

**Which?**

**POLICY REPORT JUNE 2018**

# **Control, Alt or Delete?**

**The future of consumer data**





# Contents

<b>Executive summary</b>	<b>5</b>
<b>Introduction</b>	<b>13</b>
<b>1. The transparency deficit in consumer data</b>	<b>16</b>
Attitudes to data versus commercial reality	16
Baseline attitudes to data	16
Consumer knowledge versus commercial reality	18
Post-knowledge attitudes to data	24
Why this matters for business	24
Consumer self-help: dirtying data, ad blockers and more	24
Inhibited adoption of services	24
Corporate PR disasters	26
The commercial context to opacity	27
The scale and structure of the digital advertising market	27
The role of data in the competition for ad budgets	29
How data reinforces ‘winner take all’ market dynamics	32
Revealed preferences towards commercial transparency on data	32
Impact on policymakers: intractable open questions	34
<b>2. The control deficit in consumer data</b>	<b>37</b>
Google and Facebook: a consumer choice?	37
The unknowable ‘adtech’ and ‘martech’ industries	39
The unknowable black market in data	39
Losing grip on data in motion	41
<b>3. Where do we go from here?</b>	<b>43</b>
Reframing transparency in data	43
Controlling dominant players	45
Scalable governance of data flows	46



# Executive summary

**Digitisation is remodelling consumer markets, and the use of data about our consumer lives has already brought huge benefits and great potential for empowerment. However many people feel powerless to understand either the growing commercial observations or its effects.**

The application of personal data can significantly benefit consumers, for example when it is used to provide personalisation and new innovative products, or when it empowers people to drive businesses to provide the types of products and services that they want, often reshaping whole markets to meet consumer demands. Consumers recognise and value many of these benefits.

Our mission at Which? is to make consumers as powerful as the organisations they deal with in their everyday lives, but those organisations are now processing unprecedented amounts of consumer data.

Commercial data collection and use has exploded in every consumer market, with more and more companies striving to get information about their target market at an individual level. Much of this trend has been driven by Facebook and Google: two of the most powerful organisations consumers deal with every day.

Businesses' knowing more about their customers is often beneficial for people, but our new research has found a widespread sense of disempowerment, with many people unsure of either the impact that data use has on them or whether it is even worth trying to take any action about practices that concern them.

**Our goal in this work was to understand how far consumers may require further support to rebalance power over use of their data.** Consumers may report feeling disempowered on a number of issues, but in a context where *their own behaviour* may be inadvertently causing themselves harm, and where the reach of the issue is huge, their sense of disempowerment is of particular concern to us. We scrutinised the following questions:

- 1. Do people understand how their data is collected, traded, and used, and how this may affect their choices? And if not, why not?**
- 2. Is it possible for consumers to take control of their data, once they understand more?**
- 3. How far will recent and upcoming policy changes rebalance the playing field in consumers' favour?**

The introduction of the General Data Protection Regulation (GDPR) is an important landmark, and we think our work complements its introduction. GDPR is based on key principles and their interpretation by companies and regulators needs to be built on an understanding of people's behaviour within relevant consumer markets and other settings.

We partnered with the research agency BritainThinks to conduct major new primary research, finding out consumer views over the length and breadth of the country through:

- Constructing one of the most detailed segmentation analyses of the UK population's data attitudes and behaviour.
- Four one-and-a-half-day-long deliberative face-to-face workshops in Leeds, Newport, Perth and St Albans.
- In-depth interviews with more vulnerable consumers.

This approach has allowed us to understand consumers' unprompted attitudes and behaviour, and then to systematically explore their perceptions of the data ecosystem after explaining how it works and how it could affect them. We have published this research alongside this report as *Control, Alt or Delete? Consumer research on attitudes to data collection and use*.

**We have identified various sources of potential and actual harms from commercial uses of data.** When people are increasingly profiled at an individual level, and 'micro-targeted' on that data, this creates potential for various types of consumer detriment, including

- *Financial harms*: Losses that arise from breaches and growing potential for the use of data to price discriminate in unfair or hidden ways.
- *Non-financial harms*: Other potential harms of 'micro-targeting', such as exacerbating addiction, or discriminatory access to information or services.
- *Foregone benefits*: Lower uptake of digital services due to consumer concerns, or the market power that access to consumer data gives big tech companies.

This project's primary goal was to provide a detailed evidence base, but we have also identified priority areas for action:

**1. Consumers and their advocates need more transparency about the impact that personal data has on their lives.**

Our research found that consumers usually judge the acceptability of data collection and use by what impact it has on their lives, as opposed to information about the collection and purpose. Therefore:

- a. Companies need to consider how they can ensure people understand the impact of the use of their data, at the time they are transacting with them.
- b. Government, regulators, businesses and consumer advocates must come together to understand the impacts of data usage. We see a strong role for the new Centre for Data Ethics and Innovation to coordinate action here.

**2. The Competition and Markets Authority (CMA) should conduct a market study in to the digital advertising industry as a matter of urgency.**

'People-based marketing' has become a feature of the digital advertising market, but

its impact and consequences are widespread and poorly understood, and the concentration of the digital advertising industry in Facebook and Google's hands could be harming consumers through supply-chain impacts. The Lords Communications Select Committee has recently called for the CMA to conduct a market study following its report on UK advertising in a digital age<sup>1</sup> and we strongly support this.

**3. It is time for a thoroughgoing review of governance of data in motion, with due attention given to creative ways to provide improve oversight and enforcement.**

Data portability is an important new right in the GDPR, with significant potential to empower consumers, and the government is soon to look at this in a 'smart data' review. However, we are concerned that take-up will be limited if people do not trust the data ecosystem sufficiently. A way needs to be found that allows innovation but also improves the ability to provide oversight and enforcement. This is likely to mean understanding the forefront of potential technological solutions that could provide truly decentralised and scalable accountability for how data flows. We therefore think that a review of the governance of data in motion should be a priority for the new Centre for Data Ethics and Innovation, alongside the Information Commissioner's Office's (ICO's) planned work on reference agencies and data brokers.

We provide the reasoning behind these recommendations below.

### *The transparency deficit in consumer data*

**There is an unsustainable lack of transparency on the trading and application of consumer data, meaning even extensive efforts cannot give consumers straight answers to the questions that need answering.**

**Consumer attitudes to data are often pragmatic, but the reality of the data ecosystem is an unpleasant surprise.** Our research found that consumers recognise the greater convenience and choice that the digital revolution has brought them. Their attitudes to the commercial uses of data are finely balanced, and they accept the need to share data in return for a clear benefit. However, we also found that people have low awareness of the full spectrum of ways in which data is collected about them, and how that data may affect what they see and the choices they have. In particular:

- The fact that unknown companies 'profile' them as an individual is a significant surprise to most consumers.
- The use of data science to 'infer' aspects of their profile makes many consumers uneasy, as they don't know how it could affect them.

---

<sup>1</sup> *Lords call for action on digital advertising*, (2018), available at: <https://www.parliament.uk/business/committees/committees-a-z/lords-select/communications-committee/news-parliament-2017/uk-advertising-report-publication>

- The vast majority of consumers dislike the fact their data may be sold to support and monetise this process.
- There is widespread dislike of the fact that personal data and profiles could be used to target them in potentially harmful ways without their knowledge.
- Many people are surprised by the scale and scope of data theft.

**This growing, untapped potential for ‘nasty shocks’ risks destroying value for business.**

Consumers get more nervous the more they learn, and they have plenty still to learn. Indicators are that this is a dangerous combination for business. We observe value being destroyed for business via:

- consumer behaviour that either restricts data collection or results in errors;
- inhibited adoption of new data-driven services;
- corporate public relations disasters.

**Companies have allowed – even encouraged – this disengagement and confusion**

**among consumers.** Personal data is commercialised in a number of ways, but the commercial application that is shaping the ecosystem is the targeting of brand and marketing messages. This market is huge, its growth continues to accelerate, and the competitive stakes are high. Google and Facebook dominate the market more each year (they are estimated to have commanded 54% of the UK digital advertising market in 2017, and 59% of the global market), and the harvesting and use of consumer data is central to the competitive strategies of all players. In a situation where the financial rewards to success are so high, and consumer cooperation is vital to that success, companies that make money through digital advertising have had every incentive to allow consumers to stay uninformed.

**And this meant we were unable to get clear and comprehensive answers on some of the important questions about harm.** The lack of transparency in the data industry makes it difficult for consumers, their advocates and policy makers to understand the scale of potential detriment or provide good public information about what is happening. There are material open questions that are still far too difficult to answer systematically. For example, we could not fully answer the following questions:

- What proportion of consumers have had their sensitive data breached, how much of this has been exposed and/ or sold, and are any legitimate companies buying it?
- How widely are consumer profiles originally compiled for the targeting of marketing messages, also used beyond this purpose? For example, to personalise prices?
- What is the full extent of data sharing and selling about individuals?



## *The control deficit in consumer data*

### **Meaningful consumer control over data is currently unfeasible, and this is getting worse**

**Google and Facebook have access to the widest range of online data, and use it to further consolidate their grip on their markets, limiting the prospect of meaningful consumer choice in some services.** People may dislike what they see when they gain understanding, but they often feel resigned and powerless to do anything about it. In particular, Google and Facebook have substantial market power in the provision of some of the most important data-dependent services that many consumers view as essential to their daily lives. They dominate time spent online enabling them to collect most data, and we are concerned that they could use that data to strengthen their position and inhibit the development of viable alternatives. Our research finds that many people feel they have little choice but to accept their practices, particularly because they did not think they could take effective action without disconnecting from technology.

**Beyond these powerful but well-known brands, consumers face a data trading ecosystem of thousands of actors that are unknowable, and significantly limit the effectiveness of data subject rights.** Very few people know about the broader data ecosystem, consisting of thousands of companies, and therefore it is not feasible to expect people to spend the time required to exert even their enhanced data subject rights under GDPR.

The illegal market in data may be as complicated, if not more so, than the legitimate market in data. When a consumer's data leaks into this world, they truly have no control. Many consumers appear to view it as inevitable that their data will be stolen. In many ways this resigned attitude seems to be realistic, as corporate and law enforcement actors are losing the battle against breaches and UK consumers' exposure is significant. This is why Which? has already:

- Worked with government and industry to create a new code of practice for Security by Design.
- Raised a super-complaint with the Payment Systems Regulator, advocating placing more liability on banks for consumers' losses from 'authorised push-payment' fraud. The regulator has now committed to introducing a contingent reimbursement scheme to help compensate some victims.
- Advocated for the Data Protection Bill to include powers for qualified non-profit bodies to pursue collective redress on behalf of consumers who have suffered a data protection infringement.

**The inherent nature of data means there are serious feasibility issues in controlling the ever increasing data flows between companies.** The volume of data flows between companies has vastly increased as consumers generate more data and that data is exchanged automatically.

These flows of data have transformed the economics of starting and running a business, fostering massive innovation that has benefitted millions of consumers and businesses.

There are feasibility issues in ensuring compliant use of data through a supply chain. The main channel for accountability for data protection breaches in data supply chains comes from resource-intensive, painstaking investigations by either the Information Commissioner's Office or by the press. This is far from scalable; but scalable accountability is what is required when data flows are as complex as they are today.

### *Where do we go from here?*

#### **The policy framework adapts too slowly to the realities of the way digital markets work.**

**There is not enough transparency for consumers or their advocates on what they care about – namely, the impact that personal data has on their lives.** People want greater transparency over how their lives are influenced by the use of personal data, and organisations like Which? need much better information in order to determine where harm is occurring and empower consumers to stop it. Most of data protection regulation, even after GDPR, focuses on greater transparency about the collection and purpose of personal-data use. While these rights are important, our research suggests they will not be enough to tackle widespread feelings of disempowerment.

Many people choose not to learn more as they are too uncertain about the impact on their lives to justify the extensive time and effort that would be required. Our research found that people usually judge the acceptability of data collection and use by what impact it has on their lives, as opposed to information about collection and purpose. People do not and should not want to understand the data ecosystem, but they do want to understand why their insurance quote has changed, why they are seeing certain product suggestions, or how the data held about them might change a credit decision.

Therefore, where feasible, we think that companies should build consumer trust by giving transparency in context to their customers. Companies have figured out how to target individuals at the right moment for advertising – now they need to use the same ingenuity to allow people to understand how the data held on them affects their lives.

Where the impacts are harder to understand, we want to see government, regulators, businesses and consumer advocates working together to understand the impacts of data usage. There are several forms this could take:

- A summit brought together by the Department for Digital, Culture, Media and Sport to agree how the range of organisations with an interest in this space should take action. This should involve a broad range of stakeholders, including ourselves.

- Understanding the impacts of personal data use should be a priority for the new Centre for Data Ethics and Innovation<sup>2</sup> and the CMA's new data and technology unit.
- The CMA and the Department for Business, Energy and Industrial Strategy should explore a programme of work to investigate the impacts of data use on consumer markets.
- The ICO should ensure that its regulatory work explores the impacts of data usage, as well as the legality and processes.

