“FOR THE PUBLIC BENEFIT”: WHO SHOULD CONTROL OUR DATA?

BY SARIT MARKOVICH & YARON YEHEZKEL[1]

1 Kellogg School of Management, Northwestern University (s-markovich@kellogg.northwestern.edu). and Coller School of Management, Tel Aviv University (yehezkel@tauex.tau.ac.il)., respectively.

CPI ANTITRUST CHRONICLE JULY 2022

Visit www.competitionpolicyinternational.com for access to these articles and more!

"FOR THE PUBLIC BENEFIT": WHO SHOULD CONTROL OUR DATA?

By Sarit Markovich & Yaron Yehezkel

Data-driven platforms collect and commercialize users' data through advertisements or by selling the data to other third part providers. This practice raises the question of who should have the right to decide which data to collect and commercialize, users or the platform? In this article, we describe the results of our research concerning regulating platforms' data collection. The main feature of our research is the distinction between the user's private benefit from data and the data's public benefit—i.e., the benefit a user's data provide to other users on the platform, regardless of whether they are sharing their data. We find that when users differ in their disutility from the commercialization of their data and the public benefit of data is high (low), it is welfare enhancing to let the platform (users) control the data. In contrast, when heterogeneity is in the disutility from the commercialization of different data items, it is welfare enhancing to let users (the platform) control the data when the public benefit of data is high (low). Furthermore, we find that an entrant platform may choose to give users control over their data as doing so can help it overcome the advantage the incumbent enjoys.

I. INTRODUCTION

Many platforms base their business model on the commercialization of their users' data. For example, search engines, such as Google, can collect data on users' location and keyword search. Navigation apps, such as Waze, can collect data on users' preferred routes and other driving habits. Media streaming platforms, such as Spotify, Pandora, and Deezer can collect data on users' music preferences and listening habits. Wearables, such as Fitbit, Garmin, and Samsung Watch can collect data on users' sport activities and performances. These platforms can then use the data to improve their services, but at the same time, the data can also be used for commercial purposes such as selling it to advertisers or to other third-party providers.

This raises the question of who should own the property rights over users' data? Specifically, who should have the right to decide which data items to collect and which to commercialize? On the one hand, the platform is the party that collects and analyzes the data, and users give their consent to data collection when joining the platform. In fact, it is the platform that turns the data into a valuable resource. On the other hand, users are the party that generates the data, and in many cases, bear a disutility from having their data shared. Furthermore, users typically do not have the choice to join the platform without agreeing to give away the rights over their own data.

This question has important implications for the ongoing debate on the need for data regulation. Existing U.S. laws give the property right over data to the entity that collects it. Platforms can collect and own users' data on the basis of users' consent to join the platform.[2]

Yet, when platforms have strong market power, users' voluntary consent to the platform's data policy is controversial. For example, in 2020, the U.S. Department of Justice filed a suit against Google, claiming (among other things) that "American consumers are forced to accept Google's privacy practices, and use of personal data..."[3]

Another case in point is Facebook's questionable announcement in 2021, that its users must agree to let Facebook and its subsidiaries collect their personal data on WhatsApp, including phone numbers and locations. In an extension to competing platform, preliminary results show that platforms may choose different data policies. The platform that benefits from a leading position in the market chooses to control the data while the new platform enables users that join it to control their data.

If users don't accept the new terms and conditions, they will be forced out of the app.[4] This is especially interesting given that WhatsApp has always positioned itself as a privacy focused service – encrypting all users' messages. Indeed, WhatsApp potentially has access to a lot of its users' data – phone number, contact lists, messages' content. Its intention to keep encrypting messages and not sharing this data while sharing other data items, like phone number and location, suggests that WhatsApp believes that users' disutility from sharing phone number information with Facebook is lower than their disutility from sharing messages content.[5]

In contrast to the U.S., the EU General Data Protection Regulation ("GDPR") is designed to provide users with the choice to share data; a choice that does not discriminate those that choose not to share data. The GDPR aims to move platforms from a regime that provides the platform with full control over users' data, to a regime that enables users to join a platform and enjoy, at least part of, its services without being required to give their consent to share specific data.

