CPI's Europe Column Presents:

Big Data, Consumers' Privacy, and Competition in Online Markets

By Guillaume Thébaudin (Télécom Paris)[1]

Edited by Anna Tzanaki (Competition Policy International) & Juan Delgado (Global Economics Group)

CPI COMPETITION POLICY INTERNATIONAL

July 2019

At the dawn of the Internet of Things, consumers are increasingly required to disclose their private information to online firms. With the use of data analytics, these firms are able to increase their knowledge about the preferences and characteristics of their users. This knowledge is highly valuable for them as it generates revenues through disclosure to third parties (e.g. advertisers) as part of their business models and enables the delivering of more personalized and valuable products to users.

Online behavioral research pioneered by Miyazaki & Fernandez (2001) showed that consumers heterogeneously provide their private information to online firms due to different perceptions of the risk of privacy breaches.[2] Recent scandals such as *Cambridge Analytica* as well as the increasing number of cyberattacks have shown that these concerns are justified. Barnes (2006) has emphasized a "privacy paradox": consumers concerned about their online privacy are increasingly engaged in data disclosure activities.[3] This privacy paradox is largely explained by the increasing personalization of online services. Chellappa & Sin (2005) pointed out a trade-off faced by users between their value for personalization and concerns for privacy.[4]

Competition authorities (*Autorité de la Concurrence* and *Bundeskartellamt*, 2016) argued that existing firms' access to users' data can represent a source of increasing market power if these data are hardly replicable by potential entrants.[5] Competition concerns arise as online markets are highly concentrated and consequently, only a small number of firms are able to engage in such massive personal data collection process. These data enable large online firms to offer valuable and personalized features of their service which are likely to increase consumers' lock-in. Consumers may indeed find it, psychologically and timely, too costly to re-enter the same amount of data to obtain a similar degree of personalization with another firm. The recent implementation of the General Data Protection Rule ("GDPR") in Europe is an illustration of attempts to regulate big data activities and moderate this ongoing process of ever-increasing market power.

This note develops a dynamic framework aiming at better understanding how online firms are able to incentivize consumers to disclose more of their data, despite their privacy concerns, and gain market power. It focuses on the interactions between online firms offering "free-of-charge" services, which are able to collect, analyze, and sell data, and a continuum of consumers who heterogeneously care about their privacy. This modeling framework appears to be relevant to assess the difficulty to introduce innovation in data-driven markets and challenge current dominant players. It is also found to be useful to evaluate the effects of a new regulatory instrument implemented recently by the GDPR: the right to data portability. It can further be used to rationalize the establishment of data sharing contracts between competitors, a growing phenomenon occurring online.

Benchmark Model

To study the interactions between a firm and its user base, consider an online monopolist whose business model is based on revenues from the disclosure to third parties, such as advertisers or data brokers, of consumers' data it is able to collect. In order to subscribe to the service it offers, users are required to provide a "fixed"

amount of basic information about themselves. Once they have provided these basic data, they can enjoy different features of the service. But in order to enjoy such features, they have to provide additional data of an amount that varies with the usage intensity of the service. If a consumer wishes to have a deeper usage of the service, she would need to provide more data compared to one who decides to have a more moderate usage. For instance, Facebook requires basic information such as gender and age in order to subscribe to the platform. Then, in order to consume the different personalized functionalities of the platform, such as sharing photos, today's mood, or outside activities, consumers need to provide additional information about them to the platform. Therefore, a deep user of Facebook will have provided more of her data compared to a more moderate user. In light of the view of data as a currency, this feature is equivalent to the monopolist charging a two-part tariff to its consumers.

The monopolist has some skills in data analytics, which enables it to have a better knowledge of its consumers. It uses it for two purposes. The first one is to disclose meaningful information to third parties, enabling them to make targeted advertising for instance, so as to generate revenues. The second is to develop new personalized features which will be available to users in the future.

On the demand side, consumers are heterogeneous in their willingness to disclose private information online. In other words, they have different perceptions of the risk of data breaches, such as cyberattacks, which could harm their privacy, irrespective of the websites they patronize. This risk perception increases with the amount of data online firms disclose to third parties. Users all have the same initial preference for the service but depending on their degree of privacy concerns and the amount of data the service requires and discloses to third parties, some users will decide to subscribe and some will not. They may find it hard to anticipate that by disclosing their information, the value of the service could increase for them in the next period, through new personalized features of the service, as online firms often maintain a culture of secrecy over their R&D activities.

The equilibrium in the initial period is characterized by a level of required and disclosed data to third parties and a privacy threshold above which consumers choose not to subscribe. Consumers balance the benefit they get from subscription and the cost they incur from disclosing their data. Those who decided to provide the amount of data required for subscription, i.e. the fixed cost, but located close to the privacy threshold, will have a moderate use of the service compared to others who do not care much about their online privacy.

At the beginning of the second period, the monopolist has been able to analyze the data consumers have heterogeneously been providing it in the first period. If the amount of data collected is large enough and provided consumers are relatively homogeneous beside their privacy concerns, the monopolist is able to make inferences about each consumer; even those that did not provide much data initially. The monopolist is now able to offer additional personalized features that increase every consumer's valuation for the service. In the second period, each consumer is therefore incentivized to disclose more data in order to enjoy the new features of the service. This could even incentivize consumers who decided in first period not to subscribe to do so. The interesting feature is the externality that consumers who have a low

valuation for privacy exert on the others, incentivizing them to disclose more data than they would have initially delivered. This externality could be stronger in the presence of direct network effects.