**The invention and development of people-based marketing, driven by Facebook and Google, raises two key risks for consumer harm. The CMA should conduct a market study on digital advertising, alongside the ICO.**

People-based marketing is an approach to advertising that targets an individual, regardless of what device they are using or what media they are consuming, and monitors that person's response to the advertisement. Its invention and development, driven by Facebook and Google, raises two key risks for consumer harm:

1. People-based marketing has become a feature of the digital advertising market, but its impact and consequences are poorly understood.
2. There is significant horizontal and vertical concentration of the digital advertising industry in Facebook and Google's hands, which could lead to higher prices for consumers through supply chain impacts.

Current empirical evidence on harm for consumers through higher prices for advertised goods is limited, but we think the risk of harm is great enough that the CMA should conduct a market study on digital advertising, in conjunction with the ICO. We have shown evidence in this report that many people are unaware of the breadth and depth of the practices involved and have significant concerns when they find out about them, but do not feel they can realistically take action to avoid these practices. Together with the potential for structural detriment through the cost of digital advertising, this should be a priority area for the CMA.

**In a digital economy, data flows propel innovation, but it is practically difficult to monitor and govern these flows. It is time for a thoroughgoing review of governance of data in motion, with due attention given to creative ways to improve oversight and enforcement.**

Consumers dislike their data being sold and bought. It is difficult to provide good public information on where this is happening, and it is difficult to reassure them the system is accountable. However, in our digital economy, data flows propel competition and innovation and allow for the development of many of the goods and services that consumers use.

Data portability is also an important new right in GDPR, with significant potential to empower consumers. The government is soon to look at this in a 'smart data' review, but we are concerned that take-up will be limited if people

---

2 <https://www.gov.uk/government/news/search-for-leader-of-centre-for-data-ethics-and-innovation-launched>

do not trust the data ecosystem sufficiently. A way needs to be found that allows innovation and also improves the ability to provide oversight and enforcement.

We believe it is necessary to take a first-principles look at the accountability structures for the data supply chain. The constant automated and manual trading of data is here to stay and there are major feasibility issues in ensuring compliant use of data through a supply chain. Data protection legislation has failed to govern this in the past, and it is unclear that the GDPR represents a step change in approach to this issue. This is a challenging problem that may benefit from a mix of policy and technical solutions, and it would benefit from thoroughgoing review.

We therefore think that a review of the governance of data in motion should be a priority for the new Centre for Data Ethics, alongside the ICO's planned work on credit reference agencies and data brokers. The review should consider the forefront of technological solutions and include the following:

- Measures to foster more seller due diligence on the buyers of data, so that brokers cannot sell data without satisfying more strenuous conditions on its onward use.
- Stronger measures to ensure platform accountability for third-party access to data via APIs.
- The potential solutions to be found in the nascent market of personal identity and data management providers to give truly decentralised and scalable accountability for how data flows, and whether interventions are required to enable these innovations to achieve full potential benefit for consumers.

# Introduction

**Our mission at Which? is to make consumers as powerful as the organisations they deal with in their everyday lives and those organisations are now processing unprecedented amounts of consumer data.**

Companies have never known so much about people and consumers are at the centre of this data collection. Whether it is online searches for services, detailed purchase histories from credit cards, location information from phone tracking, health data from fitness watches or inferred data from social media interactions, almost all these consumer activities result in some form of information being collected.

Commercial data collection and use has exploded in every consumer market, with more and more companies striving to get beyond consumer segments, and ‘know’ their target market at an individual level, and in real time. The growth of the smartphone (73% of households now access the internet through a smart phone, up from 36% in 2011<sup>3</sup>) provides an ever increasing volume, velocity and variety of data on our lives.

**Digitisation is remodelling consumer markets, with huge benefit already realised, and significant potential for empowerment based on the use of data about our consumer lives.**

The impact of digitisation is seismic for consumers, and holds great potential to benefit all our lives. The use of personal data can benefit consumers when used to provide personalisation and significantly greater choice as companies can better meet consumers needs and wants. Innovative companies using personal data can increase competition by ensuring their products are seen by the right potential customers and tailoring personalised offers and discounts.

The use of personal data can also empower people to drive businesses to provide the types of products and services that they want, reshaping whole markets to meet consumer demands. Data on supplier and consumer reputation has allowed whole new business models to flourish that could not exist were it not for the use of personal data, notably in the collaborative or sharing economy.

---

<sup>3</sup> Office for National Statistics, (2017); *Internet access – households and individuals*, (2017), available at <https://www.ons.gov.uk/peoplepopulationandcommunity/householdcharacteristics/homeinternetandsocialmediausage/bulletins/internetaccesshouseholdsandindividuals/2017>

Personal data can also be used to produce entirely new products and services. Examples that are already here are a new mobile app that can help people with diabetes to manage their condition,<sup>4</sup> or a prototype developed by a blind Microsoft engineer that allows him to see what's going on around him through the use of data-driven artificial intelligence.<sup>5</sup> People recognise both the greater convenience and choice that the digital revolution is bringing them, and their purchasing behaviour shows that they value this.

**However, there is growing scrutiny on potential and real harms from commercial users of data.**

There are many reasons to make use of personal data on consumers. However, our current data ecosystem has been shaped by the invention of people-based marketing; an approach to advertising that targets an individual, regardless of what device they are using or what media they are consuming, and monitors that person's response to the advertisement. Although this practice can increase competition and allow personalisation that consumers like, it also creates potential for various types of consumer detriment.

The state of knowledge on these harms varies: some are systematically scrutinised by the courts and regulators; some are surfaced in press investigations; others are more hypothetical and speculated by campaigners and commentators.

- *Financial harms*: The losses that can arise from breaches (which have been growing in number and profile) are relatively well regulated; although much remains unknown on the potential future monetisation of stolen data. There is also growing potential for the use of data to price discriminate and affect access to products and information in hidden ways.
- *Non-financial harms*: The courts are now willing to recognise distress and other non-pecuniary losses arising from breaches of sensitive data. There is also growing interest in the potential harms of being 'micro-targeted', whether that harm takes effect at an individual level (eg, addiction, psychological effects of being aware of surveillance or of losing control as data is sold), or at a social level (eg, encouragement of siloes, discriminatory access to information or services).
- *Foregone benefits*: Innovation may be stifled by the market power of the largest technology companies, as they increasingly watch consumers' every move and leverage that data to increase their power in the distribution of goods, services and information online. Consumer concerns over data security and the use of personal data may also be inhibiting the growth of new digital industries.

---

4 WellDoc homepage: <https://www.welldoc.com/product>

5 Seeing AI: Talking Camera app for those with a visual impairment: <https://www.microsoft.com/en-us/seeing-ai>

**Which?'s goal in this work was to understand how far consumers may require further support to rebalance power over use of their data.** Consumers may report feeling disempowered on a number of issues, but in a context where *their own behaviour* may be inadvertently causing themselves harm, and where the reach of the issue is huge, their sense of disempowerment is of particular concern to us. We scrutinised the following questions:

**1. Do people understand how their data is collected, traded, and used, and how this may affect their choices? And if not, why not?**

We partnered with research agency BritainThinks to conduct major new primary research with consumers to understand awareness and perceptions of data collection and the potential impacts it has on them. Where there were gaps in knowledge, we looked at why businesses might have allowed this to happen, and what might be the unforeseen implications of this for them.

**2. Is it possible for consumers to take control of their data, once they understand more?**

To answer this question, we looked at the commercial drivers of data collection and use, and the market structure in data commercialisation, to understand more about the practices of the organisations consumers are dealing with and the motivations for those practices.

**3. How far will recent and upcoming policy changes rebalance the playing field in the consumer's favour?**

In today's context of major regulatory change in the form of the GDPR and EU e-Privacy Directive, we interviewed data protection and competition policy experts to tentatively explore how far those policy frameworks respond to the changing way consumer data is commercialised, and address any transparency and control shortfalls.

We set out to understand the reality of consumer-data collection and use, and to compare this to consumer knowledge and perceptions of the data ecosystem. We have found out consumer views across the length and breadth of the country through:

- Constructing one of the most detailed segmentation analyses of the UK population's data attitudes and behaviour.
- Four in-depth deliberative face-to-face workshops with in Leeds, Newport, Perth and St Albans.
- In-depth interviews with more vulnerable consumers.<sup>6</sup>

This approach has allowed us to understand consumers unprompted attitudes and behaviour, and then to systematically explore their perceptions of the data ecosystem after explaining how it works and how it could affect them. We believe this is the most wide-ranging and systematic exploration of consumer views on this subject that has been done to date. We have published this research alongside this report as *Control, Alt or Delete? Consumer research on attitudes to data collection and use*.

---

<sup>6</sup> We define 'more vulnerable' in this research as (i) older people aged 80 and over; (ii) people belonging to a lower socio-economic group (DE consumers); (iii) people with a long-term physical or mental health condition/disability; and (iv) people who do not feel confident speaking, reading or writing in English.



# 1. The transparency deficit in consumer data

- **Consumers' attitudes to data are often pragmatic, but the reality of the data ecosystem is an unpleasant surprise.**
- **This growing, untapped potential for 'nasty shocks' risks destroying value for business.**
- **Companies have allowed – even encouraged – this ignorance and confusion among consumers**
- **This meant that we were also unable to get clear and comprehensive answers on some of the important questions about harm.**

## *Attitudes to data versus commercial reality*

### **Baseline attitudes to data**

People recognise the wide range of potential benefits that data-dependent services bring. We can see this every day through their use of new services and innovation. Research on this topic<sup>7</sup> finds that sharing data has become normalised and that consumers (albeit to differing extents) are willing to share their data if they can see a direct benefit to them, a societal benefit, or if it's required for the product or service to function.

However, people's initial reactions to data collection, when given no information about it, are mostly negative. Our research found that 67% of people are not comfortable with the idea of organisations using information that an individual has shared publicly (for example on social media) and 68% aren't comfortable with organisations using information they have gathered from observational methods, such as tracking browsing history.

To navigate this range of views we have used the statistical technique of segmentation, which works by revealing natural groupings within an overall dataset of respondents. Using our nationally representative survey of more than 2,000 consumers<sup>8</sup> we have constructed a segmentation that reflects both consumers attitudes to data use and their behaviours. We have published a digital dashboard showing our segmentation on our website at <https://consumerinsight.which.co.uk/data-dozen>.

---

<sup>7</sup> See *Consumers and their data: Research review*, Which? (2018)

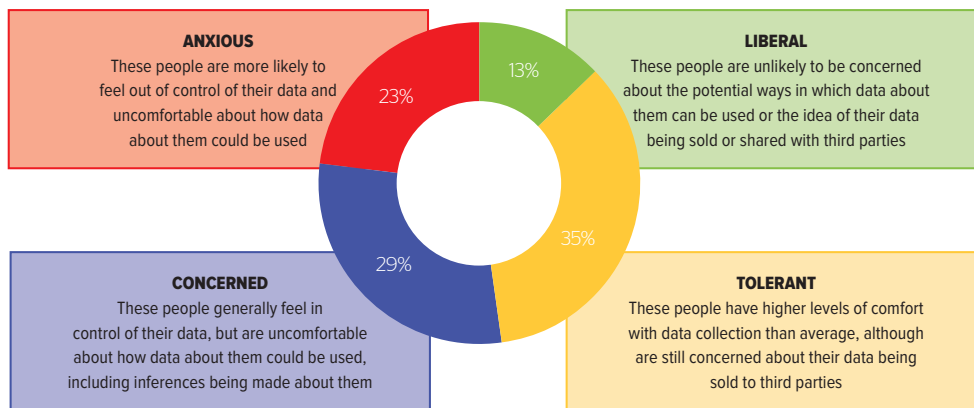
<sup>8</sup> Populus, on behalf of Which?, interviewed 2,064 UK adults by telephone, between 18-28 January 2018. Data was weighted to be representative of the UK population



When it comes to attitudes to data collection and use, we found that:

- Only a small proportion of the population are genuine ‘Liberals’, ie not concerned with almost all practices relating to their data, including the idea of their data being sold or shared with third parties.
- About half of the population are uncomfortable about the use of their data, and are split between those who we have defined as ‘Concerned’ but who still feel in control, and those who we have defined as ‘Anxious’ as they feel they are out of control of their data.
- Just over a third of the population are ‘Tolerant’; they are concerned about their data being sold to third parties but have higher levels of comfort with data collection and use other than by third parties.

Figure 1: Consumer Data segmentation attitudes



People exhibit a wide range of behaviours when it comes to data. For example:

- 19% of the population are taking considerably more action than others to restrict what data can be observed about them and “dirtying” their data by putting incorrect information in forms and using separate email addresses for organisations they do not want to receive communications from.
- However, 24% of the population are characterised by how much more they take advantage of the shortcuts afforded to them online than others (for example saving their bank details in forms and logging into other services using their social media).

