

“The Economics of Social Data” by Dirk Bergemann, Alessandro Bonatti, and Tan Gan

Summary by Michelle Fang

Big tech companies like Google, Amazon, and Facebook have an unprecedented collection of individual data. The constantly expanding user base has allowed these digital platforms to acquire massive amounts of data about consumer preferences, locations, friends, political views, and almost every facet of their life. This data is now beginning to receive considerable attention from consumers, media, and authorities alike. Germany’s top court ruled that Facebook violated competition laws by combining data collected about users across different Facebook platforms, such as Instagram and WhatsApp, and other third-party apps. Google announced a data auto-delete feature added in a user’s default setting shortly after rival Apple released new privacy features to hide users’ data from third-party trackers.

The services on these digital platforms rely critically on individual-level data to provide refined search results, personalized product recommendations, and targeted advertisements. Consequently, most data regulation on Internet platforms focuses on ensuring consumer control over their individual data. Regulators hope that ownership and control over one’s own data will result in appropriate compensation of the data one chooses to reveal. However, economists need to consider the social aspect of data collection, in which an individual user’s data is predictive of the behavior of others. Therefore, individual data is in reality social data, and the social nature of data leads to an externality. For example, in the context of shopping data, an individual’s purchase on Amazon will convey information about the likelihood of purchasing a certain product among consumers who have similar purchase histories.

The authors dispel the idea that empowering a consumer to take control of her data will prevent negative consequences since she can demand compensation from the intermediary. A consumer’s choice to release data takes into account her private benefits and costs only, and not the externality generated by the data she provides. Moreover, data externalities in the form of diminishing marginal returns to individuals’ information reduce the intermediary costs of acquiring information. While consumers can experience positive externalities like real-time traffic information, very little checks the platform from trading data for profit in ways that harm consumers. The paper provides a novel explanation for the digital privacy paradox, in which, small monetary incentive have large effects on subjects’ willingness to sell their private data (Athey, Catalini, and Tucker 2017). Specifically, aggregate data collection causes consumers to consent to any use of their data for very little aggregate compensation. However, it is possible that data is substituted and the data market is not open when socially optimal to do so.

The paper demonstrates how data ownership is insufficient to bring about the efficient use of information since arbitrarily small levels of compensation can induce a consumer to relinquish her personal data. What barriers or guarantees in terms of privacy does consumer control then provide?

The authors analyze three critical aspects of the economics of social data: how the collection of individual data changes terms of trade among consumers and digital

platforms, how the social dimension of data magnifies the value of individual data, and how data intermediaries change the level of aggregation and precision of the information they provide. A data intermediary acquires information of consumer demand and sells some version of that data to the producer. An example of a data intermediary with significant market power is Google, a large internet platform that sells targeted advertising spaces.

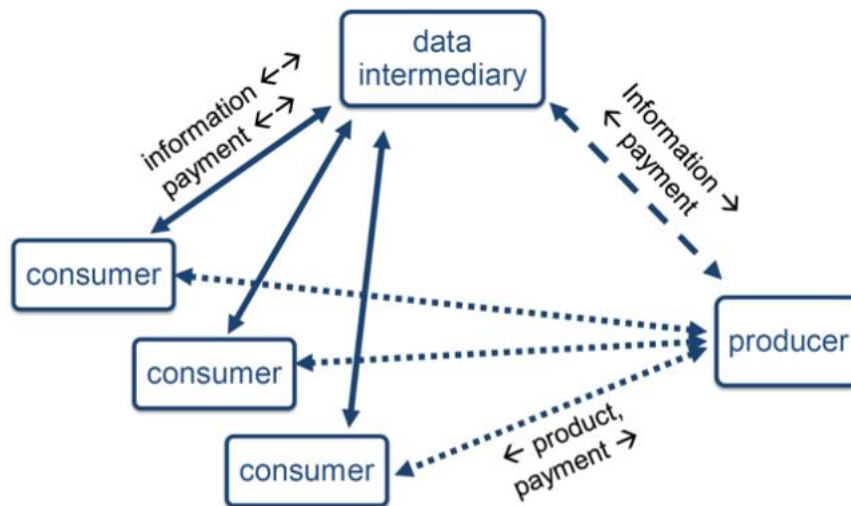


Figure 1: Data and Value Flows

In the model the paper uses, the data market has a single intermediary, a single producer, and many consumers. The data intermediary offers bilateral contracts, which determine data price and data policy, ex ante or before the consumer realizes the demand shock. The ex-ante constraint illustrates the prevailing condition where Facebook and Amazon require an account to be established where consumers and platforms accept the “terms of use/service” agreements before the realization of any event. In this example, let us set S as the information structure that includes data about demand or a consumer’s willingness to pay. The individual’s willingness to sell their personal data depends on the producer’s response in pricing to the acquired data. Perfect information of demand allows the producer to engage in price discrimination on the individual or market level since the producer gains revenue by tailoring his price to demand. As the producer personalizes the pricing policy, the producer also increases his monopoly price. The equilibrium quantity increases with the consumer’s willingness to pay, but only at half the rate it would if the producer had common information about demand and provided consumers with a constant price. Thus, consumer surplus decreases due to the responsive price.

