HOUSE OF LORDS

Revised transcript of evidence taken before

The Select Committee on the European Union

Sub-Committee on Internal Market, Infrastructure and Employment

Inquiry on

ONLINE PLATFORMS AND THE EU DIGITAL SINGLE MARKET

Evidence Session No. 2 *Heard in Public* *Questions 11 - 22*

MONDAY 19 OCTOBER 2015

4.15 pm

Witnesses: Steve Chester, Steve Wood and Des Higham

USE OF THE TRANSCRIPT

Members present

Lord Whitty (Chairman)
Lord Aberdare
Baroness Donaghy
Lord Freeman
Lord German
Lord Green of Hurstpierpoint
Lord Liddle
Lord Mawson
Lord Rees of Ludlow
Lord Wei

Examination of Witnesses

Steve Chester, Director of Data, Internet Advertising Bureau, **Steve Wood**, Head of Policy Delivery, Information Commission's Office, and **Des Higham**, Engineering and Physical Sciences Research Council Career Fellow in Data Analytics/Internet of Things

Q11 The Chairman: Welcome. As you know, we have decided to concentrate on the data dimension of platforms this week. We are regarding each of you, in your different perspectives, as being experts within that field. It is possible that some questions may stray more fully, but we are basically focused on data—its use and misuse, et cetera.

I need to point out two things. It is quite conceivable that we will be interrupted by a division bell very soon, in which case we will probably disappear to vote and then resume. It could happen twice within an hour or so. This is a formal evidence session. We will take a full note. You will get a copy of the transcript and can revise it, if you wish to. It is on the record and is being webcast. After that, it is accessible via the parliamentary website. If the sound is all right, we are ready to begin.

I am going to ask the opening question, but when each of you answers could you say a word or two about yourself? This is really a pretty general question. What do online platforms do with the data they collect? Can you describe the data chains that lie behind that, the different processes and organisations involved with those data chains and the use of the data? Who wants to start?

Steve Chester: It depends on what sort of platforms you are looking at. Amazon, for instance, will use registered data when you register on the site to look at things such as offering you like products. They might then take some of the data about you to offer you those, looking at your browsing behaviour and the things you have purchased. Essentially, it is to try to improve your experience by saying, "You might like these things". Amazon did that first when it started off by selling books, and it has obviously progressed as it expanded into more products.

Facebook's primary model is to use advertising. It will then use data about the person it to sell profile data to advertisers. The advertisers will not be able to see who that person is or any personal details, but they are actually looking at passion points and interests and being able to sell advertising based on interest levels. Somebody might be interested in a car or in a particular sport, perhaps rowing. Based on people's interests and the data they have registered on Facebook, you could then use that data to target interests. Perhaps a company selling rowing boats would then want to target those people interested in rowing, so it uses data in that respect.

Des Higham: I am Des Higham. You asked us to say a little about ourselves. I am an academic and my background is in mathematics, statistics and computing, so that is the algorithmic side that I come from. From my perspective, what happens to the data that adds value is the summarising. You have a huge amount of information, but there is literally too much for a single person to absorb, so the real key is clever algorithms that can extract useful information. They often reduce the size of the data so that you can visualise it and categorise individuals, saying which box they fit into in terms of advertising to them or understanding their behaviour. That is the sense in which the value gets added to the data, from my perspective.

Steve Wood: Hello, my name is Steve Wood. I am head of policy delivery at the Information Commissioner's Office. The Information Commissioner is the UK's regulator of the Data Protection Act and the Freedom of Information Act. Our interest in this space is focused on the concept of personal data and how online platforms use it within the framework of the Data Protection Act. We are interested in the concept of personal data, which is data that individuals can be identified from, either directly or indirectly. When online platforms are using data, as I think has already been said, they may collect it from an individual filling out a form at the start of the process. Equally, data will be generated as an individual moves through that platform's services, so they are developing transactional data. The individual is leaving a footprint as they move throughout that service, and the data can often be aggregated and analysed. It can be used by the platform to understand that individual, often with the aim of personalising the content.

In terms of online platforms' interest in personal data, the space has become more complex as platforms have often gathered together data from different sources. An online platform might have a search engine service; it might then develop a social media part of its platform; it might also have another element related to online video. As all those different elements can be aggregated together, quite a detailed picture can be developed of an individual. So we are interested in the interaction between all those data elements and how they are combined.

Q12 The Chairman: Part of my question was also: what is the chain? Are intermediaries involved in it? I do not know if Mr Chester wants to talk about that. I also understand from what Mr Higham just said that this is data that can add value. What are the benefits to the consumer or the small business that is using a platform? Perhaps you could have a go at the first part of that, Mr Chester.

Steve Chester: It was remiss of me not to introduce myself. My name is Steve Chester and I am the director of data and programmes at the Internet Advertising Bureau, which is a trade body for digital advertising.

To answer the question, it depends, but we talked just now about what we call first-party data, which is where you register for a service and that data is then used. There is also third-party data collection, involving many companies beyond online platforms such as Facebook and Google, which are well-known consumer brands. But there are other businesses in the supply chain that help with the targeting of advertising, making it more relevant. There is a whole supply chain here, from the advertiser who buys the advertising to the publisher who then sells it, which could be Facebook or Google, but there could be many businesses in between that transact various forms of data to make the advertising more targeted and relevant to the audience they are selling to. There may be just one broker, if you will, or there could be many in between, with each of them offering a different level of service.