Surprisingly, there is a relative lack of a relationship between attitudes and behaviour, which is important when thinking about how to approach communicating with consumers in these segments.

All of these attitudes and behaviours need to be understood in the context of consumers’ knowledge. In particular, our qualitative research found that people generally perceive data collection as a series of single bounded transactions, where individual pieces of data are “given” to an organisation in order to receive a specific product or service. They are mostly not aware of the extent of third-party sharing, or that their data can be amalgamated to form an individual-level profile.

## *Consumer knowledge versus commercial reality*

Over the past few years, people have increasingly seen advertisements and other messages ‘just for them’, primarily online. This has driven up awareness of the fact that their data is used to target them at an individual level, and they correctly and spontaneously identify ‘advertising’ as the main commercial use of their data. However, beyond this, consumer knowledge is limited. People become more concerned as they learn about the other uses of data, how targeting happens and how the use of the data could affect them.

### **The fact that unknown companies ‘profile’ individuals is a significant surprise to most**

Behind the scenes, individual dossiers of information, or ‘profiles’, about almost all consumers are compiled to enable individual-level targeting. Most of this targeting is of advertising and marketing messages, but the profiles can also be used to target prices and other types of information.

There are many sources of this information, but so called ‘third-party cookies’ are a major source. The third parties (data brokers, analytics firms, website optimisation specialists, or other advertising technology firms) secure agreement from the website owner to monitor browsing. Cookies allow companies to profile consumers and follow their responses to adverts to measure effectiveness and improve targeting in the future. The information is stored against a unique identifier for a consumer.

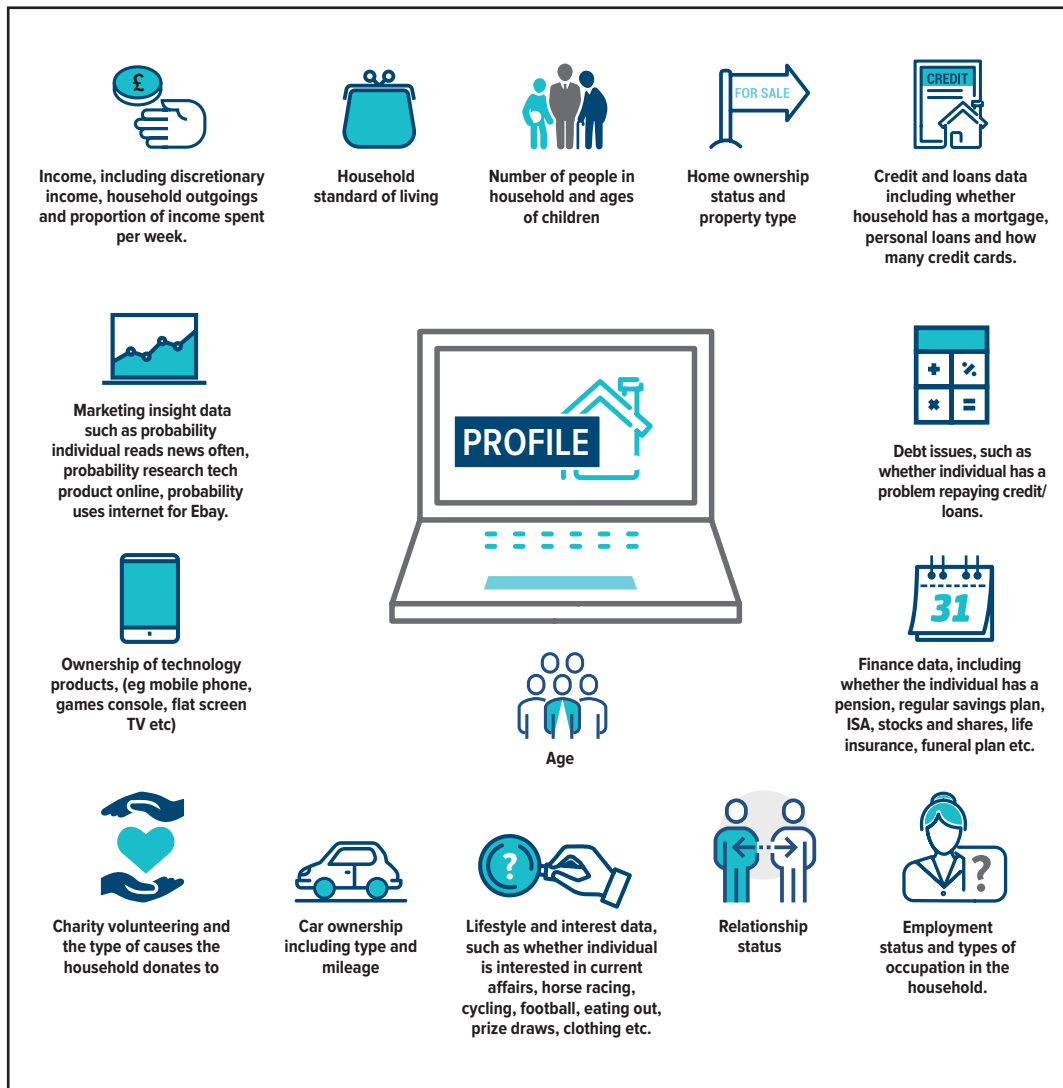
Google and Facebook have a pre-eminent position in the number and depth of consumer profiles. They own billions of ‘persistent identifiers’ (email addresses or accounts) that follow consumer behaviour, not only on their own sites and services, but across the internet, as people tend to stay logged in.

Other firms, often collectively referred to as ‘data brokers’, profile consumers. Sometimes this happens ‘on the fly’ as the crumbs of information held by different brokers are combined in milliseconds to target an advert effectively. But larger brokers store very detailed profiles of information about most individual consumers. They look to combine information on individuals from a variety of sources in order to build profiles that can be sold to advertisers. Many of these companies are little known in the UK, such as Axciom and Oracle – although some companies such as Experian and Equifax, which are relatively widely known as credit-referencing agencies, have similar capabilities.

These profiles include a wide range of information, from information about income and home ownership to relationship status and leisure interests. Examples of the full range are shown in the graphic on the following page.

In non-European countries this information gathering goes even further, including everything from health interests to indicators of religion or sexuality. The profiles do not usually contain an individual’s name, seemingly making them anonymous.

Figure 2: Typical information in a consumer profile compiled by a data broker



However, companies use email addresses and telephone numbers to link data together and then assign profiles unique identifiers. These profiles can then often be re-identified with a minimal number of variables, even if their email address is hashed.<sup>9</sup>

We shared details about the information that data brokers put together with the participants in our workshops. As well as UK practice, we included types of data that are only collected in non-EU countries as well, which allowed us to understand where consumers draw the line between acceptable and unacceptable practices.





<sup>9</sup> The book *Networks of Control* (Christl and Spiekermann, 2016), (<http://crackedlabs.org/en/networksofcontrol>), points out that: 'Apparently, hashing is in fact pseudonymisation rather than anonymisation.' In CMO, Adobe's digital marketing magazine, 'leading privacy lawyer' Ruth Boardman suggests that 'marketers should stop trying to convince themselves they are working with anonymised data, rather than personal information'

The majority of consumers envisage that the use of data about them is relatively generic, anonymised, and specific to a single transaction with a product or service. When learning more about data profiling, most are surprised about the extent and detail of their ‘digital self’.

For some, this crosses the line of acceptability by making them feel they are not in control of information about themselves. Many participants reported feeling uncertain about how this amount of information could be being used in their best interests, and that they felt their privacy has been invaded.

At this point in the workshops, many of those whose starting perceptions fitted with the ‘Tolerant’ and ‘Concerned’ segments started to become more negative.

Figure 3: Typical reactions of our research participants to finding out more about data profiling

 <p><b>Control</b></p>	<p>Many struggle to identify how these different types of information have been collected, causing them to feel out of control and that this happened without their consent.</p>
 <p><b>Relevance</b></p>	<p>The level of detail included in many data profiles goes far beyond what most consider to be relevant to the functioning of a specific product or service.</p>
 <p><b>Benefits</b></p>	<p>Beyond basic demographics and 'consumption' information which might be used for marketing, advertising and tailoring, people question how this information can be used in their best interests.</p>
 <p><b>Harms</b></p>	<p>Some see potential for discrimination on the basis of 'sensitive' information including ethnicity, religion, sexuality and health data when this is de-anonymised and appended to other data.</p>

*‘A key turning point on the acceptability of sharing my data in my mind are things that are attributable – so where I am identified. But where it is not attributable and identifies me as part of a wider population, that is acceptable. It becomes questionable when it identifies the individual, as it becomes intrusive, and can be used inappropriately and illegally.’*

Workshop participant, Perth

**Targeting of information and prices beyond adverts is unexpected**

Most people have some awareness that their data is used for targeting and tailoring, as the outcomes of this are visible to them in marketing and advertising through targeted adverts, personalised recommendations on services such as Netflix or Spotify and personalised vouchers through loyalty cards.

Spontaneously, most people feel that the targeting and tailoring of adverts and recommendations is positive, as it enhances services by increasing the relevance of the content they are shown. At worst, targeted adverts were seen as ‘irritating’, especially when they were irrelevant or were outdated.

Overall, however, it was felt that targeted adverts and recommendations were innocuous, and if a person didn't like them they could just be ignored. Additionally, there is little concern that targeted adverts and recommendations will restrict choice, as people felt that online adverts and recommendations weren't their only source of insight into the products and services available to them – they would also receive recommendations from family, friends and the media.

However, data could be used for targeting in other ways too, such as:

- tailoring information: such as different search engine results depending on previous search history or location.
- micro-segmentation: such as an insurance company providing a different motor insurance quote based on data as widely varying as data gained from a social media profile<sup>10</sup> to driving behaviour data from an in-car device.
- personalised prices: for example in-app purchases in mobile computer games, that can be varied based on a player's behaviour in the game or their characteristics.

When participants were introduced to examples of personalised information and pricing they had fundamental and deeply held concerns. The vast majority felt that targeting information deliberately denies them access to other information which may be relevant to them, and which they consider to be their “democratic right” to access, such as opposing views from political parties. There was concern that personalised information could lead to the potential for powerful groups and organisations to manipulate public opinion, particularly in relation to news, politics and elections.

Similar to concerns about personalised information, people believe that the ‘stakes’ are higher in relation to personalised pricing. For many, personalised pricing undermines their sense of control as information is used without their knowledge, and there is concern that the information used would be irrelevant. For example, they felt that information from their social media account or about their email provider would not be an accurate predictor of their risk profile.

They were particularly concerned with examples where they were unaware of personalised pricing, because they were unable to take action against it. They also felt that it undermined their consumer choice, as it is impossible to act as a savvy consumer – eg by shopping around and looking for the best prices and deals – if you have no way of knowing whether you are being shown the lowest prices.

---

<sup>10</sup> This has been proposed by a UK based insurer, but was not implemented. ‘Admiral to price car insurance based on Facebook posts’: <https://www.theguardian.com/technology/2016/nov/02/admiral-to-price-car-insurance-based-on-facebook-posts>

**The use of data science to ‘infer’ aspects of their profile makes people uneasy**

Although a lot of data exists about consumers, companies also try to fill the gaps in their knowledge by using statistical modelling techniques to infer likely consumer characteristics. For example, a 2014 report<sup>11</sup> by the US Federal Trade Commission found that data brokers collect offline and online consumer data from many sources in order to build marketable profiles of individuals that make inferences about factors such as a person’s age, income or interests.

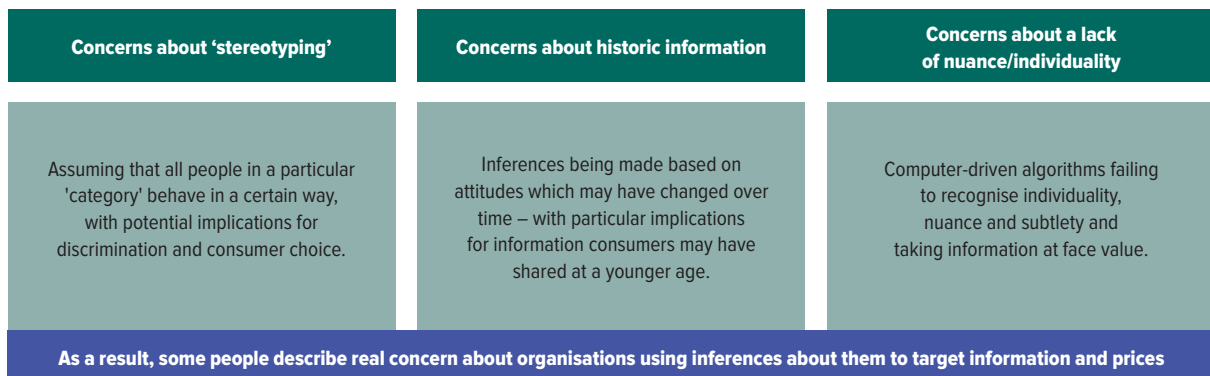
This process can be done through statistical modelling. While this can in some cases be surprisingly accurate,<sup>12</sup> the data is often wrong. Deloitte<sup>13</sup> asked 107 of its US professionals to privately and anonymously review their personal data made available by a leading consumer data broker. More than two-thirds of them stated that the third-party data about them was less than 50% correct as a whole.