This paper examines the above research question using a theoretical model. We find that whether a regime that gives the platform control or a regime that gives users control over their data is welfare enhancing depends on market conditions such as the type of consumers' heterogeneity, as well as what we refer to as the public benefit of data (on which we elaborate below). We therefore argue that it is not necessarily the case that giving users control over their data is beneficial. Instead, regulating data-driven platforms should be on a one-to-one basis, depending on the dominant platform and market conditions. Below, we describe our theoretical model and findings which identify the market conditions under which it is beneficial to impose a regime that provides users control over their data. We further explain the results and intuition behind them. We conclude with some policy implications.

---

2 See Economides, Nicholas, & Ioannis Lianos. "Restrictions on Privacy and Exploitation in the Digital Economy: A Market Failure Perspective." *Journal of Competition Law and Economics* (forthcoming).

3 See: *The Verge*, Oct 20, 2020. Available at https://www.theverge.com/2020/10/20/21454192/google-monopoly-antitrust-case-lawsuit-filed-us-doj-department-of-justice.

4 See, for example, *The Verge*, Feb 22, 2021. Available at https://www.theverge.com/2021/2/22/22294919/whatsapp-privacy-policy-may-15th-messaging-calls-limited-functionality.

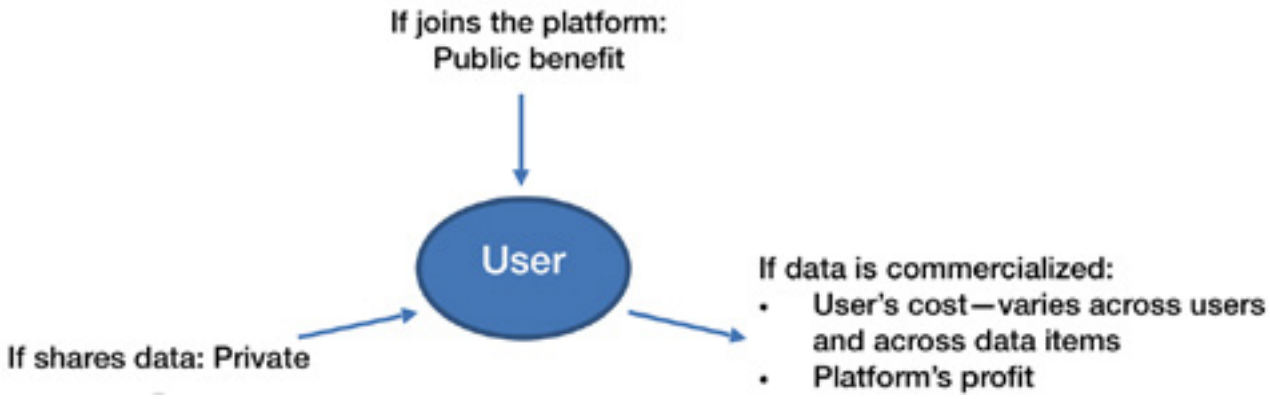
5 See: *The Verge*, Oct 20, 2020. Available at https://www.theverge.com/2020/10/20/21454192/google-monopoly-antitrust-case-lawsuit-filed-us-doj-department-of-justice.

II. THE BENEFITS AND COST OF DATA

Our research considers a game between users and a platform (or competing platforms) that collect and can commercialize users' data. The platform collects multiple *data items*. For example, Waze collects data on location, time, and route that users take; Fitbit collects data on the number of steps its users take and their heart rate; and Facebook collects data on text and photos users upload as well as posts they read, the people and groups they follow, etc. Each data item provides three potential benefits:

1. **Private benefit for users**. For example, if users share data on their number of steps and heart rate with a fitness tracker, the platform can help these users to monitor their training and provide them with recommendations concerning more effective training. Likewise, when a driver uses a navigation app and agrees to let the app track their route, the data collected can help direct the driver to un-congested routes.
2. **Commercial benefits**. The same data provides the platform with *commercial benefit*. The fitness tracker or navigation app, in our example, can sell the user's data to advertisers. Users, however, bear disutility from having their data shared for commercial benefits. This disutility may differ across users. For example, some users are more sensitive to their privacy than others. Moreover, this disutility may differ across data items. For example, users may not care about Waze sharing information about the route they take but suffer disutility from Waze sharing their exact location at a specific point in time. Similarly, users' disutility from Fitbit sharing one's number of daily steps may be lower than that of sharing their heart rate.
3. **Public benefits**. The novel feature of our model is that the data may also benefit all other users that join the platform, regardless of whether they share their data — i.e. provide a *public benefit*. For example, the data collected by a fitness tracker from an individual user can help the fitness tracker provide better training recommendations to all other users. Fitbit's uses its heart rate data to identify episodes of irregular heart rhythm suggestive of atrial fibrillation ("AFib"), the most common form of heart rhythm irregularity. Fitbit intends to use this information to alert users about an irregular heart rhythm so that notified individual would connect with a doctor. Likewise, data collected from a driver can benefit other drivers that consider taking the same route. Other relevant examples are users that provide their location data on a contact-tracing app benefit others who now know they were in proximity of someone who tested positive for COVID-19. Contact tracing apps use one's phone, or other mobile device, to track and alert individual if they'd crossed paths with someone who within a certain window of time tested positive to COVID-19.[6] This third public benefit of data is the most important one for innovation and product improvement, as it implies that data creates positive externalities where users can benefit from other users' data, regardless of whether they share data themselves. Figure 1 below summarizes the different benefits and costs to the user and the platform:

Figure 1: Benefits and Costs of Data



To study who should control users' data, we study three extreme data regimes. In the first regime (hereafter, "regime 1"), the platform has the right to decide which data items to collect and commercialize. Users who want to join the platform must consent to sharing the data items the platform chooses to collect and commercialize. That is, users can only decide whether to join the platform (and agree to its data policy), or stay out. The second regime (hereafter, "regime 2") does not allow the platform to contingent users' participation in the platform with their consent to collect their data. The third regime (hereafter, "regime 3") does not allow the platform to contingent users' participation or data collection on their consent to the commercialization of their data. In this case, in order to incentivize users to allow the platform to commercialize their data, we allow the platform to compensate users for selling their data.

6 Contact tracing apps use one's phone, or other mobile device, to track and alert individual if they'd crossed paths with someone who within a certain window of time tested positive to COVID-19.

We find that the different benefits of data create market inefficiencies. The platform only cares about the commercial benefit, and will thus collect data as to maximize this benefit, subject to the constraint that users agree to join it. Users only care about their own private benefit. If given the opportunity to decide which data to provide the platform, users would only provide data that offers them private benefit, as they enjoy the public benefit regardless of their data contribution. Most ill-considered, however, is the public benefit of data. Although it provides benefits to all on the platform, the public benefit is, at least partially, ignored by both the platform and the users. That is, both parties ignore that while data collected on an individual user may create a disutility for this user, it may benefit the platform's entire user-base. These market inefficiencies raise the question of which regime achieves the best balance between the benefits of data (public, personal, and commercial) and disutility to users, as well as whether competition can mitigate these market inefficiencies. We find that giving users full control over their data is not always welfare enhancing, as it may result in too little data collected for the public benefit.

III. USER OR DATA COVERAGE?

In general, the platform's optimal strategy can take one of three possible outcomes: all data is commercialized but not all users join (i.e. full data coverage but partial user coverage) as some users' disutility from the commercialization of their data is higher than the benefits from joining the platform ; all users join but not all data is commercialized (full user coverage and partial data coverage) as some data items exhibit high commercialization disutility; or partial user and data coverage. As it turns out, our results and intuition crucially depend on whether the market is mostly characterized by data coverage or user coverage, which further depend on whether the market is mostly characterized by users with different disutility from the commercialization of their data (hereafter, "heterogeneous users"), or by data items that differ in the disutility that commercializing them inflicts on users (hereafter, "heterogeneous data items").

