Mobile apps can track daily habits such as exercise & lifestyle routines, and sleeping & eating patterns.

However when it comes to personal data, if they know what data is being gathered, the public can become offended, shocked and scared of what can be achieved and learnt by gathering their seemingly private information.

Why is this?

Alexander Tkaczyk-Harrison

The General Data Protection Regulation (GDPR) is a legal framework that sets guidelines for the collection and processing of personal information of individuals within the European Union (EU). Furthermore, it extends to countries outside of the EU. Companies which have operations within the EU have to sign up to the GDPR rules, along with companies who store data about EU citizens. [1]

The GDPR describes 'Personal data' as:

"any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier [etc…]" [2]

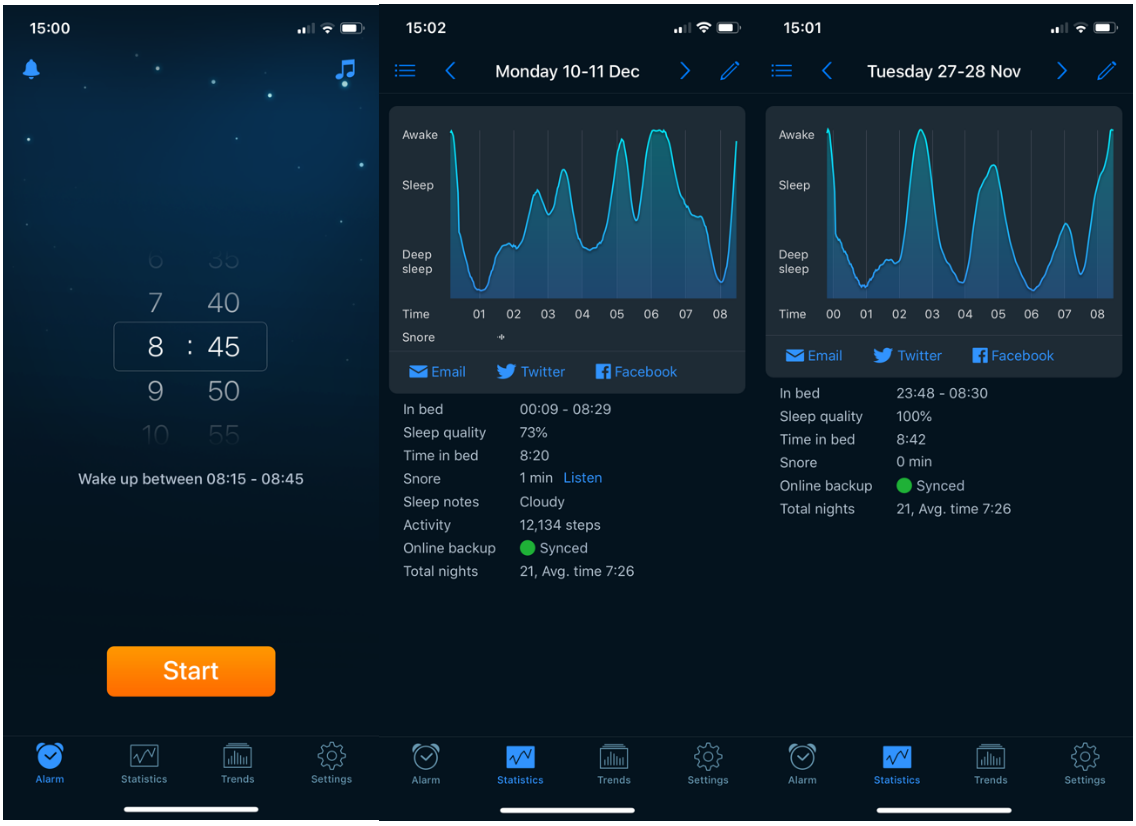
The GDPR then goes on to describe that:

"Processing shall be lawful only if and to the extent that at least one of the following applies:

(a) The data subject has given consent to the processing of his or her personal data for one or more specific purposes" [3]

On Monday 26th of November, I downloaded an app which I assumed to be a simple alarm clock. However, it turns out that it is not just an alarm clock, it is a sleep tracker. Sleep Cycle is an app which, after the alarm has been set, listens to your breathing patterns while you sleep.

It is a very well tailored app to meet the needs of the user. For example, the more you snooze the alarm, the shorter the "snoozes" become, waking you up gently and effectively. You can see your "sleep efficiency" percentage, graphs of your sleep cycle over the past weeks, months and years. You can even see if you snore, how long for, and listen back to it.



Wake up between 08:15 - 08:45

In bed	00:09 - 08:29
Sleep quality	73%
Time in bed	8:20
Snore	1 min Listen
Sleep notes	Cloudy
Activity	12,134 steps
Online backup	● Synced
Total nights	21, Avg. time 7:26

In bed	23:48 - 08:30
Sleep quality	100%
Time in bed	8:42
Snore	0 min
Online backup	● Synced
Total nights	21, Avg. time 7:26

Pictured above are 3 screenshots I captured from the app, on the left you can see how you set the alarm and the given wake up time frame (when you're in light sleep), in the middle an OK nights sleep, including 1 minute of snoring at roughly 01:40, and on the right a very good nights sleep.

It does just about everything one could want an alarm clock to do, and more. However, bearing in mind that it logs your data indefinitely, and backs it up to a server, did I pay any attention to the terms and conditions I agreed to when downloading the app? Absolutely not. This unconscious disregard of the terms and conditions (T&C's) is the basis of this paper.

When actually reading through the privacy policy of Sleep Cycle, it can be found out that they collect information about you, including information that directly or indirectly identifies you. This information includes IP addresses, time stamps, log files and operating system information [4]. The privacy policy argues that the information is needed to operate and improve Sleep Cycle, and to, for example provide you with “big data” statistics such as average sleep times in certain countries, at which point it is de-identified from any personal data [5]. The T&C’s go on to say: “To analyse, gain insight and improve Sleep Cycle we use third-party services that might receive and process de-identified information on our behalf” [6]. They also “work with partners who provide us with analytics and advertising services”, and that these companies “may use device-specific advertiser identifiers, cookies and similar technologies to collect information about your interactions with the Services and other websites and applications” [7].