As an example, advertising agencies are increasingly investing more in their own infrastructures to use data—and to use their clients' data, with permission—to then target advertising and buy from websites that sell it. But they are buying audiences, or people who are more relevant to their clients. P&G, for instance, might want to sell Persil to housewives and not necessarily to sell it or appeal to a 13 year-old schoolchild, who would not necessarily be interested in it. Each of those businesses will then offer a service in between, so agencies are using their own data platforms, and there are then various other intermediaries. Increasingly, rather than buying advertising in chunks or upfront, there is a move towards buying it in real time; you choose advertising that is relevant to your client, and you can do that in real time, or certainly in microseconds, and then purchase only the advertising that is relevant to you. Both the advertiser and the agency, negotiating on behalf of the advertiser, can do that. Essentially, there can be one or many different businesses within that supply chain.

The Chairman: Okay. Thanks very much.

Q13 Lord Liddle: Sticking on this point about advertising, how does competition take place in this market? Do the different platforms compete for advertisers? How does that work?

Steve Chester: Absolutely. When internet advertising first came about, in 1994 or so, it was simply the brand itself buying directly from the person who created the website and sold advertising against it. Advertising agencies increasingly represented their clients so that the clients would not buy directly from the person selling the advertising but use an agency as their intermediary, if you will. Over the last 20 years, we have gone from buying the advertising in advance—essentially by buying an area on a site, perhaps on Yahoo, and saying, "We are going to put some advertising there, so that everybody who lands on or goes to Yahoo on that particular day sees that advertising"—so that, having moved from that model, it is now just targeting people who are relevant to that brand. An example would be housewives and showing it just to housewives.

In that space, there are many different people selling advertising. Facebook and Google sell advertising and they are just two of the prominent brands, but there are of course hundreds of thousands of websites all competing. The advertising agencies or the brands who buy it all need to choose who they do business with. There is no doubt that Facebook and Google are among the biggest—with Twitter also and Yahoo—but there are many different people who sell advertising, and they will all compete to sell the advertising on their site against their competitors.

Lord Liddle: This is a follow-up question about the dominant platforms. Are they unreasonably exploiting their position with advertisers?

Steve Chester: I think they are trying to innovate and to offer innovation to people buying advertising in a compelling way compared with others that they are competing with. It depends what you mean by "using their dominance". Often they try to innovate and through innovation try to attract advertisers who want to spend money with them. Of course, Facebook has been a big success story over the past nine years. It has grown from being a very small start-up to the entity that it is today.

On the question of whether it exploits that, it has a free market opportunity. You can essentially buy Facebook directly; we can buy Facebook through third parties. It does not stop others or an advertiser of a brand buying from a very small website, but of course advertisers in general want to attract large audiences and to get to large audiences. Facebook has close to 1.5 billion people on its platform, and that in itself is very compelling in wanting to get to the end audience. It can almost get to any audience that you would want to get to if you were a brand wanting to get into contact with people who might be interested in your product.

Those platforms attract spend towards their platforms through competition and innovation, and obviously the bigger they are the more investment they have to invest in further innovation.

Lord Liddle: Does that mean that they inevitably become more and more dominant as time goes on?

Steve Chester: There is no doubt about it that a significant amount of spend goes to those platforms through innovation offering more compelling services, so yes, they could become more dominant over time. Certainly the buyers, the P&Gs and Unilevers of this world, are entirely free to buy from much smaller publishers and not from Facebook and Google, but they choose to buy on those because they offer compelling tools and services, which means that they get in touch with just the audience and the people who are relevant and interested in their products.

Des Higham: I can chip in with a bit more information from my perspective, having worked alongside an advertising agency looking at social media data. One question that it was asked by its clients was, "Given a specific budget, which portion should we spend on online advertising versus newspaper or television advertising?". It is not always clear that online is the best, so looking at the data gives you some perspective on the potential for gathering an audience in that way. There are certain types of product where the online community is not sufficiently closely knit—they do not talk to each other, they are not pushing your own message across—so in some instances it is much more effective just to use more traditional means of advertising.

Steve Chester: Absolutely. Television can reach a large number of people at any time. It is a fantastic medium for doing that, particularly if you pick very prominent programmes such as "Downton Abbey", which attracts large audiences. If you are a brand that wants to get in touch with a lot of people at any time, TV can be a better reach medium in that respect.

Q14 Lord Green of Hurstpierpoint: Are there issues with the online platforms to do with the transparency of the costs to an advertiser: you pay for your clicks if you are a recognised brand, you have no real idea of whether what you are paying for your click compares favourably with what others might pay for a click on the same access area? At the other end of traditional advertising, you know what it costs to buy some square centimetres of a newspaper, and TV sits somewhere in between. There is certainly perceived to be a lack of transparency. Is that an issue that is made worse by the strongly monopolistic positions of a couple of key platforms?

Steve Chester: With television, for instance, the price will vary depending on supply and demand; it fluctuates and is not a consistent price. With digital advertising, there tends to be what is referred to as a rate card price, which you offer as a business, which is a flat cost. If you are a large advertiser such as Unilever and P&G, you may say, "Well, actually we are going to spend a lot with you", so you will demand a lower price. But it tends to be more fixed pricing, which remains the case today.

On the point about clarity, there is auditing in digital advertising as there is in traditional advertising. Brands will ask an auditor such as Berlitz or Accenture to come in and audit their media. Essentially the auditor will look at the media plan, at where the advertising is placed, at the rates which the advertising agency secures on behalf of the brand, and say, "Is this against benchmark? Is this what you would expect against your competitors, against a typical rate?". That exists, and if anything it leads a lot of conversations when it comes to pitching for clients. They know that a lot of rigour is going to be brought to bear in the pricing that they are getting on behalf of their clients. There is a high level of transparency in that respect.

Do Google, Facebook or others publish exactly what they charge every client? No, but we do not find that in traditional media either. But there is that high level of auditing and accountability in digital as well.

Lord Mawson: Specifically on Google and advertising, what do the numbers look like? What are they actually making from it? What is the profit in all that?

