

Dr. Ansgar Koene, Senior Research Fellow, Horizon Digital Economy Research Institute, University of Nottingham, Nottingham, UK—Written evidence (OPL0079)

Response to “The House of Lords SELECT COMMITTEE ON THE EUROPEAN UNION Internal Market Sub-Committee CALL FOR EVIDENCE on Online Platforms and the EU Digital Single Market”

Questions:

Section 1: Online platforms, consumers, suppliers

Defining online platforms

Q1. Do you agree with the Commission’s definition of online platforms? What are the key common features of online platforms and how they operate? What are the main types of online platform? Are there significant differences between them?

Many similar, but mildly different, definitions of online platforms exist and the Commission’s definition is as good as any.

The common feature is that they are multi-sided markets, while the significant differences arise from the data types that underpin their operation and the primary sources of income from the different sides of the market. Sales platforms (marketplaces, ‘sharing’ economy, app stores, payment systems) have a direct revenue stream from the sale of goods (percentage of sales price, cost of listing products charged to sellers); social media & information platforms (communication platforms, social networks, search engines, maps, news aggregators) rely more heavily on revenue from advertising and market intelligence delivered by the platform.

Q2. How and to what extent do online platforms shape and control the online environment and the experience of those using them?

Service ‘personalization’ is frequently used to ‘optimize’ the customer interaction, this involves filtering/recommending the products/services the customer is presented with. Facebook newsfeed for example selects the stories that are presented based on a complex, hidden, algorithm that includes information about past ‘likes’ by the customers, but also many other parameters. Even though the ‘personalization’ of services may results in an improved customer experience, it raises a number of concerns. 1. In order to do the personalization the platforms collect a wide variety of information about the customer, including past behaviour on the platform, location tracking (Facebook location tracking was recently blamed as possible cause for large power drain in iPhones), scanning of content posted by the user (Google scans content of Gmail, Facebook traces everything a user does, including messages that were deleted prior to posting). 2. The algorithms act as editors, selecting which information is presented and how visible it is to the customers, for instance the ranking/positioning in a list of search results. This editorial power of the online experience of customers has potentially far reaching consequence. During the 2012 US

presidential election Facebook ran an experiment which showed that people who had been notified when their friends mentioned that they'd just voted were significantly more likely to have also voted during the election. One concern we have is that the filtering/recommendation algorithms used by online platforms are becoming increasingly complex and may not even be understood by the platform providers themselves, especially if adaptive systems are used. One consequence of this is that platform providers may not be able to guarantee that they are compliant with regulations.

Effects on consumers, suppliers (including SMEs), competitors and society

Q3. What benefits have online platforms brought consumers and businesses that rely on platforms to sell their goods and services, as well as the wider economy?

The clear benefits include: access to a much wider, international market, as buyer or seller; "free stuff" based on the widespread freemium model; and cost reductions due to both benefits of scale and their use of "big data" to drive the business – that is the organisations have competitive advantage by being data led and analytic in their management.

Q4. What problems, if any, do online platforms cause for you or others, and how can these be addressed? If you wish to describe a particular experience, please do so here.

One of the main problems with online platforms arises from the centralized architecture where data from/about the user is transferred to the platform provider, resulting in a loss of control over the data by the user and a strong power imbalance in favour of platform providers. Users are often confronted with an all-or-nothing choice in which they must accept complete surrender of control over their data, even if they wish to use only certain parts of the platform services. This can result in discontent and/or suspicion by the users, who might nevertheless feel compelled to use the service due to peer-pressure (fear of missing out) or lack of alternatives (for many online services there is only a single large player in the market; closed systems make it impossible to interact with users of the platform without buying in to the platform as well). The problem is often confounded by the use of an advertising based revenue model where consumer data becomes the 'gold' that is mined by the platform.

An alternative approach to online platforms, that would address the issue of loss of control over personal data by the users, would be a decentralized data ecosystem where users retain control over their data. Instead of uploading their data to a centralized system under the control of the platform provider, users could keep their data on a personal 'databox' and provide monitored/controlled temporary access to those parts of their data that are required for a service. By retaining control over the data, users could monitor who is requesting access to the data, withdraw access rights when the user no longer desires the service. Prototype systems of this type have been developed by various research groups, including Horizon.

Q5. In addition to concerns for consumers and businesses, do online platforms raise wider social and political concerns?

The fact that online platforms are increasingly becoming the information gateway for people, especially younger generations who get much of their news from online platforms via mobile devices, raises social and political concerns similar to traditional news media. Concerns about media empires with too much dominance in newspapers or TV coverage, should equally apply to online platforms where it is now common for a single provider to dominate a service sector (Facebook for social networks, Google for search). As shown by Facebook's own study (2012 US elections impact on likelihood to cast a vote [see Q2]), they have the power to influence voting behaviour.

