy 2002, p. 379.

[277] Siegel 2013, p. 11.

[278] Hildebrandt et al. 2008, p. 241, which refers to Hildebrandt 2008, p. 19. See in detail on defining profiling Bosco et al. 2013.

With profiling, information about a group of people can be applied to a person who isn't part of that group. To illustrate, the American retail store Target wanted to reach people with advertising during moments in life when they're more likely to change their shopping habits, as usually it's hard to make people change their habits. Therefore, Target wanted to know when female customers were going to give birth. "We knew that if we could identify them in their second trimester, there's a good chance we could capture them for years."[279] By analysing the shopping behaviour of customers, Target was able to construct a "pregnancy prediction" score, based on 25 products. If a woman buys those products, Target can predict with reasonable accuracy that she's pregnant.[280] Hence, Target uses data from a group of people to predict something about a person who wasn't part of that group.

Calo suggests firms might soon be able to analyse large amounts of data in order to find the characteristics and weaknesses of individuals. Is a person easier to persude with an ad in orange colours, or on rainy afternoons? "Firms will increasingly be able to trigger irrationality or vulnerability in consumers," says Calo.[281] "A firm with the resources and inclination will be in a position to surface and exploit how consumers tend to deviate from rational decision making on a previously unimaginable scale."[282] A press release of a marketing firm suggests that Calo's worries may not be completely unfounded. "New beauty study reveals days, times and occasions when US women feel least attractive."[283] The firm suggests advertising beauty products on Mondays. A Dutch firm is doing research on "persuasion profiling", which "lets you gain insight into your customer's psychological patterns (…)."[284] A firm could add to a profile what kinds of arguments convince a person to buy a product, rather than the

[279] Duhigg 2012, quoting the statistician of Target.

[280] Duhigg 2012. See on the Target case also Siegel 2013, chapter 2.

[281] Calo 2013, p. 5.

[282] Calo 2013, p. 22. See on the risk of manipulation through behavioural targeting chapter 3, section 3. See on "biases", deviations from rational decision making chapter 7, section 4.

[283] PHD Media 2013.

[284] PersuasionAPI. See also Kaptein 2011; Kaptein 2012; Kaptein & Eckles 2010; Groot 2012.

person's interests. Does he or she react to discounts, or to phrases such as "special offer, only today"?

Schedule 2.5. Examples of predictive models

- There is a 0.4% chance that a person who visits websites about consumer electronics, clicks on ads for phones.

- There's an 80% chance that a person who lives in neighbourhood X, has an income that is lower than 1500 euro a month.

2.6 Phase 4, data disclosure

The fourth behavioural targeting phase concerns data disclosure. Firms make data available to other firms. Two kinds of data disclosure can be distinguished. First, a firm might sell copies of data to other firms.[285] For example, data brokerage is a large industry in the US. Data brokers are "companies that collect consumers' personal information and resell or share that information with others."[286] Firms can buy data to tie them to online profiles. For instance, a firm called Collective enriches online profiles with off-line consumer data from more than "35 world-class data providers such as Polk, Nielsen and eXelate, integrated into profiles representing the most desirable segments of the US online audience."[287] The American firm CampaignGrid merges data from its database with registered voters with cookie-based profiles for political campaigns. The firm deletes the name from the profile after it merges the

[285] From a legal perspective, data may not be goods that can be "sold." We'll leave this issue aside.
[286] Federal Trade Commission 2014, p. 1. See also Federal Trade Commission 2013a.
[287] Collective 2011.

different data sets, and suggests that this makes the profiles "de-identified."[288] However, chapter 5 shows that European data protection law usually applies to pseudonymous individual profiles. Deleting the name from a profile is not enough to remain outside the scope of data protection law, and is not enough to make information anonymous.[289]

A second type of data disclosure doesn't involve selling copies of the data. For example, an ad network can allow an advertiser to target individuals based on their characteristics. The ad network shows the ad on behalf of the advertiser. The advertiser usually doesn't receive a copy of the data in a profile. This type of data disclosure could be seen as a modern version of list rental. With list rental, a list broker sends leaflets to a set of people, based on what it knows about those people. The advertiser doesn't receive a copy of the list .[290]

Another example of data disclosure is cookie matching, or cookie synching, "linking the profiles of a single user in databases of two independent companies."[291] Cookie synching happens routinely. For instance, researchers found that the cookies of Google's DoubleClick ad network are synched with cookies of at least 125 other firms.[292] Depending on the design of the system, cookie synching may or may not involve disclosing copies of data to others.