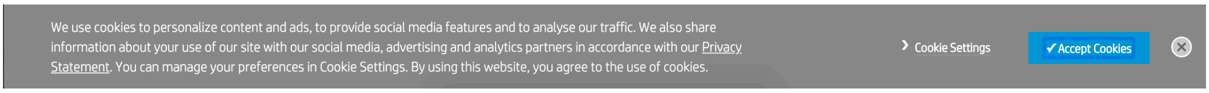
This seems like an awful lot to agree to by ticking a box, however, according to the GDPR it is all legal.

There are many apps like Sleep Cycle, which are used to monitor daily life. These apps include:

- MyFitnessPal, which is used as a food diary and to count calories.
- Activity, used as a step counter to track daily, weekly and monthly activity.
- Elite HRV, used to track heart rate variability and stress levels.
- Happy Scale, where you enter your body weight and it helps you to set goals for weight loss, and understand weight fluctuations etc.

Imagine the digital image which could be depicted of you if, for example, you are using all of the apps mentioned above as well as Sleep Cycle, and they all share their data with the same advertising companies. They would know what you eat, how far you walk, how stressed you are, how much you weigh, weather or not you snore, how much you sleep, how regularly you exercise, and so forth.

It is not only mobile apps which have the ability to track you. Websites also have this ability, and is mostly done through the use of Cookies. Google defines a cookie as “a packet of data sent by an Internet server to a browser, which is returned by the browser each time it subsequently accesses the same server, used to identify the user or track their access to the server" [8].

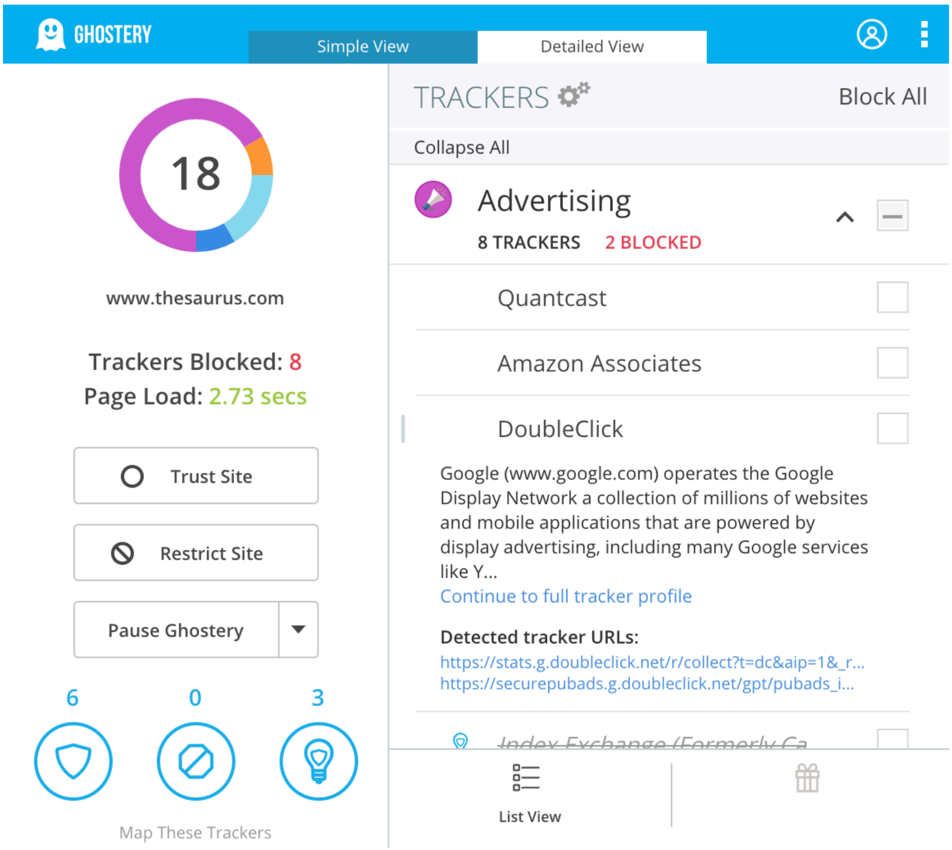


We use cookies to personalize content and ads, to provide social media features and to analyse our traffic. We also share information about your use of our site with our social media, advertising and analytics partners in accordance with our Privacy Statement. You can manage your preferences in Cookie Settings. By using this website, you agree to the use of cookies.

[9]

Have you noticed those annoying banners which take up half of the screen prompting you to "accept cookies"? If you agree, you are legally giving the website your permission to share your data. The best thing you can do in this case is to press the "x" button, which does not give the site your permission to share your information.

To see what trackers are surreptitiously integrated into websites, there is a Chrome extension called Ghostery which can be installed to block data being sent to nonessential third parties. Unfortunately, many trackers are "essential" to website functionality and cannot be blocked. One of which, which can be seen below being used on the website www.thesaurus.com, is called Doubleclick.



Trackers Blocked: 8

Page Load: 2.73 secs

Google (www.google.com) operates the Google Display Network a collection of millions of websites and mobile applications that are powered by display advertising, including many Google services like Y...

Continue to full tracker profile

Detected tracker URLs:

https://stats.g.doubleclick.net/r/collect?t=dc&aip=1&_r...

