Data access, consumer interests and social welfare – An economic perspective on data

Bertin Martens

A. Introduction

This chapter presents an economic introduction to digital data, data access and the well-being of consumers and society at large. 'Data access' may cover a variety of modalities of data exchange between two or more parties, from monetised trade to free access or exchange of data in return for a service. Any voluntary data exchange is a market-based transaction. A key question for data policy makers is whether private voluntary data access decisions maximise the welfare of society as a whole. Economists define market failures as situations where the aggregate private welfare of firms and consumers remains below the total welfare that society as a whole could achieve. This occurs when the incentives of private firms and/or consumers make them behave in ways that diminish overall social welfare. This may justify regulatory intervention in data markets and the imposition of mandatory access conditions that overrule private decisions. The focus of this chapter is therefore on data market failures.

Following the European Commission's 'Better Regulation Guidelines'[1] we take a broad approach to possible regulatory intervention in data markets and data-driven services markets. It includes monopolistic market failures that are usually handled by competition law and extends to other sources of market failures such as externalities, asymmetric information and missing markets because of high transaction costs and risks. We also point out potential regulatory failures and social concerns, such as welfare distribution and discrimination that could motivate regulatory interven-

---

[1] In European Commission 'Better Regulation Guidelines' (2017) <https://ec.europa.eu/info/law/law-making-process/planning-and-proposing-law/better-regulation-why-and-how/better-regulation-guidelines-and-toolbox_en> accessed on 31 August 2020 and Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions – 'A European strategy for data' COM (2020) 66 final, 13 note 39, where the Commission also advocates a market-failure based approach to regulatory intervention in data markets.

tion. While mainstream competition law takes consumer welfare as the policy objective,[2] this chapter takes a wider public policy economics view and focuses on the overall social welfare of society as a policy objective, the combined welfare of firms and consumers. This distinction may become important for example in data-driven online platforms where an exclusive focus on the consumer side may have unintended negative effects on the supply side, and vice versa.

Furthermore, this chapter goes beyond markets and looks at the impact of data on institutions and organisational arrangements in the economy. Digital technology helped to overcome pre-digital information constraints that prevented the realisation of higher levels of social welfare. Digital data contributed to the emergence of new markets for goods and services. These markets often require new ways of organising economic exchange and new types of firms to do this, which are generically labelled as 'platforms', and they give rise to new sources of market failures that need new regulatory interventions.

This chapter is structured as follows. Section B discusses the specific economic characteristics of data that are in several respects different from ordinary goods and services. We explore how these characteristics affect data collection and data use markets. Section C brings digital platforms into the picture, a new type of firms that leverage some of the economic characteristics of data to create new markets. Section D broadens the perspective on data market failures and possible regulatory solutions. Section E adds some concluding observations.

B. The economic characteristics of data

I. Data as intermediary input

Data are usually an intermediary input, not a final consumer good. For example, unless they are aviation aficionados, consumers do not search for flight schedules on Google or Skyscanner because they enjoy looking at these schedules but because they want to buy an air transport service. Data

---

2 See Jason Furman, Diane Coyle, Amelia Fletcher, Derek McAuley and Philip Marsden, 'Unlocking Digital Competition – Report of the Digital Competition Expert Panel' (UK Government 2019) <https://assets.publishing.service.gov.uk/governmen t/uploads/system/uploads/attachment_data/file/785547/unlocking_digital_competi tion_furman_review_web.pdf> accessed 31 August 2020 (so-called 'Furman Report').

are not created ex nihilo. They are collected by firms from observing the behaviour of people, machines and nature – the data originators. Data exchange involves at least two markets, an upstream data collection market and a downstream data use market for the production of goods and services. These two markets can be vertically integrated in a single firm or they can be carried out by different firms that trade data between them. Data collection can happen prior to their use in services, or it can be a by-product of services. For example, the Google Search ranking depends on data collected from users of the search engine. There are many data exchange modalities. They can be traded for a monetary compensation or in exchange for a service, sharing can be for free or subject to conditions in other markets, etc. Data can be traded directly – when they are effectively transmitted between parties – or indirectly – when parties do not transmit data but only a data-driven service. For example, Google online advertising services do not transmit consumer data to advertisers. They sell a targeted advertising service based on consumer data which they keep in-house.

II. Data collection has an economic cost

The data collector needs a financial incentive to invest in data infrastructure, for example because it offers the prospect of monetising data. Data originators also need incentives to share their data with a collecting firm. A frequently observed business model in data collection markets is to offer originators a free service in return for sharing their personal or industrial data. The willingness of data sources to share data with collectors will not only depend on conditions in the data market but also on subsequent use of the data in services markets. For example, the willingness of consumers to share their data with a website will depend on the quality of services offered by that website as well as subsequent use of the data by the website, for instance for online advertising. Firms that offer free services need to find a way to cover the cost of producing these services. Google and Facebook offer free services in return for the ability to monetise user data in online targeted advertising. There are no free lunches in the data economy, though the party that pays for the lunch may be different from the party that enjoys the lunch. Any change in the cost of data collection and in benefits for data users will affect the volume and possibly the quality of data collected.

III. *The value of data depends on their use*

Data have no value on their own; they become valuable only to the extent that consumers and firms can use them to improve their position in data-driven services markets. Economists have tried to get a better understanding of these services markets' effects and the impact on stakeholders. There is no coherent framework yet for the economic analysis of data, though research focuses on the welfare and revenue-shifting potential of data.[3] Pro-competitive data uses imply that both firm revenue and user benefits from a data-driven service increase with additional data. For example, more data collection and more efficient use of the data in hotel booking platforms can simultaneously improve the user experience, revenue for hotels and platform revenue. Competitive use may still cause welfare shifts between firms and their customers and trigger equity and welfare distribution concerns. For example, firms can use data for price discrimination or other forms of discrimination strategies that increase the welfare of the firm but not of all users. All these generic statements on data impact are subject to empirical evidence. This may be easy to obtain for firms that collect large amounts of user data and run behavioural experiments with their online users to decide on their profit-maximising commercial strategies. A wide data-access gap between firms and policy makers often inhibits the design of policies to improve social welfare.[4]

IV. *Excludability and monopolistic data trade*

In contrast to physical goods, data are not excludable by nature. They can easily be copied and disseminated. The law can assign exclusive rights to

---

3 Alexandre de Cornière and Greg Taylor, 'Data and Competition: A General Framework with Applications to Mergers, Market Structure and Privacy Policy' (2020) CEPR Discussion Paper No. DP14446, <https://ssrn.com/abstract=3547379> accessed on 31 August 2020.

4 Business-to-government data sharing initiatives in several EU Member States and by the European Commission seek to bridge this gap. See for example European Commission, 'Towards a European Strategy on Business-to-Government Data Sharing for the Public Interest: Final Report prepared by the High-Level Expert Group on Business-to-Government Data Sharing' (2020) <www.euractiv.com/wp-content/uploads/sites/2/2020/02/B2GDataSharingExpertGroupReport-1.pdf> accessed 31 August 2020. This subject is discussed in detail by Heiko Richter, 'The law and policy of government access to private sector data ('B2G data sharing')', in this volume.

data originators and/or collectors. So far, there are no general data owner-ship rights in the EU or elsewhere.[5] In a few cases, the law grants erga omnes exclusive rights. For example, the EU Database Directive[6] grants, under restrictive conditions, sui generis ownership rights to producers of databases. The EU General Data Protection Regulation (GDPR)[7] grants some exclusive and inalienable control rights to natural persons as data subjects, including the right to give consent to access to personal data, and rights to data access, portability and deletion. The data subject is unam-biguously defined as the rights holder over personal data. This is not neces-sarily the case for non-personal machine-generated data that may be co-generated by several parties. Assigning exclusive rights to one party may af-fect the entire industry value chain.[8] Attempts at assigning exclusive pri-vate rights over data are inspired by the Coase Theorem, which hypothesis-es that markets will work efficiently when ownership rights are well-de-fined and transaction costs are low or zero. However, the intrinsic social value of many data generates externalities.[9] In these circumstances, private rights cannot bridge the gap between the private and the social value of da-ta.

   In the absence of legal private ownership rights, a data-holding firm can apply technical protection measures to protect its de facto exclusive con-trol and access to the data. This enables the firm to raise revenue from sell-ing the data or data-driven services to users, with bilateral contracts that benefit from legal protection under commercial law. However, they can-

---

5 Néstor Duch-Brown, Bertin Martens and Frank Mueller-Langer, 'The Economics of Ownership, Access and Trade in Digital Data' (2017) Joint Research Centre Working Papers on Digital Economy <https://ssrn.com/abstract=2914144> accessed 31 August 2020.