Real time bidding

Ad networks can bid on automated auctions for the chance to show an ad to a person, a process which is referred to as "real time bidding", "audience selling", or "audience buying."[293] Ad exchanges are automated market places where advertisers can trade

[288] CampaignGrid 2012. See also Kreiss 2012.
[289] See section 4 of this chapter.
[290] Under data protection law, list rental should probably be seen as a type of data disclosure. See chapter 6, section 2.
[291] Castellucia et al. 2013, p. 1.
[292] Castellucia et al. 2013, p. 7.
[293] See for example Pubmatic 2011.

with multiple ad networks in one place.[294] Ad exchanges owned by Google, Yahoo, and Microsoft are among the largest.[295] Real time bidding "creates a data market where users' browsing data are sold at auctions to advertisers."[296]

In brief, real time bidding works as follows. A website has an empty spot for a banner ad. Somebody visits the website. An ad network that works with the website recognises this person as the cookie with, for instance, number 22be6e056ca010062|lt=1392841778|cs=002213fd48e6bd6f7bf8d99065. For ease of reading, this study speaks of ID *xyz*. When the website is loaded in the user's browser, the ad network offers the empty banner spot on the advertising exchange (the auction). The ad network can include information about the person behind ID *xyz*, such as the person's inferred interests and location, and the time of day.

Other ad networks bid to reach a person who is, for instance, interested in cars, just visited a website with information about loans, and as been visiting websites with reviews of a certain car type for the past three weeks. The ad network that submits the highest bid obtains the right to target an ad to this specific group. Then, the winning bidder (for instance another ad network) can display an ad on the website for an advertiser. This process happens automatically and within a few milliseconds. (For more information on targeted advertising, see the next section, on phase 5 of the behavioural targeting process.) Researchers conclude that "user's browsing history elements are routinely being sold off for less than $0.0005."[297] Billions of such auctions take place per day.[298] "We are not buying content as a proxy for audience", explains one marketing firm. "We are just buying who the audience is."[299]

[294] Turow 2011 p. 79; Evans 2009. The Interactive Advertising Bureau provides a definition of advertising exchanges (Interactive Advertising Bureau United States 2010).
[295] Turow 2011, p. 79. In 2007, Google, Microsoft and Yahoo each acquired a firm running an ad exchange (Google 2011, p. 3).
[296] Castellucia et al. 2013, p. 14.
[297] Castellucia et al. 2013, p. 1.
[298] Econsultancy 2011, p. 6; Turow 2012, p. 69.
[299] Quoted in Singer 2012.

If a website publisher contracts with an ad network, and that ad network sells part of its inventory through an advertising exchange, the publisher doesn't always know in advance who will display the ads on its site. Therefore, sometimes publishers don't know which firms are collecting data on their websites.[300] "As a publisher we feel we've been raided by the ad industry," says the chairman of the Association of Online Publishers. "We've done site audits and been flabbergasted by how many third party cookies have been dropped on our site by commercial partners – they were stealing our data."[301] Some firms offer a service that wesbite publishers can use to monitor their own websites, to reduce such "data theft."[302]

The Interactive Advertising Bureau US claims that "virtually every publisher site, advertiser, ad network, or analytics firm collects or shares data with other parties in order to make the digital economy work."[303] Behavioural targeting can seem more complicated than it is, because firms tend to introduce new phrases, such as "data driven marketing,"[304] and "programmatic buying."[305] Notwithstanding this, the data flows behind a behaviourally targeted ad can be extremely complicated. Many different types of firms are involved in serving a behaviourally targeted ad, and many firms disclose information to each other. LUMA Partners, an investment bank for the media and technology sector, provides an infographic with an overview of the types of players involved in display advertising, which includes many types of firms, such as "demand side platforms", "agency trading desks", "data suppliers", and firms involved in "tag management", and "measurement and analytics."[306] It would go beyond the scope of this study to discuss each type.