Consider first the case of heterogeneous users. In this case, if data does not have any public benefit, regime 1 and regime 2 are identical. When the public benefit of data is positive, in comparison with regime 2 where users control their data, regime 1 that gives the platform control over data has the disadvantage that fewer users join the platform. This is because users that are sensitive to their privacy prefer not join the platform over the possibility of joining and adhering to the platform's strict data sharing policy. At the same time, regime 1 has the advantage that all users who join the platform give all data requested by the platform, which then provides public benefit. That is, there is a tradeoff between the number of users that join the platform and enjoy the public benefit and the amount of data collected which thereby provides public benefit.

We find that when the public benefit per data collected is small, the first effect dominates and since more users join the platform under regime 2, it is welfare enhancing to give users control over their data. As the public benefit of data increases, more users join the platform under regime 1, because they would like to enjoy the public benefit of data. That is, users compare their costs to the sum of the private benefit and the public benefit rather than just to the private benefit (see Figure 1). As a result, the disadvantage of regime 1 decreases while the advantage of regime 1 becomes stronger. Consequently, when the public benefit of data is high, it is welfare enhancing to give the platform control over data. These results highlight the important role the public benefit of data plays when evaluating data regulation.

Interestingly, the opposite conclusion emerges in the case of homogeneous users and heterogeneous data items. In this case, it is welfare enhancing to let the platform control the data when the public benefit of data is low, while giving the users control on their data is welfare enhancing only when the public benefit of data is high. The intuition for this result is that when users are homogeneous, under both regimes 1 and 2 all users join the platform. Recall that under regime 2 users can enjoy the public benefit regardless of whether they share their data. It follows that, under regime 2, the platform can only commercialize data items that provide users with high private benefit, because otherwise users will not agree to share these data items. Yet, under regime 1, the platform can "bundle" data items with low disutility and some data items with high disutility because data items with disutility that is smaller than the private benefit leave positive surplus for users. The platform, then, can force users to agree that the entire "bundle" of data items is commercialized, or they stay out of the platform. That is, relative to regime 2, under regime 1, the platform can collect more data for commercial benefit. As the public benefit of data increases, regime 1 enables the platform to bundle even more data items with high disutility, which makes regime 1 less beneficial to welfare in comparison with regime 2. Notice that this is in contrast to the case of heterogeneous users, where welfare is higher under regime 1 when the public benefit of data is high. Table 1 summarizes the comparison between the two regimes under heterogeneous users and data, and the differences in intuition between the two cases.

Table 1: Comparison Between Heterogeneous Users and Data

Type of Heterogeneity	The main problem with regime 1 (in comparison with 2)	Effect of an increase in the public benefit of data	Result
Heterogeneous Users	Partial user coverage: "Sensitive" users do not join	An increase in the public benefit mitigates this problem because it attracts more users to join	For high public benefit, welfare in regime 1 is higher than in regime 2
Heterogeneous data	Partial data coverage: too much data is commercialized: the platform "bundles" data	An increase in public benefit exacerbates this problem because the platform can bundle more costly data items	For high public benefit, welfare in regime 1 is lower than in regime 2

We also examine the case where the market exhibits both heterogeneities together — i.e. users differ in their disutility from the commercialization of their data and data items differ in the disutility their commercialization imposes. We find that, if the public benefit of data is small, the platform focuses on user coverage. Once the user market is fully covered, the platform turns its focus to commercializing more data items. That is, with both heterogeneities, if the public benefit of data is low, welfare dynamics follows the dynamics in the "heterogenous users" case. As the public benefit of data increases, all users join the platform and welfare dynamics follows the dynamics in our heterogenous data case.

IV. COMPETITION

To study whether competition motivates platforms to adopt the welfare maximizing data regime as well as whether regulating data collection can facilitate entry, we extend our analysis to competition between two platforms: an incumbent platform that benefits from a "focality" advantage — users expect other users to join the incumbent. The second platform is an entrant that suffers from a non-focal position, yet can offer a better base quality due to innovative new features. Our results show that competition does not necessarily motivate platforms to give users control over their data. Furthermore, competition does not necessarily motivate either platform to choose the welfare-maximizing regime. Nevertheless, the entrant has stronger incentives than the incumbent to give users control over data and may choose to do so when doing this enables it to overcome its non-focal position. Specifically, an entrant may choose to give users control over their data as a tool to differentiate itself and strengthen its market position.