Most of the people in our deliberative groups initially saw it as relatively acceptable and unavoidable that organisations might make assumptions about them based on their information, largely reflecting the perception that this can result in benefits such as targeted recommendations for products.

However, on learning more, the extent to which these inferences are made is a surprise that crosses a line for many. Significant proportions express concerns about organisations inferring things about them that they would not want to share or to be collected.

Underlying these concerns is scepticism about the extent to which inferences are likely to be accurate. For some people this is a real concern, while for others it is a hopeful sign that organisations don’t know all that much about them. People identified a number of reasons why inferences may be reductive or inaccurate, often relating to concerns about the principle of being stereotyped or ‘put into a box’.

**Figure 4: Consumer concerns about inferences made about them by organisations**



11 Federal Trade Commission, (2014), *Data brokers: A call for transparency and accountability*, available at: <https://www.ftc.gov/system/files/documents/reports/data-brokers-call-transparency-accountability-report-federal-trade-commission-may-2014/140527databrokerreport.pdf>

12 W Youyou, M Kosinski and D Stillwell, (2015), *Computer-based personality judgments are more accurate than those made by humans*, PNAS, 112(4), pp1036-1040, 12 Jan, available at: <http://www.pnas.org/content/112/4/1036>

13 Deloitte, (2017), *Predictably inaccurate: The prevalence and perils of bad big data*, available at: <https://www2.deloitte.com/insights/us/en/deloitte-review/issue-21/analytics-bad-data-quality.html>

Underlying many people's concerns of personalised pricing and information was the accuracy of the information being used to tailor and target such prices and information. Those who have limited faith in the accuracy of algorithmic inferences, and feel that demographic information is likely to stereotype them, are often concerned that these forms of personalisation will lead to unjust outcomes that they are unable to challenge or rectify, because they don't know what assumptions organisations are making about them.

**A large majority of the population are concerned about selling data to third parties**

Third-party selling is a concern across the majority of the population. In our nationally representative survey (see footnote 8 for details) we found that 81% of consumers would be concerned if organisations are selling anonymised information about an individual to a third party. In our segmentation, only those who hold 'Liberal' attitudes are not concerned, consisting of just 13% of the population.

Most of our research participants' negative perceptions were intensified when they were told about the extent of sharing within the data ecosystem, leaving them feeling that:

- They don't have control over where their data goes.
- It is made purposefully difficult for them to opt out of their data being shared.
- Data they consented to give in one context is being used in another, which they wouldn't have given consent for if asked.

*'I think it's unacceptable that organisations are able to sell my information on to third parties. I don't mind organisations having it when I've agreed to it, but I don't trust them when they start selling it on to other organisations.'*

Workshop participant, Leeds

**People are surprised by the scale of data theft, and the scope of data monetised by criminals**

Many people appear to view it as inevitable that their data will be stolen. Unlike data sharing, security of information was rarely front of mind when we asked people to think about their data. This partly reflects a broader belief that the wider benefits far outweigh any potential risks to their security, but also growing fatalism among consumers and a perception that securing their data is beyond their control. Many particularly struggled to engage with the security risks beyond the risks to their financial information, and even this feels abstract and low priority in a context where consumers believe they will be reimbursed by their bank, building society or payment provider for any financial losses.

In many ways this resigned attitude seems to be realistic, as corporate and law enforcement actors are losing the battle against breaches and UK consumers' exposure to breaches is significant (see our later section on the unknowable black market).



When people understood more about their own personal exposure to breaches during our deliberative workshops, the sources surprised them in two ways:

- Information had been stolen from ‘reputable’ providers who hadn’t told their customers that they had suffered a breach (either directly, or indirectly, through the media) – for example, there was some surprise in seeing references to accounts held with organisations such as LinkedIn and Yahoo being compromised.
- Information about them had been stolen from unfamiliar organisations that they had not heard of before and who they had not consented to hold their data – this included data brokers.

### ***Post-knowledge attitudes to data***

When the data ecosystem is explained to people it becomes clear to them that their information travels beyond the bounded transactions that they imagine existed. There was surprise that this was allowed to happen, and some people assumed that regulations would not allow such practices. Particularly concerning was the role of data brokers who are seen to typify the murky world of data collection and use in the fact that they are ‘hidden’ from consumers, and that consumers don’t have control over their data. Almost all people have concerns about third-party selling and some have concerns about assumptions being made about them and invasion of privacy.

### ***Why this matters for business***

People get more nervous the more they learn, and they have plenty still to learn. Indicators are that this is a dangerous combination for business. The combination of anxiety and a large potential for ‘nasty shocks’ is destroying value, and having a negative effect on even those businesses that do not engage in bad practice. There are at least three ways that the current environment hurts businesses as well as consumers:

- consumer self-help
- stifling the adoption of new services
- corporate public relations disasters.

### ***Consumer self-help: dirtying data, ad blockers and more***

Consumer anxiety and responses to shocks lead to an environment where it is harder for businesses to get and trust data. Many behaviours illustrate this:

- 27% of consumers say that, in the last three months, they have at least sometimes used a different email account for websites that they do not want to receive communication from.



- 14% say that, in the last three months, they have at least sometimes deliberately given incorrect information on a form (thereby ‘dirtying’ their data).<sup>14</sup>
- More than one in five of the UK population have installed ad blockers<sup>15</sup> and other software that tries to stop data collection.
- There are a range of services that offer privacy as one of their main selling points. This includes the Signal and Telegram messaging apps,<sup>16</sup> the search engine DuckDuckGo,<sup>17</sup> the browser Firefox<sup>18</sup> and the new Brave browser<sup>19</sup> which explicitly states that it allows you to ‘browse faster by blocking ads and trackers that violate your privacy and cost you time and money’.

It is unsurprising in this context that the data businesses hold on consumers is often incomplete and incorrect.

### *Inhibited adoption of services*

People appear to be resigned to the collection of their data as the necessary cost of getting access to the goods and services they want, and are not opting out en masse. However, there are several signs that consumer concern is inhibiting the adoption of new services and stifling the growth of whole new business models.

For example, in January 2018 Open Banking was launched. The CMA hopes that this will ‘harness the technological changes which we have seen transform other markets, [with customers] able to access new and innovative apps which will tailor services, information and advice to their individual needs’.<sup>20</sup> However, Mintel reports that 87% of consumers are concerned about sharing their financial data<sup>21</sup> and a recent report by the Financial Services Consumer Panel found that people who did not use third-party service providers were concerned about the potential consequences of losing control of their data, in particular the risk of security breaches, malicious activity or it ending up in many different hands.<sup>22</sup>

In the energy sector, the roll-out of smart meters is also facing consumer opposition. In 2016 an academic paper on the British public’s perception of the UK smart-metering initiative<sup>23</sup> lists privacy concerns as one of the threats to the roll-out. A recent survey suggests that 53% of people without a smart

---

14 Populus, on behalf of Which?, interviewed 2,064 UK adults by telephone, between 18 and 28 January 2018. Data was weighted to be representative of the UK population

15 The Drum, (2017), *Adblocking problem ‘yet to be solved’ as one-fifth of UK internet browsers install software this year*: <http://www.thedrum.com/news/2017/06/08/adbblocking-problem-yet-be-solved-one-fifth-uk-internet-browsers-install-software>

16 Signal (2018), website homepage; Telegram (2018), website homepage

17 <https://duckduckgo.com>

18 <https://www.mozilla.org/en-GB/firefox>

19 <https://brave.com>

20 Gov.uk, (2016), *CMA paves the way for open banking revolution*, available at: <https://www.gov.uk/government/news/cma-paves-the-way-for-open-banking-revolution>

21 Mintel, (2018), *Attitudes toward data sharing, UK*

22 Consumer Panel Position Paper: *Consenting adults? - consumers sharing their financial data*, available at: [https://www.fs-cp.org.uk/sites/default/files/final\\_position\\_paper\\_-\\_consenting\\_adults\\_-\\_20180419\\_0.pdf](https://www.fs-cp.org.uk/sites/default/files/final_position_paper_-_consenting_adults_-_20180419_0.pdf)

23 Buchanan, Banks, Preston, Russo, (2016), *The British public’s perception of the UK smart metering initiative: Threats and opportunities*, available at: <https://www.sciencedirect.com/science/article/pii/S0301421516300039>

meter have no desire to get one in the next year,<sup>24</sup> and regular newspaper stories report both privacy and security fears<sup>25</sup> with their installation.

## Corporate PR disasters

The lack of consumer knowledge about data collection and use also creates a big risk of a public relations disaster for companies when consumers are surprised by how their data is used or treated.

As we were finalising this project the revelations that the data of up to 87m Facebook users (including 1.1m in the UK) was improperly shared with political consultancy Cambridge Analytica came to light.<sup>26</sup> This damaged Facebook in a range of ways. Following the story:

- Facebook's stock price dropped almost \$100bn in the two weeks following the breaking of the story on 17 March 2018.<sup>27</sup>
- In a nationally representative survey<sup>28</sup> that we conducted in the aftermath, we found that:
  - 89% of the UK population said they had concerns about it.
  - 60% were concerned that there seemed to be little effective regulation over what Facebook or Cambridge Analytica were doing with their information.
  - 6% of those who used Facebook say they deleted or deactivated their accounts shortly after the story broke, with a further 24% saying they were thinking about it, but hadn't yet followed through.
  - 34% of those who used Facebook said they had altered their privacy settings, and 33% said they had restricted their permissions.
  - 21% of those who used Facebook said they are using Facebook less.

In his book *How to fix the future*,<sup>29</sup> author and tech entrepreneur Andrew Keen examines the case of the American car industry in the 1950s and 60s. In the mid-fifties, the big three American carmakers controlled 96% of the American market. However their intense competition drove the production of cars that would look good but wear out quickly, and safety was not a priority – between 1961 and 1966, fatalities on American roads increased by 38%. However, the car companies' dominance looked assured.

This was punctured by the publication of *Unsafe at Any Speed* by Ralph Nader, which drew public attention to the terrible safety record of the cars they were driving and led to increased automotive regulation. Unfortunately

24 CityAM, (2018), *Half of UK consumers without a smart meter say they don't want one – despite high levels of satisfaction*, available at: <http://www.cityam.com/285156/half-uk-consumers-without-smart-meter-say-they-dont-want>

25 For example, <https://www.thetimes.co.uk/article/millions-of-homeowners-reject-smart-meters-over-hacking-fear-rhbm98ps2>; <http://www.thisismoney.co.uk/money/bills/article-5600925/Stop-smart-meter-bullying-tricks-power-firms-use-force-switch-digital-meters.html>

26 *Facebook scandal 'hit 87 million users'*: <http://www.bbc.co.uk/news/technology-43649018>

27 [https://ycharts.com/companies/FB/market\\_cap](https://ycharts.com/companies/FB/market_cap)

28 Populus, on behalf of Which?, surveyed 2,068 UK adults online, between 26 and 27 March 2018. The data was weighted to be representative of the UK population

29 *How to fix the future: Staying human in the digital age* (Andrew Keen, Atlantic Books, 2018)

the American car industry was not ready for this, but other countries such as German carmakers were. By 2017, Keen points out that the big three US carmakers now make up only 45% of the US market.

Consumer behaviour can be difficult to shift. However, the combination of scandal, regulation and consumer-friendly innovations can drive huge changes in value. Taking measures to respect customers' privacy and handle their data in thoughtful ways benefits consumers and businesses alike.

### ***The commercial context to opacity***

Businesses have allowed, even encouraged, consumer ignorance and confusion about how their data is collected and used, as their business practices have quietly tested the boundaries of ethics, legality and consumer acceptance.

The digital advertising market is important context to understand for several reasons: it has driven much of the innovation in the use of consumer data; it has allowed consumers unprecedented access to free content and services; and the high stakes competition for advertising budgets sets the scene for corporate behaviour with data.

### ***The scale and structure of the digital advertising market***

Individual-level consumer data is commercialised in a number of ways, including risk-mitigation products like fraud detection. But the commercial application that is shaping the ecosystem is the targeting of brand and marketing messages, and it is the advertising and marketing technology industries (referred to as 'adtech' and 'martech') that have been enabling the ever-finer profiling, targeting and measurement of consumers.