6 Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases [1996] OJ L77/20. See also Matthias Leist-ner 'The existing european IP rights system and the data economy: An overview with particular focus on data access and portability', in this volume.

7 Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) [2018] OJ L127/2.

8 For an example from the agricultural sector, see Can Atik and Bertin Martens, 'Governing Agricultural Data and Competition in Data-driven Agricultural Ser-vices: A Farmer's Perspective', (2020), <www.researchgate.net/publication/3421058 35_Governing_Agricultural_Data_and_Competition_in_Data-driven_Agricultural _Services_A_Farmer%27s_Perspective> accessed on 31 August 2020.

9 See section B.VIII. below.

not be enforced against third parties. In case of data leaks, firms have no recourse against third parties that benefit from these leaks.

Data use markets require monopolistic conditions to generate revenue. If more parties have access to the same dataset, or to close substitutes, competition will drive prices down to the marginal cost of reproduction, which is usually close to zero for digital data. That eliminates opportunities to generate revenue and incentives to invest in data collection and processing. Monopolistic data pricing above marginal cost requires rationing or reducing the quantity (and possibly the quality) of data that can be accessed. Not all demand will be satisfied, unless perfect price discrimination is feasible. Monopolistic trade does not maximise social welfare. It increases the welfare of the data holding firm at the expense of data users. Data access policies require careful balancing between monopolistic and open data markets.

V. Data are not a homogeneous product

Data are subject to quality differentiation and can be traded in various levels of fine graining and information content. Quality differentiation may be necessary to avoid falling into the Arrow Paradox: once data are revealed to a potential buyer there is no point in trading them anymore because the buyer already has the information he wanted to buy. There are many strategies that a potential data seller can apply to reduce the information content of a 'demonstration' dataset to entice a potential buyer while avoiding this paradox.[10] The seller can offer a reduced sample of the data, or a coarse-grained or aggregated version that does not reveal details, or an anonymised version etc. The seller can also refrain from sharing data directly with a buyer and deliver an indirect data-based service only, as in the Google advertising example. Data quality differentiation may facilitate price discrimination between buyers. Trading detailed consumer data with data users may reduce the willingness of consumers to share data with the collecting firm. Data buyers on the other hand will prefer more detailed data because it enables them to price discriminate in services sales. The data intermediary will adjust the quality of the data that he collects from consumers and sells to users in order to maximise his profits.

---

10 For an overview, see Dirk Bergemann and Alessandro Bonatti, 'Markets for Information: An Introduction' (2019) 11 Annual Review of Economics 85.

VI. Non-rivalry and economies of scope in data re-use

Data are non-rival. Many parties can use the same dataset at the same time for a variety of purposes without functional loss to the original data collector. Rival goods can only be used by one party at a time. For example, a car is a rival physical good and can only be used by one driver at the time. If a car were non-rival, all drivers could use the same car at the same time to drive to different destinations. The economic welfare gains would be enormous: it would suffice to invest in the production of a single car to cater to the needs of all drivers. This promise of substantial welfare gains from exploiting non-rivalry in data re-use constitutes the foundation stone of the data access and sharing debates.[11] Data collected by one firm can be reused for other purposes, either by the same firm or by other firms provided they can access the data. The primary data collection effort is a sunk cost that can be amortised across many uses, rather than remaining confined to a single user. It can boost innovation and enable the production of new and innovative data services that the original data collector did not envisage.

Economies of scope in re-use were originally defined in the context of joint production and (re-)use of the same product or asset to produce other outputs.[12] For example, a car manufacturer can re-use the same engines in different car models. Re-use of the same non-rival engine design entails zero marginal re-design costs. However, there is a positive marginal cost for physical re-production of additional engines. Non-rival digital data have quasi-zero marginal reproduction costs because it involves only copying an electronic data file. Still, data re-use by other firms may create interoperability problems and important fixed costs for the design of a data transmission interface.

Data re-use and access by other parties also has a cost side. All digital data can, in principle, be made interoperable and shared for the benefit of so-

---

11 OECD Directorate for Science and Technology, 'Maximizing the Economic and Social Value of Data – Understanding the Benefits and Challenges of Enhanced Data Access' (OECD 2016) DSTI/CDEP(2016)4 <https://one.oecd.org/document/DSTI/CDEP(2016)4/en/pdf> accessed 31 August 2020; Charles I. Jones and Christopher Tonetti, 'Nonrivalry and the Economics of Data' (2020) 110 American Economic Review 2819.

12 David J. Teece, 'Economies of Scope and the Scope of the Enterprise' (1980) 1 Journal of Economic Behavior and Organization 223; David J. Teece, 'Towards an Economic Theory of the Multiproduct Firm' (1982) 3 Journal of Economic Behavior and Organization 39; John C. Panzar and Robert. D. Willig, 'Economies of Scope' (1981) 71 The American Economic Review 268.

ciety.[13] However, neither firms nor individuals want their private data to be widely available. Privacy and commercial confidentiality are important for the autonomy of private decision-making and for extracting private value from these decisions. While non-rival data can be shared by firms and individuals without functional losses, sharing may entail an economic opportunity cost and losses for the original data holder. Other firms may reuse the data in service applications that compete with those of the original data holding firm and undermine the latter's market position.[14] The data holder may also want to produce these alternative services in-house and appropriate the benefits, rather than leaving it to another firm.

Firms and persons will trade off the expected benefits from data sharing against the expected costs and risks that they might incur from doing so. These private cost-benefit perceptions may limit the extent of data exchange, sharing and re-use. The question for policy makers is whether private data decisions by consumers and firms maximise the welfare that society as whole could derive from the data. If not, there is a market failure that may require policy intervention. Data sharing is not an objective in its own right but a means to achieve higher social welfare for society.

VII. *Economies of scope in data aggregation*

A second, and often neglected, source of economies of scope in data comes from data aggregation. Merging two complementary datasets can generate more insights and economic value compared to keeping them in separate data silos. This insight can be traced back to the economics of learning and division of labour.[15] When two datasets are complementary and not entirely separable, applying data analytics – the equivalent of learning – to the merged set will yield more insights and be more productive than applying it to each set separately, especially when the marginal cost of applying analytics to a more complex dataset is relatively small.

Economies of scope in data are controversial in economics, in part because they are misunderstood. Authors usually do not distinguish between

---

13 John Palfrey and Urs Gasser, *Interop: The Promise and Perils of Highly Interconnected Systems* (Basic Books 2012).

14 Hongwei Zhu, Stuart E. Madnick and Michael D. Siegel, 'An Economic Analysis of Policies for the Protection and Reuse of Noncopyrightable Database Contents' (2008) 25 Journal of Management Information Systems 199.

15 Sherwin Rosen, 'Specialization and Human Capital' (1983) 1 Journal of Labour Economics 43.

economies of scale and scope, especially not scope in data aggregation. For example, one author defines economies of scope somewhat ambiguously as cost savings relative to an 'increased level of production of multiple products'.[16] 'Increased level of production' implies economies of scale; 'multiple products' refers to economies of scope in re-use, not in aggregation. A useful way to distinguish economies of scale and scope is to consider a dataset as a two-dimensional spreadsheet, with the number of columns representing the number of variables and the number of rows the number of observations on these variables. Economies of scale refer to increased prediction accuracy due to an increase in the number of rows. Economies of scope refer to increased prediction accuracy due to an increase in the number of columns or explanatory variables. Adding more columns (variables) is not helpful when they are highly correlated or when they are not related at all. A number of empirical studies claim that economies of scope in data are weak or non-existent.[17] All these studies are more about economies of scale rather than scope. Bajari and others[18] come closest to economies of scope in aggregation. They find that product sales forecasts do not become more accurate when historical data from several products markets are aggregated. However, this is explained by weak complementarity among product markets that result in separable datasets and thus in weak economies of scope. The absence of empirical studies on economies

---

16 Catherine Tucker, 'Digital Data, Platforms and the Usual [Antitrust] Suspects: Network Effects, Switching Costs, Essential Facility' (2019) 54 Review of Industrial Organization 683.

