5

Regulating the Feedback Effect

Viktor Mayer-Schönberger

A century and a half ago, Karl Marx wrote about markets becoming concentrated over time – the fundamental genetic defect that would eventually lead to capitalism's demise and spur the proletarian revolution.[1] Today, his dire prediction seems to ring true. While business dynamism has been stagnant for years, a few firms grow fast, amass astonishing profits, and capture a huge share of the market.[2]

Just in the fall of 2017, Apple – already the world's most valuable company by market capitalization – announced an increase in profits of 19 percent. Google/Alphabet's profits grew by an even stronger 33 percent, and Facebook topped both with an increase in profits of 79 percent. Each of these three has also cornered an enormous portion of a market. Apple is the world's largest smartphone producer by revenue, Google dominates online search with a market share of well over 80 percent globally, and Facebook rules the social media platform market, with over 2 billion users worldwide in 2017.

They are not the only ones. Digital markets, in particular, seem to move swiftly towards concentration. The market for domain names is cornered by GoDaddy, Netflix controls more than 75 percent of the US video streaming market, and Amazon accounts for 43 percent of US online retail sales.[3]

[1] Karl Marx, *Capital* (Penguin Classic 1992); but see also Joseph A. Schumpeter, *Capitalism, Socialism, and Democracy*, 3rd ed. (Harper Perennial Modern Classics 2008).

[2] On stagnant business dynamics, see Ryan A. Decker et al., *Declining Dynamism, Allocative Efficiency, and the Productivity Slowdown*, Finance and Economics Discussion Series 2017–019 (Board of Governors of the Federal Reserve System 2017); on profit concentration see David Autor et al., *The Fall of Labor Share and the Rise of Superstar Firms*, NBER Research Working Paper 23396 (May 2017); David Autor et al., *Concentrating on the Fall of the Labor Share*, NBER Working Paper 23108 (January 2017), Simcha Barkai, *Declining Labor and Capital Shares*, http://home.uchicago.edu/~barkai/doc/BarkaiDecliningLaborCapital.pdf.

[3] On GoDaddy's market share see Andrew Allemann, *Go Daddy Marches Toward $1 Billion*, *DomainName Wire*, August 17, 2010; on Amazon's share of online retail sales in the US, see *Amazon Accounts for 43 Percent of US Online Retail Sales*, Business Insider, February 2, 2017; on Netflix market share see Sara Perez, *Netflix Reaches 75 percent of US Streaming Service Viewers*, TechCrunch, April 20, 2017.

Unsurprisingly, a growing number of commentators has called for strengthening existing antitrust and competition laws to ensure that the concentration process is slowed, or at the very least does not lead to uncompetitive behavior by market-dominating firms.[4]

In this chapter, I suggest that these concerns are well-warranted, but that the remedies proposed are ill-suited to address the challenge we face. This is because while we see market concentration, the reason for that concentration has shifted, rendering much of the existing regulatory toolkit to ensure competitive markets essentially ineffective. In its place, I suggest a novel regulatory measure – the progressive data-sharing mandate – that is specifically designed to tackle the very concentration dynamic we witness.

The first part of this chapter describes the concentration process in greater detail, and how the dynamic is fueled by a new driver. I then explain why this novel driver – the feedback effect – cripples market competition differently than existing drivers, and thus requires novel regulatory measures. In the second part, I put forward such a measure, carefully tailored to address the crippling dynamic that leads to market concentration, and map out the consequences of such a measure should it be enacted.[5]

A CLOSER LOOK AT MARKET CONCENTRATION

Markets become concentrated when one or a small number of market participants enjoy a significant cost differential in offering products or services. There are many drivers for such a concentration dynamic, but the most prominent and well-analyzed are so-called scale effects. Scale effects signify the downward slope of marginal costs to produce or sell a particular good the more a particular actor sells.