Social concerns also arise from the fact that the majority of online platforms are developed in the US (Silicon Valley) and therefore operate under US (Silicon Valley) oriented social values which can differ significantly from EU/UK values, as for example with attitude towards the precautionary principle for consumer products.

Platforms as part of the Digital Single Market Strategy

Q6. Is the European Commission right to be concerned about online platforms? Will other initiatives in the Digital Single Market Strategy have a positive or negative impact on online platforms?

Yes, the European Commission is right to be concerned about online platforms. They are dominated by a small number of virtual monopolies and due to their cross-border operating reach are best regulated at the EU level, rather than individual country level.

"Internet of Things" and "Smart cities" have the potential to further expand the influence of online platforms. Services like Uber are positioning themselves to mine information from social media communications and mobile device location tracking to identify the time and places where there is likely to be high demand for drivers. The functioning of the online platform service, and the algorithms that drive them, is set to have direct implications for 'real world' phenomena like city traffic.

Section 2: Competition, data, collaborative economy

Competition and dominance

Q7. Is there evidence that some online platforms have excessive market power? Do they abuse this power? If so, how does this happen and how does it affect you or others?

When considering social media platforms it is important to consider that 'market dominance' is strongly linked to the number of peers who are using the platform. Anecdotaly, many people who would like to quit Facebook and move to a different platform ultimately continue to use Facebook because that is where their peers are. Potential competitors (e.g. the open source social network platform Diaspora) who try to move into the social media market often struggle because of this. No users wants to move to a platform if there aren't already many other users there to interact with.

[Q8. No response.]

Collection and use of data

Q9. What role do data play in the business model of online platforms? How are data gathered, stored and used by online platforms and what control and access do consumers have to data concerning them?

For many online platforms the default business model has become the 'freemium' / free to use model that is supported by advertising revenue. While the obvious side of the advertising revenue are the ads that are shown on an online platform, a second source of income is often the sale of platform user behaviour statistics. Data are commonly gathered through multiple sources, including: storing of the information that is posted to the platform (e.g. product reviews), tracking of user behaviour on the site (tracking-cookies track behaviours like, where the users has clicked on a site and the amount of time between clicks), purchasing of data about behaviour/interest of demographic classes of users. The data is used to sell targeted ad space to advertisers and to feed into the filtering/recommender algorithms that 'personalize' the user experience. Users typically have very little control over any of this data collection. Privacy settings on sites like Facebook primarily stipulate how information is shared between users, not how the platform provider gathers and uses the data. Terms & Conditions of online platforms are usually formulated to give maximum freedom to the platform provider to use the data as they wish. For example T&Cs often include vague, broad-stroke, clauses such as 'data may be used for research purposes', where the research question is not specified to the user. Users usually have no options to control how their data is used, if they want to use the services, or even just part of the services, of the platform provider, they have to consent to handing over full control of their data to the platform. Various platforms do provide users with comprehensive access to the content that the user contributed to the platform, such as a download of the posts that were made to G+, but do not provide access to the tracking data that was collected about the user.

Q10. Is consumer and government understanding and oversight of the collection and use of data by online platforms sufficient? If not, why not? Will the proposed General Data Protection Regulation adequately address these concerns? Are further changes required and what should they be?

Consumer understanding of the collection and use of data is severely hindered by the overly long and complicated Terms & Conditions, Privacy statements and similar text that they are presented with at the time of signing up to an online platform. Faced with these long texts that are written to be understood and used in US court rather than by ordinary consumers, as well as having no recourse to accepting the T&Cs other than not being able to use any part of the platform services, consumer commonly default to a behaviour of 'click-signing' the agreements without even trying to read the content.

Information concerning the collection and use of data should be presented to the consumer in clear language, preferably supported by kite-marks, as proposed/concluded in the House of Commons Science and Technology Committee report on Responsible Use of Data (Fourth Report of Session 2014-15).

On the topic of automated decision making by algorithms, the UK Data Protection Act's 'principle 6' specifies that "the right of subject access allows an individual access to information about the reasoning behind any decisions taken by automated means. The Act complements this provision by including rights that relate to automated decision taking. Consequently:

- an individual can give written notice requiring you not to take any automated decisions using their personal data;
- even if they have not given notice, an individual should be informed when such a decision has been taken; and
- an individual can ask you to reconsider a decision taken by automated means.

These rights can be seen as safeguards against the risk that a potentially damaging decision is taken without human intervention." Importantly, however, these rights arise only if "the decision has a significant effect on the individual concerned". This caveat means that virtually all of the filtering/recommendations made by online platforms are exempt.

An important factor to consider is the magnitude of exposure to algorithm decisions that citizens are confronted with. Even if no single decision by any of the algorithms is violating the protection against having significant decisions made about an individual by wholly automated means, the accumulated effect can be difficult to estimate.

With regards to the EU GDPR, it is not yet complete nor ratified, and there will be a 2 year implementation phase only after which we will truly know how it will operate. As such it would be premature to suggest it is not fit for purpose or indeed jump to further data protection legislation before the existing plans execute.