17 Lesley Chiou and Catherine Tucker find no decrease in search engine accuracy when time series of consumers' historical searches are shortened because of EU privacy regulation. Nico Neumann, Catherine E. Tucker, Timothy Whitfield, (2019) 'Frontiers: How Effective Is Third-Party Consumer Profiling? Evidence from Field Studies' (2019) 38 Marketing Science 918 show that large data brokers do not necessarily perform better in consumer profiling than data brokers with fewer consumer profile data. Jörg Claussen, Christian Peukert and Ananya Sen, 'The Editor vs. the Algorithm: Targeting, Data and Externalities in Online News' (2019) <https://ssrn.com/abstract=3399947> accessed 31 August 2020, find that more individual user data help algorithms to outperform human news editors but decreasing returns to user engagement set in rapidly. Preston McAfee, 'Measuring Scale Economies in Search' (Lear conference, Rome, 2015) slides available <www.learconference2015.com/wp-content/uploads/2014/11/McAfee-slides.pdf> accessed 31 August 2020, finds that Google Search outperforms Microsoft Bing in long-tail searches because of a higher number of users.

18 Patrick Bajari, Victor Chernozhukov, Ali Hortaçsu and Junichi Suzuki, 'The Impact of Big Data on Firm Performance: An Empirical Investigation' (2019) 109 AEA Papers and Proceedings 33.

of scope in data aggregation is a major gap in data economics. There is some supportive anecdotal evidence. Google gradually improved its targeted advertising by combining personal data from several sources, starting from web searches and adding email and maps (location) data.[19] Navigation apps like Waze and Tom-Tom combine real time GPS location data with maps that are populated with data from a wide range of public and private sources including road and traffic authorities, municipalities, firms and in-map advertisers. These public sector data may have little commercial value on their own but contribute to a valuable service when aggregated with private data.

Economies of scale and scope in data aggregation are a source of positive externalities. In the age of artificial intelligence and machine learning, personal data collected on the behaviour of one set of consumers has predictive value for the behaviour of other consumers.[20] Once a firm has accumulated a critical mass of consumer data, the additional insights obtained from adding another consumer's personal data are small. Acemoglu and others[21] argue that this diminishes the value of individual personal data. Consumers cannot prevent this negative externality and market failure for their personal data. Their best deal is to harvest some consumer surplus by trading their data for an online service that has a higher marginal use value than the depressed market value of their personal data. This could explain the privacy paradox:[22] consumers value their privacy but do not invest in protecting it. They understand the low value of their personal data and the futility of investing in privacy protection in the presence of negative externalities from other consumers' data.

The re-use and aggregation interpretations of economies of scope in data may lead to very different policy implications. Economies of scope in reuse are an argument in favour of data dissemination and de-concentration. Economies of scope in aggregation, by contrast, favour data concentration in large pools. They are not mutually exclusive. Non-rival data can be stored at the same time in concentrated pools and in distributed settings.

---

19 Roger McNamee, *Zucked: Waking Up to the Facebook Catastrophe* (Penguin Random House 2019).

20 Ajay Agrawal, Joshua Gans and Avi Goldfarb, *Prediction Machines: The Simple Economics of Artificial Intelligence* (Harvard Business Review Press 2018).

21 Daron Acemoglu, Ali Makhdoumi, Azarakhsh Malekian and Asuman Ozdaglar, 'Too Much Data: Prices and Inefficiencies in Data Markets' (2019) NBER Working Paper No. 26296 <www.nber.org/papers/w26296> accessed 31 August 2020.

22 Alessandro Acquisti, Curtis Taylor and Liad Wagman, 'The Economics of Privacy' (2016) 54 Journal of Economic Literature 442.

Both concentration and de-concentration can result in market failures that undermine social welfare.[23]

VIII. *The social value of data*

A peculiar characteristic of many[24] data is their social value. Economies of scope in aggregation add a first social dimension to the value of data. Two owners of separate but complementary datasets can only achieve a higher value from their data if they collaborate and pool the two sets. A second source of social value comes from economies of scale. Once a sufficiently large sample of behavioural observations has been compiled to produce robust predictions, those data can be used to predict the behaviour of agents outside the sample.[25] This implies that collecting more data about other agents with similar characteristics has zero marginal value because the existing dataset is sufficiently representative.

These externalities imply an inherent market failure in exclusive private control over data. The party that does (not) provide the data to a collector is not necessarily (may still be) the party that is affected by their use. The de facto exclusive data holder is not necessarily the party that maximises benefits from the data. Pooling data can generate the full social value. However, coordination costs and risks may undermine spontaneous pooling. An intermediary agent may be required in order to realise the social externalities from data pooling and turn them into benefits that pay for the coordination costs and incentivise individuals to participate in the pool. With this, we reach the world of data platforms in the next section.

---

23 Economies of scope in aggregation and re-use exist in intellectual property rights. The market value of a set of complementary patents may be higher than the sum of their separate values. Hence the practice of patent bundling and thickets, and the bundling of standard-essential patents (SEPs) to facilitate re-use of technical standards. Bundling strengthens the monopolistic position of patent holders. Fair, reasonable and non-discriminatory (FRAND) licensing seeks to avoid abusive behaviour.

24 Some types of data may have little or no social value as it remains situation, person or firm-specific and cannot be used to infer something about other agents or situations, or has no complementarity with other datasets.

25 Dirk Bergemann, Alessandro Bonatti, and Tan Gan, 'The Economics of Social Data' (2020) Cowles Foundation Discussion Paper No. 2203R, revised version March 2020 <https://cowles.yale.edu/sites/default/files/files/pub/d22/d2203-r.pdf> accessed 31 August 2020.; Acemoglu and others (n. 21).

C. Platforms and data-driven network effects

Much of the contemporary policy debate on data access is still set in the context of data trade between traditional firms, consumers and data re-users. However, a substantial volume of data exchanges and data-driven services trade takes place in a new type of firms that are usually classified under the generic label of 'platforms'. In this section we explore the crucial role of platforms in the data economy and the benefits and problems that they generate. We start with network effects and then explain the positive and negative roles that they play in platforms.

I. Data-driven network effects

Network effects occur when bringing more users into a group increases the value of the group for all users. For example, when more users join a telephone or social media network, it becomes more valuable to all users and thereby attracts even more users. Data play a role in generating network effects. In some cases that role is very minimal and static. For example, users in a telephone network differ only by their telephone number, a unique lexicographic address. Users can be unambiguously matched by combining two lexicographic addresses. The only dataset required to make the telephone network operate optimally is a telephone directory. Matching between telephone users cannot be improved by observing the behaviour of the users. Similarly, in simple online e-commerce stores, a targeted search for a well-defined product may just require a catalogue of unambiguously defined products. For example, search for a book title in the Amazon book store. In these cases, network effects are mainly driven by the variety of products and users and their unique identification. Data on user behaviour or product quality play no role in networks with an unambiguous matching process. However, in more complex networks matching is not unambiguous and becomes probabilistic. This requires more data than a simple catalogue of lexicographic addresses. For example, matching in search engines and targeted advertising markets requires more data on the characteristics of users and products, beyond a lexicographic identifier, in order to select the most likely and optimal matches. It collects contents of webpages and tallies user clicks on pages in the search ranking in order to better understand the relevance of pages for a specific search term. It will then carry out a probabilistic matching between users and pages, with a ranking of the most likely matches. More precise data on user prefer-

ences will increase the efficiency of probabilistic matching and generate data-driven network effects.[26]

When the quantity and quality of data play an essential role in probabilistic matching we come back to economies of scale and scope in data aggregation.[27] For example, it has been shown that the larger number of users in Google Search make it more efficient in rare search terms than Microsoft Bing, which has a much smaller number of users.[28] That difference in efficiency, in turn, motivates users to shift to Google. Economies of scale mean more observations on similar search terms while economies of scope in aggregation imply collecting search results from a wider variety of search terms. The two may reinforce each other. The rise of artificial intelligence and machine learning has further amplified economies of scale and scope in data aggregation. While human learners can learn a behavioural response from a few observations, machine learning algorithms often require huge numbers of observations to learn an appropriate response.

II. *The role of platforms in the data economy*

Platforms are well-placed to realise the benefits from economies of scale and scope in data aggregation. There are many definitions of platforms, or multi-sided markets in economic jargon, in the economics literature, and there is no consensus among economists on these definitions.[29] Data played no role in the first generation of multi-sided market models,[30] which were an extension of the economics of infrastructure networks.

---

26 Data-driven network effects were first analysed by Jens Prüfer and Christoph Schottmüller, 'Competing with Big Data' (2017) TILEC Discussion Paper No. 2017–006 <https://ssrn.com/abstract=2918726> accessed 31 August 2020.

27 Maurice E. Stucke and Allen P. Grunes, *Big data and competition policy* (Oxford University Press 2016) already speculated that there is a link between economies of scope and network effects or network externalities. Tucker is not convinced. See Tucker (n. 16).