Scale effects manifest themselves in various forms. For instance, buyers of large volumes can often negotiate discounts and thus get better deals than those purchasing smaller quantities. Fixed costs, too, are less onerous when they are spread across many goods rather than just a few. That's why it is costlier per good sold for Microsoft to operate its retail stores than for Apple. Many scale effects aren't consumed once a company has reached a certain size, but continue to be present as they grow and deliver improvements to a company's bottom line. As Ford Model Ts came off the assembly line almost a century ago and sales increased steadily over time, Ford was able to continue to lower cost year after year, largely because of the focus on exploiting scale effects in every aspect of manufacturing and sales.

[4] For instance, Columbia cyberlaw professor Tim Wu (*The Curse of Bigness: Antitrust in the New Gilded Age* (Columbia 2018)) and Harvard economist Kenneth Rogoff (*Has Big Tech Gotten Too Big for Our Own Good?*, *MarketWatch*, July 11, 2018) see the digital superstars as clearly being too big and too powerful.

[5] This is based on Viktor Mayer-Schönberger and Thomas Ramge, *Reinventing Capitalism in the Age of Big Data* (Basic Books 2018); see also Viktor Mayer-Schönberger and Thomas Ramge, *A Big Choice For Big Tech*, 97(5) *Foreign Affairs* 48 (September 2018).

Because smaller competitors do not reap similar benefits of scale, they have to produce at higher cost. This provides large players with an advantage in the market that they can convert into higher profits or, through lower prices, an increase in demand and thus market share. The resulting dynamic, at least in principle, is that through scale effects bigger players grow bigger, while smaller competitors wither away.

Of course, this theory does not always pan out in practice. Some large players are badly managed, become complacent, and lose their ability to produce the goods the market wants. Large firms have the scale advantage, but it often comes with a lack of flexibility and an inability to adjust swiftly to changes in the marketplace. In general, however, scale effects do lead to market concentration. Economists have been worrying about the negative effects for competitive markets for decades.

More recently, a further effect has led to much discussion. Termed the network effect, it is present whenever additional customers joining a particular service increase the utility of that service to everyone else using it.[6] It is often associated with digital networks and the Internet. Every additional person using Facebook, for example, improves the utility Facebook has for other users as they can reach more people through the service. But network effects are not an invention of the digital age. They already were a key driver leading to the consolidation of telephone networks (and thus concentration of network operators) in the early twentieth century, and even before that in the concentration of railroad networks.[7]

Network effects are particularly beneficial for the companies that provide the respective service, because apart from accommodating additional users, they need not actively do anything for the utility of their service to improve. Unlike with scale effects, for instance, firms do not have to renegotiate sourcing contracts to reap the benefits.

Network effects offer another advantage to scale effects, at least in theory: They grow more quickly with an increase in size than the often more linear scale effects. This is true even though in practice the increases in utility are not evenly spread, and certainly do not simply grow by the square of the number of users, as a particularly popular formulation of a certain type of network effect, "Metcalfe's law" seems to suggest.[8]

These qualities have made network effects the focus of many startups, and an important driver for substantial entrepreneurial successes. Particularly social

[6] See e.g. Oz Shy, The Economics of Network Industries (Cambridge University Press 2001); David Easley and Jon Kleinberg, Networks, Crowds and Markets (Cambridge University Press 2010), 449–478.

[7] On the evolution of US telephone networks, see Richard R. John, Network Nation: Inventing American Telecommunications (Harvard University Press 2010).

[8] For a description of Metcalfe's law see Carl Shapiro and Hal R. Varian, Information Rules (1999); on why this is flawed, see, e.g., Bob Briscoe, Andrew Odlyzko and Benjamin Tilly, *Metcalfe's Law is Wrong*, IEEE Spectrum (July 2006), https://spectrum.ieee.org/computing/networks/metcalfes-law-is-wrong.

network services, such as Facebook, Twitter, and LinkedIn have greatly benefitted from them, much as in the 1990s telecom providers and mobile phone equipment producers profited from the network effects of the GSM mobile phone standard. This has propelled network effects into public prominence.