Q11. Should online platforms have to explain the inferences of their data-driven algorithms, and should they be made accountable for them? If so, how?

Data-driven algorithms are an increasingly important element in determining the customer experience when using online platforms. The algorithms filter and rank which information is presented to the user and where it is presented, which affects the likelihood that a customer will notice and interact with the data. The high volumes of data available online means these algorithms are vital for enabling users to find the relevant information, be it search results, news stories or product offers. Consumer decision, ranging from who to vote for, down to choices in music to listen to are all influenced by the information people are exposed to. This is the basis of advertising and propaganda. Lack of transparency about the way in which algorithms manage this information introduced the potential for abusive manipulation. This can take the form of censorship, such as suppressing negative comments about the platform provider, or anti-competitive business practices such as the alleged manipulation by Google of ranking their own products higher in search results.

Accountability or algorithm inferences, or lack thereof, affects the development process behind the creation of the algorithms. In the current environment where the platforms are not accountable for algorithm behaviour, there is little incentive to focus on the interpretability of algorithmic processes. Due to the large number of parameters that are used by the algorithms, even the engineers who constructed the system are often not able

to explain why the algorithms made specific decisions. This is even more so in the case of adaptive systems that learn from continuously evolving example data sets, as is the case with deep-learning and similar systems. We do know however, that all data-driven systems are susceptible to bias based on factors such as the choice of training data set. Since the dominant online platforms are US based, it is likely that training data sets will contain biases that reflect US culture. For example, it was shown that Google's deep-learning based image classification system relied heavily on the presence of black and yellow colours to identify an image as containing a school bus, because the training dataset was based on US images, where school busses were always yellow and black. Another example comes from a study that compared Google Ads which were presented when using the US version of Google to search based on either typically White-American names or African-American names. In the case of African-American names, the Ad Words algorithm produces significantly more ads for criminal background checks. Since this kind of a search based on a person's name is a common practice when evaluating job applicants, the Ad Words results have potential to subconsciously promote racial discrimination in employment practices. There is no reason to assume any deliberate discriminatory intentions by the developers of the Ad Words algorithm. The algorithm is probably data-driven, based on statistics of Google searchers correlated with the names. The resulting recommendations generated by the algorithm, however, are likely to influence user choices. In this case further increasing the probability of requesting a criminal background check when searching an African-American name. This in turn reinforces the correlation that caused the algorithm to make the discriminatory recommendation. By this way, small initial biases can become self-reinforcing and magnify themselves. As demonstrated by the racial-discriminatory behaviour of the Ad Words algorithm, even supposedly neutral algorithms that are based purely on observations of internet usage statistics are not value-neutral. Rather they tend to reinforce an existing status-quo which might not be in the interest of the values that the UK society is striving for.

However, the topic of data driven algorithms and their interpretability is currently in the realm of research, so through Research Councils and initiatives such as the Alana Turing Institute, this should be made a priority for "big data" related research funding.

The collaborative economy

Q12. Can you describe the challenges that the collaborative economy brings? What possible solutions, regulatory or otherwise, do you propose?

One of the largest challenges from the collaborative economy is the power relationship between the platform provider and the people who perform the services that are being offered, such as the drivers for Uber. Contrary to the image that these platforms advertise, of people signing up to the service to 'share' their skills during their free-time, the collaborative economy platforms are likely to attract people who are in low or precarious employment and in need of additional income. It is therefore important that these workers are given the rights and protections of employees, and not treated like 'holiday workers'.

The current regulatory environment and possible interventions

[Q13. No response]

Q14. Should online platforms be more transparent about how they work? If so, how?

Yes, online platforms should be more transparent about how they work. They should provide clearer insight into the kind of data they collect about users, including behaviour tracking, as outlined in the House of Commons Science and Technology Committee report on Responsible Use of Data (Fourth Report of Session 2014-15).

The platforms also need to provide more clarity about the level to which the user experience is manipulated by filtering/recommender systems, preferably giving users more control over the level of filtering that is applied.

Q15. What regulatory changes, if any, do you suggest in relation to online platforms? Why are they required and how would they work in practice? What would be the risks and benefits of these changes? Would the changes apply equally to all online platforms, regardless of type or size?

As noted on data protection previously, the implications of the impending EU GDPR cannot be predicted and further regulation in this area would be counter productive to the effective implementation of GDPR. However, the need for interpretability of algorithms should be reinforced as a requirement of GDPR and work undertaken to make it possible.

As we are not market and competition specialists, we could only comment that until actual market harm has been demonstrated, it would be premature to engage in further competition regulation.

Q16. Are these issues best dealt with at EU or member state level?

Due to the cross-border operating nature of the large online platforms, and the desire for a digital single market, whenever possible, preference should be given to dealing with these issues at EU level.

December 2015