The technology giants are arguably the most data-generative companies in the world, and advertising comprises roughly one quarter of their combined annual revenues. This revenue stream is overwhelmingly concentrated in Google and Facebook.<sup>30</sup>

It is important to understand the scale and growth of the digital advertising market to appreciate the commercial context for many of the firms that collect, trade and process consumer data. The global digital advertising market is worth around \$230bn,<sup>31</sup> and in 2017 analysts found it had overtaken TV advertising.<sup>32</sup> Growth is still accelerating, even in the most mature markets, like the US where annual growth is currently at around 20% a year.<sup>33</sup>

---

30 UK Business Insider, (2017), *The tech industry is dominated by 5 big companies – here's how each makes its money*: <http://uk.businessinsider.com/how-google-apple-facebook-amazon-microsoft-make-money-chart-2017-5>

31 eMarketer, (2017), *Worldwide ad spending: eMarketer's updated estimates and forecast for 2016-2021*, available at: <https://www.emarketer.com/Report/Worldwide-Ad-Spending-eMarketers-Updated-Estimates-Forecast-20162021/2002145>

32 eMarketer, (2016), *Digital ad spending to surpass TV next year*, available at: <https://www.emarketer.com/Article/Digital-Ad-Spending-Surpass-TV-Next-Year/1013671>; Recode, (2017), *2017 was the year digital ad spending finally beat TV*, available at: <https://www.recode.net/2017/12/4/16733460/2017-digital-ad-spend-advertising-beat-tv>

33 Mary Meeker's *Internet trends* report, (2017), available at: <http://www.kpcb.com/file/2017-internet-trends-report>

There are numerous economies of scale for technology firms operating in this business, which means high margins for the winners. This, combined with unabated growth, has created excellent returns for the winning firms, for the firms they buy, and for their investors.

As discussed in the next section, data science has played a big role in persuading advertisers to switch their budgets to digital. To contextualise the scale of this market, it is illustrative to compare digital advertising with another market where growth attracted data scientists and begat complexity and opacity. At the height of investor interest in the ‘adtech’ boom, Martin Kihn, vice president at Gartner Research, reflected on comparisons between ‘real-time bidding’ of advert placements and high frequency trading on Wall Street:<sup>34</sup>

*‘There’s a perception that Wall Street is trading real money while advertising is in the kids’ room using Monopoly cash... A recent estimate of the total commissions paid to Wall Street banks last year for stock market trades was about \$10 billion. So RTB [Real-Time Bid] trading is only one-fifth the size of stock trading? Maybe. But the ultimate prize is much, much bigger. Total global media spending is approximately \$450 billion. If any significant portion of that ends up on real-time bid exchanges – as seems all but inevitable – then you’re not doing too much fancy math to project that in a very few years ad tech brokers will be making more funny money than Wall Street brokers.’*

*‘Yes, I said more. That’s the prize here, my friends. That’s why all the data scientists and enterprise data warehouse barons and salesforce automation women and electric gangsters are rushing – rushing – into the arms of advertising. It’s not because they thrive on understanding customers.’*

The scale of this market, augmented by capital subsidies as investors pursue the potential returns, has enabled consumers to enjoy unprecedented access to free content and services.

The structure of this market is consolidating. Of a global digital advertising market of \$228bn, we estimate that Google and Facebook’s advertising revenues comprised 59% of advertisers’ total spending in 2017, and eMarketer has estimated that they commanded 54% of the UK digital advertising market.<sup>35</sup> They made 88% and 97% respectively of their revenues from advertising in 2016,<sup>36</sup> and account for the majority of the growth in the industry.<sup>37</sup>

34 Kihn, Martin, (2014), *Why ad tech is more complicated than Wall Street*, available at: <https://blogs.gartner.com/martin-kihn/ad-tech-worse-wall-street>

35 Which? calculations, based on: global digital advertising market 2017: \$228B (<https://www.emarketer.com/Report/Worldwide-Ad-Spending-eMarketers-Updated-Estimates-Forecast-20162021/2002145>); Google 2017 global ad revenues (p28 of *Form 10K*, [https://abc.xyz/investor/pdf/20171231\\_alphabet\\_10K.pdf](https://abc.xyz/investor/pdf/20171231_alphabet_10K.pdf)); Facebook 2017 global ad revenues (<https://investor.fb.com/investor-news/press-release-details/2018/Facebook-Reports-Fourth-Quarter-and-Full-Year-2017-Results/default.aspx>). UK figure from Emarketer: *Digital duopoly to remain dominant in UK ad race*: <https://www.emarketer.com/Article/Digital-Duopoly-Remain-Dominant-UK-Ad-Race/1016481>

36 UK Business Insider, (2017), *The tech industry is dominated by 5 big companies – here’s how each makes its money*: <http://uk.businessinsider.com/how-google-apple-facebook-amazon-microsoft-make-money-chart-2017-5>

37 Estimates vary; examples are 74% (<https://www.fool.com/investing/2017/03/17/google-and-facebook-will-account-for-74-of-digital.aspx>) or 85% of US growth (Mary Meeker’s Internet Trends report, (2017), <http://www.kpcb.com/file/2017-internet-trends-report>)

This trend has caught even leading analysts by surprise. As recently as 2015, eMarketer reports significantly underestimated Google's and Facebook's growth potential.<sup>38</sup> The competitive dynamics around what is commonly referred to as the 'digital ads duopoly' set the scene for how consumer data gets used to attract advertising budgets.

### *The role of data in the competition for advertising budgets*

From the earliest days of the digital advertising market, consumer data has been vital to its growth. As digital sales teams competed with their 'offline' counterparts to attract advertising budgets online, the value proposition centred on the superior targeting and measurement possible in a digital environment: an advertiser cannot easily confirm how many people noticed a message on a billboard, but they can easily monitor how many people clicked on a banner ad.

However, for a long time, much targeting was still done on the basis of crude inference based on the context (eg an audience reading sports content web pages should be shown adverts for men's shaving products).

From 2014, individual-level consumer data took centre stage, with the advent of 'people-based marketing'. This term was popularised by Facebook in 2014<sup>39</sup> and has since become part of the advertising industry lexicon. This approach to marketing promises to:

- Target on the basis of an individual's profile: In this case, the most important attribute is often an explicit signal of current interest in a product, for example, visiting a product page.
- Show them the advert wherever they are browsing (ie, a woman reading the sports website would no longer be shown the men's shaving products, but products relevant to her).
- Follow that person's response all the way to their purchase.

Google and Facebook have a strong advantage in the delivery of one-to-one targeted advertisements. Google has over a billion Gmail accounts,<sup>40</sup> and Facebook has two billion registered users. The identifiers help Google and Facebook profile users on basis of behavioural data across any of the millions of websites containing Google or Facebook code,<sup>41</sup> and also via the large amounts of various types of mobile data that they collect.<sup>42</sup>

Their advertising platforms serve ads in many third-party websites too, so they can reach that person with the advertisement, no matter where they happen to be browsing.

---

38 Fast Company, (2017), *Witness Google's and Facebook's insane digital ad dominance in these two charts*, available at: <https://www.fastcompany.com/40512141/witness-google-and-facebooks-insane-digital-ad-dominance-in-these-two-charts>

39 [https://digiday.com/wp-content/uploads/2016/12/WTF\\_PeopleBasedMarketing\\_Final\\_V4.pdf](https://digiday.com/wp-content/uploads/2016/12/WTF_PeopleBasedMarketing_Final_V4.pdf)

40 <https://techcrunch.com/2016/02/01/gmail-now-has-more-than-1b-monthly-active-users>

41 S Englehardt and A Narayanan, (2016), *Online tracking: A 1-million-site measurement and analysis*, available at: [http://randomwalker.info/publications/OpenWPM\\_1\\_million\\_site\\_tracking\\_measurement.pdf](http://randomwalker.info/publications/OpenWPM_1_million_site_tracking_measurement.pdf)

42 The Guardian, (2017), *Are you ready? Here is all the data Facebook and Google have on you*, available at: <https://www.theguardian.com/commentisfree/2018/mar/28/all-the-data-facebook-google-has-on-you-privacy>



Google, historically, had lower commercial imperative to leverage these identifiers for targeting and measurement. The explicit intent data harvested when consumers search (for example, for ‘Hotel in Paris’) is more than good enough to enable one-to-one targeting. In many cases online advertisers only pay when someone clicks or ‘converts’ in some other way, and consumers are also more ready to ‘click’ when they are searching.<sup>43</sup>

It was more challenging for Facebook to demonstrate to investors that it could create an advertising business with comparable economics because:

- It targets with less explicit and more fragmented signals. It matches advertisers with its target customers via thousands of proxy signals that they might be ‘in market’ for that service. So, a potential customer for ‘Hotel in Paris’ might be targeted on basis of visits to travel blogs, the likelihood of having sufficient disposable income, or being in the appropriate demographic. Those signals will be created across more than one device, and in the offline world too.
- Its audience is not shopping or booking a service when they are in Facebook. Like all display advertisements, Facebook’s newsfeed and other display ads are served at moments when a person might not be thinking about shopping, or booking a service – the advert merely plants a seed for later. So, unless Facebook follows that consumer’s response to an advert for a while, it can be difficult to demonstrate ROI to advertisers in the way Google can. It also makes it more difficult to make click-based revenues.<sup>44</sup>

Facebook’s answer was people-based marketing, and the growing prevalence of this approach is in many ways a story of strategic move and counter-move between Google and Facebook (see the timeline on the next page).<sup>45</sup>

The remaining 41% of global digital advertising budgets are increasingly fought over by a more fragmented market of advertising technology and data vendors, including advertising agencies, demand-side platforms, advertising exchanges and supply-side platforms that serve adverts to consumers on websites including publishers.

These companies collectively purport to offer advertisers the same sophistication of targeting that can be achieved on Google and Facebook, at a lower price<sup>46</sup> and/ or with greater transparency. To achieve this, they

---

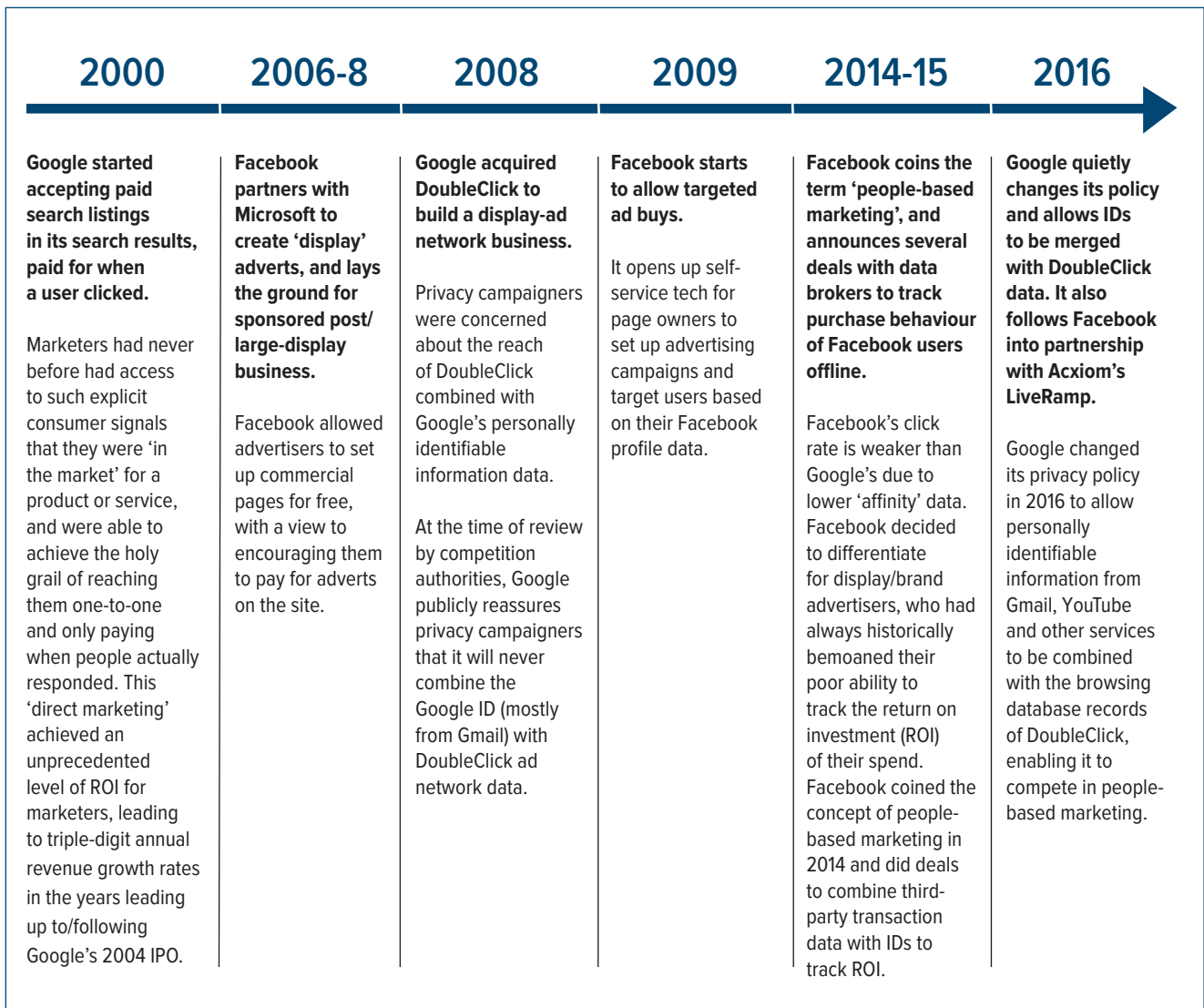
43 HubSpot, (2018), *What’s a good clickthrough rate? New benchmark data for Google AdWords*, available at: <https://blog.hubspot.com/agency/google-adwords-benchmark-data>