Scale and network effects are not exclusive. In digital markets, often both of them are at play. Amazon profits from scale effects thanks to the huge volume of orders it fulfills, but also from network effects as it offers its customers a community to review and share information about products. Google can spread its indexing cost for its search engine across billions of users every day, but its shared document standards and interfaces (e.g., Google Docs) also create a sticky global network of activity.

For consumers, scale and network effects often have positive consequences. In competitive markets, lowering the cost of production per unit will result in lower prices, saving consumers a bundle. And an improving utility of a particular service, thanks to network effects, is a benefit for all of its customers.

Unfortunately, scale and network effects also help companies to increase their market shares, as they can produce more cheaply than smaller competitors or, thanks to network effects, offer a superior service. Combined, scale and network effects have facilitated a concentration process in countless markets around the world, reducing, at times even eliminating robust competition. Without sufficient competition, large firms can extract extra rents from customers.

Over time, this dynamic would cripple markets if there weren't a powerful counterforce that enabled new entrants to successfully compete with large incumbents and at times even to topple them. This counterforce is innovation.

Three quarters of a century ago, economist Joseph Schumpeter emphasized the disruptive force of innovation – of new products or production processes (or markets).[9] Innovative companies with ground-breaking ideas are able to overtake existing players, and in doing so preserve competition in the marketplace. Economies of scale, as well as network effects, are kept in check by the might of innovation, resulting in a dynamic balance that enhances consumer welfare.

Even (or perhaps particularly) in the context of digital markets, there is ample evidence that innovation has been the key driver that enabled new entrants to unseat dominating incumbents. 3G and LTE pushed aside GSM, Facebook dethroned MySpace, Google replaced Yahoo! as the online search champion, and Apple's operating system now runs on more digital devices than Microsoft's Windows. In our digital economy, too, innovation has provided the much needed antidote to market concentration.

Schumpeter feared that eventually innovation would become concentrated in a few large companies, because of their ability to attract talent and create a suitable environment for stimulating and bringing to market new ideas.[10] So far, fortunately, this has

[9] Schumpeter, supra note 1.
[10] Id.

not happened. Quite the contrary: Large companies often become unwieldy, with many hierarchies and entrenched processes that are difficult to adapt, and with ossified structures that stunt innovation and scare away talent, leaving new entrants with ample opportunities to compete against established players.[11]

Quite obviously, the counterforce of innovation is strongest in sectors that experience quick and substantial technical change. Unsurprisingly therefore, the Internet seemed to usher in an especially disruptive phase of innovation, with lowered cost of entry and an amazing variety of new ideas. Unlike Henry Ford, new data startups do not require huge amounts of capital to build factories, neither do they today even need huge server farms (as Google and Amazon did) to build out a new innovative idea. Digital disruptors can use digital platforms and infrastructures available to them often at commodity price. Until 2010, Twitter was renting cloud assets for its service rather than having built its own. And Uber, Lyft, and Didi Chuxing did not have to design and build digital devices to connect their drivers to their services; they are smartly piggybacking on existing smart phone infrastructure.[12]

Away from the realm of the digital economy, however, the innovation story has become a bit less impressive in recent years. Innovation economists have been pointing to the great deceleration in business dynamism outside of a small number of high-tech fields as an indication that our ability to disrupt through innovation in many sectors is actually stalling.[13] But at least in high-tech, encompassing an ever-increasing share of the economy, the innovation engine seemed as capable of disruption as ever. Unfortunately, and dramatically, this is changing, and the change is linked to a shift in the nature of innovation.

Since the beginning of humanity, innovation has been tied to the inventiveness of the human spirit. The ability to innovate rested on new ideas generated by human creativity and ingenuity – ideas that could then be translated into superior products and services or production processes. The romantic image of the single idea that changes the world overnight might rarely have been true, but persistent entrepreneurs have been able to shape markets and society. Because, at least in principle, such novel and valuable ideas can come from any human being, the capacity to innovate is not limited to large organizations with deep pockets. Sometimes, the tinkerer in the garage outsmarts armies of corporate researchers, whose

[11] A useful overview is Jan Fagerberg, David C. Mowery and Richard R. Nelson, The Oxford Handbook of Innovation (Oxford University Press 2005).