28 McAfee (n. 17).

29 For an overview of the (fairly recent) history of economic thinking on platforms, see for example Bertin Martens, 'An Economic Policy Perspective on Online Platforms' (2016) Institute for Prospective Technological
Studies Digital Economy Working Paper 2016/05 JRC101501 <https://ssrn.com/abstract=2783656> accessed 31 August 2020.

30 Bernard Caillaud and Bruno Jullien, 'Chicken & Egg: Competition among Intermediation Service Providers' (2003) 34 RAND Journal of Economics 309; Geoffrey Parker and Marshall W. Van Alstyne, 'Two-Sided Network Effects: A Theory

They focused on markets with at least two types of users, for instance buyers and sellers. Platforms are faced with a 'chicken and egg' problem: they need many users in order to attract many users. They can solve this problem by charging a very low or zero price to attract many users on one side of the market and charging a higher price to the other side to pay for the cost of the platform. Users with a high price elasticity of demand pay low or zero entry costs while users with low price elasticity pay a higher price. This explains why advertisers pay for ads while users get free access to search and social media services: advertisers have no choice but to advertise in the platform where users with specific profiles are looking for goods or services that the advertiser sells. Users can however multi-home between many platforms to find what they are looking for. These first-generation models ran into problems distinguishing between intermediary platform and ordinary retailers and defining the type of interaction between two sides.[31] To overcome these problems, recent models have broadened the definition of platforms to firms that bring economic agents together and actively promote network externalities between them.[32] In other words, platforms are firms that seek to maximise the social value of data. Economies of scale and scope in data aggregation in a platform ensure that the collective social value of data exceeds the sum of their individual private values.[33] Individuals cannot realise this social value on their own; only platforms can do this through their data aggregation role. Creating a searchable catalogue of products or a directory of users is a first step in generating that social value. For more efficient matching in ambiguous search settings, the platform operator collects more detailed data on buyer preferences and product characteristics. For example, Netflix can improve its

---

of Information Product Design'(2005) 51 Management Science 1494; Jean-Charles Rochet and Jean Tirole, 'Platform Competition in Two-Sided Markets' (2003) 1 Journal of the European Economic Association 990; Jean-Charles Rochet and Jean Tirole, 'Two-Sided Markets: A Progress Report' (2006) 37 RAND Journal of Economics 645.

31 Andrei Hagiu and Julian Wright, 'Marketplace or Reseller?' (2015) 61 Management Science 184.

32 Jens-Uwe Franck and Martin Peitz, 'Market Definition and Market Power in the Platform Economy' (2019) Center on Regulation in Europe Report <https://cerre.net/publications/market-definition-and-market-power-platform-economy/> accessed 31 August 2020. This definition does avoid the problem of setting a minimum number of market sides; one is enough.

33 Bergemann, Bonatti and Gan (n. 25).

film title search engine when it learns more about user preferences and film characteristics.[34]

A comparison with traditional offline markets illustrates the importance of online platforms as data collectors and producers of data-driven externalities. In a traditional town market, buyers walk around and collect information on what is on sale, and sales conditions, and make their choices. The town authority as market organiser has hardly any information on sellers' offers, buyer preferences and actual transactions. Each user has to collect this information separately; there is no common information pool. This is costly for users and socially inefficient. Costs increase with market size. In online markets, the platform operator collects an aggregated view of supply and demand and actual transactions. Users can benefit from this aggregated information. It would be impossible for users in large online platforms with millions of product entries to collect all the information on their own. Platforms are in a unique position as third-party data aggregators to realise economies of scale and scope in data aggregation across many users. Individual users cannot realise these benefits.[35]

Platforms are new types of firms that emerged in the wake of digital data. The traditional view of the firm goes back to Ronald Coase.[36] Coase wondered what makes firms an efficient arrangement between workers who divide tasks and exchange intermediate goods between each other within a firm rather than going through the market for these exchanges. He argued that firms reduce transaction costs compared to going through the market. By implication, the borderline of the firm, between in-house production and external trade, depends on transaction costs. Digital data and online platforms have dramatically reduced transaction costs to quasi-zero in many cases. As a result, some firms stop in-house production altogether, delegate production to external agents and transform themselves into market places. In contrast to traditional firms that keep the market

---

34 Marco Iansiti and Karim R. Lakhani, *Competing in the Age of AI: Strategy and Leadership When Algorithms and Networks Run the World* (Harvard Business Review Press 2020) Ch. 6.

35 For example, Imke C. Reimers and Joel Waldfogel estimate that the welfare effect of aggregated consumer book review data on the Amazon book sales platform is about 15 times larger than the welfare effects from a single-authored book review in a newspaper. Imke C. Reimers and Joel Waldfogel, 'Digitization and Pre-Purchase Information: The Causal and Welfare Impacts of Reviews and Crowd Ratings' (2020) NBER Working Paper No. 26776 <www.nber.org/papers/w26776> accessed 31 August 2020. This informational advantage puts platforms in a strong bargaining position vis-à-vis individual user data.

36 Ronald H. Coase, 'The Nature of the Firm' (1937) 4 Economica 386.

outside, these 'inverted' firms[37] become market organisers rather than production organisers. They organise a market platform where different types of users, for instance buyers and sellers, can trade goods and services. Iansiti and Lakhani[38] show that data-driven platforms are not subject to diminishing returns to scale. Human labour is replaced by data-driven algorithmic procedures with high fixed set-up costs but nearly zero marginal costs. Non-rival data and algorithms make these platforms infinitely scalable. This leads to huge productivity and efficiency gains but also to increased market power and monopolisation.

Many of today's largest online platforms are probabilistic data-driven matching services: Google Search, Facebook, Amazon, Netflix, Uber, e-scooter platforms, etc. They put data at the core of their business model and specialise in transactions that require substantial datasets for efficient matching between users. They compete on increasing matching efficiency. Platforms help to create new markets that were missing in the pre-digital economy because information-related transaction costs were too high. For example, finding a hotel was costly in the analogue economy and required intermediation from travel agencies that offered a limited choice to consumers. Finding 'information' in general was costly. These missing information markets were not a market failure because the technology to overcome them was not available at the time, or remained very imperfect. Digital data technology has dramatically reduced information cost and thereby expanded user choices. However, users require third-party intermediary platforms to collect and classify the avalanche of digital information in order to make efficient use of it.

III. Monopolistic market failures in platforms

In the traditional platform economics model, network effects incentivise users to congregate together on the largest platforms. This strengthens the position of the incumbent platform at the expense of potential new entrants into the market. The latter will have to overcome network effects to compete with incumbent platforms. Ordinary network effects require only simple data, a lexicographic directory of addresses of users. More complex networks require more elaborate data for efficient matching between

––––––––––––––––––––

37 See Geoffrey Parker, Marshall Van Alstyne and Xiaoyue Jiang, 'Platform Ecosystems: How Developers Invert the Firm' (2017) 41 MIS Quarterly 255.
38 Iansiti and Lakhani (n. 34).

users. Large platforms can compile aggregated datasets on user behaviour and product characteristics. They achieve positive network externalities that enable them to provide more efficient matching services and further amplify network effects. The downside of these positive network externalities is that it reinforces platforms' monopolistic market position and may lead to abuse of dominant market positions.

The strength of data-driven network effects plays a key role in tipping[39] and varies by type of platforms and the relevant data in these platforms. For example, in ride-hailing and e-mobility platforms, network effects are very local. The platform may be organised on a global basis but network effects depend on local supply and users in cities. Expanding the supply in city A has no benefits for users located in city B, unless they happen to travel frequently between the two cities. This makes it easier for smaller local platforms to compete in local markets with global platforms. Hotel booking platforms are global however. Users search for hotels in many cities and platforms have to ensure a wide geographical variety of offers. This makes competition more difficult. Platforms can pursue deliberate strategies to tip the market in their favour, for example by increasing the costs of multi-homing or switching to other platforms. For example, drivers can easily switch between ride-hailing platforms with little costs. To discourage drivers from switching, platforms may offer them an uninterrupted sequence of rides, with advance notice of the next ride before the on-going ride is completed.

Hagiu and Wright[40] illustrate how the value that platforms can extract from data is conditional on several factors. Improving the quality of insights and the matching efficiency of data can be subject to economies of scale. In some cases, a few observations are sufficient to make an accurate prediction, while in other cases millions of observations are required to reach a reasonably accurate prediction. For example, automated driving algorithms are still far from perfect despite millions of miles of accumulated driving data by leading firms such as Google for its Waymo project. This is often true for artificial intelligence-based applications in platforms that depend on large numbers of observations. Insights that can be extrapolated to a wide number of users have high value. For example, personalised music recommendations in Pandora, based on cumulative learning from individual users, cannot easily be applied to other users. Spotify's shared music

---

39 See Iansiti and Lakhani (n. 34) sec. 6.
40 Andrei Hagiu and Julian Wright, 'When Data Creates Competitive Advantage' (2020) 1 (Jan.-Feb.) Harvard Business Review 94.

recommendations by contrast benefit from strong network externalities because they are useful to many users.