Steve Chester: At the moment, Google, as part of the SEC, is required in its figures to break out any markets above 10%. They do not actually publish their figures wholesale by market, but they do publish UK figures, which is true of any corporate company in America.

In terms of the advertising market as a whole, we have just published our view of the market for half-yearly figures. That totals about £4 billion, of which just over half is from search, and obviously Google is the majority player in that. So we can make certain summations about the size, but we do not know definitely how big Google or Facebook are because they do not necessarily have to declare absolute revenues and split them by YouTube, for instance, or by their search products. So in the publication of accounts in America, if something falls below 10% we cannot see it; if it is above 10%, we can.

Baroness Donaghy: On the consumer side of things, the terms and conditions can very often include rather vague clauses, which state that the online platform will use your data to improve the service when this actually means that they will sell your data to advertisers to generate profits that will be reinvested in the service. How we can we make these vague terms and conditions clear, or at least protect the consumer from them?

Steve Wood: I am happy to respond on that. As online platforms have grown, concern has increasingly been expressed about how users interact with terms and conditions. As we all use more and more services, the amount of time that it will take an individual to read standard terms and conditions means that it can be challenging for that individual to meaningfully read and understand them. The oft-quoted statement is that many privacy policies are longer than "Hamlet". There is certainly a challenge there, and we, as a data protection regulator, did not want there to be a reliance just on a very long notice at the start of the process. There needs to be a focus on really meaningful and effective transparency rather than a reliance on a legal notice. That means what we call just-in-time information, which means that information can be layered for the individual as they move throughout the service, and perhaps a greater use of icons, which could perhaps be standardised across the industry and which people understand and recognise. We could, for example, have a standard icon for areas such as third-party data-sharing.

We have to expect innovation from the online platforms, too, in how they deliver this privacy information to individuals. I think that some online platforms are starting to do more on this, but more can be done, and the new proposed data protection regulation that is being negotiated in Brussels at the moment places greater emphasis on trying to tackle some of these problems. Article 14 of that regulation puts greater emphasis on the clarity and transparency of the information, and on a concept called data protection by design: the idea that when a service is being designed, consideration should be given to designing in a focus on the consumer at the outset. More needs to be done to tackle this issue, and as a regulator we want to be able to take enforcement action. This is about a focus on an individual's consent. If privacy information is genuinely misleading so that an individual could not understand that the data was being sold to a third party, we want to be able to take action. We recently took action in a case involving a company called Optical Express. It had bought in data that had been harvested from a survey that was given to holidaymakers on a plane. The survey just said, “The information will be used by a third party”. That led to individuals getting text messages from Optical Express. We took enforcement action in that case to tackle the issue.

Q15 Lord Aberdare: Thank you. Putting that in a slightly broader context, at our last session, the Competition and Markets Authority suggested that there were three ways of improving the market in data. One was to have greater transparency. The second was giving consumers more choice about how their data was used and collected. The third was giving them more control over its usage. Would you agree that those three aspects are key? How would one realistically go about achieving them?

Steve Wood: Again, I am happy to answer that one. We have spoken to the CMA about the work it has been doing and we are quite closely aligned in our thinking and focus in these areas. The focus on control and choice is important so that the user can easily make those choices online. In practical terms, it really means things like online dashboards—having a layout with a way of controlling settings that a user can easily recognise for that level of control. It is also about the level of information that a consumer will feel content with. As I mentioned earlier, it is about the layers of the information. With the basic, key information in the top layer, we may increasingly need to standardise the use of icons. They would really help the consumer to work their way through the data to be able to make their choice, so the control issue is important.

Des Higham: If I could chip in on control, a colleague of mine in academia has a useful model in certain circumstances, in particular where you are generating constant streams of data—for example, on electricity usage in the home. A novel idea is that the data would stay in the house, and if a third party or provider wishes to access the data it gets processed by being averaged so that as little as possible is being given away. You have given them what they asked for, if you think it is appropriate, but not the full stream of data. That is a localised form of control that, in certain circumstances, would be very effective.

Lord Aberdare: Are there examples where platforms are offering something other than a one-size-fits-all approach, or where users can select different options through a dashboard or the kind of system that Mr Higham has been talking about?

Des Higham: I am aware of social media sites where you can turn privacy on or have certain levels of privacy, so that is one.

Steve Chester: There is also a programme called Your Online Choices, which you may have seen, where some online ads have a blue triangle in the top left-hand corner. It is a consumer-facing campaign initiative that gives people the ability to control the cookies that allow ads to be targeted to them. Essentially, there is a very consumer-facing site called youronlinechoices.eu, which tells people about how advertising works and how adverts are targeted to people. It then gives people a view of current businesses that are collecting their cookie data and targeting ads towards them, and gives them control about which ones they will continue to allow. They can either switch them all off or pick and choose the businesses that they are happy to allow to continue doing that. That is already in existence on desktops.

We are trying to find the right mechanism for mobiles because the adverts are much smaller there, having a very small triangle on an advert does not necessarily work. We are having a discussion about whether we could roll out a similar initiative for all ads, so that for all ads we could have some sort of feedback link where people could not only select but find out more information about ads that were targeted to them—and, perhaps more widely, even about the types of ads they actually like and are willing to accept.

Steve Wood: I think it is fair to say that most of the online platforms have some form of data centre where you can log on and control your settings. We tend to find variable quality and accessibility of that information. How easily is it made prominent from a link on the homepage or when the user is signing up to the service, so that they know they can change those settings? Could they also know what the default settings are? Is it automatically turned on so that the search history is aggregating with social media data, for example, across a large platform? It is about those sorts of issues. Often, the service itself is available; it is about making sure that it is effective in practice.

Q16 Lord Wei: To build on that, there has been talk of things like personal servers, or perhaps apps, which would allow one to fine-tune and know exactly what was being collected so that you could control it to a granular degree. How realistic is that? What is the progress in trying to get that kind of technology embedded, particularly as we move towards the internet of things where the level of personal information that could be shared, even anonymously, starts to get huge?