The paper generates insights and presents additional findings on data flow, aggregation, and intermediation. First, a consumer’s decision to share data depends on her anticipation of how the intermediary uses the newly gained information. The intermediary implements data outflow policies that maximize the producer’s surplus. As

the intermediary maintains complete control over the acquired data and the way it is used, an opacity in how data inflow is linked to data outflow results.

Moreover, the collection of aggregate data does not prevent the occurrence of data externalities, but it allows consumers to retain some form of privacy. This further reduces the intermediary's cost of acquiring their information, more so than it does the value of the data for the producer downstream. If one consumer does not participate in the contract, the producer will optimally aggregate all available data to form the best predictor of the missing data point. Thus, aggregate data policies maximize intermediary profits by minimizing loss of surplus.

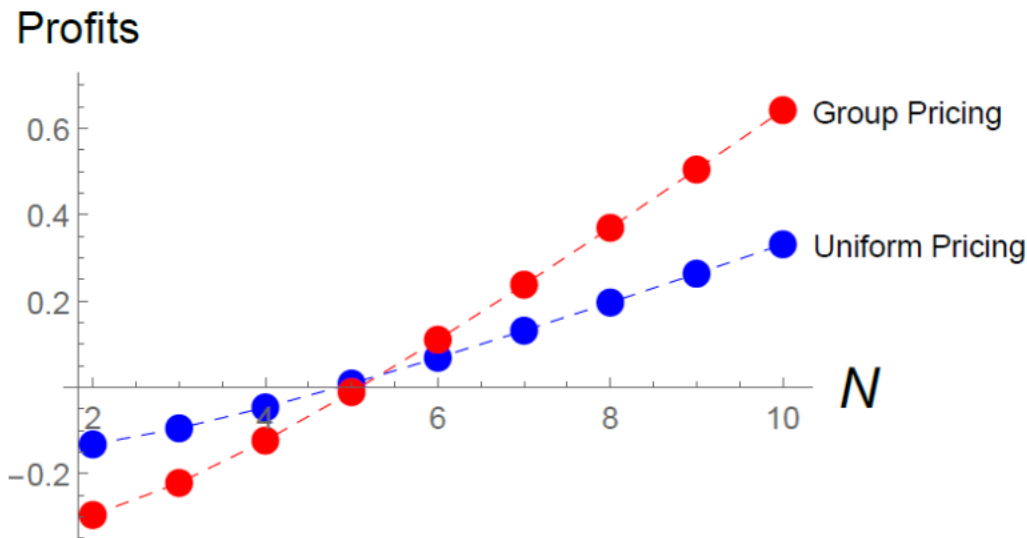


Figure 8: Marginal Value of an Additional Consumer ($\sigma_w = 1, \sigma_e = 0$)

Finally, if the data externality is significant enough, the intermediary can acquire the consumers' information for sufficiently cheap that it can generate positive profits even if information reduces total surplus. Information structures that maximize the intermediary's profits leverage demand data with optimal noise levels. The key elements of information design are aggregation and noise. There are two main types of noise: idiosyncratic (individual taste) and common (all consumer's taste). Whether adding idiosyncratic or common noise is optimal depends on if anonymization is applied. If the anonymization is applied, then adding common noise is optimal while adding the idiosyncratic noise is not. And the result is reverse if the anonymization is not applied.

However, the paper finds that intermediation becomes increasingly profitable in larger markets since consumers impose severe data externalities on one another as the value of information for the producer increases boundlessly. Moreover, if the market is sufficiently large, the intermediary will transmit group characteristics that allow the producer to discriminate across but not within groups. For instance, Uber and Amazon claim they do not discriminate on the individual level but price discriminate based on location and time.

Finally, the paper concludes that the data intermediary could acquire consumer data and commit to not passing the information to the producer in exchange for compensation from the consumer. The authors establish that the revenue-maximizing data policy is to acquire all consumer data but not to forward it downstream to the producer. Moreover, competition might exacerbate consumer surplus and not effectively correct for data externalities. A potential solution is consumer unionization to internalize the data externality and earn bargaining power against powerful intermediaries. However, there exists theoretical and practical implementation challenges to consumer unions.