[12] This strategy of repurposing the existing smart phone infrastructure is not limited to ride-hailing and ride-sharing; it plays a central role in the ongoing displacement of existing payment infrastructures such as conventional credit cards with smart phone-based systems such as Apple Pay, Google Pay or Alipay; the same strategy is at play when it comes to health applications that innovatively utilize some of the many sensors found on smart phones to detect anything from irregular sleeping patterns to early signs of Parkinsons and heart problems.

[13] See, e.g., Ryan A. Decker et al., Declining Dynamism, Allocative Efficiency, and the Productivity Slowdown, FEDS Working Paper No 2017–019 (2017); David Autor et al., The Fall of Labor Share and the Rise of Superstar Firms (2019), https://economics.mit.edu/files/12979.

preconceived notions of what works and how may cloud their ability to – in the words of Steve Jobs – see the world differently.

The story of innovation has been tied to human originality, and – at least in more recent decades – to the small startup Davids upending slow and stodgy Goliaths. Although this narrative may be mythical and naïve, innovation unquestionably has acted as a powerful counterforce to scale-based concentration effects. But innovation's positive role in ensuring competitiveness hinges on it being equally available (at least in principle) to organizations and players of any size.

The world is abuzz with talk about artificial intelligence and machine learning. The labels may conjure up alarming visions of humanoid robots roaming the world. The reality, however, is more sanguine – and more troubling. What is termed AI or machine learning isn't a system that acquires abstract knowledge by being fed general rules. Machine learning denotes the ability to analyze massive amounts of data and uncover statistical patterns in them, devising algorithms that capture and replicate these patterns. Importantly, as new data points become available, the analysis is redone, and the resulting algorithm adjusted. Incrementally, the system "learns." It captures an ever more comprehensive slice of reality, and thus more accurately reflects the essence of the phenomenon in question.

We already see such systems popping up everywhere in the digital realm. Google's services – from online search and spell-check to language translation, from voice recognition all the way to autonomous driving – are continuously improving because of the gigantic stream of data available to Google. And Google isn't alone. Apple's Siri and Amazon's Alexa are improving because of data-driven machine learning, and so are the music recommendations presented by Spotify and the products recommended on Alibaba. Every additional data point so gathered is an opportunity to automatically learn from and adapt the system. And the more data points gathered the faster machines learn, producing innovation.

Whether it is a data-driven system defeating some of the world's best poker players, or a system diagnosing skin cancer with precision equal or better than the average dermatologist, the principle is always the same: Automated learning from the analysis of huge amounts of training and feedback data.[14] As the ability of machines to collect and analyze data comprehensively and increasingly unsupervised has grown dramatically, the source of innovation has been shifting from humans to systems of data and software. Yesterday's innovator was a human with bold ideas; tomorrow's equivalent are those capturing and feeding the most data into their learning systems.

Of course, humans will continue to invent, and human-based innovation is far from over. But data-driven innovation offers a few advantages that human

[14] On AI playing poker Olivia Solon, *Oh the Humanity! Poker Computer Trounces Humans in Big Step for AI*, Guardian, January 30, 2017; see also Mayer-Schönberger and Ramge, Reinventing Capitalism, 59–62; on IBM Watson diagnosing types of skin cancer better than the average dermatologist, *Computer Learns to Detect Skin Cancer More Accurately Than Doctors*, Guardian, May 29, 2018.

innovation can't easily replicate. It scales well, while human innovation doesn't so easily. It can be formalized and thus incorporated in organizational processes while standardizing the human innovation process is far more difficult; human ingenuity is unpredictable. And, arguably most importantly, data-driven machine-based innovation is less inhibited by human imagination – by the questions we humans ask as much as by the questions we don't ask; by the tensions we sense, and by what we seem deaf to.