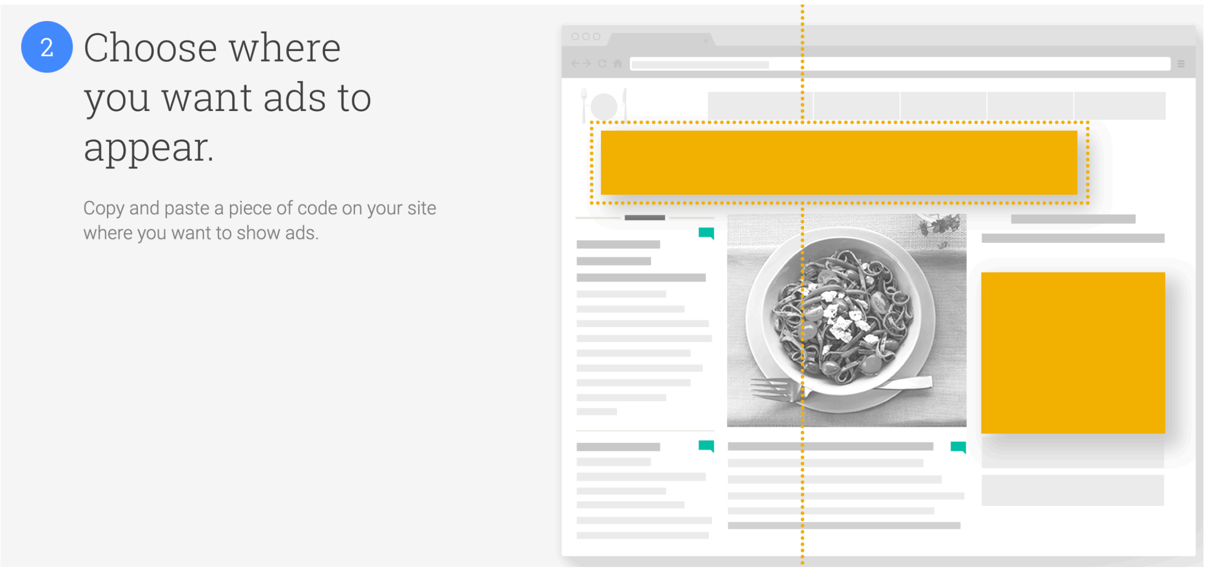
https://securepubads.g.doubleclick.net/gpt/pubads_i...

Doubleclick, founded in 1995, was bought by Google in 2008 for $3.1bn [10] and uses cookies to gather data about the particular websites traffic. In Google's privacy policy, Google explains how data is recorded from a generic Doubleclick cookie. An example of a Doubleclick cookie looks like this [11]:

time: 06/Aug/2008 12:01:32
ad_placement_id: 105
ad_id: 1003
userid: 0000000000000001
client_ip: 123.45.67.89
referral_url: http://youtube.com/categories

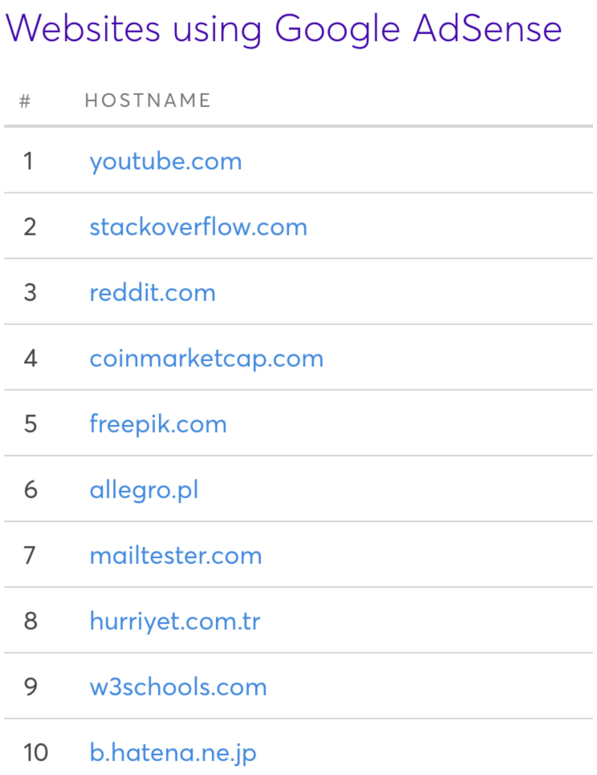
Adsense, another branch of Google, allows website creators to put advertising code directly into their websites. They choose where the adverts go, and Google handles the rest. Google fits adverts appropriate to the website, and to the user viewing the website (so different users see different adverts).

[12]



2
Choose where you want ads to appear.
Copy and paste a piece of code on your site where you want to show ads.

[13]