44 Slate, (2013), *Facebook followed you to the supermarket*, available at; [http://www.slate.com/articles/technology/technology/2013/03/facebook\\_advertisement\\_studies\\_their\\_ads\\_are\\_more\\_like\\_tv\\_ads\\_than\\_google.single.html](http://www.slate.com/articles/technology/technology/2013/03/facebook_advertisement_studies_their_ads_are_more_like_tv_ads_than_google.single.html)

45 Sources for timeline on p31: <https://www.statista.com/statistics/266249/advertising-revenue-of-google>; <https://blog.hubspot.com/marketing/history-facebook-adtips-slideshare>; [https://digiday.com/wp-content/uploads/2016/12/WTF\\_PeopleBasedMarketing\\_Final\\_V4.pdf](https://digiday.com/wp-content/uploads/2016/12/WTF_PeopleBasedMarketing_Final_V4.pdf); <https://www.cnet.com/news/privacy-concerns-dog-google-doubleclick-deal>; <https://www.smartinsights.com/internet-advertising/internet-advertising-analytics/display-advertising-clickthrough-rates>; <https://adexchanger.com/online-advertising/google-adds-cross-device-metrics-to-doubleclick-partially-answers-facebooks-people-power>; <https://www.propublica.org/article/google-has-quietly-dropped-ban-on-personally-identifiable-web-tracking>

46 NewCoShift, (2017), *Lost context: How did we end up here?*, available at: <https://shift.newco.co/lost-context-how-did-we-end-up-here-fd680c0cb6da>

Figure 5: The development of people-based marketing



trade in discrete bits of data known as ‘keys’ to build ‘identity graphs’ or ‘device graphs’ (for example, associating a mobile phone device ID with a desktop computer ID, a phone number and an email address), which are then used to build profiles of the consumer.

Having attracted billions in venture funding for years, this sector has fallen out of investor favour as the Google-Facebook duopoly has consolidated its hold on the market.<sup>47</sup> This has driven consolidation – much of it driven by larger ad tech and data brokers like Oracle, Acxiom and Adobe – with the remainder facing an increasingly hostile competitive environment.

47 Ad Tech funding drops in face of Facebook-Google duopoly, available at <https://www.ft.com/content/c4c358ca-c6af-11e6-8f29-9445cac8966f>

## ***How data reinforces ‘winner takes all’ market dynamics***

Google and Facebook already have substantial power in the digital advertising industry, and gain more each year. Consumer data is important to this trend. They entered the digital advertising market with a vertically integrated position across:

- Data assets, with detailed and long historical profiles reliably compiled around a persistent individual identifier.
- Advertising technology, that enables them to own both the publisher and advertiser relationship.
- Consumer ‘attention’/ media assets. Although both companies dispute the contention that they are media companies,<sup>48</sup> they are the dominant channels to information in the digital age. The average US adult spends over 40 minutes per day in the Facebook newsfeed,<sup>49</sup> and YouTube has now overtaken television in total viewed minutes in the US.<sup>50</sup>
- Third-party media, via other content companies embedding Google and Facebook code in their sites (whether to show adverts or offer other consumer-facing functionality, such as logins).

This combines to offer advertisers one-stop access to massive audiences that can be finely targeted, with granular post hoc measurement. This ability to micro-target almost the entire world’s online population has proven to be an irresistible value proposition to advertisers who put more and more of their budgets through the platforms every year.

It appears to be a self-reinforcing combination in several ways. The more advertising they serve, the smarter their data assets and targeting technology become, the more bidders they attract to their platforms, and the more their business efficiency improve. This enables them to win greater distribution of their advertising platform across third party properties, which allows them to track more consumer behaviour online, further improving their data assets.

## ***Revealed preferences towards commercial transparency on data***

In a situation where the competitive stakes are high and consumer cooperation is vital to winning the prize, businesses have faced a choice: either take the high road and secure informed consent to their handling of data, or keep consumers uninformed in the hope that they stay docile while the race is on.

It is undeniably difficult to achieve useful transparency on a topic as complex as data processing. That the data and advertising technology industries are

48 CNBC, (2018), (<https://www.cnbc.com/2018/04/11/mark-zuckerberg-facebook-is-a-technology-company-not-media-company.html>) and BBC, (2017) <http://www.bbc.co.uk/news/av/technology-41251246/google-news-boss-we-re-not-a-media-company>)

49 eMarketer, (2017), (<http://www.emarketer.com/Chart/Average-Time-Spent-per-Day-with-Facebook-Instagram-Snapchat-by-US-Adult-Users-of-Each-Platform-2014-2019-minutes/211521>)

50 Fortune, (2017), *YouTube could be about to overtake TV as America’s most watched platform*, available at: <http://fortune.com/2017/02/28/youtube-1-billion-hours-television>



unintelligible has now become a commonplace observation, even among industry insiders. An editorial in trade publication *AdExchanger*<sup>51</sup> comments:

*'If policymakers, academics, privacy researchers and technologists have trouble making sense of the cross-device landscape, what chance does the average consumer have? ... The industry has a serious problem with opacity. Making more information available – what amounts to either meaningless marketing-speak or reams of befuddling techno-speak on how specific ad networks and exchanges operate – arguably obfuscates rather than educates.'*

But companies have had few incentives to try: many of the practices which concern people have not been strongly tackled by regulators; and consumers can respond anxiously to further information about how their data is collected and used. It is unsurprising that the vast majority of firms have avoided proactively informing consumers.

This stance is most starkly exemplified in corporate responses to security breaches. Forrester Research provides a series of analyses of the world's highest profile data breaches. In our view, the research finds corporate transparency and support of consumers to be severely wanting, even when those consumers have been exposed to potentially severe harm.

High-profile consumer brands, like Yahoo! (0.5bn records breached) have good opportunities to inform and educate their users on the implications and handling of the breach. However, they rarely do so. In the case of Yahoo!, perceived corporate interest in downplaying the situation led to it failing to use its numerous direct channels and its considerable reach to inform consumers.<sup>52</sup>

In the case of data aggregators and brokers, Forrester identifies a 'huge systemic risk' in the growing market for data, in which 'humans are treated as a commodity' and countless numbers of companies are entering the business of aggregating and selling data as an incremental revenue stream. It comments strongly on the Equifax breach:<sup>53</sup>

*'Every individual affected by the Equifax breach made a false assumption when they first heard about it: they thought it was all about them. When the company's response showed little remorse, let alone an attempt to make things right, most were shocked to realise that Equifax didn't care about them. It turns out that the individuals affected are not Equifax's customers: they were simply the raw materials used to manufacture products. The disdain that Equifax executives showed for the victims of its breach reflects much larger systemic problems in our data economy.'*

---

51 Adexchanger, (2015), *The FTC seeks clarity on cross-device tracking – But opt-out remains a murky mire*, available at: <https://adexchanger.com/data-exchanges/the-ftc-seeks-clarity-on-cross-device-tracking-but-opt-out-remains-a-murky-mire>

52 Balouras, Pollard, Shey, Cser, Hayes, September 2016, Forrester report, *Quick take: Lessons for security and risk pros from the Yahoo breach*

53 De Martine, Pollard, Shey, December 2017, Forrester report, *Equifax Exposed Two Massive Systemic Risks*

This analysis is reinforced by findings from our deliberative workshops. In these workshops, participants were given the opportunity to experiment with an online tool called Have I Been Pwned? to see if their data had been breached. Despite their sense of fatalism, most found that they had underestimated the amount of information about them that had been breached, and were surprised to see both familiar and unfamiliar names in the list of organisations that their data had been stolen from.

### ***Impact on policymakers: intractable open questions***

The lack of transparency in the data industry makes it difficult for consumers, their advocates and policymakers to get beyond isolated anecdotes and (i) understand the scale of potential detriment or (ii) provide good public information about what is happening.

There are a number of open questions, of potentially material impact, that are too difficult to answer systematically:

#### **1. What proportion of consumers have their sensitive data breached?**

Breach statistics are notoriously difficult to compile. The Crime Survey for England and Wales 2016-17 (CSEW16-17<sup>54</sup>) indicated that there were 468,000 victims of unauthorised access to personal information (including hacking) over the year (October 2016 to September 2017). This equates to 1.2% of the adult population of England and Wales.

However, in November 2017, Uber revealed that it had suffered a data breach in 2016 that involved the compromise of data of approximately 2.7m people,<sup>55</sup> involving information that included names, email addresses and mobile phone numbers. Thus this single breach led to criminals having access to personal information on 5.1 % of all adults in the UK, much higher than the estimate from the CSEW16-17. This case is an excellent example to illustrate the issues that make it difficult to estimate the prevalence and extent of breaches. Uber not only failed to disclose the breach, but also admitted paying \$100,000 to hackers.<sup>56</sup>

#### **2. How prevalent is the use of profiles of individual consumers to personalise offers and prices?**

Many consumers *believe* that they have been exposed to personalised prices on, for example, airline websites. While there is no evidence that airlines

---

54 Office for National Statistics, (2018), *Crime in England and Wales: Year ending September 2017*, available at: <https://www.ons.gov.uk/releases/crimeinenglandandwalesyearendingseptember2017>

55 Information Commissioner's Office, (2017), *Latest ICO statement on Uber data breach*, available at: <https://ico.org.uk/about-the-ico/news-and-events/news-and-blogs/2017/11/latest-ico-statement-on-uber-data-breach>; Reuters, (2017), *Uber says 2016 data breach hit 2.7 million UK users, or most of its base*, available at: <https://uk.reuters.com/article/uk-uber-cyberattack-uk/uber-says-2016-data-breach-hit-2-7-million-uk-users-or-most-of-its-base-idUKKBN1DT1Z7>

56 New York Times, (2018), *Inside Uber's \$100,000 payment to a hacker, and the fallout*, available at: <https://www.nytimes.com/2018/01/12/technology/uber-hacker-payment-100000.html>

have personalised prices based on personal data, we have found evidence that the mobile game industry has tested the approach,<sup>57</sup> and insurance companies now use a much larger range of data to personalise quotes.<sup>58</sup> Many other companies have the ability to act in this way too, without consumers being aware.

The practice of price discrimination can be good for consumers when it allows more people to access products and services or it helps to drive competition. However, it can harm consumers when they are not aware of the practices and when it undermines their trust. Personalised pricing has the potential to cause individual consumers measurable detriment if it exploits them, but – if it is occurring – there is currently a complete asymmetry of information between the consumer and their supplier, and watchdog organisations like Which? are not able to rebalance this for them.

### **3. Do the detailed profiles compiled for advertising purposes create a broader ‘digital reputation’ that can affect access to services?**

When so much information has been compiled about consumers to target advertising, it seems implausible that the other potential commercial applications of that information do not get monetised – for instance, to affect access to services like credit or insurance.

There are anecdotal reports of this marshalled by data and privacy commentators,<sup>59</sup> but it is difficult to gauge how often profiles initially compiled for advertising-targeting purposes (containing much incorrect data) are used to affect more material decisions affecting consumers.

### **4. Who is selling data?**

Consumers want to know the answer to this question, and it is surprisingly difficult to offer them a clear answer. In 2010, focused investigations in the *Wall Street Journal*'s ‘What they know’ series exposed that many popular mobile apps were sending or selling user data (mostly device identifiers and location data) to third-party advertising platforms.<sup>60</sup> Similarly, many websites are embedded with functionality that appears to have some consumer utility, but is supplied by vendors whose primary business model is to harvest and sell data. The AddThis plugin (now owned by Oracle) is one example: this embeds social media functionality to websites and monetises information on the browsing behaviour that it captures.<sup>61</sup>

---

57 VentureBeat, (2015), *Cut the rope 2 gets more in-app spenders by using Gondola's dynamic pricing*: <https://venturebeat.com/2015/12/15/cut-the-rope-2-gets-more-in-app-spenders-by-using-gondolas-dynamic-pricing>

58 Financial Conduct Authority, (2016), *Feedback statement: Call for inputs on big data in retail general insurance*, available at: <https://www.fca.org.uk/publication/feedback/fs16-05.pdf>

59 In China this scenario is already advanced, Wired, (2017), *Big data meets Big Brother as China moves to rate its citizens*, available at: <http://www.wired.co.uk/article/chinese-government-social-credit-score-privacy-invasion>

60 The Wall Street Journal, (2010), *Your apps are watching you*, available at: <https://www.wsj.com/articles/SB10001424052748704694004576020083703574602>

61 Oracle, (2010), *Oracle buys audience tracking firm AddThis for around \$200m*, available at: <https://techcrunch.com/2016/01/05/oracle-addthis>

Some large data brokers do offer directories of where they source data,<sup>62</sup> though it is difficult, if not impossible, to go one layer beneath these lists and find out where the sources are sourcing their data. In our interviews with service providers to companies likely to be selling data, they were unable to disclose their knowledge due to client NDA (non-disclosure agreement).