Of course, the data that AI systems use to learn from isn't devoid of biases and errors, leading to biased decisions.[15] But the hope is that with comprehensive data from diverse sources, some of these distortions are flattened out.[16] At the very least, they are less constraining than the biases of a small number of humans in a conventional innovation lab, subject to "group-think."[17] With fewer constraints due to human preconceptions, data-driven machine-based innovation is less encumbered by traditional thinking and conventional beliefs. This is not to suggest that data-driven innovation is always better than human innovation; only that in the future human innovation will no longer be always the most relevant game in town.

As the source of innovation shifts at least partially from human ingenuity to data analyzed by machine learning systems, those with access to much data have most of the raw material to translate into innovative insights. Here, too, having the raw material does not necessarily equate with the ability to employ it successfully. Some companies will fail in the innovation drive despite having all the data, because of their shortcomings in setting up and maintaining appropriate systems and processes to learn from it.[18] But the reverse is even more true: Without data, even the best AI company will falter. In short, having access to data is *the crucial* (albeit not sufficient) condition for innovative success in the data age.

Due to this change in the nature of innovation, the emerging market dynamic pits large incumbent companies with substantial scale economies and network effects on the one hand against smaller startups with needs for lots of data to learn from. This in itself is already problematic: Companies with lots of data will hardly let entrepreneurs have access to their data troves, when they realize that access to data is a source of innovation (and thus competitive success). This likely inhibits small startups to turn themselves into disruptive innovators. The result is a market that may still be competitive and innovative among the larger firms, but no longer easy to enter for newer, smaller players.

[15] For a particularly pessimistic view, see, e.g., Cathy O'Neil, Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy (Broadway Books 2016).

[16] See Viktor Mayer-Schönberger and Kenneth Cukier, Big Data (HMH 2013).

[17] There is a rich literature on group think, with the term itself made popular by Irving Janis, Groupthink: Psychological studies of policy decisions and fiascoes (HMH 1983).

[18] It is said that up to 85 percent of data being collected in North America and Europe is shockingly underutilized; see, e.g., Pedro Hernandez, *Enterprises are Hoarding 'Dark' Data: Veritas*, Datamation, October 30, 2015, https://www.datamation.com/storage/enterprises-are-hoarding-dark-data-veritas.html

The situation is far more damning, however, when we look at the kind of data that will lead to innovation advantages. Most of the digital superstar firms aren't conventional manufacturing behemoths, neither are they exploiters or traders of natural resources. They are data-rich marketplaces with smart recommendation and other decision-assistance systems to aid market participants in their transaction decisions.[19] For these marketplaces, innovation lies in the ability to better match transaction partners: to help them in discovery and selection, as well as the process of decision making. If these companies want to innovate a better decision assistant, they need not any kind of data, but data about preferences and transactions – precisely the kind of data available to these digital superstars. It is data that offers feedback to the superstars about market activities from which to learn. Amazon's recommendation engine is getting better and better with each additional data point collected through a customer's interaction on Amazon's website. Spotify's choice of music gets better with each feedback from one of its listeners, much as each time you let Siri proceed with what she understood to be your request, Siri's underlying voice recognition system learns to understand human speech.

Not simply any kind of data, but *feedback* data is the raw material that lets these superstar companies innovate and offers customers a superior experience on their marketplaces. It's no surprise that eBay, a major long-established marketplace, is investing heavily into data and machine learning, because it, too, needs to tap into the innovative power of the feedback data streams it has access to if it wants to compete successfully against the other superstars.[20] Neither is it astonishing that shopping curation service Stitch Fix, one of the latest digital unicorns, employs many dozens of data analysts to mature its machine-learning systems that translate rich feedback data from multiple channels, including through photos posted online, into its innovative curation service.[21]

These digital superstars are innovating, but such innovation is no longer acting as a counterforce to market concentration. Instead, it's fed by valuable feedback data that only large data-rich markets collect in huge quantities. This makes firms that operate data-rich marketplaces and smart decision assistants (and who already benefit from scale and network effects) the most innovative as well. The *feedback effect* aligns innovation with market concentration, and shapes markets and the economy much as scale and network effects already do. The result is an unprecedented impulse towards larger and more powerful firms that could lead to a dramatic restructuring of our economy.