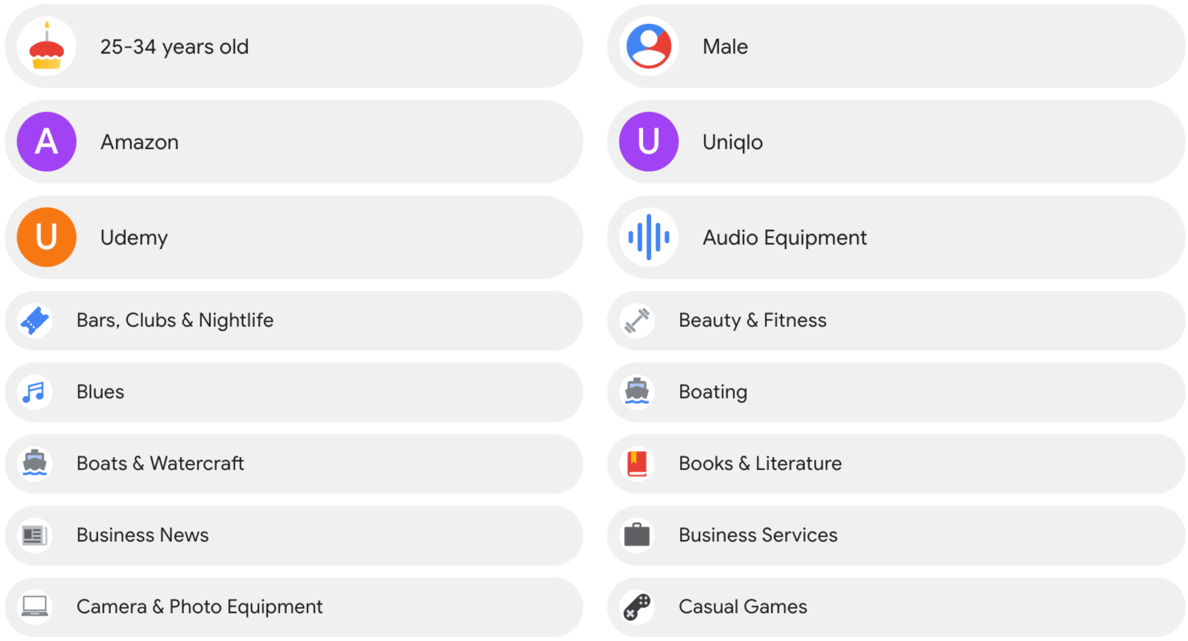
Websites using Google AdSense

#	HOSTNAME
1	youtube.com
2	stackoverflow.com
3	reddit.com
4	coinmarketcap.com
5	freepik.com
6	allegro.pl
7	mailtester.com
8	hurriyet.com.tr
9	w3schools.com
10	b.hatena.ne.jp

This is all very clever, but when Adsense is paired with Doubleclick, Google can do some amazingly scary things. If a Doubleclick cookie is set on a website which is also using Adsense, and then you navigate to another site, also using Adsense, the a collection of data will be pooled. Over time, Google can make predictions about the interests of the user, for example, weather they prefer football to rugby, or if they shop at Waitrose, rather than Tesco. The users browsing habits are then put into 'segments', such as 'football lovers'. These segments become labels which Doubleclick allows advertisers to choose from when selecting who they want their adverts to be targeted at. [14]

To see what segments you are part of, navigate to:
https://adssettings.google.com/authenticated

Here are some of the 92 segments Google has me listed in:



25-34 years old

Male

Amazon

Uniqlo

Udemy

Audio Equipment

Bars, Clubs & Nightlife

Beauty & Fitness

Blues

Boating

Boats & Watercraft

Books & Literature

Business News

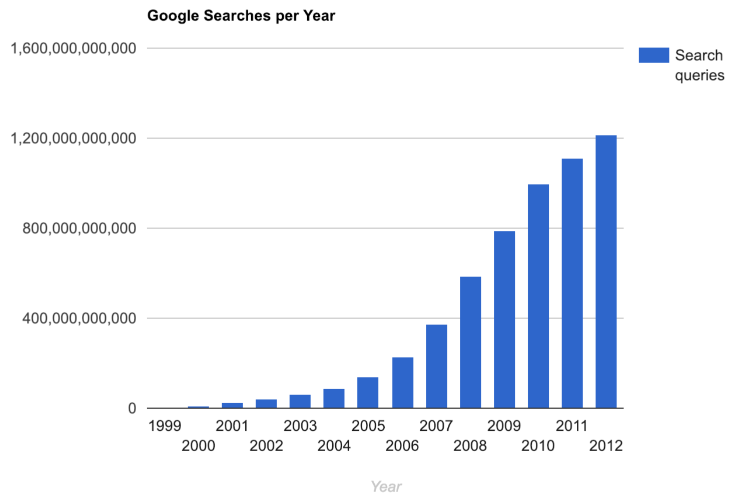
Business Services

Camera & Photo Equipment

Casual Games

Google provides the world with the top search engine, web browser, and mobile platform. As a search engine, Google handles 3.5 billion searches per day [15], and has a 90.6% search engine market share [16]. Google Chrome has had over 2 billion active installs world wide [17]. Android has over 2 billion monthly users [18]. YouTube (also owned by Google) has 1.8 billion logged in users per month. Gmail, and Google Maps, each draw over 1 billion users per month [19].

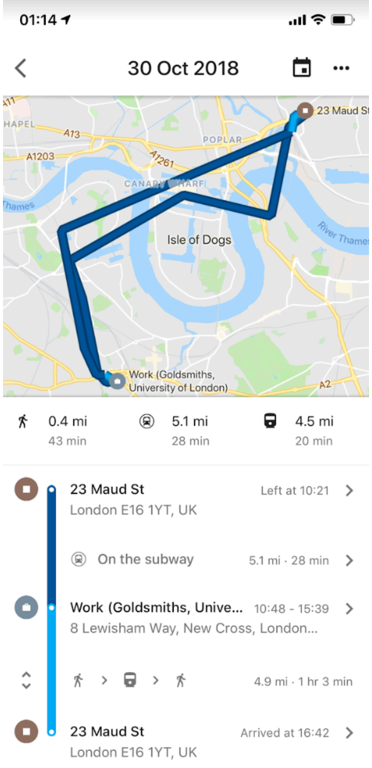
[20]



Data is collected on these platforms both actively, and passively. Actively is usually in the form of search queries such as a destination entered into Google Maps, or a query entered into Google's search engine. Passively can be through the form of anonymised cookies on 3rd party websites with Googles trackers installed. These cookies can then be linked back to your Google account, if you have a Google application open in the same browser, thus un-anonymising your anonymous data [21].

Here is an example of active vs passive collection [22]:

Activity: Searches for cold medicine while on the subway.
Active collection: Records search query.
Passive collection: Travelling on the subway.



0.4 mi 43 min
5.1 mi 28 min
4.5 mi 20 min

23 Maud St
London E16 1YT, UK
Left at 10:21

On the subway
5.1 mi · 28 min

Work (Goldsmiths, Unive...
8 Lewisham Way, New Cross, London...
10:48 - 15:39

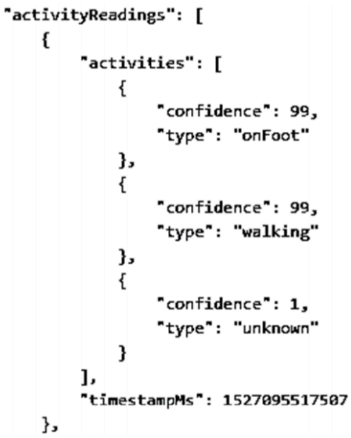
4.9 mi · 1 hr 3 min

23 Maud St
London E16 1YT, UK
Arrived at 16:42

Above is a screen shot from my own "Google Timeline" (which can be found within Google Maps), confirming the previously mentioned passive data gathering methods.