In spite of the fact that this information is subject to legal requirements for transparency at an individual level (via subject access rights), it is not possible for comprehensive transparency to be achieved in the interests of public information. To build even a partial picture of which popular consumer services are making revenue from selling their customers' data requires strenuous effort.

### 5. Who is buying data, and for what purpose?

People are as, if not more, interested in the other side of the data-trading transaction: who is buying information about me, and why? This 'buy side' of the data industry is even more opaque.

As Forrester Research comments:

*'Third-party aggregators sell personal data injudiciously. The data brokerage market has long existed on the basis of capturing data from myriad sources – public records and retail transactions, for example – and then enriching that consumer or household record with attributes such as “presence of child” and “lifestyle cohort”.*

*'Not only is this data often incorrect, but it's also frequently sold on the open market, largely without limit. In other words, anyone, from Guns & Ammo magazine to the Sierra Club, can purchase exactly the same bits of data about an individual, for very different – and often unwanted – purposes.'*<sup>63</sup>

Again, focused press investigations shine a light on how 'open' the open market in data really is, how injudicious the sale can be, and how riddled these transactions are with vulnerabilities.<sup>64</sup> It is impossible for an individual consumer to get a comprehensive list of who has purchased data about them. It is similarly impossible to build a comprehensive guide on the buyers of data, and their objectives from these transactions.

---

62 Oracle, (2018), *Oracle Data Cloud Data directory*, available at: <http://www.oracle.com/us/solutions/cloud/data-directory-2810741.pdf>

63 Khatibloo, Fatemeh, January 2015, Forrester Report, *Make sense of a fractured consumer data ecosystem*

64 <https://www.theguardian.com/technology/2017/aug/01/data-browsing-habits-brokers>

## 2. The control deficit in consumer data

**Transparency is a necessary, but not sufficient, condition for consumer empowerment when it comes to their data. While people may dislike what they see when they gain understanding, they feel resigned and powerless to do anything about it. This resignation appears justified. Meaningful consumer control over data is currently unfeasible, and this is getting worse:**

- **Google and Facebook have access to the widest range of online data, and use it to further consolidate their grip on their markets, limiting the prospect of meaningful consumer choice in some services.**
- **Beyond these powerful but well-known brands, consumers face a data trading ecosystem of thousands of actors that are practically unknowable, and significantly limit the effectiveness of data subject rights.**
- **The inherent nature of data means there are serious feasibility issues in controlling data as it flows to other companies, and such data flows are exploding.**

### *Google and Facebook: a consumer choice?*

Google dominates its core market of search, but is also powerful in mapping, browsers and web-based email. Facebook dominates social media and messaging through its own branded applications and those it has acquired. A range of observers now ascribe ‘utility’ status to these services. Calls to regulate the companies as utilities emanate from a perhaps unexpected coalition that spans from press outlets like *The Economist*<sup>65</sup> to then-members of the Trump White House.<sup>66</sup>

From a consumer perspective, if they are dealing with a provider of a service they can’t do without, and acceptance of data terms and conditions is a requirement of accessing that service, they cannot be said to have a choice. This was the preliminary view of German competition authorities in 2017, when they concluded a 20 month review into whether Facebook’s privacy policies constituted an abuse of market power.<sup>67</sup>

---

65 The Economist, (2017), *What if large tech firms were regulated like sewage companies?:* <https://www.economist.com/news/business/21729455-being-treated-utilities-big-techs-biggest-long-term-threat-what-if-large-tech-firms-were>

66 Fortune, (2017), *Steve Bannon wants to regulate Facebook and Google as utilities*, available at: <http://fortune.com/2017/07/29/bannon-facebook-google-monopoly>

67 Bundeskartellamt, ( 2017), *Preliminary assessment in Facebook proceeding: Facebook’s collection and use of data from third-party sources is abusive*, available at: [https://www.bundeskartellamt.de/SharedDocs/Meldung/EN/Pressemitteilungen/2017/19\\_12\\_2017\\_Facebook.html](https://www.bundeskartellamt.de/SharedDocs/Meldung/EN/Pressemitteilungen/2017/19_12_2017_Facebook.html)

In April 2018, Facebook asked UK Facebook users to review and consent to its first major terms and conditions update in three years. Users were told: ‘If you don’t want to accept the terms, see your options’. A user following the link to their ‘options’ was told ‘if you don’t accept these, you can’t continue to use Facebook’, and offered a choice to accept the terms, or delete their account.

The combination of data and advertising businesses with dominant consumer services has been increasingly troubling the range of commentators and regulators that scrutinise the companies. Two types of issues are typically raised:

1. Consumers are signing away far more data than they would do if their consent was freely given. Where there are few practical alternative options to a service, because for example of network effects inherent in assets like the Facebook social graph,<sup>68</sup> people face a ‘Hobson’s choice’ with many thinking they cannot object to unreasonable terms and conditions other than by opting out of the service.
2. Some allege that people are being manipulated into excessive consumption of media,<sup>69</sup> and sometimes the content served will be misinformation.<sup>70</sup> This is said to be achieved by the combination of the company’s data and targeting assets with an almost endless supply of ‘niche’ content.

The second issue above further demonstrates the way in which data makes the vertical integration of Google and Facebook distinct to the many other examples of media companies selling advertising too. Aspects of this vertical integration today that are distinctive are:

1. The level of consolidation, and its global scale.
2. The importance of individual-level data to achieving and securing this consolidation: advertisers are increasingly buying ‘people’ instead of buying placements in media context likely to attract a target audience. This further weakens the position of other media outlets.<sup>71</sup>
3. The endless availability to these platforms of an extremely long tail of niche content uploaded by users and companies, as production and distribution costs for content have plummeted. This niche content, combined with data, means that users’ attention can be captured with increasingly manipulative methods, meaning that more and more media attention accrues to Google and Facebook.
4. The fact that this content is often produced for marketing purposes (whether political or commercial) and – unless the placement of the content has actually been paid for – the consumer may not be made aware of this, so may be unknowingly consuming misinformation.

---

68 Techcrunch, (2018), *Don’t break big tech, fix it*: <https://techcrunch.com/2018/04/11/dont-break-big-tech-fix-it>

69 For example, see <https://www.theguardian.com/technology/2017/dec/11/facebook-former-executive-ripping-society-apart>; Tim Wu, 2017, *The Attention Merchants: The Epic Struggle To Get Inside Our Heads*

70 Nieman Lab, (2018), *Has Facebook’s algorithm change hurt hyperpartisan sites? According to this data, nope*: <http://www.niemanlab.org/2018/03/has-facebooks-algorithm-change-hurt-hyperpartisan-sites-according-to-this-data-nope/>; Wall Street Journal, (2018), *How YouTube drives people to the internet’s darkest corners*, available at: <https://www.wsj.com/articles/how-youtube-drives-viewers-to-the-internets-darkest-corners-1518020478>

71 NewCoShift, (2017), *Lost context: How did we end up here?*, available at: <https://shift.newco.co/lost-context-how-did-we-end-up-here-fd680c0cb6da>



## *The unknowable ‘adtech’ and ‘martech’ industries*

Data protection law has, for decades, given people rights to access the data organisations hold on them and to demand that it is erased. The GDPR expands and strengthens these rights.<sup>72</sup> The definition of personal data is expanded to include the identifiers of browsers, devices and individuals, and potentially other pseudonymised data, that the advertising industry trades in every millisecond.<sup>73</sup> These rights must be placed in context of an industry where the number of companies is in the thousands, and has roughly doubled each year since 2011.<sup>74</sup>

The fragmented industry of companies commercialising consumer data makes data subject rights essentially notional, as consumers do not know about these companies, and could not keep up with them if they tried.

Figure 6: The complex marketing technology landscape



## *The unknowable black market in data*

Consumer resignation extends to security breaches. As discussed earlier, in a single data breach in 2016, 2.7m people in the UK (around 5% of the population) were affected by the data breach from Uber, and ‘mega-breaches’ (which constitute the overwhelming majority of breached records) rose every year from 2014 to 2016.<sup>75</sup>

People do not understand what happens to breached data, and there is precious little assistance from either companies or government to help them find out if their data has not only been stolen, but also sold or otherwise publicly exposed. What is clear is that, once publicly exposed, the data is out of anyone’s control.

<sup>72</sup> <https://ico.org.uk/for-organisations/guide-to-the-general-data-protection-regulation-gdpr/individual-rights>

<sup>73</sup> <https://ico.org.uk/for-organisations/guide-to-the-general-data-protection-regulation-gdpr/key-definitions>

<sup>74</sup> Martechtoday, (2017), *Infographic: The 2017 ‘Martech 5000’ marketing technology landscape*, available at: <https://martechtoday.com/infographic-marketing-technology-landscape-113956>

<sup>75</sup> Symantec, 2017, *2017 Internet Security Threat Report*

Much of the monetisation of identity and other sensitive stolen data is in supporting and executing credit card fraud.<sup>76</sup> To support this market, there is a well-developed and sophisticated ecosystem of black-market suppliers of cyber-crime products and services. This includes ‘cyber-crime as a service’ (or the software tools that enable criminals to steal data), as well as files of information about consumers. The table below indicates ‘going rates’ for some of these products and services in dark-web marketplaces.<sup>77</sup>

**Figure 7: Pricelist for various black market goods and services.**  
 Extracted from Symantec Internet Security Threat Report 2018

Credit cards	Price
Single credit card	\$0.50-25
Single credit card with full details (Fullz)	\$1-40
Dump of magnetic strip track 1 and 2 data (e.g from skimming)	\$20-60
500 credit cards already used for fraud in the last week	\$1
Services	Price
DDoS service, short duration <1 hour, medium protected targets	\$5-20
DDoS service, duration >24 hours, medium and strong protected targets	\$10-1000
Hacker for hire	\$100+
Credit score repair	\$50
Messing up people's online presence	\$500
Aeroplane ticket and hotel bookings	10% of value
Accounts (username and password)	Price
Video and sound streaming accounts	\$0.10-10
Various services, more than 120+ available (gaming, food, shopping etc)	\$0.5-10
Online banking accounts	0.5-10% of value
Online money accounts (depending on value and verification)	\$10-\$100
Retail shopping account	\$5-10
Cloud accounts	\$5-10
Hacked Gmail accounts	\$0.1-5
500,000 email accounts with passwords from data breaches	\$90
Hotel loyalty/reward program accounts with 100,000 points	\$10-20
Shopping loyalty accounts with cash points	\$2-7
VPN services	\$1-10
Online retailer gift cards	15-50% of face value
Restaurant gift cards	15-40% of face value

According to Symantec, the prices above are taken from publicly accessible underground forums and dark web TOR sites. Closed, private forums tend to have even lower prices. They cannot verify if the goods are genuinely sold for the asked price; some of them might be fake offers. Payments are also often made using cryptocurrencies, such as Bitcoin or Monero, at the daily exchange rate equivalent values.

<sup>76</sup> <https://techbeacon.com/resources/hpe-monetizing-stolen-credit-card-data>

<sup>77</sup> Adapted from Symantec 2018. Internet Security Threat Report, available at: [https://img03.en25.com/Web/Symantec/%7Bf8ca1fb2-b73c-46d0-ae9e-17456c45df87%7D\\_ISTR23-FINAL.pdf?aid=elq\\_16379](https://img03.en25.com/Web/Symantec/%7Bf8ca1fb2-b73c-46d0-ae9e-17456c45df87%7D_ISTR23-FINAL.pdf?aid=elq_16379)



The exposure will be global, and not time-bound. In his article *A day of the life of a stolen healthcare record*, an internet security journalist and commentator cites a 2015 experiment:<sup>78</sup>

*In an experiment conducted earlier this month, security firm Bitglass synthesised 1,568 fake names, social security numbers, credit card numbers, addresses and phone numbers that were saved in an Excel spreadsheet. The spreadsheet was then transmitted through the company's proxy, which automatically watermarked the file. The researchers set it up so that each time the file was opened, the persistent watermark (which Bitglass says survives copy, paste and other file manipulations), "called home" to record view information such as IP address, geographic location and device type.*

*'The company posted the spreadsheet of manufactured identities anonymously to cyber-crime marketplaces on the dark web. The result was that in less than two weeks, the file had travelled to 22 countries on five continents and was accessed more than 1,100 times.'*

The illegal market in data may be as complicated, if not more so, than the legitimate market in data. By its nature, the buyers and sellers of data cannot be identified, and the data they trade cannot be traced. When someone's data leaks into this world, they truly have no prospect of control, and there is a dearth of guidelines on actions to take.