V. COMPENSATING USERS FOR THEIR DATA

Finally, we analyze the third regime which in essence provides users with all the control over their data: the platform needs to ask users for their consent to collect and to commercialize the data. A user may agree to collecting the data for private and public benefit, yet require monetary compensation in order to agree to have the data commercialize. While one would expect such a regime to lead to the first best, we find that this regime is not always welfare enhancing. While under this regime all users join and share data for public and private benefits, since the platform needs to pay users for agreeing to share data for commercial benefit, the platform may choose to collect too little data.

The advantage of requiring the platform to compensate users for their data is that the platform internalizes users' disutility from the commercialization of their data. Under heterogeneous data, this leads to the first best. Under heterogeneous users, however, since the platform cannot discriminate across users — i.e. pay more to users with higher disutility — the platform under-collects data for commercial benefit. In this case, this regime may underperform the first two regimes that we consider, especially when the commercial and public benefits of data are high.

VI. POLICY IMPLICATIONS

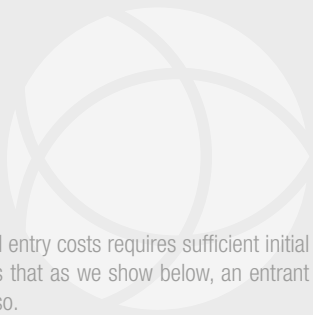
Our results suggest that whether the EU's firmer approach to data regulation as compared to the U.S. enhances welfare, depends on the magnitude of the public benefit of data and the type of heterogeneity in the market. More generally, our paper provides specific conclusions on how to regulate dominant data-driven platforms. When data have significant public benefits and the market is characterized by heterogeneous users, such that users that are relatively sensitive to privacy prefer to stay out, the regulator should not intervene in the platform's data policy. In this case, regulation will result in fewer users giving data for public benefit and may eventually reduce consumer surplus as well as social welfare.

When the market is characterized by mostly homogeneous users and is almost fully covered, regulation that requires the dominant platform to give users control over data can enhance social welfare.[7]

We focus on platforms that do not have high fixed entry costs into a new market. Naturally, a new platform that needs to cover its fixed entry costs requires sufficient initial profits. Hence, regulating the data policy of such new platforms may deter entry. Another argument against regulating a new platform is that as we show below, an entrant platform may independently choose to give users control over data in order to gain a foothold in the market, if the incumbent does not do so.

Lastly, we want to emphasize that the question of who should control our data is also – perhaps foremost – an ethical question of social morality. Is it ethical to allow a platform to collect our personal data items? The moral aspects of this question are important but are beyond the scope of our theoretical model. The goal of our research is to contribute to the debate on data regulation by highlighting some economic forces, specifically, with regards to the public benefit of data. Our results and potential policy implications cannot be placed in isolation from a discussion on the moral aspects of privacy and data protection.

In a somewhat related moral debate in Israel, the question is whether to allow public authorities share information concerning the identity of civilians that did not receive the COVID vaccine. Such data may have valuable public benefit in fighting COVID, yet may violate civilians' privacy rights.



7 We focus on platforms that do not have high fixed entry costs into a new market. Naturally, a new platform that needs to cover its fixed entry costs requires sufficient initial profits. Hence, regulating the data policy of such new platforms may deter entry. Another argument against regulating a new platform is that as we show below, an entrant platform may independently choose to give users control over data in order to gain a foothold in the market, if the incumbent does not do so.

CPI COMPETITION POLICY™ INTERNATIONAL

CPI Subscriptions

CPI reaches more than 35,000 readers in over 150 countries every day. Our online library houses over 23,000 papers, articles and interviews.

Visit competitionpolicyinternational.com today to see our available plans and join CPI's global community of antitrust experts.