Steve Wood: There have certainly been some initiatives recently about what we call personal data stores and having ways for individuals to download datasets, such as those of their banking or utility data, which could increasingly be linked to the internet of things and devices that they have in the home. There is the Midata initiative, which has been developed by the government department BIS. There is also a provision in the proposed European Data Protection Regulation, in Article 18, on data portability. When individuals request data it would enhance the right, which has always been there in data protection law, to request it in an open format to enable them to use it in that way. This area has potential, but we are probably quite a way off from consumers being empowered to take advantage of these services. There is a risk that we put too much emphasis on it. We can see that it could help people to assert their rights over such data, but sometimes people are struggling with the basics online, when doing things with settings, so it is about how we evolve to that point where it can be really valuable for people.

Steve Chester: Yes, to build on that further, part of this is personal data management platforms. These are often referred to as personal information management systems, or PIMS. There is discussion about the consumers having control of that data, which is perhaps some way off, so that they could have a whole range of datasets. They could even have health data, or essentially any data that is controlled by them. They would then decide on how that data is transacted, if at all. Perhaps somebody wants to know how much heating they use; perhaps EDF would want to get in contact with them. They would decide after asking, "Am I willing to share that data about my heating and its usage, so that EDF could give me a compelling offer?". If it is to be that universal or big, it is going to be some time off. It would obviously need to be very secure and we would need to decide who was going to house that data—how it would be looked after and administered—but I believe that there are already in the region of 250 companies looking at personal information management systems. There is a lot of work going into this, but it is unclear at the moment just how far away it is.

Lord Wei: Is there a standard that the industry itself will generate, or will it require top-down regulation to force that standardisation?

Steve Chester: I think it is going to cross many industries. It will obviously require a great deal of thought and application. Ideally, you would try to find a mechanism that is universal—

that works for all—and has those businesses coming together. Whether it needs any formal regulation in the future is a bit of an open question at the moment. That due diligence is certainly being done by many companies, and I believe that a number of people are already working on this, making business cases to the Government by stating the likelihood of these things happening and how they would be administered. But it is not clear at the moment how long it would take to create those things and curate them properly.

Q17 Lord Mawson: Transparency is an interesting word that we all probably trot out too easily. In offline environments, what one is seeing out there is that while we talk transparency, the behaviour is actually a lot more opaque. Some very interesting and strange things are going on out there that are costing us many millions of pounds but that may not be making the world more transparent at all. They may be making it more confusing and human relationships in large parts of the public sector quite complicated. If that is true offline, what are the unintended consequences that we are noting online of this transparency agenda?

The Chairman: That is indeed the bell, so you have a little time to think about the answer. We shall return in 10 minutes.

The Committee suspended for a Division in the House.

The Chairman: Right, could we go back into session? It was Lord Green's question—no, sorry, Lord Mawson was half way through his. Sorry—I beg your pardon. Do the witnesses recall the question after this long time?

Lord Mawson: It was about transparency and the difficulty of transparent behaviour off-line, which appears to talk about transparency but actually produces opaque behaviour in all sorts of complicated ways, and what the unintended consequences of all this are online. It can seem straightforward, but in practice it is not.

Steve Wood: I was not sure whether your question was really focused on the impact on individuals, and the fact that they are being more transparent about themselves online.

Lord Mawson: It is on individuals but also about organisations and what happens in the dynamics. I am more familiar with it offline because I watch what happens in the behaviour of organisations, which, in my view, behave in ways that are not totally honest because they have to find ways around all this stuff. So if those things go on, I just wonder what ways around all this stuff happen in an online environment. As I tried to say, transparency seems a straightforward word, but actually it is not.

Steve Wood: In terms of transparency, particularly when someone is said to be consenting to something, we focus on that consent having to be specific. It has to be informed and freely given. Sometimes it is important to drill down into all those elements, because there are those out there who will sometimes wish to stretch or get around some of those different elements. For us, it is about focusing on all those three elements that particularly relate to consent to understand whether an organisation is effectively being transparent to the consumer. But there is no doubt that in some areas, particularly where there is an incentive to sell data—in our work at the moment, we are looking at the data brokerage industry and what we call list brokers in relation to nuisance calls—there is a large market for data that has probably been gathered in circumstances that have not been completely transparent to the end user. That is still of concern to us.

Des Higham: There is a related point. It is possibly more about accountability, but a lot of these data-processing algorithms have to make predictions or give you decisions. If you train these algorithms on data, they can become very effective but they are really just a black box. You give them the data and they categorise it: "This user is in category A"; "This user is in category B". If you tried to interrogate the black box, it would be difficult to know on what basis that decision is being made. So we are building machines or algorithms and do not really understand why they are effective, and in that sense it is difficult to account for the decisions that are being made.

The Chairman: That is something we may return to. Can we now go to Lord Green's question?

Q18 Lord Green of Hurstpierpoint: Do you think there are concerns about how online platforms handle other businesses' data? There are two versions of this question. One has to do with the platform that is not in direct competition with the business concerned but where, for instance, a free click that leads to a business selling a product to a client is a piece of information that the platform has. My understanding is that it has become more difficult for the business to access information about free-click usage than it used to be, and that in fact Google has started to block that. I may be wrong with my facts, in which case correct me, but there is clearly an opacity about what platforms do with that information anyway.

The second and perhaps more obvious variant of the problem is where the platform is itself in direct competition with the supplier of the product. Amazon is the obvious example in that not only is it supplying books and other things, but acting as a platform for other suppliers of those same sorts of product. While one cannot necessarily put cause and effect together, it is quite clear that gross margins have gone down on a lot of the kinds of products that Amazon deals with. What issues arise, in your views, either under the heading of opacity/transparency or under the heading of monopoly power?