[300] See for instance Martijn 2013, who interviews Dutch publishers who say they don't know what happens on their sites.
[301] Barnes, chairman of the Association of Online Publishers, quoted in Hall 2013.
[302] See for instance Krux 2014. See also Vascellaro 2010.
[303] Zaneis 2012.
[304] Data-Driven Marketing Institute 2014.
[305] Interactive Advertising Bureau United States 2014a
[306] Luma Partners 2014.

2.7 Phase 5, targeting

In the fifth phase of behavioural targeting, a firm targets a person with an ad, based on information about that person. Any kind of digital advertising can be based on behavioural profiles, such as display ads, ads shown by search engines, and marketing emails. Two people simultaneously visiting a website may each see a different ad, because they have different profiles. Firms can adapt ads in real time, and can serve each user a unique personalised ad.[307] The advertiser's goal is to "reach the right person with the right message at the right time."[308] A firm might also refrain from showing an ad to certain people, based on their profile.

Advertising can be defined as a "paid, mediated form of communication from an identifiable source, designed to persuade the receiver to take some action, now or in the future."[309] On the internet, the boundaries between brand advertising and direct-response advertising are blurry, because most ads enable people to click on ads to interact with advertisers.[310] The lines between behavioural targeting and other types of online advertising are blurry as well.[311] For example, nowadays ads that are shown by a search engine are often behaviourally targeted. In principle search ads don't have to be based on analysing people's behaviour over time. To illustrate, until around 2009

[307] Personalisation can be defined as the "use of information about a particular user that provides tailored or personalized services for the user" (Serino et al. 2005, p. 1). Some authors distinguish "system-initiated personalisation" from "user-initiated customisation" (Marathe & Sundar 2010, p. 300).

[308] TRUSTe (Drawbridge) 2013.

[309] Curran & Richards 2002, p. 74. The word mediated in this definition means "conveyed to an audience through print, electronics, or any method other than direct person-to-person contact." See for a EU legal definition of advertising: article 2(a) of Directive 2006/114/EC on misleading and comparative advertising. "Commercial communication" is defined in article 2(f) of the e-Commerce Directive 2000/31/EC.

[310] See McStay 2009, p. 7. Brand advertising aims at making a brand more famous, rather than at enticing the recipient to take action immediately. "Direct response advertising" is "[a]n approach to the advertising message that includes a method of response such as an address or telephone number whereby members of the audience can respond directly to the advertiser in order to purchase a product or service offered in the advertising message (…)" (American Marketing Association dictionary).

[311] Strandburg 2013, p. 99.

Google had refrained from behavioural targeting.[312] Now Google ties the profile of a searcher to the other data it has about that person.[313]

A category of behavioural targeting that is particularly notable for users is retargeting. Sometimes ads for a product appear to follow somebody around the web. Retargeting allows a firm to show potential customers personalised ads, based on earlier behaviour that the firm interprets as an intention to buy. Retargeted ads aim to remind the potential customer of a product. Google explains retargeting as follows to advertisers:

> Let's say you're a basketball team with tickets that you want to sell. You can put a piece of code on the tickets page of your website, which will let you later show relevant ticket ads (such as last minute discounts) to everyone who has visited that page, as they subsequently browse sites in the Google Content Network. In addition to your own site, you can also remarket to users who visited your YouTube brand channel or clicked your YouTube homepage ad.[314]

Retargeting is easy to notice. If somebody looks at red shoes in an online shop, and keeps seeing ads for those same shoes elsewhere on the web, it's obvious that the ads are tailored to the individual. Other kinds of behaviourally targeted ads can be harder to recognise. For somebody who visits the literature section of an online newspaper, it's not always clear whether an ad for a book is based on his or her earlier surfing behaviour or not. A behaviourally targeted ad might be mistaken for a contextual ad, or vice versa.

[312] Hoofnagle 2009.

[313] See Article 29 Working Party 2013 (Google letter), and chapter 8, section 1.

[314] Weinberg 2010.

In principle, few data are needed for retargeting, because there's no need to build a detailed profile of somebody's tastes and behaviour. A firm drops a cookie on a user's device, and the firm only needs to store the information that the person behind ID *xyz* looked at a certain product. In practice, a firm might also store the user's IP address, and the list of all websites where the firm showed the user the retargeted ad.