Thereby, in a dynamic setting, consumers have heterogeneous and increasing valuations for the service due to personalization, enabled by data analytics. Consumers find themselves increasingly locked-in, as the costs of switching while preserving an equivalent level of personalization increases, and the monopolist gets escalating revenue from disclosing information as the amount of data it collects increases over time. Therefore, this model could explain the ever-increasing market power of online firms, the tendency for market tipping, and the data disclosure behavior of the most privacy concerned users.

Market Entry and the Right to Data Portability

Laid down in the European Union's GDPR passed in April 2016, the right to data portability allows internet users to obtain personal data they had transmitted to an online service and transfer them to another data controller. The right to data portability aimed at reducing consumers' lock-in by reducing the switching costs related to re-entering the same amount of data they already have provided in order to obtain a similar degree of personalization and therefore value of the new service. The implementation of the right to data portability by the GDPR distinguishes between data provided by users, which are portable, and data derived and inferred by the firm through data analytics, which users are not able to obtain.

Consider an innovative entrant in the monopolistic market developed in the baseline model where, at first, consumers are not able to port their data. It is able to provide a greater initial value than the incumbent at subscription to all users, besides the fact that it has no access to users' data. However, not all consumers will switch because some, who have a deeper usage of the incumbent's service, derive from it a higher utility than they would initially get from the entrant. Some consumers may have a higher preference for the new service, but have to incur switching costs related to getting used to and re-entering data to the new service which deters them from switching. Only the most privacy-concerned users, which are less locked-in and have a relatively lower valuation for the incumbent's service, will end up switching, and deliver the fixed amount of data required by the entrant. Users who will switch are the ones that care the most about their privacy, and the ability of the entrant to collect data, and therefore to increase the value of its service and generate revenue via disclosure, appears to be limited as compared to the incumbent. Imbalances between the incumbent and the entrant are likely to persist in a dynamic setting as the incumbent may be able to increase the value of its service at a higher rate than the entrant due to differences in the composition of their respective consumer base. Thereby, some users who previously switched may decide to switch back if the value offered by the previous monopolist is higher than the entrant's. Incentives for switching back are facilitated by lower switching costs as the incumbent's service is already known and the previous degree of personalization can be recovered if the incumbent has kept their data.

If users are able to port their data, switching costs are reduced as they do not have to re-enter data they have already provided to obtain a similar level of personalization. More consumers, who on average care less about their privacy, will end up switching, which increases the entrant's data collection ability. However, not all consumers will switch as the incumbent is able to deliver greater value to least privacy-concerned, through data it has inferred of them which are not portable, compared to the entrant. A possible strategy for the entrant would be to pay these consumers in exchange of porting their data, so as to attract them and increase the probability of successful entry. In anticipating entry, the incumbent also faces an *ex ante* trade-off between lowering its data collection so as to limit the amount of data which will be available to the entrant, and increasing it, in order to increase the consumer's preference for its service thereby deterring them from switching.

Data Sharing Contracts between Differentiated Competitors

Online platforms are observed to establish data sharing contracts with their competitors. The "Facebook login" API is an example of such contracts. It enables Facebook users to login on other platforms which will receive some data they have provided to Facebook. In exchange, these platforms share with Facebook data users provide on their website. Such conduct of Facebook has recently been investigated by the *Bundeskartellamt* which concluded that it constitutes an abuse of dominance, enabling it to limitlessly amass consumers' data from other sources. The framework previously developed is found to be useful to give a rationale to such contracts in two respects, and to assess whether they should be allowed under competition law.

First, a dominant firm could find it profitable to offer a data sharing contract to a smaller differentiated competitor with some users multi-homing the two platforms. Consider a contract which grants the smaller firm access to the dominant firms' database in exchange of the small firm continuously sharing some of the data she collects over time with the dominant one. This contract can be in the interest of the smaller competitor, which will be able to increase the value of its service and therefore attract more consumers. The dominant firm could also be interested in this contract as it will be able to continuously amass consumers' data of a different scope than it is able to collect if the two firms are horizontally differentiated, and consequently increase the output of its data analytics activities. From the smaller competitor's perspective, an increase in its consumer base enabled by this contract comes at the cost of a restricted privacy policy. Such contract may constitute an abuse of dominance, aiming at increasing the dominant player's data collection possibilities and consequently its market power.

Second, still in a horizontally-differentiated framework with two or more competitors of similar size, this model enables to study the establishment of data sharing contracts from a collusive perspective. Users, who can multi-home, provide data to each of the competitors individually, according to the disclosure strategy each of them has set. By contracting on continuous data sharing, along with an increasing complementarity and interoperability between the different services, firms could collectively acquire more data of different scopes, thereby increasing consumers' valuation for all services. As

data dissemination across entities increases the risk of privacy breaches, consumers highly concerned with their privacy could decide to stop patronizing the services, but new entry, which could best suit their privacy concerns, is likely to be deterred as consumers find themselves increasingly locked-in with the existing services.

[1] Upcoming PhD student at Télécom Paris.

[2] Miyazaki, A. D. & Fernandez, A. (2001), "Consumer Perceptions of Privacy and Security Risks for Online Shopping," *Journal of Consumer Affairs*, Vol. 35, 27-44.

[3] Barnes, S. B. (2006), "A privacy paradox: Social networking in the United States," *First Monday*.

[4] Chellappa, R. K. & Sin, R. G. (2005), "Personalization versus Privacy: An Empirical Examination of the Online Consumer's Dilemma," *Information Technology and Management*, Vol. 6, 2-3.

[5] *Autorité de la Concurrence* and *Bundeskartellamt*, (2016), "Competition law and data."