But the feedback effect isn't only pushing market concentration, undermining competition, and ultimately leading to higher prices and dangerously powerful oligopolies. It, perhaps surprisingly, also makes the superstar firms that use the

[19] On such data-rich markets, see Mayer-Schönberger and Ramge, Reinventing Capitalism, 59–85.
[20] Mayer-Schönberger and Ramge, Reinventing Capitalism, 69–70.
[21] See, e.g., Ryan Mac, *Stitch Fix: The $250 Million Startup Playing Fashionista Moneyball*, Forbes, June 1, 2016; see also https://algorithms-tour.stitchfix.com.

feedback effect extensively for innovation, as well as their customers, shockingly vulnerable to systemic failure. As AI systems learn from data and get better and better in assisting participants in a superstar's marketplace, more and more customers will be attracted to such a market. As a result, everyone may be using the same decision assistant. In such a context, any hidden errors and biases in the decision assistant will affect and potentially cripple every participant's decision making. It's akin to the danger of a brake in one's car malfunctioning due to a manufacturing defect – and the realization that everyone else's car uses the same faulty brake, too. This single point of failure undermines decentralized decision making, the very quality of markets that makes them so resilient and successful. In turn, this structural defect in the marketplace will make the superstar firm that runs the marketplace deeply vulnerable as well, because an error in decision assistance may now bring down the entire company.

Thus, superstar firms utilizing data-driven machine learning find themselves in a highly uncomfortable position. On the one hand, their exploitation of the feedback effect creates a unique and dangerous vulnerability. On the other hand, without data-driven innovation they will no longer be able to compete. Faced with the imminent threat of competition versus the potential danger of a systemic vulnerability, most firms will likely opt to tackle the former rather than focus on the later. Long-term, this will lead to a problematic situation, not just for their customers but for the companies as well.

POLICY RESPONSES

Market concentrations aren't new. By themselves, in most instances they aren't seen as problematic, at least from a consumer welfare perspective. As long as players even with very high market share don't exploit their position – for example by raising prices beyond what would be acceptable in a competitive market – regulatory measures are usually deemed unnecessary. Consequently, in competition law, the focus in recent years has been on constraining anticompetitive *behavior*.[22]

This poses a unique challenge in the context of the feedback effect. Companies exploiting the feedback effect, especially as detailed in the previous section, aren't behaving illicitly as defined by conventional competition regulation. They are simply utilizing the resources available to them, in the form of feedback data, to advance innovation. Companies have no obligation under existing competition law to make it easy for their rivals – large or small, established or new – to compete

[22] The focus on behavior in competition law in the US is evident even in works critical of the rise of digital superstars, such as Ezrachi and Stucke, Virtual Competition: The Promise and Perils of the Algorithm-Driven Economy (Harvard University Press 2016); one may believe that European competition law may be different, given its distinct historical trajectory, but as David J. Gerber (Law and Competition in Twentieth Century Europe (Oxford University Press 2001) points out, at least on this issue it is not; it, too, is focused on behavior.

against them. Only when they cross a red line into illegal behavior will they be stopped. Hence, one could argue that feedback effects even if resulting in market concentration aren't problematic in themselves, and do not require regulatory attention.

The flaw in this argument is the focus on consumer welfare, narrowly defined. Until now, competition law could be focused on such a narrow scope of consumer welfare, because innovation, which acted as a crucial counterforce to scale and network effects, was founded on human ingenuity. Nobody, not even the largest player in the market, could hope to have a monopoly on the human mind to have new ideas. At least in theory, therefore, the biggest incumbent could be dethroned by an innovative startup with an ingenious idea, so long as the incumbent did not engage in deliberately anticompetitive manipulation of the market. In practice, of course, large players have been able to use their market power to capture a disproportionate slice of human talent. But that has not precluded others from having great ideas, and bringing them to market, at least in sectors with relatively low barriers to entry.