Which? is working hard to reduce the risk and impact of data breaches by:

- Working with government and industry to create a new code of practice for Security by Design<sup>79</sup>.
- Raising a super-complaint with the Payment Systems Regulator<sup>80</sup>, advocating placing more liability on banks for consumers' losses from 'authorised push-payment' fraud. The regulator has now committed to introduce a contingent reimbursement scheme to help compensate some victims.
- Advocating for the Data Protection Bill to include powers for qualified non-profit bodies to pursue collective redress on behalf of consumers who have suffered a data breach.

## ***Losing grip on data in motion***

Consumer data has been bought and sold for decades. What's different now is that the flows of data are many orders of magnitude larger: first, because the volume of consumer data has exploded, and second, because these exchanges of data between parties are often happening programmatically – that is, companies' technologies are constantly interfacing with each other to automatically gather data and process for their own purposes.

---

78 Krebsonsecurity, (2015), *A day in the life of a stolen healthcare record*, available at: <https://krebsonsecurity.com/2015/04/a-day-in-the-life-of-a-stolen-healthcare-record>

79 Which?, (2018), *Easy-to-hack smart devices targeted by government*, available at: <https://www.which.co.uk/news/2018/03/easy-to-hack-smart-devices-targeted-by-government>

80 PSR, (2016), *Super-complaint from Which?*, available at: <https://www.psr.org.uk/psr-publications/news-announcements/Which-super-complaint>

These programmatic flows of data transform the economics of starting and running a business, and have enabled massive innovation. Leveraging the data assets of the largest consumer technology platforms via their APIs,<sup>81</sup> hundreds of thousands of applications have been built that offer fun or productivity for consumers, or efficiency for small businesses. But at the core of the 2018 Facebook-Cambridge Analytica scandal was a failure of effective governance over third-party use of data, obtained via one of these interfaces. As a case study, this shows the two core problems in the policy framework to govern data in motion:

**1. Many data controllers have not to date taken their responsibility for third parties seriously.<sup>82</sup>**

Facebook apparently had no meaningful process, beyond terms and conditions, to monitor and enforce use of that data in line with terms and conditions. Upon being ‘found out’ by forensic press scrutiny, the company’s first public response was to pass the blame to the partner that failed to comply with terms and conditions. It was only as the scandal escalated that Facebook admitted it had failed in its oversight of partner use of data, albeit that in Senate hearings Zuckerberg repeated the defence that it had been assured the offending data had been ‘deleted’.<sup>83</sup>

**2. There are significant feasibility issues in enforcing compliant use of data through a supply chain.** Once issues have been uncovered, assurances that data have been ‘deleted’ are extremely difficult to verify. Moreover, processes to identify those issues, or to execute partner due diligence to prevent them in the first place, is undeveloped in many organisations handling data. One of the world’s largest companies, making troves of personal data available via a highly permissive API, admitted to needing to do more thinking on its accountability structures for partner access to data. It appears unrealistic to expect the tens of thousands of other businesses daily transmitting data to be able to do better, whether this is the tangled web of the adtech industry, or the tens of thousands of small businesses interfacing with dozens of ‘software-as-a-service’ providers to enable marketing, accounting and other operations for their business.

Much of the effective accountability for the many-tiered data supply chain is brought to bear by targeted press investigations. The Cambridge Analytica revelations have been perhaps the highest profile case in point, but there are other isolated examples of this. German journalist Svea Eckert and data scientist Andreas Dewes saw the potential vulnerability in governance of the ‘martech’ industry, and set out to obtain the browsing histories of their fellow citizens. They found they could purchase and de-anonymise the browsing history of three million Germans, unchallenged.<sup>84</sup> Both of these investigatory efforts were painstaking and resource-intensive. The press cannot be relied upon alone as the main means to keep data flows safe, and buyers and sellers honest.

81 API stands for Application Programming Interface: software that opens up some element of the platform and allows other applications to work with it

82 Another case study was TomTom’s sale of driver data to the Dutch police in order to improve enforcement against speeding offences, available at: <https://www.ft.com/content/3f80e432-7199-11e0-9b7a-00144feabdc0>

83 The Washington Post, (2018), *Transcript of Mark Zuckerberg’s Senate hearing*, available at: <https://www.washingtonpost.com/news/the-switch/wp/2018/04/10/transcript-of-mark-zuckerbergs-senate-hearing>

84 The Guardian, (2017), *Anonymous’ browsing data can be easily exposed, researchers reveal*, available at: <https://www.theguardian.com/technology/2017/aug/01/data-browsing-habits-brokers>

# 3. Where do we go from here?

**The industries that catalyse the collection and processing of consumer data move fast, are structurally diffuse and their workings technically complex. This is a challenging context for policy makers, and the policy framework continues to adapt too slowly to the reality of the way digital markets work.**

**The GDPR is a major step in the right direction, and we welcome it. But it may not be enough to keep the digital market thriving and working for everyone. In light of our investigations, we think that the priorities for attention are:**

- **Consumers and their advocates need more transparency about the impacts of the use of data.**
- **The potential for consumer harm from the digital advertising market:**
  - **The CMA should conduct a market study into the digital advertising market, working in concert with the ICO.**
  - **The ICO should publish further details on its regulatory priorities in this area<sup>85</sup> and invite input on the scope and focus of its work. In particular we welcome its priorities on web and cross-device tracking and credit reference agencies and data broking.**
- **In a digital economy, data flows propel innovation, but it is practically difficult to monitor and govern these flows. It is time for a thoroughgoing review of governance of data in motion, with due attention given to creative ways to regulate use of data.**

## *Reframing transparency in data*

People want greater transparency over how their lives are influenced by the use of personal data, and organisations like Which? need much better information in order to determine where harm is occurring and empower consumers to stop it. But the countervailing forces to this transparency are powerful.

Most of data protection regulation, even after GDPR, focuses on greater transparency about the collection and storage of personal data; for example clearer consent requirements and stronger subject access rights. However, whilst these rights are important, our research suggests they will not be enough to tackle widespread feelings of disempowerment. We have found that many people do not feel it is worthwhile taking action because they:

---

<sup>85</sup> Information Commissioner's Office, (2018), *Regulatory Action Policy*, available at: <https://ico.org.uk/media/about-the-ico/consultations/2258810/ico-draft-regulatory-action-policy.pdf>

- do not know what they could do to take action, without disconnecting from technology (as mentioned above);
- find it hard to understand the impact of the data ecosystem on their lives;
- feel it is too late – organisations already have their data.

Although some consumers are trying to regain control by taking action and restricting what data can be observed about them, many others are disengaging, as they understand that they could learn more, but they choose not to as they are too uncertain about the impact on their lives to justify the time and effort.

People do not and should not want to understand the data ecosystem, but they do want to understand why their insurance quote has changed, why they are seeing certain product suggestions, or how the data held about them might change a credit decision.

Where it is possible, we want to see companies giving transparency in context to their customers. Companies have figured out how to target individuals at the right moment for advertising – now they need to use the same ingenuity to allow people to understand how the data held on them affects their lives.

Where the impacts are harder to understand, we want to see government, regulators, businesses and consumer advocates working together to understand the impacts of data usage. There are several forms this could take:

- Understanding the impacts of personal data use should be a priority for the Centre for Data Ethics and Innovation to coordinate action from others, as the ethics of an action cannot be judged without information about the impacts.
- The CMA and the Department for Business, Energy and Industrial Strategy should explore a programme of work to investigate the impacts of data use on consumer markets. This programme should involve academics who can advise and design approaches that involve the use of simulated or real individual profiles. This should be a priority for the CMA's new data and technology unit.
- The ICO should ensure that its regulatory work seeks to explore the impacts of the data usage as well as the legality and process. In particular, the ICO should quickly monitor how the GDPR provisions on profiling are being put into practice to see whether they could realistically tackle the lack of consumer knowledge that we have identified.
- All of the above organisations need to work together on this complex and important issue, and involve external stakeholders including Which? in coordinating action. We propose that the Department for Digital, Culture, Media and Sport should hold a summit to bring all of the relevant organisations together.

## *Controlling dominant players*

Use of consumer data is not just a matter for privacy regulation. Oxera Consulting, in its paper for us that accompanies this report, notes that consumers show a wide variety of willingness to share information, and that therefore there is an appropriate role for consumer choice. However, consumer choice can only occur where there is vigorous competition.

In many areas of the digital world this is true, but the strong position and vertically integrated nature of Google and Facebook in these markets has raised two types of concerns:

1. A dominant position may allow a large vertically integrated platform to set higher prices or offer lower-quality services to advertisers, for example by artificially creating scarcity of possible advertising placements. Consumers would be affected if higher advertising costs were passed on to them. However, there may be limits to any potential anti-competitive behaviour – for example, large advertisers can insist on certain quality standards being met, and other advertising channels (such as TV) may be converging with online advertising in the medium term, creating more competition.
2. A dominant position may allow a vertically integrated platform to foreclose competitors by refusing them access to their systems – for example, by limiting inter-operability with upstream or downstream ‘partial’ competitors, thereby hindering competition. This could then allow the platform to set higher prices, again potentially increasing final prices for consumers. Platforms would have the ability to foreclose competitors if:
  - a. Competitors have to incur (prohibitively) high costs to obtain the same consumer data, or cannot update it with the same frequency. However, efforts by publishers to create similar datasets suggest that such concerns may be limited.
  - b. There are ‘captive’ consumers who can be identified or reached only through certain advertising platforms.

Current empirical evidence on harm for consumers through higher prices for advertised goods or lower-quality advertising is limited. However, we are concerned that the privacy impacts themselves need to be explored. We have shown evidence in this report that:

- Many people are concerned or anxious about the use of their data.
- The primary driver of the collection and use of this data is digital advertising.
- Google and Facebook have a large and increasing market power in digital advertising.
- People do not feel they have realistic alternatives (for example, 24% of those who we sampled and who used Facebook said they considered leaving the site following the Cambridge Analytica revelations, but didn’t. Only 6% actually say that they deactivated or deleted their account).

People-based marketing has become a feature of the digital advertising market, but its impact and consequences are poorly understood. Together with the significant horizontal and vertical concentration of the digital advertising industry in Facebook and Google's hands, we think the risk of harm is great enough that the CMA should conduct a market study on digital advertising, in concert with the ICO.

The Lords Communications Select Committee has recently called for the CMA to conduct a market study following their report on UK advertising in a digital age<sup>86</sup> and we strongly support this call.

### *Scalable governance of data flows*

Consumers dislike their data being sold and bought. It is difficult to provide good public information on where this is happening, and it is difficult to reassure them the system is accountable.

However, in our digital economy, data flows propel competition and innovation and allow for the development of many of the goods and services that consumers use. The constant manual and automatic trading of data is here to stay, but it is practically difficult to monitor and govern these flows.

Data portability is also an important new right in GDPR with significant potential to empower consumers, and the government is soon to look at this in a 'smart data' review. However, we are concerned that take-up will be limited if people do not trust the data ecosystem sufficiently. A way needs to be found that not only allows innovation but also improves the ability to provide oversight and enforcement.

We and other commentators observe that traditional data protection policy levers have not kept up with this reality.<sup>87</sup> Data protection legislation has failed to govern this in the past, and it is unclear that the GDPR represents a step change in approach to this issue, albeit that organisations are encouraged to map their data flows in their self assessment process.<sup>88</sup>

As discussed in this report, resource-intensive press investigations have been responsible for exposing some of the highest profile data protection breaches that can arise from data in motion. With the volume of data flows growing so rapidly, a more scalable approach is required to create accountability for these flows and ensure that both regulations and consumers' preferences are being followed. This is a challenging problem that may benefit from a mix of policy and technical solutions, and it would benefit from thoroughgoing review.

---

86 Lords call for action on digital advertising, (2018), available at: <https://www.parliament.uk/business/committees/committees-a-z/lords-select/communications-committee/news-parliament-2017/uk-advertising-report-publication>

87 Harvard Business Review, (2018), *The US needs a new paradigm for data governance*, available at: <https://hbr.org/2018/04/the-u-s-needs-a-new-paradigm-for-data-governance>

88 <https://ico.org.uk/for-organisations/resources-and-support/data-protection-self-assessment>

It is time for a first-principles look at the governance of data in motion, with due attention given to creative ways to understand and regulate the use of data. We therefore think that the new Centre for Data Ethics should make it a priority to review the governance of data in motion. The Centre should work alongside the ICO (in particular taking account of its work on credit reference agencies and data brokers) and ensure that there is a good understanding of the forefront of technological solutions in this area.

The review should consider the following:

- Measures to foster more seller due diligence on the buyers of data, so that brokers cannot sell data without satisfying more strenuous conditions on its onward use.
- Stronger measures to ensure platform accountability for third-party access to data via APIs.
- The potential solutions to be found in the nascent market of personal identity and data management providers to give truly decentralised and scalable accountability for how data flows, and whether interventions are required to enable these innovations to achieve full potential benefit for consumers.





Which?, 2 Marylebone Road,  
London NW1 4DF  
Phone +44 (0)20 7770 7000  
Fax +44 (0)20 7770 7600