Steve Chester: Certainly, from an advertising perspective we do not get to see that. I think that in the example you used of Amazon, you were talking about its marketplace. It essentially offers the ability for a range of products to be sold using Amazon's logistics through its marketplace. We do not have a view on that. We do not get to see that e-commerce data, and certainly from an advertising point of view we have not had any third parties saying that they were unhappy or making complaints as if to say, "We feel that one party"—such as Amazon or Apple—"is exerting a dominance and not giving us access to our data". From an advertising perspective, it is something that we certainly do not see. From an e-commerce perspective, looking at general commerce and whether there is monopolistic behaviour or data is being withheld, that is not something that we have access to, so we cannot directly comment on it.

The Chairman: Thank you. Lord Freeman has the next question.

Q19 Lord Freeman: So far, not entirely but in general, our questions and your answers have concentrated on protection regulation. Do you think there is now a case for moving towards more proscriptive regulation? That is to say, what platforms can and cannot do with the use of data? There is of course a parallel here with the telecoms industry**.**

Steve Wood: Our view at the moment is that very extensive data protection regulation is obviously still being negotiated in Brussels, which was certainly motivated in part by the concerns of the European Commission about the activities of online platforms. If you read the statements of ex-Commissioner Viviane Reding, she certainly had big multinational internet companies in mind as some of the drivers for bringing in the regulation. The new regulation is complex legislation. We are also moving from a Directive to a Regulation, so we would urge a bit of a stepping back when the new text of the Regulation is agreed to see what it adds to being able to regulate the personal data that online platforms will use. We are already in the position where it would be hard to imagine a situation where an online platform is not classed as what is called a data controller under the Data Protection Act or the new Data Protection Regulation. So most online platforms will be regulated by data protection law.

As I mentioned, there are a lot of new provisions in that legislation on transparency: there are areas on the right to be forgotten and the right to erasure, and concepts such as data protection design and data protection impact assessments. All these are relevant to the activities of online platforms and should assist in improving the situation in the data protection rights of individuals online. It is important that we take a step back to see what further regulation is needed. When we have been observing the data protection regulation and how the text has evolved, we see that there is a balance between over prescription in the text and what is allowed as a risk-based approach. As a regulator, we very much want to be able to take action in the cases where there is the greatest harm to the citizen or consumer, not just to have to sanction in the case of every violation of the rules. We want a text that focuses on that approach, so there needs to be a stepping back before a new focus is made on online platforms.

Lord Freeman: Can you help the Committee by indicating the sort of timeframe that the new Regulation might take?

Steve Wood: Yes, it is currently in trialogue stage at the moment. It is certainly difficult legislation to negotiate; I think that it has been called one of the most lobbied pieces of legislation in European history. We think that we are nearing the end. The text was first published in 2012 and the official line is that agreement should be reached by the end of this year. That is still challenging—it might go into next year. There would then be a two-year implementation period, so I suspect that we would be looking at a new regulation being in force for businesses sometime in 2018. If the text is agreed in 2016, two years would take us to 2018 but it is still possible to change that.

Steve Chester: If I could comment on that, the data protection regulation has not been updated in 20 years. It clearly needs updating; I think we are all agreed on that, as the world is very different. The one concern we have is about implementation. For instance, with music downloads today, where terms and conditions are served people sometimes click yes or okay just to get through them and click ahead. So what will the mechanism be for explicit consent? Does it mean that a notice will have to be delivered at every website you go to and every point of data collection? How much notice will people take of that? Our concern is whether for the most important data, such as health data, if that is treated in the same respect as your cookie data, will people make a distinction when a notice comes up, or will they become slightly blasé and just click to go to the next page without really understanding how that data is used? Should we use it just for the most important data and not have to use it for every piece of data collected, because there is a risk that it will just merge into this idea of data being collected?

The Chairman: It is highly probable, I would say.

Steve Chester: Yes. We definitely think that the reforms need to come and that there needs to be an update; we are just concerned about the mechanism and that people might lose sight of different forms of data and their importance, because there are varying levels. I am very happy with my cookie data being shared, for example, but not necessarily my health data. We need to make sure that that is transparent and clear to people: that it can deliver the service that they want but gives them clarity about how their data is managed.

Lord Rees of Ludlow: Do you think that by 2017 we will be happy with the regulations that have then gone through, or do you think there will be a need for third-party, what ifs or other ways of reassuring the public? Do you have any other ideas that you think ought to be discussed as a follow-up to what Mr Wood said was already going through?

Steve Wood: On the framework that we need, the position that we have set out is that we need to have a strong and clear regulation that enhances data subjects' rights but that works in practice for businesses. That means making sure that data protection authorities such as the Information Commissioner's Office have the right powers to be able to fine and to enforce in the most serious cases. There is also a space for what we call co-regulation: how do we start to raise the bar for good practice, and how do we give consumers a way of recognising the businesses that have done that extra work on good practice so that they can start to differentiate between one company's practices and another? We are interested in the concept of what we call a privacy seal, a mechanism for which companies would be able to apply that would probably be awarded by a third party, but the third party would have to be endorsed by us, the regulator, so we would play a role in that process. It is really a way of scaling out the role of the regulator, because we cannot do all the work ourselves, to raise awareness among the companies involved. It is matter of making sure that we start to raise the bar and that consumers have a way to understand difference. There is a space for co-regulation. We would not necessarily say that it would be self-certification, but there would probably be a role for third parties auditing and awarding a seal.

Lord Rees of Ludlow: Do you have any other comments on this?