Several competition policy reports investigate the link between data and platform market power.[41] They suggest some re-thinking of competition policy tools to take into account that data-driven network effects are often the cause of competition problems. The reports pay attention to data policy tools as a means to attenuate data-driven monopolistic behaviour, for example by opening access to exclusive datasets, or a variety of data pooling and data sharing modalities. Data sharing with competitors may prevent an upstream monopolistic data collector from foreclosing downstream services markets. For example, car manufacturers design the car data architecture to retain exclusive access to car data, which they can leverage to increase their share in aftersales services markets. Mandatory data access for other aftersales service providers can prevent this competition problem.[42] Opening data access may backfire however. It may reduce rather than increase competition when data from small competitors are aggregated by large platforms that can offer users additional advantages, based on economies of scope in re-use and aggregation with other data sources. For example, payment services offered by Apple and Google, or payment services on the WeChat social media app in China and perhaps in future on Facebook, compete with local banks. Google Android and Apple iOS are increasingly present in cars and may offer a wide variety of aftermarket services that compete with smaller service providers. Since data are not a homogeneous product, data access and sharing can be restricted to a degree of coarseness that preserves some incentives and advantages for the original data collector while still broadening competition in the market

---

41 Jacques Crémer, Yves-Alexandre de Montjoye and Heike Schweitzer, 'Competition Policy for the Digital Era – Final Report' (European Union 2019) <https://ec.europa.eu/competition/publications/reports/kd0419345enn.pdf> accessed 31 August 2020; Furman and others (n.2); Fiona Scott Morton and others, 'Committee for the Study of Digital Platforms – Market Structure and Antitrust Subcommittee Report' (2019) George J. Stigler Center for the Study of the Economy and the State and the University of Chicago Booth School of Business <www.chicagobooth.edu/research/stigler/events/antitrust-competition-conference> accessed 31 August 2020.

42 Bertin Martens and Frank Mueller-Langer, 'Access to Digital Car Data and Competition in Aftermarket Maintenance Market' (2020) 16 Journal of Competition Law and Economics 116; Wolfgang Kerber, 'Data Governance in Connected Cars: The Problem of Access to In-Vehicle Data' (2018) 9 Journal of Intellectual Property and Information Technology and Electronic Commerce 310, para 1.

for data-driven services. That would require a careful balancing act and constant market and technology monitoring by regulators.

Data sharing with potential competitors will erode firms' data aggregation monopoly,[43] lower the value of the data and undermine their ability to monetise the data and invest in data collection. In a multi-sided market, modifying access conditions on one side of the market will have implications for other sides. For example, forcing a search or social media platform to share consumer data with competitors may not only affect consumer privacy. It lowers entry costs into advertising and will force platforms to increase entry costs on the consumer side, or integrate new money-raising sides into the platform to compensate the lost revenue.

Platforms are both a blessing and a curse in the digital data economy. They are necessary intermediaries to generate benefits from data aggregation, realise data-driven positive network externalities and enable the emergence of new markets that were not feasible prior to the arrival of digital data. At the same time, data aggregation generates new sources of market failures that did not exist in the pre-digital economy. In the next section we discuss non-monopolistic market failures induced by data-driven platforms.

D. Other data-driven market failures

The European Commission's 'Better Regulation Guidelines' distinguish between several types of market failures that may require regulatory intervention to maximise welfare for society. Besides monopolistic market failures, other sources of failure include externalities, information asymmetries and missing markets because of high transaction costs and risks. Regulators may also intervene in the case of social concerns such as discrimination and unequal distribution of welfare. In this section we discuss three types of data-driven non-monopolistic market failures: negative externalities from data aggregation, asymmetric information problems that distort decision making by data users, and newly missing markets that emerge in the wake of the data economy because of high data transaction costs and new sources of data-related risks.

---

43 Competition policy issues in data-driven platforms are discussed by Heike Schweitzer and Robert Welker, 'A legal framework for access to data: A competition policy perspective', in this volume.

I. *Information externalities*

In Section C we discussed the crucial role that platforms play in capturing data-driven positive network externalities, turning them into benefits for users and monetising them to their own benefit. In this section we turn to negative data-driven externalities caused by data aggregation in platforms, with examples on the consumer side in personal data markets, and on the producer side in commercial data markets. While positive externalities increase social welfare, negative externalities should be avoided or internalised by the party that causes them.

A first example of the negative impact of consumer platforms on the value of personal data is mentioned above[44] on economies of scope in data aggregation. Data collected on the behaviour of one set of users has predictive value for the behaviour of other users.[45] Once a firm has accumulated a critical mass of consumer data, the marginal return in terms of improved insights and additional value in the secondary re-use market – for example for advertising purposes – from adding another consumer's personal data is close to zero. This reduces the marginal value of a single person's dataset. It also reduces incentives for consumers to protect their privacy since their profile can be assembled from data collected from other persons. Consumers may not understand the low market value of their personal data and continue to invest in privacy protection. That in itself may have signal value that can be exploited against consumer interests.[46] An empirical study on the use of personal data for advertising in the travel industry[47] finds that, since the entry of the EU GDPR, 12 percent of consumers withhold consent to collect their personal data. The study also finds that the reduction in the supply of available data increases the value of the remaining advertising data and, because of externalities, does not negatively affect the predictability of consumer responses to advertising.

Is this negative externality a market failure that requires regulatory intervention to be corrected? Individuals have no better alternative option to

---

44 See section B.VII.
45 Bergemann, Bonatti and Gan, (n. 25).
46 Sebastian Dengler and Jens Prüfer, 'Consumers' Privacy Choices in the Era of Big Data' (2018) TILEC Discussion Paper No. 2018–014 <https://ssrn.com/abstract=3159028> accessed 31 August 2020.
47 Guy Aridor, Yeon-Koo Che and Tobias Salz, 'The Economic Consequences of Data Privacy Regulation: Empirical Evidence from GDPR' (2020) NBER Working Paper No. 26900 <www.nber.org/papers/w26900> accessed 31 August 2020.

realise a higher value for their personal data. Brynjolfsson and others[48] present empirical evidence that at least some 'free' services platforms actually compensate the negative externality and generate a large consumer surplus. Consumers trade personal data at nearly zero value for valuable online services. That suggests that the positive network externalities produced by platforms outstrip the negative externality on personal data. Consumers get more value out of the trade than they put into it. Zero prices are often seen as a market distortion from a traditional competition policy perspective.[49] However, trying to correct this may reduce overall social welfare because it would reduce the number of consumers and the volume of data and make the platform less attractive for advertisers and for other consumers. Public opinion often goes in the other direction, as the quip 'if you are not paying you are the product' suggests. Some authors suggest that consumers should be paid for the 'data labour' that they contribute to platforms.[50]

A similar phenomenon of data value depreciation because of externalities takes place on the firm or supply side of platforms. Suppliers sell their goods and services through online platforms like Amazon, eBay or Netflix. Platform operators collect and aggregate data on product characteristics, sales and consumer choices across many users. Once sufficient data are collected, the operators can predict market responses to changes in product characteristics and prices. This reduces the marginal prediction value of individual supplier data.

Newspapers are an example of negative externalities between two types of service suppliers on a platform. In the pre-digital era, printed newspapers had a strong market position in advertising, both commercial ads and classifieds. That revenue cross-subsidised news production and kept printed newspaper prices low to maximise consumption. In the digital era, consumers moved online to search engines and social media platforms, and so did advertising, which followed consumers to Google and Facebook. This blew a big hole in newspaper revenue. The revenue from remaining online

---

48 Erik Brynjolfsson, Avinash Collis, W. Erwin Diewert, Felix Eggers and Kevin J. Fox, 'GDP-B: Accounting for the Value of New and Free Goods in the Digital Economy' (2019) NBER Working Paper No. 25695 <www.nber.org/papers/w25695> accessed 31 August 2020.

49 Joshua S. Gans, 'The Specialness of Zero' (2020): <https://ssrn.com/abstract=3486964> accessed on 03/11/2020.

50 Eric A. Posner and E. Glen Weyl, *Radical Markets: Uprooting Capitalism and Democracy for a Just Society* (Princeton University Press 2018).

ads on newspaper webpages does not compensate the losses from print advertising.