In a paper entitled Google Data Collection, it was revealed that a dormant, stationary Android phone with chrome active in the background communicated location information to Google 340 times during a 24 hour period, or an average of 14 data communications per hour. [23]

This is how both Chrome and an Andoid device communicate passive data with google [24]:



```
"activityReadings": [
    {
        "activities": [
            {
                "confidence": 99,
                "type": "onFoot"
            },
            {
                "confidence": 99,
                "type": "walking"
            },
            {
                "confidence": 1,
                "type": "unknown"
            }
        ],
        "timestampMs": 1527095517507
    },
```

The shear scale of data harvesting and profiling conducted by Google on a global scale, is quite frankly – unbelievable.

You'd assume data harvesting on this scale would require top level security. However, in March 2018 Google discovered a bug in the API for Google+. Up to 438, 3rd party apps were using the offending part of the API, potentially leaking private information of up to 500,000 users [25]. A second breach in the Google+ API, discovered in December 2018, allowed developers to access users personal information, even if the information was set to private. The second leak, had the power to share names, email addresses, genders, birthdays, and more of over 52 million Google+ accounts [26]. Google has since announced that the Google+ platform will be shut down in April 2019, and that they do not know how much information may have been accessed.

Security breaches on this scale are, unfortunately, becoming more and more frequent.

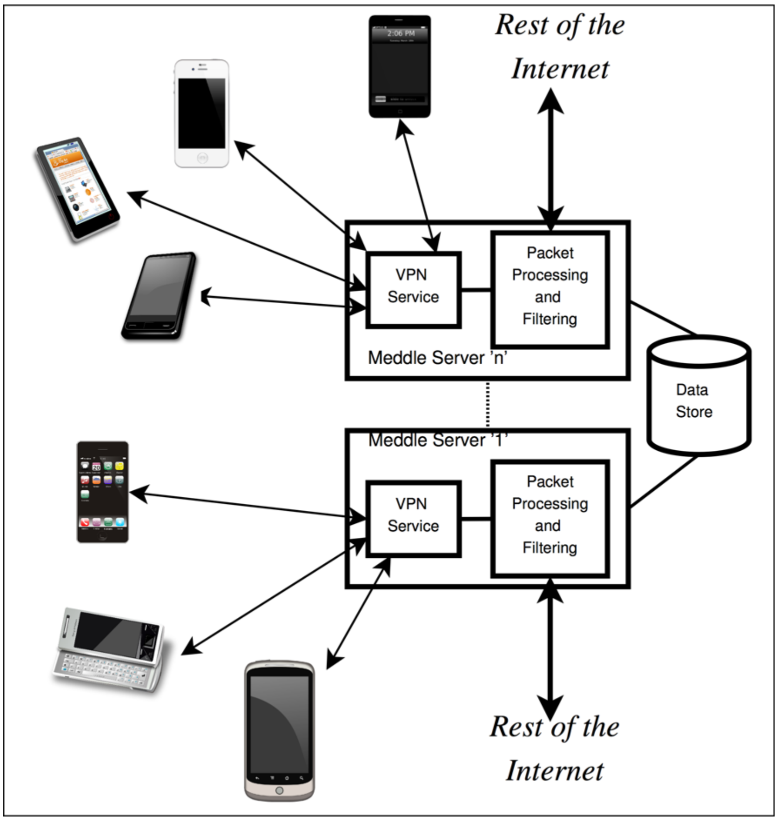
ReCon

A team led by a man called Dave Choffnes from Northeastern University in Boston MA, developed ReCon. ReCon is a Virtual Private Network (VPN), developed on Meddle servers, for personally identifiable data which can be identified through a smartphone.

https://www.meddle.mobi/details.html

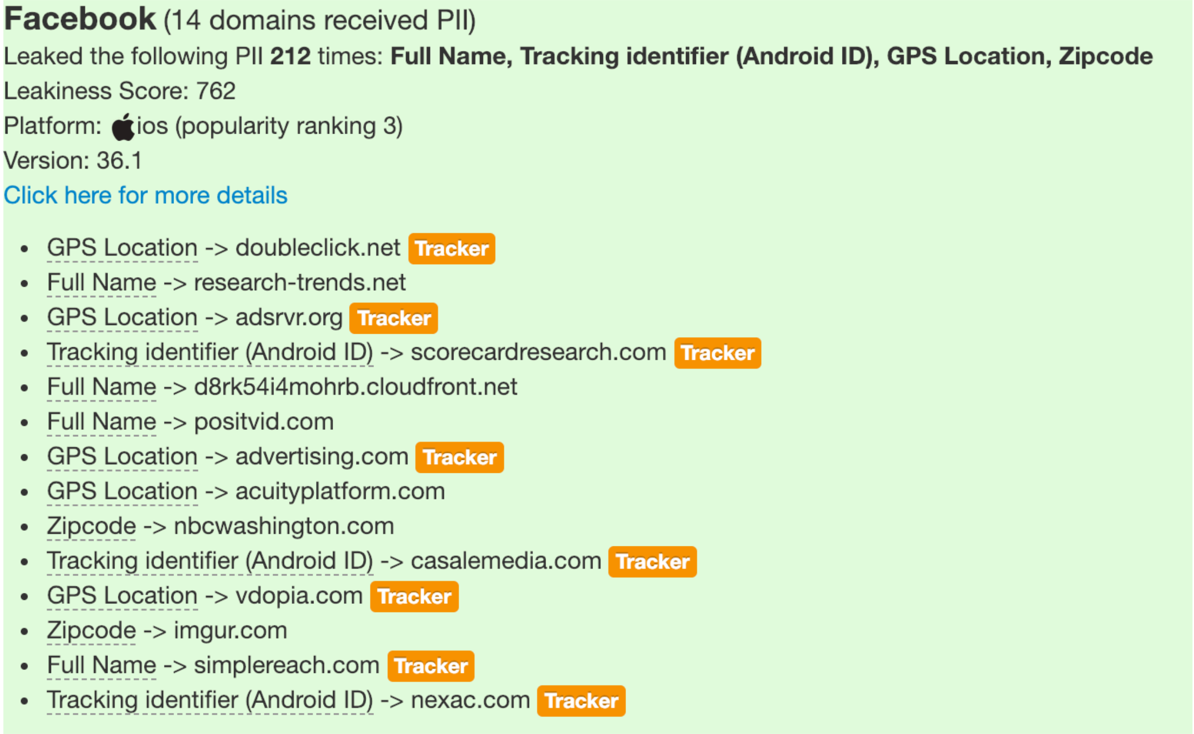
ReCon uses the VPN connection as a tunnel to send all data sent from your phone to the Meddle servers. Here, ReCon can fully examine the contents of the unencrypted connections.