Therefore, policymakers have not felt the need to worry about market concentration inhibiting innovation, at least not in the highly dynamic sector of digital technologies with its fast pace of innovation and comparatively low barriers to entry. With data-driven machine learning turning into a prime engine for innovation, this is no longer the case. Those with access to data in general, and feedback data in particular, now enjoy a massive advantage over others. This concentrates innovative activity among the largest players in a market and reduces the breadth and diversity of innovation.

Some may argue that a concentration of innovation activity isn't bad in itself. High-tech areas such as chip manufacturing, despite being highly concentrated, have been subject to continued innovation in recent years. The truth is, however, that low barriers to entry have always been the hallmark of strong and sustained innovation. It may be true that chip manufacturing is concentrated, but the advances in computer chips over the past two decades (if not longer) do not primarily stem from advances in chip manufacturing coming from the few remaining large chip fabs, but from innovation in tooling for chip manufacturing, and even more importantly from chip design – a far less concentrated area with far lower financial barriers to entry. A similar argument could be made for the innovation dynamic in biotech, and the crucial role of CRISPR in enabling genetic engineering at relatively low cost.[23] As it turns out, the most recent trajectories of high-tech innovation offer ample evidence for diversity and against concentration.

There is a decisive further reason for being concerned about the concentration of innovative activity. As detailed in the previous section, feedback-data-driven

[23] See, e.g., Juan Enriquez and Steve Gullans, Evolving Ourselves: Redesigning the Future of Humanity–One Gene at a Time (Portfolio 2016).

machine learning concentrates innovation among superstar firms that operate data-rich markets and offer smart decision assistance. The resulting single points of failure create vulnerabilities for the entire market, including market participants. Just consider a failure of the recommendation system in an app store market for one of the two big smart phone ecosystems: It could prompt hundreds of millions of consumers worldwide to download apps they don't need or want, and that perhaps are even nefarious, irrespective of any illegal or uncompetitive behavior of the market provider.

Massive market failures and the resulting potential for huge losses in consumer welfare, could, perhaps should turn into a valid concern for policymakers. But because this dynamic involves no illicit behavior by market participants, current competition law fails to protect against it.[24] Worse, even if competition law were triggered by data-driven innovation, it offers no suitable remedies. Behavioral remedies fail when the problem lies in the very dynamic of data-driven innovation. For instance, prohibiting or greatly constraining data-rich companies to utilize their data troves to gain novel insights, makes little sense: it would stifle innovation, limit the insights gleaned from data, and reduce data's overall utility, which likely translates into a reduction in overall consumer welfare and market efficiency.

In light of the limitations of behavioral remedies and the huge power of the most well-known superstar firms, the so-called GAFA – Google (technically a unit of the holding company Alphabet), Amazon, Facebook, and Apple – some have suggested that these firms should be broken up, much like AT&T, to prevent their controlling such large shares of the market.[25] This seems a blunt and draconian remedy, and it's unclear how such a remedy could be anything more than a temporary fix. Given the underlying drivers of market power (scale, network, and feedback effects), market concentration after such a breakup would likely continue anew.

This calls for different and novel policy measures – both new triggers for regulatory action, and new remedies – that are more carefully crafted to address the root cause of the problem: the shift in the source of innovation. Expanding on work by Jens Prüfer and colleagues on search engines, I suggest a progressive data-sharing mandate.[26]

The principle of such a mandate is straightforward: Every company with a share above a certain threshold – for instance 10 percent – in a market has to let other market participants have access to a subset of the data it has collected and uses in its

[24] See, e.g., Ariel Ezrachi and Maurice Stucke, Virtual Competition.

[25] For example, US Senator Elizabeth Warren has called for GAFAs to be broken up (Elizabeth Warren, *Here's How We Can Break Up Big Tech*, Medium, March 8, 2019, https://medium.com/@teamwarren/heres-how-we-can-break-up-big-tech-9ad9e0da324c).