Data-driven market failures may also occur in the presence of positive data externalities when these externalities cannot be captured or monetised by a party, or when contributing parties cannot agree on the distribution of the benefits from these positive externalities. A typical case is the private production of public goods, for example public health. Public goods are non-rival and non-excludable. Their use value cannot be captured and monetised by an agent. As a result, private agents have no incentive to invest in the production of these goods and the production is sub-optimal. This occurs for example when pooling of personal health data would create a dataset that can be used to discover innovative medicines, treatments and therapies or combat viral diseases. However, individuals and medical service providers have no incentive to contribute their data to the pool, unless they could expect direct benefits from new treatments. In some cases, innovative firms can grant direct benefits to individuals. But this is not always feasible and may be costly to achieve. Alternatively, governments can make data pooling mandatory and facilitate open access to the data for health researchers.[51] This imposes costs on contributors and leaves all the benefits to innovators and consumers who benefit from the innovations.

II. Asymmetric information

Asymmetric information between individual users and data-collecting platforms is an almost natural state in a data-abundant digital world. Platforms as data aggregators will always have more and better information on the data collection and use markets that they cover than do individual platform users (persons and firms). Users' willingness to share information with the platform depends on the level of detail and the use of the data.[52] Conversely, platforms will manipulate and may degrade the information that they share with users in order to segment markets and maximise rev-

---

51 For example, in Finland the government adopted an act that makes health data pooling on a government server mandatory for all private and public health service providers. See Finland Ministry of Social Affairs and Health 'Secondary Use of Health and Social Data' <https://stm.fi/en/secondary-use-of-health-and-social-data> accessed 31 August 2020.

52 Bergemann, Bonatti and Gan (n. 25); In this model, data collection for advertising has a negative effect on the welfare of data originators because there is no compensatory service offered in return for the data.

enue from their data intermediation role. Users may take sub-optimal decisions because of imperfect information signals received from platforms.

An extreme form of information manipulation by platforms is 'self-referencing'. For example, in July 2019 the European Commission opened an investigation into Amazon.[53] Amazon combines the roles of online retailer on its own account and market place for independent sellers. The platform allegedly used data that it collects about the activities of independent sellers to engage in anticompetitive practices and degrade the quality of information signals to consumer search results to favour Amazon sales and reduce the prominence of sales by independent sellers.

The market-distorting effects of asymmetric information in favour of the platform operator is well-documented in empirical studies on all kinds of search engines.[54] Platforms apply business models that may be based on sales margins (for retailers), commissions on sales (for market places) or advertising revenue (pure information matchmakers). The incentives embedded in the business models affect search rankings and drive a wedge between user preferences and the financial interests of platforms. For example, hotel booking platforms can manipulate search rankings towards price offers that increase their fee revenue. Another example of self-referencing occurs in the automotive industry, where car manufacturers have exclusive access to all data collected by connected cars.[55] Manufacturers can give preferential access to their own network of accredited dealers and after-market service providers. That distorts competition with independent service providers. Competition policy tools, such as the pre-digital EU Block Exemption on Vertical Restraints and the EU Motor Vehicle Type Approval Regulation, can force manufacturers to share maintenance information with independent repair shops. Industry self-regulation has failed because of weak incentives for industry players to come to an agreement.

---

53 See European Commission, 'Antitrust: Commission Opens Investigation into Possible Anti-Competitive Conduct of Amazon (Press Release, 17 July 2019) <https://ec.europa.eu/commission/presscorner/detail/en/IP_19_4291> accessed 31 August 2020.

54 See for example Babur de los Santos and Sergei Koulayev, 'Optimizing Click-Through in Online Rankings with Endogenous Search Refinement' (2017) 36 Marketing Science 542.

55 See Martens and Mueller-Langer (n. 42).

There is considerable debate on what an unbiased 'neutral' search engine in an inherently information-asymmetric world would look like.[56] The 'conduit' theory sees search engines as passive intermediaries that make an 'objective' selection of relevant search results in response to a user's search query. The ideal consumer-focused search engine would be a 'trusted advisor' that presents results that match consumer preferences. That search engine is not achievable, but would frustrate the preferences of service suppliers as well as the platform's own profit-maximising objective. At the other extreme, the 'editor' theory sees the search outcome as a subjectively curated ranking of results in response to a query, with the search engine as an active editor. Any ranking would represent the search engine operator's profit-maximising view. In reality, search results are necessarily a combination of objective conduit and subjective editing. Search operators are squeezed between the wishes of different types of platform users and carve out a profit margin while keeping all parties reasonably but not entirely satisfied.[57] The stronger their market position, the more they may distort the information picture. Locked-in users have no choice to go elsewhere for their services. Competitive pressure may sometimes limit platforms' margin for manoeuvre.[58] These models show how ranking bias is inherent to the platform's use of asymmetric information. Platforms need to drive a wedge between the preferences of users on different sides of the market in order to extract a profit margin to ensure the sustainability of their business model. More recent information theory models expand this insight from rankings to the quality of information collected and shared by platforms.[59]

Note that not-for-profit platforms would not perform better in this respect. They have no profit motive and could limit their financial needs to cost recovery by charging users a fixed fee, possibly as a function of their intensity of use. The market side that pays the fee would receive the most optimal information to match their preferences. Other sides may still suffer from bias in the collection and use of information. A platform cannot use its data to simultaneously maximise the welfare of all users on all sides

---

56 James Grimmelmann, 'Some Skepticism About Search Neutrality' in Berin Szoka and Adam Marcus (eds) *The Next Digital Decade: Essays on the Future of the Internet* (Tech Freedom 2010) 435; James Grimmelmann, 'Speech Engines' (2014) 98 Minnesota Law Review 868.

57 De los Santos and Koulayev (n. 54).

58 Maurice E. Stucke and Ariel Ezrachi, 'When Competition Fails to Optimize Quality: A Look at Search Engines' (2017) 18 Yale Journal of Law and Technology 70.

59 Bergemann, Bonatti and Gan (n. 25).

of the market, unless their preferences are perfectly aligned. Information asymmetry is a fact of life in digital platform economies.

Data sharing is often touted as a means to overcome information asymmetry and maximise social welfare benefits for society[60] because it generates economies of scope in re-use. Data sharing markets may fail however when the data originator or collector perceives a risk of negative repercussions on his private welfare. Data-driven platforms may offer compensation for this perceived risk, for instance by offering consumers a free service in return for sharing their data, or offering firms enhanced market access in return for sharing their data. Alternatively, platforms can modulate the degree of fine-graining and segmentation of the data they collect and share. Mandatory data-sharing obligations upset these platform strategies, both on the data collection and on the data use side of the platform. This may result in less data collection and undermine the positive externalities from data aggregation. Data policy makers need to carefully balance these positive and negative aspects of data-driven platforms.

III. Missing markets because of high transaction costs and risks

High transaction costs in the analogue economy prevented the emergence of many types of markets. Digital data massively reduce information costs and thereby facilitate market entry for consumers and small suppliers, from small hotels and bed & breakfasts that can now compete with large hotel chains on accommodation booking platforms, to independent taxi drivers who can offer their services on Uber and Lyft, and workers entering the online labour market, or staying in touch with a large number of family, friends and professional contacts on social media. All this is made possible by intermediary online data-aggregating platforms. Markets that were 'missing' in the pre-digital era suddenly emerge as a result of the drop in market-entry and transaction costs. However, even in the digital data economy some markets still remain blocked due to high transaction costs. Moreover, new services are required in order to keep digital markets running but they may not appear autonomously because of high transaction costs and risks. In this section we present a few examples of such missing

---

60 OECD Directorate for Science and Technology (n. 11); OECD, *Enhancing Access to and Sharing of Data: Reconciling Risks and Benefits for Data Re-Use across Societies* (OECD 2019).

markets and explore how these market failures may be addressed by a mixture of regulatory intervention and private third-party intermediation.

1. Transaction costs in personal data markets

Under the EU GDPR, data subjects have the right to consent to the use of their personal data before a firm can collect it. Consent notices pop up when consumers browse the internet. Consumers rarely read these notices. Even when they do, personal data consent notices are difficult to read and uninformative about possible data re-use.[61] The cost of time invested in reading these notices is too high compared to their informative value. A consumer survey confirms consumers' ambiguous attitudes towards privacy notices.[62] Another recent consumer survey[63] illustrates how risk assessments about sharing personal data on the internet vary widely according to type of data. Financial and biometric information commands high subjective opportunity costs. Data use for advertising is not perceived as entailing a significant privacy cost. Location and social network data are somewhere in the middle. The use of personal data has ambiguous welfare effects.[64] It can increase personal welfare when the data are used in an informative way, for example by search engines to reduce search costs and provide better search results that are more in line with consumer preferences. It may reduce welfare when data are used for targeted advertising that is more persuasive than informative and drives consumers away from their original preferences.