Steve Chester: Yes. There is already a form of auditing. We talked about the blue triangle icon called Your Online Choices. That gives people the facility to look at companies that are using their data to target ads at them. Currently, it is a programme whereby companies sign up to it, and within six months they need to self-certify and say that they will abide by it. Within one month of self-certifying, they need to have an audit from a third-party company, which then audits them for their compliance. They will receive the EDAA seal only once they pass it and are seen to pass it. That is independently audited. At the moment, 91 companies have passed it. There are many other signatories in the same chain that are yet to be audited. Companies that have passed include P&G, Yahoo, AOL, Microsoft and Adobe. There are many others in the pipeline waiting to join. We are talking potentially about the further rolling out of that to all forms of advertising and then having an audit of all of them, so that all companies would sign up. We would need to work with various partners in the industry and the regulators and talk about how we extend that, but there is already some auditing in place and we view that auditing as potentially expanding to all advertisers.

Lord Rees of Ludlow: What sort of organisations do the auditing?

Steve Chester: The third-party auditors are independent auditors. They are not employed or controlled by any one of those companies, and they have to be certified themselves; they have to be agreed upon. There is already some due diligence and auditing for ads that specifically use your behaviour to retarget you. We are talking about potentially then rolling that out to all ads, and companies that wish to engage in advertising, such as P&G and Unilever, would then have to be audited.

Des Higham: If I could just chip in with a lay comment from my own experiences and from talking to colleagues, the younger generation do seem to be more accepting of giving away aspects of their privacy and more savvy and aware of what they are doing. They appreciate, for example, the free nature of some of the services that they get in return. There is a sense that, as time moves forward, the bar may be getting a bit lower.

Lord Rees of Ludlow: Yes, but some have had nasty surprises, have they not?

Des Higham: Yes, of course. I am not saying that it is correct; I am saying that the general perception is that that is what is happening.

Baroness Donaghy: I just wanted to follow that up before Lord German asks his question. You might feel at a certain age that it is not important for this information, particularly embarrassing stuff on Facebook, to be available, but of course this stays in the ether, does it not, until one is a respectable married person with school-age children, and all of a sudden these things can pop up. I realise that this feature of youth is something that you may not be able to do anything about, but is there not some developmental thing whereby at some stage somebody can review the information and say, "Oh boy, let's get rid of that lot"?

Steve Wood: It is important that people have the ability to control their data. We have talked about that today, and it still applies in those situations. The most notable instance of this has been since the Google Spain judgment last year, which was about the right to have certain links removed from a search made against an individual's name.

Certainly that judgment has clarified the position now, and individuals can request search engines to remove certain search results against their name. They do not always have to remove them in all instances. There has been a significant debate about how that interacts with freedom of expression, so it has to be adjudicated on carefully, but it is an important issue: people may move on in life and may want to control their data. The issue on the internet is that it is very difficult for the internet to forget you completely, but we can provide people with tools so that at least the information is less available; it is not on the first page of the search result or the top search result. Say you are going out on a date with someone for the first time. That can have a privacy-enhancing effect, even if all instances of the information are not removed from the internet, which would practically be very difficult to achieve. California, for example, has now introduced into state law the concept of a delete button that social media sites should have for all users under 18. Again, the idea is about introducing control, and ultimately we have to get more discussions in schools. We have done some work as a regulator on teaching materials to get people at a young age to understand how to control the information. There is a lot of information about being safe online and the dangers of online predators, but it is also the slightly more mundane thing of getting people to learn and feel comfortable about controlling their information online. It is all those areas.

Q20 Lord German: Mr Wood, you anticipated this question earlier; it is about the balance to be drawn. You talked about the balance between data protection and making data work for business. Clearly Mr Chester on one side and you on the other represent both parts of the equation. But in matters of pure principle, where should we draw the line to balance your two positions and say, "These are the principles on which we should operate", in order to achieve satisfaction for both of you?

Steve Wood: With big data, as a regulator we try to make sure that there is scope for a framework that allows the risks to be assessed. Big data is a large, sprawling concept. You can have a range of societal uses for big data through to very commercial uses. We have advocated a risk-based approach. Some have advocated that we need to change the rules for big data and we have drawn up the classic eight data protection principles around issues such as fairness. They are still very useful and still hold firm in that environment. We probably need to make sure that people are skilled in the impact assessments and in understanding what the impact on the individual is going to be in processing data as part of a big data project.

That probably means moving beyond having just a data protection officer in the organisation who is responsible for legal compliance; it is about the data scientists and the people who are designing the algorithms and the people who are saying, “Let’s combine these data sources”. They should have a real process—what we call a privacy impact assessment—to work through what impact that will have from different angles, including the consumer angle and the role of different third parties and stakeholders, to reach a rounded view. The way that is approached at the moment is still patchy in terms of the awareness within organisations of how to properly assess the impacts, but that is how the balance has to be struck, with people being more skilled in doing these impact assessments. We are also quite happy for data protection to be linked to the broader concept that is being discussed at the moment of an ethical impact assessment. There may be other impacts that go beyond privacy and we are quite happy for data protection to be packaged up alongside an ethical impact assessment. That is how the balance should be struck.

Steve Chester: Yes, we believe in and focus on this throughout the work we are doing on data protection and online choices, to give people an informed view on how data are being used and to ensure transparency. It is absolutely key to let them know how their data are being used. We believe in self-regulation, and we work very closely with government and the Department for Culture, Media and Sport to make sure that what we are implementing is fit for purpose and agreed by government. We work with the ICO and others to make sure that whenever we implement any kind of self-regulation, it is scrutinised and fit for purpose. We try to make sure that we have a broad consensus on that, but there is more work to be done. The icon that you talked about is a good start. We need to implement on mobile and make sure that people are more informed about how their data is used, in order to give them more control. That is what we are heading towards—a world where people have more control over how their data is used and in choosing who they let in and how they let them in. The exchange needs to be very clear in saying, “Okay, if you are going to do this, this is how your data is being used. These are the services and benefits that you can expect”, and in simply allowing people to make a choice.