[27]



Choffnes and his team found some surprising results. For example, GrubHub, a food delivery app, shared users passwords with Googles Crashlytics. When notified about this, GrubHub revised their code, and made Crashlytics detele all information containing passwords.

[28]



Facebook (14 domains received PII)

Leaked the following PII **212** times: **Full Name, Tracking identifier (Android ID), GPS Location, Zipcode**

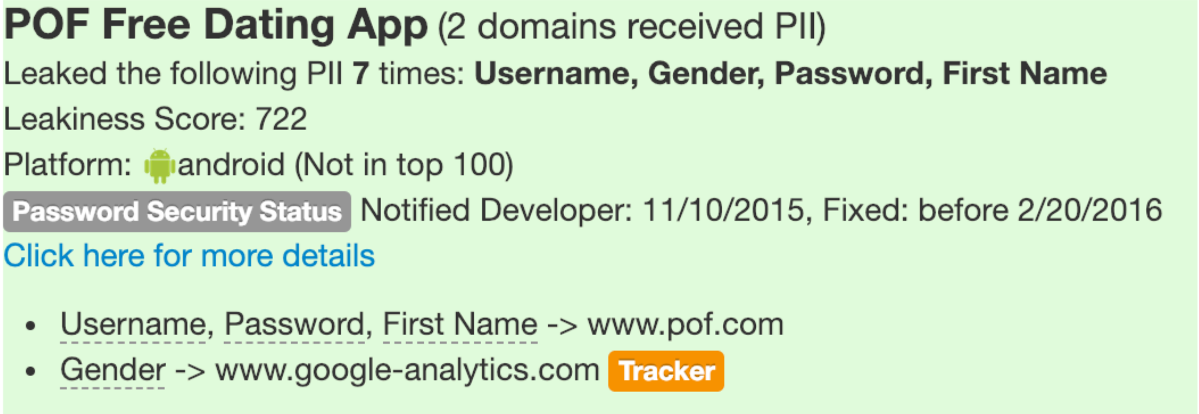
Leakiness Score: 762

Platform: ios (popularity ranking 3)

Version: 36.1

Click here for more details

- GPS Location -> doubleclick.net Tracker
- Full Name -> research-trends.net
- GPS Location -> adsrvr.org Tracker
- Tracking identifier (Android ID) -> scorecardresearch.com Tracker
- Full Name -> d8rk54i4mohrb.cloudfront.net
- Full Name -> positvid.com
- GPS Location -> advertising.com Tracker
- GPS Location -> acuityplatform.com
- Zipcode -> nbcwashington.com
- Tracking identifier (Android ID) -> casalemedia.com Tracker
- GPS Location -> vdopia.com Tracker
- Zipcode -> imgur.com
- Full Name -> simplereach.com Tracker
- Tracking identifier (Android ID) -> nexac.com Tracker



POF Free Dating App (2 domains received PII)

Leaked the following PII **7** times: **Username, Gender, Password, First Name**

Leakiness Score: 722

Platform: android (Not in top 100)

Password Security Status Notified Developer: 11/10/2015, Fixed: before 2/20/2016

Click here for more details

- Username, Password, First Name -> www.pof.com
- Gender -> www.google-analytics.com Tracker

Please note: PII in the above screenshots refers to Personally Identifiable Data.

A full list of leaks they have discovered, follow this link: https://recon.meddle.mobi/app-report.html

Cambridge Analytica

Founded in 2013, and closed down in 2018, Cambridge Analytica (CA) is regarded as a full service propaganda machine. Described as a British Political Consulting firm, they famously worked for Donald Trumps presidential campaign, and allegedly Leave.EU (one of the organisations campaigning for the UK to leave the EU).

CA developed Facebook apps, one of which was called "thisisyourdigitallife", which had special permissions to harvest data. Not only from the person using the app, but their entire friends list as well. CA's apps only had to touch a few hundred thousand people, for them to be able to gain an insight into the Facebook profiles of most of America. The information they harvested included status updates, likes and private messages. Over a 2-3 month period, they pulled data from roughly 60 million profiles. From this information, they would know what kind of messaging you would be susceptible to, including the framing of the message, tone, topics, content, weather its scary or not, etc. They knew where you would consume the messages, how far down a particular news article or post, and how many times the would have to touch you to change the way you think about something. This process is also known as "psychographic profiling". CA had an entire team of creatives, designers, photographers and videographers who would create content and "inject" it into the internet for you, not as a voter, but as a personality, to find. [29]

In an article titled "Leaked: Cambridge Analytica's blueprint for Trump Victory", The Guardian discloses a leaked presentation, used by used by CA to attract new clients. In it, you can see the reality of how they would manipulate the public [30].