High transaction costs make the current system of consent notices dysfunctional. Many private start-ups have tried to enter the market for per-

---

61 See for example Fred H. Cate, Viktor Mayer-Schönberger, 'Notice and Consent in a World of Big Data' (2013) 3 International Data Privacy Law 67.

62 The survey confirms that nearly two-thirds of consumers would appreciate government intervention in setting privacy rules but only about 20 % of consumers bother to regularly read privacy notices. Results from the Brookings survey can be found here: Darrell M. West 'Brookings Survey finds Three-Quarters of Online Users Rarely Read Business Terms of Service (TechTank, 21 May 2019) <www.brookings.edu/blog/techtank/2019/05/21/brookings-survey-finds-three-quarters-of-online-users-rarely-read-business-terms-of-service/> accessed 31 August2020.

63 Jeffrey Prince and Scott Wallsten, 'How Much is Privacy Worth Around the World and Across Platforms?' (2020) <https://ssrn.com/abstract=3528386> accessed 31 August 2020.

64 Alessandro Acquisti, Curtis Taylor and Liad Wagman,'The Economics of Privacy' (2016) 54 Journal of Economic Literature 442.

sonal information management services (PIMS).[65] They offer an intermediary platform to handle personal data exchanges with commercial platforms. However, none of these have scaled up to become significant market players in personal data markets. The reason is clear: they do not really reduce high individual transaction costs. Management costs are still relatively high, at least in time spent on the platform, compared to the depressed value of individual personal data. Economically feasible personal data management would require technology that substantially lowers transaction costs. This could happen for example when consent notices become standardised and machine-readable so that they can be processed by AI-driven machines. Standardisation could include the identity of the data collector, the purpose for which it is collected, the level of fine-graining in use of the data and third-party commercial partners that may access the data. A privacy service provider could machine-read the consent notices, estimate possible risks for the data subject as a function of his or her pre-set preferences and use of the internet, and machine-grant or -deny consent. Machine learning could gradually become more efficient by learning from individual consumer behaviour as well as aggregated data across individuals and websites and collecting evidence on data sharing practices between firms and websites. It could suggest alternative service providers with lower privacy costs. Automation of the consent process would complete it in milliseconds, saving data subjects a substantial amount of time. The bottleneck lies in the standardisation process however. Platforms can produce their own standardised consent notice but without interoperability the system would run into high obstacles. Collective action seems to be required and that requires regulatory intervention.[66]

---

65 Mydata.org <https://mydata.org/> is one example among many initiatives to help individuals manage their personal data. The European Data Protection Supervisor has advocated the use of Personal Information Management Systems (PIMS). For a detailed economic discussion of PIMS, see Jan Krämer, Pierre Senellart and Alexandre de Streel, 'Making Data Portability More Effective for the Digital Economy: Economic Implications and Regulatory Challenges – Report' (Centre of Regulation in Europe 2020) <https://cerre.eu/publications/report-making-data-portability-more-effective-digital-economy/> accessed 31 August 2020.

66 Posner and Weyl (n. 50) propose a particular variant on this theme. They suggest that data subjects should unite in unions to negotiate a higher value for their data with data collecting platforms. Automated data consent notices would reduce coordination and market entry costs for such unions. These unions would still face the problem of allocating the social value of the data between private members. See also the conclusions section in Bergemann, Bonatti and Gan (n. 25).

2. Transaction costs and lack of transparency in commercial services markets

A similar lack of transparency in management services occurs in online advertising markets. Online advertising can be split between 'walled gardens' in Search (Google) and social media (Facebook), and open-display advertising where Google holds a strong position. Advertising is a two-sided market between publishers and advertisers, with several layers of intermediary platforms that do intermediate matching and price auctions for the supply of ad publishing windows and the stock of ads produced by advertisers. For every euro spent on ads by the advertiser, only 62 cents reach the publisher; the rest remains in intermediate steps, largely dominated by Google.[67] It is challenging for advertisers to verify publishing and views of ads because of the lack of transparency in intermediate stages. Price auctions in these markets are problematic[68] because Google itself participates in the bidding while it has privileged information on the offers of its competitors. Self-(p)referencing is an issue. Data transparency and sharing through open standards and automated market tracking tools could be a solution. It could improve transparency and oversight for advertisers, publishers and content providers, increase competition and enable all participants to get a better overview of what they pay for and what they achieve.

Filling missing market gaps does not always require regulatory intervention. Entrepreneurs may propose innovative services to fill the information gap between platform operators and users. For example, data providers like AMZScout and JungleScout[69] collect and analyse data from e-commerce platforms like Amazon, eBay and Zalando and sell findings to independent sellers to help them improve their commercial strategies on the platform. The e-commerce platforms only provide data related to the seller's market.[70] The intermediary service provider aggregates data across

---

67 Damien Geradin and Dimitrios Katsifis, 'Google's (Forgotten) Monopoly: Ad Technology Services on the Open Web' (2019) 2019–3 Concurrences 1.

68 Ibid. Several EU competition authorities have launched investigations in online advertising, including those in the UK, France and Germany. A UK Competition Market Authority study is exploring potential remedies for ads markets that could be part of an ex-ante regulatory regime.

69 See <https://amzscout.net/> and <https://www.junglescout.com/> accessed 31 August2020.

70 The EU Platform-to-Business Regulation specifies the type of information that platforms have to provide to their business suppliers. See Regulation (EU) 2019/1150 of the European Parliament and of the Council of 20 June 2019 on promoting fairness and transparency for business users of online intermediation services [2019] OJ L186/57.

products on the platform and sells the joint analysis. Adjacent data services may be tolerated by platform operators, or they can be blocked.

3. Risks

There are circumstances in which potential data suppliers refrain from participating in the production of services markets because it may be costly for them. An example is pooling mobility data between transport service providers in a city. This can have positive social welfare effects by improving traffic management and reducing congestion and pollution. However, commercial transport service providers (buses, metros, taxis, e-scooter platforms etc.) may gain or lose market shares from sharing data on a common platform.[71] Competitors may use the data to improve their offers and increase their market share. Alternatively, being on the common platform may attract more users to a particular provider. The net impact is an empirical question. These risks may motivate transport providers to stay away from the platform, unless the platform is in a position to compensate losers by re-allocating part of the overall social welfare surplus to them. For example, if drivers are willing to pay a positive price for improved congestion management, some of that revenue could be re-allocated to transport service providers that lose from participation. Alternatively, regulators can intervene to make data sharing mandatory in the interest of public welfare.[72]

Another dimension of transaction costs is ex-post risk in the execution of contracts. According to incomplete contract theory, contracts of finite length inevitably come with residual uncertainties that can give rise to ex-post costs during monitoring and execution of a contract. This is especially the case for trade in non-rival and hard-to-exclude data. Some contractual provisions may be unenforceable, non-monitorable or lack a commitment device.[73] They are subject to the hold-up problem: parties will try to re-negotiate the contract when an unforeseen or non-committable event occurs.

---

71   Bruno Carballa Smichowski, 'Determinants of Coopetition through Data Sharing in MaaS (Mobility-as-a-Service)' (2018) 2 Management & Data Science <https://doi.org/10.36863/mds.a.4160> accessed 31 August 2020.

72   The European Commission's initiative to promote business-to-government data sharing 'in the public interest' should be seen in this context. See European Commission (n. 4). See also Richter (n. 4).

73   Anastasios Dosis and Wilfried Sand-Zantman, 'The Ownership of Data' (2019) <https://ssrn.com/abstract=3420680> accessed 28 August 2020.