Behavioural targeting can also be used for political advertising. A firm gives an example of the possibilities: "targeting fathers aged 35-44 in Texas who frequent gun enthusiast websites."[315] Messages can be tailored to the profile of the recipient. In 2012, campaigners for Obama divided an email list into 26 segments, in order to be able to send each segment a different message.[316] Political behavioural targeting firm CampaignGrid claims that it reaches 90 % of American internet users. The firm enables politicians to target people with ads on LinkedIn, Facebook, and elsewhere on the web.[317] An article in the magazine *Campaigns & Elections* discusses the possibilities of digital TV for political campaigns.

> While there's plenty of potential for political campaigns in set-top box targeting, mining data from television set-top boxes and pairing it to the voter file is a good starting point this [election] cycle, according to NCC Media's Tim Kay. "It's no longer hoping you're hitting the person," says Kay, the company's political director. "Now it's about knowing whether you're hitting the person and knowing how to hit the person."[318]

[315] Retargeter Blog 2012.
[316] Judd 2012.
[317] CampaignGrid 2014.
[318] Williams 2014.

Not only ads, but also other content can be personalised. Major search engines personalise search results.[319] And two people visiting the same website at the same time may see a different front page.[320] To illustrate, Yahoo shows more than thirteen million different versions of its news page each day.[321] Yahoo shows the news selection that keeps visitors on the website for as long as possible, in order to show them more advertising. Yahoo doesn't ask visitors whether they want to receive personalised news. The line between content and ads can be fuzzy on the web. For instance, advertorials and "native ads" are ads that resemble editorial content.[322]

Some firms specialise in website personalisation. A company called Personyze says: "[s]egment your visitors in real-time and serve them personalized and optimized content based on their demographic, behavioural and historical characteristics."[323] Personalisation can be "based on demographics, keywords searched, referring affiliate website, articles read, favorite categories and more."[324] A website's design can also be adapted to the visitor, called morphing. "Morphing involves automatically matching the basic 'look and feel' of a website, not just the content, to cognitive styles."[325] Research suggests that website morphing could increase online sales with approximately 20%.[326] At present, website morphing doesn't seem to be widely used. As behavioural targeting makes it possible to show each person personalised ads and other content and services, Hildebrandt has called behavioural targeting an early example of ambient intelligence, technology that senses and anticipates people's behaviour to adapt the environment to their inferred needs.[327]

[319] See Hannak et al. 2013.

[320] Turow 2011; Pariser 2011.

[321] Yahoo 2012.

[322] See Federal Trade Commission 2013. See on the blurry line between advertising and other content Van Hoboken 2012 (chapter 10, section 3).

[323] Personyze 2014.

[324] Personyze 2014b

[325] Hauser et al. 2009, p. 202.

[326] Hauser et al. 2009.

[327] Hildebrandt 2010. See also Hildebrandt 2011, p. 12. See on ambient intelligence: Philips Research 2014; see also Van Den Berg 2009.

Behavioural targeting offers more possibilities beyond personalised advertising. For instance, firms could personalise prices based on group or individual profiles – also referred to as price discrimination.[328] A user whose profile suggests that he or she often buys expensive goods without first looking for the cheapest price online could be profiled as a "big spender."[329] A Harvard Business Review article explains that a shop could charge higher prices to some people. "Just as it's easy for customers to compare prices on the Internet, so is it easy for companies to track customers' behavior and adjust prices accordingly."[330]

It's unclear to what extent firms adapt prices to people's online profiles. Perhaps firms are hesitant to personalise prices because they fear consumer backlash.[331] However, in the US, firms have adapted credit card offers to the cookie profile of website visitors, based on a person's inferred income for instance.[332] And in 2012, Soltani et al. found that the online shop Staples charged visitors from certain areas (based on their IP address) different prices than people from other areas. This had the effect, likely unintentional, that people from high-income areas tended to pay less.[333] Opinions differ on the question of whether personalised pricing is desirable. From an economic

[328] See generally Turow 2012, p. 108-110.

[329] Bluekai 2010. Marketers can buy access to "high spenders", "suburban spenders" or "big spenders" (p. 6-8). Bluekai says the profiles are "anonymous" (Bluekai 2012).

[330] Baker et al. 2001, p. 123.

[331] The English Office of Fair Trade examined whether firms raised prices based on people's online behaviour, but didn't find any evidence. The office did find that firms offer discounts based on people's profiles some cases (Office of Fair Trading 2010; Office of Fair Trading 2012). A discount for one person is a type of price discrimination, which could also be seen as a higher price for the others.