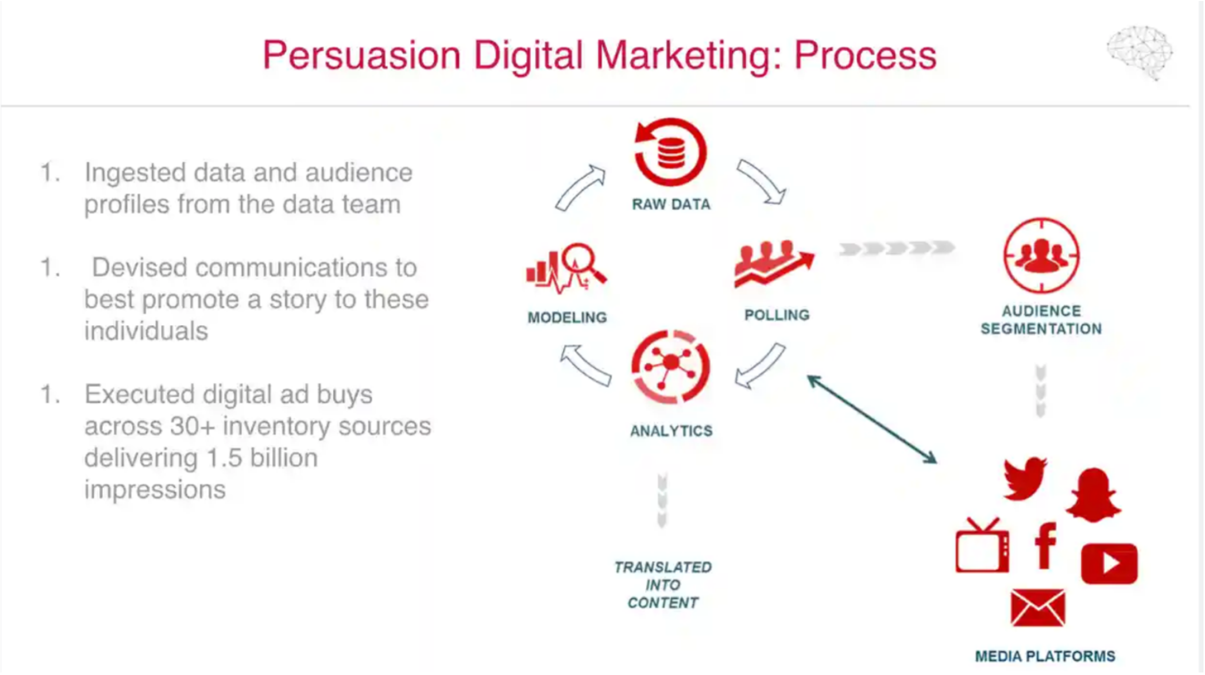
Why Cambridge Analytica

Starting from Scratch

When we started on the campaign (second week of June) the Trump Campaign had no speakable data infrastructure.

- No database of record
- Many disparate data sources
- No data science program (models)
- No proper digital marketing apparatus
- Research being done by up to 5 pollsters at one point

Any most importantly: **No unifying data, digital and tech strategy.**

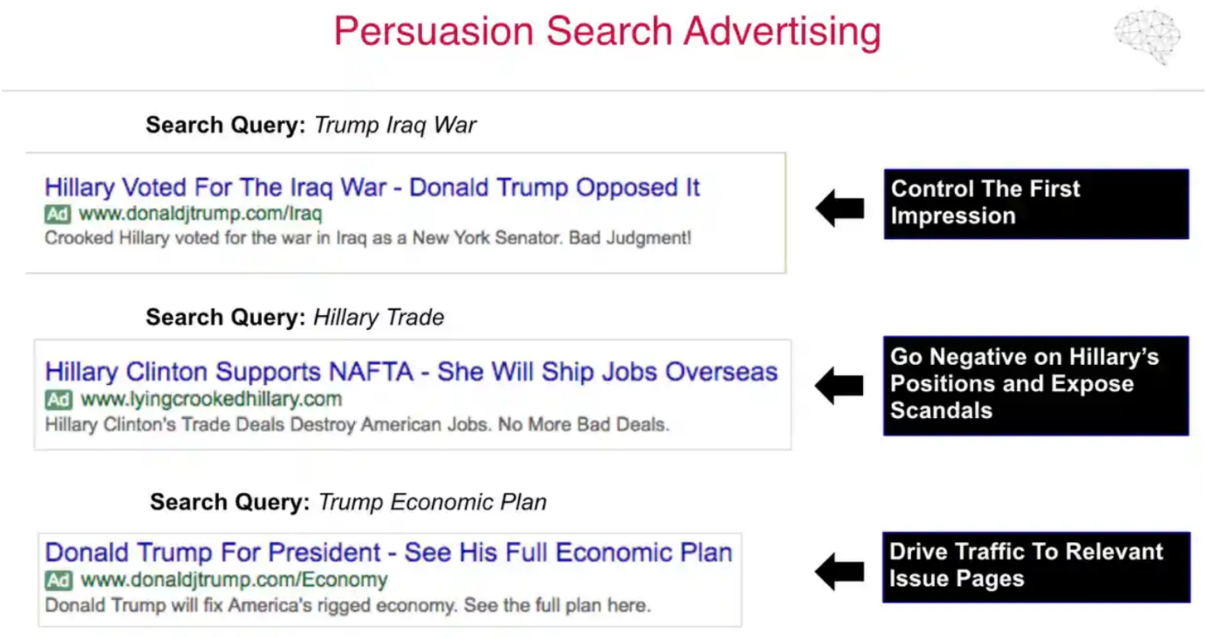


Persuasion Digital Marketing: Process

1. Ingested data and audience profiles from the data team

1. Devised communications to best promote a story to these individuals

1. Executed digital ad buys across 30+ inventory sources delivering 1.5 billion impressions



Persuasion Search Advertising

Search Query: *Trump Iraq War*

Hillary Voted For The Iraq War - Donald Trump Opposed It
Ad www.donaldjtrump.com/Iraq
Crooked Hillary voted for the war in Iraq as a New York Senator. Bad Judgment!

Control The First Impression

Search Query: *Hillary Trade*

Hillary Clinton Supports NAFTA - She Will Ship Jobs Overseas
Ad www.lyingcrookedhillary.com
Hillary Clinton's Trade Deals Destroy American Jobs. No More Bad Deals.