This includes risks from data leaks, unexpected data quality problems or processing errors. In traditional contracts, unexpected costs and benefits are assigned to the owner of the traded good or service. In the absence of legal data ownership rights[74] that is more problematic and may reduce incentives to make data available for re-use. The risks of contractual hold-up may be too big for holders of valuable or commercially sensitive datasets, as the Facebook/Cambridge Analytica case demonstrated. Some authors have suggested assigning data ownership rights to overcome this problem.[75] Debates on the possible introduction of such rights[76] have diminished and attention has now shifted to introducing data access rights.[77] Ownership and access rights are complements. Who should get such rights, if any, is not an easy question. For personal data, there is a 'natural' rights holder, the data subject. For non-personal machine-generated data that may involve several parties for the co-generation of the data, it is often hard to unambiguously identify a 'natural' rights holder. For example, in agriculture land owners, land operators, machine manufacturers, machine operators, sensor owners, data analytics providers, etc. may all claim rights over the data.[78]

A more pragmatic solution may be to appoint a neutral third-party intermediary who is tasked with managing the data exchange in accordance with an agreed protocol. For example, a city mobility service provider may require pooled data from all mobile phone operators in that city to create detailed insights on citizen mobility patterns. None of the data suppliers trust the other to handle the data pool that has strategic commercial value for competitors. Solving this coordination problem requires a trusted

---

74 Duch-Brown, Martens and Mueller-Langer (n. 5).
75 See Herbert Zech, 'Data as a Tradeable Commodity' in Alberto De Franceshi (ed.), *European Contract Law and the Digital Single Market: The Implications of the Digital Revolution* (Intersentia 2016) 51; and Andreas Wiebe, 'Protection of Industrial Data: A New Property Right for the Digital Economy?' (2017) 12 Journal of Intellectual Property Law & Practice 62.
76 Communication from the European Commission of 10 January 2017 – 'Building a European data economy', COM(2017) 2 final and An Commission Staff Working Document, 'The free flow of data and emerging issues of the European data economy' SWD (2017) 2 final.
77 Communication from the European Commission of 25 April 2018 – 'Towards a common European dataspace' COM(2018) 232 final; Josef Drexl, 'Data Access and Control in the Era of Connected Devices: Study on Behalf of the European Consumer Organisation BEUC' (BEUC 2018) <www.beuc.eu/publications/beuc-x-2018-121_data_access_and_control_in_the_area_of_connected_devices.pdf> accessed 31 August 2020.
78 Atik and Martens (n. 8).

third-party intermediary that collects the data, performs the analysis and ensures that only the processed results are shared with agreed users. This is the domain of semi-commons or governance agreements that seek to overcome the pitfalls of the commons – which lead to overutilisation and underinvestment and facilitate free-riding – and of the anti-commons – exclusive private use that leads to underutilisation and keeps data locked in silos.[79] Semi-commons are often costly to manage. They are economically feasible when the value of the agreement for the participants exceeds the costs.

Data trusts and industrial data platforms fit the neutral intermediary profile. In order to guarantee enforcement, the intermediary should have no stake in the data or the outcomes of the analysis. That avoids strategic behaviour at the expense of the participants. The intermediary should only receive a fixed remuneration to produce the desired outcome. It can enforce the commitment because it has full control over the data and access to the server. That reduces post-contractual risks and monitoring costs for participants. Commercial for-profit data platforms may also provide guarantees against data leaks but they will exploit the data in their own interest, and sometimes against the interests of the data providers. They create new sources of ex-post risks.

E. Concluding remarks

The data economics issues that we have discussed here have much in common with the law and economics of intellectual property rights (IPRs), such as patents and copyrights.[80] The economic characteristics of data, non-rivalry and no natural excludability, are similar to those of innovation. IPRs give exclusive ownership rights to innovators in order to yield a return on investment and an incentive for innovation. IPR policies struggle with the same balancing act as data policies, between the social welfare costs of monopolistic exclusive rights and the social welfare gains from the innovation incentive effects. Monopolistic IPR licence pricing, above the marginal cost of reproduction, reduces access to innovation. This is an unavoidable social harm accepted as the cost of generating dynamic innova-

---

79 Henry E. Smith, 'Governing the Tele-Semicommons' (2005) 22 Yale Journal of Regulation 289. Exclusive private property rights are cheaper to manage – the exclusive owner sets the price – and so are full commons because the price falls to zero.
80 See Leistner (n. 6).

tion benefits. IPRs compensate this by limiting the scope of exclusive rights. Similar considerations apply to data collection, access and use, including in online platforms. It took several centuries for society to develop a coherent system of IPR rights, and this system is still evolving, driven by technology that affects the cost of innovation production and dissemination and therefore the balance between protection and access. Digital data are a very new product in society. There are lively discussions between proponents of exclusive ownership rights and defendants of more open access rights.[81] A major difficulty with data is the attribution of such rights. Innovations are usually produced by a well-defined innovator or group of innovators with common interests. Data, by contrast, usually originate from a large and poorly defined group of providers, often with diverging interests. While personal data rights may be 'naturally' attributed to a data subject, attribution is more difficult for non-personal data, where many parties may be involved in origination, collection, aggregation and analysis of the data. Changes in attribution of rights may affect entire data value chains and downstream services markets. They will affect the pace of innovation that data can bring to society.

More importantly, both ownership and access rights overlook the inherent social value of data and the externalities that they entail. A single data originator or collector is usually not in a position to internalise these externalities. Market failures will remain. The discussions sometimes give the impression that the attribution of exclusive ownership, access and sharing rights are policy objectives in themselves. This data economics chapter has emphasised that such rights are only policy instruments that should be used to maximise the social welfare that society as a whole can derive from the use of data.

The title of this volume reflects the dichotomy between consumer and social welfare. In line with public policy economics, this chapter has focused mainly on social welfare as a benchmark for identifying market failures and policy intervention. Public policy economics defines the measure of social welfare as the combined welfare of all stakeholder groups in society, including consumers and producers. Mainstream competition law focuses on a narrower consumer welfare benchmark, even in the digital data economy setting with interactions between multi-sided markets.[82] These

---

81 The European Commission's Communications (n. 4, n. 76 and n. 77) on data issues over the last years reflect this societal debate. See also Wiebe (n. 75) and Zech (n. 75).
82 Furman and others (n. 2).

two measures can easily lead to contradictory conclusions. For example, regulatory intervention to open market access on one side of a platform may reduce welfare on other sides of the platform market. Classic economics rejects the comparison of welfare gains and losses between groups or individuals because consumer welfare is assumed not to be quantifiable. Alternative approaches accept quantification but open the door to measures of social welfare improvement whereby some parties gain at the expense of others. Economics distinguishes between strictly Pareto-improving welfare measures whereby no agent loses welfare and a less stringent Kaldor-Hicks[83] welfare measure whereby some agents may lose but could, in principle, be compensated by the gains that other agents make in order to avoid equity concerns. Western societies have historically put emphasis on individual wellbeing and are reluctant to impose private costs on individuals in order to achieve wider social welfare gains, unless they are compensated by transfers to ensure some degree of equity. Other societies have a more collective view of social welfare and attach less importance to individual welfare. They would find it easier to accept private costs as long as overall welfare increases. This underscores the borderline between the economics of data and cultural, social and political value judgements in society on how to maximise societal welfare from data.

Another dimension of data economics that was not discussed in this chapter is the emergence of ecosystems of bundled platforms. Our focus on individual market failures and multi-sided platforms may have given an excessively static picture of the data economy. Over the last decade, new data- and technology-driven business strategies have resulted in more complex and rapidly evolving ecosystems of interlinked platforms[84] and strong competition between major players. Data can be re-used to build service production conglomerates around a single data source or platform.[85] Platform 'envelopment' strategies[86] seek to re-bundle data-driven services in new ways in order to invade the markets of other platforms. Data-driven

---

83 For more details on Kaldor-Hicks measures, see for example <https://en.wikipedia.org/wiki/Kaldor–Hicks_efficiency> accessed 31 August 2020.

84 For a definition of ecosystems, see for example Michael G. Jacobides, Carmelo Cennamo and Annabelle Gawer, 'Towards a Theory of Ecosystems' (2018) 39 Strategic Management Journal 2255.

85 Marc Bourreau and Alexandre de Streel, 'Digital Conglomerates and EU Competition Policy' (2019) <https://ssrn.com/abstract=3350512> accessed 31 August 2020.

86 Thomas Eisenmann, Geoffrey Parker and Marshall Van Alstyne, 'Platform Envelopment' (2011) 32 Strategic Management Journal 1270.

bundling of products underpins similar competition strategies.[87] Data-based artificial intelligence technologies have greatly contributed to these market dynamics. Policy makers and regulators appear to be running behind the technology curve, among other reasons because regulatory interventions are slow political processes. The slowness of regulators compared to technological innovators has been underlined in some of the recent data and competition policy reports. Regulatory solutions often engage when the harm is already done. Catching up with the speed of technological innovation may require permanent intensive monitoring of the data economy, as proposed by some regulatory authorities.[88]

---

87 Yannis Bakos and Erik Brynjolfsson, 'Bundling and Competition on the Internet' (2000) 19 Marketing Science 63, 68.
88 Furman and others (n. 2) suggest that the UK competition authority should set up a digital markets monitoring unit, which it has since done. Scott Morton and others (n. 41) suggest likewise for the US.