[332] Steel & Angwin 2010.

[333] Valentino-Devries et al. 2012; Angwin 2014, p. 16.

perspective, price discrimination is a good thing, under certain assumptions.[334] On the other hand, many regard price discrimination as unfair or manipulative.[335]

2.8 Conclusion

This chapter described what behavioural targeting is, and how it works. Different factors can help to understand the rise of behavioural targeting. Technology has made behavioural targeting possible. Behavioural targeting fits into a trend of increasingly targeted advertising at ever-smaller audience segments. Furthermore, advertisers have always wanted information on how many people they reached with an ad, and on what kind of people they reach. Behavioural targeting provides such information, at the individual level.

Behavioural targeting is the monitoring of people's online behaviour in order to use the collected information to show people individually targeted advertisements. In this study, the behavioural targeting process is analysed by dividing it into five phases: (1) data collection, (2) data storage, (3) data analysis, (4) data disclosure, and (5) the use of data for targeted advertising.

In phase 1, firms collect information about people's online activities. People's behaviour is monitored, or tracked. In phase 2, firms store the information about individuals, usually tied to identifiers contained within cookies, or via similar technology. As discussed later in this study, article 5(3) of the e-Privacy Directive requires consent for the use of many tracking technologies, but some tracking

[334] In economics, price discrimination, or price differentiation, is used in a broader sense than personalised pricing: "the practice of a seller charging different prices to different customers, either for exactly the same good or for slightly different versions of the same good. (…) [P]rice differentiation includes not only charging different prices to different customers for the same product (group pricing), but also the less controversial strategies of product versioning, regional pricing, and channel pricing" (Phillips 2005, p. 74). See generally on price differentiation Phillips 2005, chapter 4; Shapiro & Varian 1999, chapter 2 and 3. See generally on price discrimination and behavioural targeting Zarsky 2002; Odlyzko 2003; Turow 2011; Calo 2013; Narayanan 2013; Strandburg 2013; p. 138-141; Odlyzko 2014; Miller 2014.

[335] For instance, in a nationally representative survey, Turow et al. 2005 "found that they [US adults] overwhelmingly object to most forms of behavioral targeting and all forms of price discrimination as ethically wrong" (p. 4). Klock 2002 argues (not focusing on behavioural targeting): "[a] sound policy would prohibit firms from charging different prices based solely on the identity of the customer" (p. 367).

technologies, such as passive device fingerprinting, may fall outside the scope of that provision.[336]

In phase 3 the data are analysed. For instance, a firm can construct a predictive model, along the following lines: if a person visits website A, B, C and D, there's a 0.5 % chance the person will click on ads for product E. For behavioural targeting to be useful, a predictive model doesn't have to be accurate when applied to an individual. Chapter 5 shows that predictive models are outside the scope of data protection law, as a predictive model doesn't refer to an identifiable person.[337]

Phase 4 concerns data disclosure. Firms make data available data to advertisers or other firms. For example, an ad network can enable advertisers to target individuals with ads, based on their behavioural profiles. Or a firm can sell copies of data to other firms. Many types of firms are involved in behavioural targeting, and the resulting data flows are complicated. The complicated data flows make it difficult to explain to people what happens to information about them (see chapter 7).[338]

In phase 5 data are used to target ads to specific individuals. Behavioural targeting enables advertisers to reach a user, wherever he or she is on the web. A website publisher often doesn't know in advance who will serve ads on its website. Firms can personalise ads and other website content for each visitor.

Website publishers can increase their income by allowing ad networks to track their visitors and to display behaviourally targeted ads. But in the long term behavioural targeting may decrease ad revenues for some website publishers. For example, an ad network doesn't have to buy expensive ad space on a large professional news website to advertise to an individual. The ad network can show an ad to that person when he or she visits an unknown website, where advertising space is cheaper. Chapter 7

[336] See chapter 6, section 4, and chapter 8, section 4.

[337] See chaoter 5, section 2.

[338] See in particular chapter 7, section 3 and 4,

returns to the topic of the economics of behavioural targeting.[339] But first we turn to the privacy implications of behavioural targeting, in the next chapter.

* * *

[339] See in particular: chapter 7, section 2.