Lord German: Are we looking perhaps, alongside legislation and regulation in this area, at codes of good practice? If so, could they and would they work in an environment that is global?

Steve Chester: Yes, we believe in those sorts of codes of practice—hence the Your Online Choices and others, which we believe people should follow. They should be not just self-certified but independently verified to make sure that they are meeting the codes of practice. This applies locally, as well. Some initiatives that we have in place are B2B initiatives where we look at the placement of advertising against safe content—content that the advertiser is happy to be shown against. There is a programme that is local to the UK and different from the one in the US, whereby businesses such as Google and others are independently assessed against the question, “Did you actually put the advertising against content that the advertiser is happy for their advert to be shown against?”. That is localised. It is applied differently in the US because buyers—advertising agencies and brands—said, “We want some assurances that businesses are doing this and that they are accountable for this”. So I think that locally applied solutions can definitely work, and this is something that we are looking into.

Steve Wood: Incidentally, codes of practice are an important tool that we as a regulator use. We have the power to produce codes of practice under the Data Protection Act. For example, we produced a code of practice on anonymisation, which is obviously relevant to the big data area. We also have a code of practice on privacy notices—this is relevant to the issue of transparency that we talked about earlier—that we are in the process of revising. Increasingly, the codes of practice that we produce will have to be worked on with other European colleagues and other European data protection authorities. We will see the development of more European-level guidance and codes of practice. Ultimately, we will probably have to build bridges to ensure that we have areas of common practice with the US, because there are common areas in some of these issues and challenges, although we have quite different legal systems. Particularly around the issue of transparency and how we tackle it effectively, we have certainly had some good conversations with other regulators such as the FTC[1], which is dealing with this issue. The issue has lots of the same elements to it, although the FTC is coming at it from a slightly different legal perspective because it is a consumer regulator rather than a data protection regulator.

The Chairman: We are going back to algorithms at this point

Q21 Lord Mawson: Given the inference that algorithms can have serious consequences for people and businesses, should platforms be made more accountable for the inferences of their algorithms—and, if so, how? For example, a business that plummets down the Google search ranking may lose access to the online market without knowing why. Equally, algorithms could make inferences about a person's creditworthiness based on their race or friends on social media.

Des Higham: It is a tricky issue. From my perspective, one of the difficulties is that you can indirectly infer things from other sources of data. You may not be looking directly at ethnic background or political or religious beliefs, but you may be able to look at other sources of data. Possibly even without knowing it, you are then basing your decisions on things that correlate with those direct issues. So it is difficult for me to see how you would deal with that from a regulatory point of view.

Steve Wood: In terms of algorithms and big data, obviously some of the key principles that lie behind a lot of exponents of big data concern the volume and variety of the data and the velocity at which you can analyse it. That often leads to a default setting among data scientists that is to get hold of as much data as possible and try to link as many different types of data to try to understand things that they do not know about already. That is what lies behind the power of the algorithm. Clearly that can be a big benefit from the processes in terms of new insights that may be revealed. The challenge is how to communicate a process like that to the end user. It can be very challenging in terms of understanding the concept. The user does not need to understand how the algorithm is put together, but they do need to understand what the impact of the algorithm might be for them and how they can use it to inform their choices. Certainly we have seen some innovation, such as different companies producing online videos that explain how they use data. You can then get things across in a completely different way. You cannot explain an algorithm in a long privacy notice and expect a lot of people to understand it, but there are different ways to get the impact across. As I said earlier, the other thing that we are keen to do is reach out to data scientists, who are the people who had the idea of combining the data in the first place, to make sure that they have some sort of ethical impact assessment that asks questions at the right points about why it is a good idea to combine the data in this way and what impact it would have.

Steve Chester: Specifically in this case, Google was also referenced, particularly with people and businesses plummeting down their search rankings. There are other search engines, of course, Bing being one of them. They will not necessarily reveal the exact nature of how their algorithms work, because they want to protect a certain amount of intellectual property, but they are very clear when they publish these algorithms. When they update them—Google will do one major update a year, as will Bing, which is owned by Microsoft—they will let people,

[1] Federal Trade Commission in the United States of America.

both advertisers and consumers, know what those changes mean. The most recent one was about promoting sites that have a mobile-specific site. When you are on your mobile phone, the screen is significantly smaller and you do not want the same experience or exactly the same view as you would have on a much bigger screen, as the scripts and writing would be too tiny for you to read; you want something that is optimised to your experience on your mobile. Google has recently promoted sites that have sites specifically for mobiles. Someone going on to a mobile will see a slightly different layout, but it would be just as transparent and clear. It was clear that the update promoted those sites that adhered to this, so it was clear about the impact rather than the exact nature of the change to the algorithm. The advertising agencies and the brands that are very sophisticated in the way they use search will understand that and implement changes accordingly. Another important point is that the OFT[2] did an inquiry last year, I think—I know that it is now defunct—which found no evidence of discrimination based on algorithms. It was looking a price discrimination in the market and its conclusion was that no bias was offered by algorithms.

Lord Wei: Just to follow up on that, if you are a business or consumer who feels that you have been discriminated against by algorithms, what existing rights do you have or what courses of action can you take other than just what the particular company or platform has put in place? What can force an action, or is it very much left to the platform to decide?