Go Negative on Hillary's Positions and Expose Scandals

Search Query: *Trump Economic Plan*

Donald Trump For President - See His Full Economic Plan
Ad www.donaldjtrump.com/Economy
Donald Trump will fix America's rigged economy. See the full plan here.

Drive Traffic To Relevant Issue Pages

After learning about CA's activity on Facebook, Facebook told CA to delete all data they had obtained through the violation of their rules. However reports say that CA never deleted the data, and Facebook never chased it up to make sure it had indeed been deleted.

The depth of this data breach through CA's apps on Facebook landed Facebook's CEO, Mark Zuckerberg, in front of the US senate in a 10 hour testimony [31].

To conclude

If you look at another person through a window, or from behind, they cannot see, smell or hear you, so technically they do not know you are there. However, can they feel your gaze? I believe this to be a fairly common experience, that of being watched. A lot of people have also had the opposite experience, making someone turn around by staring at them.

What we are doing by ticking that box which says “agree to terms and conditions”, “accept cookies”, or “turn location services on”, is giving the app and website developers permission to stare at us from all angles, watching and logging our every move, purchase, and search.

The exterior is harmless, however under the hood, websites and apps are money making machines, selling your private personally identifiable data to third party advertising companies. We, the public, are the product, the app developers are the vendors, and the advertising companies are the buyers.

In my opinion, the internet has almost turned into, what can only be only be described as a shopping centre for advertising companies. However, through services like ReCon, and Ghostery, the public are becoming more aware of the people looking at them through the windows of the internet. The public are beginning to smell and hear the shoppers presence, feeling their gaze more and more, and they do not like it.

References

[1] – Definition of the GDPR - https://www.investopedia.com/terms/g/general-data-protection-regulation-gdpr.asp

[2] – Description of 'Personal Data' - https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1528874672298&uri=CELEX%3A32016R0679

[3] – processing shall be lawful only if … - (article 6) - https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1528874672298&uri=CELEX%3A32016R0679

[4] – Sleep Cycle Privacy Policy – Technical Information - https://www.sleepcycle.com/privacy-policy/

[5] – Sleep Cycle Privacy Policy – How We Use information - https://www.sleepcycle.com/privacy-policy/

[6] – Sleep Cycle Privacy Policy – How Information Is Shared - https://www.sleepcycle.com/privacy-policy/

[7] – Sleep Cycle Privacy Policy – Analytics Information – https://www.sleepcycle.com/privacy-policy/

[8] – Google definition of a cookie - https://www.google.com/search?q=define+cookie&oq=define+cookie&aqs=chrome..69i57j0l5.3845j1j7&sourceid=chrome&ie=UTF-8

[9] – Screen shot of cookie agreement – https://store.hp.com/UKStore/Merch/Offer.aspx?p=c-hp-spectrex360&ocid=AID740620_FACEBOOK_oo_spl100000427361081

[10] – Google buys Doubleclick for $3.1bn - https://www.nytimes.com/2007/04/14/technology/14DoubleClick.html

[11] – Explanation of what a Doubleclick cookie looks like - https://www.theguardian.com/technology/2012/apr/23/doubleclick-tracking-trackers-cookies-web-monitoring

[12] – Adsense screenshot - https://www.google.com/intl/en_uk/adsense/start/how-it-works/#/

[13] – List of websites using Adsense - https://www.wappalyzer.com/technologies/google-adsense

[14] – Description of how Adsense works - https://www.theguardian.com/technology/2012/apr/23/doubleclick-tracking-trackers-cookies-web-monitoring

[15] – Google handles 3.5bn searches per day - https://youtu.be/8qS7eyUo_uk?t=872

[16] – Google's 90.6% market share – Page 25 - https://digitalcontentnext.org/wp-content/uploads/2018/08/DCN-Google-Data-Collection-Paper.pdf

[17] – Chrome 2 billion installs – Page 3, article C - https://digitalcontentnext.org/wp-content/uploads/2018/08/DCN-Google-Data-Collection-Paper.pdf

[18] – Android monthly users – Page 3, article B - https://digitalcontentnext.org/wp-content/uploads/2018/08/DCN-Google-Data-Collection-Paper.pdf

[19] – Youtube, Gmail and Gmaps users per month- Page 25, table 2 - https://digitalcontentnext.org/wp-content/uploads/2018/08/DCN-Google-Data-Collection-Paper.pdf

[20] – Table of Google searches per year - http://www.internetlivestats.com/google-search-statistics/

[21] – Google being able to link deidentified data back to your Google account – Page 4, section G - https://digitalcontentnext.org/wp-content/uploads/2018/08/DCN-Google-Data-Collection-Paper.pdf

[22] – Active vs Passive data collection – Page 37 - https://digitalcontentnext.org/wp-content/uploads/2018/08/DCN-Google-Data-Collection-Paper.pdf

[23] – Android phone data communication - https://youtu.be/8qS7eyUo_uk?t=759

[24] – example of passive data collection - Page 13, figure 5 - https://digitalcontentnext.org/wp-content/uploads/2018/08/DCN-Google-Data-Collection-Paper.pdf

[25] – Google security breach - https://www.theguardian.com/technology/2018/oct/08/google-plus-security-breach-wall-street-journal

[26] – Google second security breach - https://www.pocket-lint.com/apps/news/google/145970-google-is-finally-shutting-down-google-for-good-following-security-breach

[27] – How ReCon works - https://meddle.mobi/details.html

[28] – ReCon picked up on data leaks - https://recon.meddle.mobi/app-report.html

[29] – Large chunk of information about Cambridge Analytica – re-worded from an interview with Christopher Wylie - https://www.theguardian.com/uk-news/video/2018/mar/17/cambridge-analytica-whistleblower-we-spent-1m-harvesting-millions-of-facebook-profiles-video

[30] – Leaked Cambridge Analytica presentation - https://www.theguardian.com/uk-news/2018/mar/23/leaked-cambridge-analyticas-blueprint-for-trump-victory

[31] – Mark Zuckerberg / US Senate interview - https://youtu.be/6ValJMOpt7s