Steve Wood: I can answer from a data protection angle, which obviously focuses on the individual and the consumer. In that situation, you can think about the data protection principles in the Data Protection Act, which are about personal data being processed fairly. Obviously, a key part of that is transparency. Fairness is about thinking about the impact on the individual. If there is a situation where consent was involved, it would need to follow the tests that I mentioned earlier. Also, the purpose for which data is used can be incompatible with the original purpose. So there is a set of principles in the Data Protection Act that could be contravened in a scenario such as the one mentioned earlier in which someone's social media data was used to assess their creditworthiness or to look at their suitability for medical insurance, for example. An individual could complain under the Data Protection Act in that situation about the way in which their personal data had been used. The new General Data Protection Regulation will hopefully strengthen the data subject's rights in those situations and lead to more collective action at a European level when we need to tackle some of these issues. So there are some tools available to a citizen in that situation. I guess that some citizens probably are not always aware, so we need to do more work on the availability of the ability to complain to a data protection regulator. I cannot comment on the business angle, because it is outside our jurisdiction.

Lord Wei: Are there examples from the last few years of people successfully complaining or raising those kinds of issue?

Steve Wood: The closest example, which I have referenced in our written evidence, is the case that we dealt with about Google's privacy policy. It was not directly focused on the algorithms in that context but it was about transparency in the way Google was combining data across all 60 of its services. We looked at that issue at a European level. As a result, Google signed an undertaking to us to improve the transparency and accessibility of its privacy policy. Some of our European counterparts as part of the task force also took enforcement action. That is an example of quite a high-profile case that came to us.

Steve Chester: When I tried to use Google yesterday it kept showing me a notice asking, "Have you read our privacy policy?". Because I was already aware of it, I just ignored it and

[2] Office of Fair Trading (closed on 1 April 2014).

eventually I was not able to use the search facility until I had given my consent. It had gone from implicit consent to saying that I absolutely had to give my consent to go any further, so I think there are good moves afoot.

Lord Mawson: I am interested in how this stuff interfaces with a human being. How much work have you done on how much information a human being can hold in the round? I deal a lot with the NHS, which is a very big organisation with lots of codes of practice. What you see is that often people are so busy trying to keep up with the codes of practice and various things that they miss the patient. I am just wondering what you have done on that. Human beings cannot absorb exponentially this amount of information.

Steve Wood: The key area that we are starting to look at is trying to put an increased emphasis on what we call privacy by design, or what is called data protection by design in the proposed data protection regulation, which is about trying to link data protection compliance with the really good work that is often done in computer science around usability and what makes a system effective and interactive for an individual. The problems in the past have sometimes been that data protection has been closely owned by a lawyer, when really it needs to be a multidisciplinary area where a team of people work together to focus on the individual in drawing on some of the issues that you have highlighted. It is a good point that, in the end, it has to work for the individual. We are increasingly highlighting that in the guidance that we produce.

The Chairman: We have two more questions and then we will have to wind up.

Lord Green of Hurstpierpoint: This is a question born of ignorance. How does the Data Protection Act apply to business data? Does it apply at all? If so, are there differences between the way it applies to business data and an individual's rights?

Steve Wood: Essentially, if you are processing personal data—if you essentially control that data and make decisions about it—you will be a data controller under the Data Protection Act. It probably leads to most businesses being data controllers in that situation, so most businesses are covered by the Data Protection Act. Any consumer or citizen in the UK has rights under that legislation.

Lord Green of Hurstpierpoint: That was not quite what I was getting at. I was trying to get at whether the data that businesses own in respect of their own businesses is or is not protected by the Data Protection Act. It sounds as if it is not.

Steve Wood: Do you mean their ownership of it in terms of intellectual property?

Lord Green of Hurstpierpoint: Not so much intellectual property as the pricing mechanics of their business, if you will. You have a product that is being sold on a platform. Quite a lot of the financial information about that product is derivable from looking at the patterns of sale. Is that protected under the Data Protection Act?

Steve Wood: Sorry, I see your question. What constitutes personal data is a contentious issue, as is what constitutes an activity by the business when it is changing a price and whether that is an issue that impacts on the privacy of an individual and the dividing line between that. Some of the decisions about pricing et cetera in competitive decisions are not really decisions about data processing and are outside the Data Protection Act. It depends on the context, but as it becomes easier and easier to identify people by combining data sets, the concept of personal data is increasingly being challenged. It is not always the direct identifiers; it is sometimes the indirect identifiers.

Q22 Lord Aberdare: Do you have any views of what lessons we might learn from the NHS? It had the very laudable aim of collecting a lot of patient data and using that in a big data way to tackle disease and other issues in a very effective way. Was it just bad marketing? I think that patients had to sign up to say that they were happy for their data to be used, but it crashed in flames because it had not been properly explained. Are there any lessons that we should learn about how this area should be regulated and the right and wrong ways to go about it? It would be a pity if that sort of thing was barred by regulation. It seems to me to be an eminently good use of big data, but it is personal data and health data.

Steve Wood: Certainly some of the key lessons have been drawn out today for transparency and communication, but that is not an easy or simple activity when you are dealing with data relating to people's health. The other issue was the range of organisations that are able to access and use that data and there were particular concerns about the data being shared beyond the confines of what people might have normally termed the NHS family. There was also the issue of how that was all explained as a package to someone just receiving a letter from their GP. That kind of data sharing raised different concerns depending on people's different experiences. There were some debates about whether the data would in effect be anonymised, but sometimes the consumer does not make that distinction; they just say, "That's my data and I don't want it to be used in that way". That is not a technical issue we can always address but, in terms of the trust issue, people felt uncomfortable about it. It was a broader concept that went beyond data protection; it was about how people interacted with it in that way.

The Chairman: Thank you very much. I am going to wind up there. You have given us far more of your time than we advertised for. I think that Mr Chester's organisation has not yet submitted any written evidence, but that would be jolly useful—as it would from Mr Higham—as there are some things that we have not managed to cover today. Written submissions would be extremely helpful.

Steve Chester: Of course.

The Chairman: Meanwhile, thank you very much indeed. You have been very instructive in this area. We have yet to make our conclusions—we are aiming for a final report in the new year—so if you wish to submit anything else, please do so.