[26] Jens Prüfer and Christoph Schrottmüller, *Competing with Data*, TILEC Discussion Paper 2017–006 (February 16, 2017); see also Cedric Argenton and Jens Prüfer, *Search Engine Competition with Network Externalities*, 8 Journal of Competition Law 73 (2012).

data-driven machine learning systems. The higher the market share of a particular company, the larger the slice of the data that it has to share.[27]

Data would be depersonalized to avoid any undue privacy risks[28] and the data shared would be chosen randomly from the full dataset.[29] Technically, access would happen through an appropriate API – and without a regulator in the middle to eliminate any unnecessary slowdown.[30] The task of the regulator would be to ensure and enforce compliance, and to categorize companies based on market share.

Importantly, every competitor would be granted access to the appropriate slice of the data. For instance, in a market with two large players commanding 30 and 40 percent of the market and many small players below the 10 percent threshold, not only the small players would get access to a slice of the data of the big players. Each of the big players would be entitled to access a slice of the other big player.

This setup has a number of advantages. Every market participant could continue to utilize the data available to it; unlike a monetary redistribution through a tax, the progressive data-sharing mandate does not "rob" the large players of the ability to innovate. But by granting access to other especially smaller players, it enables these smaller players to amass large enough data sets to use in data-driven machine learning, and thus to stay innovative. It also facilitates competition in the market by helping smaller players without inhibiting the ability of large players to utilize data as well. Every player benefits from access to data, but smaller players benefit more so – relatively speaking – than larger players. In short, the idea builds on the unique quality of data to be used multiple times by different parties without losing its value.

The progressive data-sharing mandate is also narrowly tailored to tackle the problem in question. If the challenge is the shift in the source of innovation to data, enabling access to data spreads the raw material of innovation. The policy also

[27] The then leader of the German social democratic party SPD, Andrea Nahles, advocated a more comprehensive model of progressive data sharing that would be limited to companies above a certain market share threshold, but require all firms above a certain simple minimum threshold of revenue or customers to be included. Termed "data-for-all" law, the proposal aimed not only to curb the information power of the digital superstars, but to stimulate data-driven innovation in Germany, especially among its numerous smaller and medium-sized enterprises. See Digitaler Fortschritt durch ein Daten-für-Alle-Gesetz, SPD, https://www.spd.de/aktuelles/daten-fuer-alle-gesetz/.

[28] A lot of data is actually not personal data at all, but data gathered from sensors measuring aspects of reality, like vibrations of jet engines or temperatures of machines in a factory; and while I do not want to downplay the privacy challenge, much of the remaining personal data can be depersonalized, although this demands skills and requires effort as well as the appropriate depersonalization strategy. Recent advances in cryptography have made effective depersonalization even of large data sets or highly personal data possible, and new technologies are becoming available that expand the depersonalization toolkit, like fully homomorphic encryption (for an overview see Frederick Armknecht et al., *A Guide to Fully Homomorphic Encryption*, https://eprint.iacr.org/2015/1192.pdf).

[29] Random choice of data ensures against strategic gaming in the choice of data being shared, either from the requester or the provider of data.

[30] When the US Department of Justice demanded that Google enable access to data it got through the acquisition of travel back-office service provider ITA, Google established access through an API.

does not negate the effort expended by large data collectors, as competitors gain access to only a randomly chosen subset of the data trove and not all of it. Crucially, this measure protects not only market competition, it also ensures a diversity of players based on a diversity of data sets. As the data subset provided to competitors is randomly selected, each player will have a somewhat different data source to learn from. This means that, for instance, not only multiple recommendation engines will be possible, but that the data used to train each such engine differs from one another, preventing the likelihood of systemic weaknesses.

Won't the large data-using superstars battle such a mandate, thereby dooming its chance for legislative success? Not necessarily. To be innovative in the context of data-rich markets will require access to lots of relevant feedback data. But even though data is the prime component for success, it is not the only one. Google's chief economist Hal Varian has said as much, when he highlighted the differentiating power of algorithms, especially regarding the most appropriate machine learning tools.[31] This suggests that large companies that have superb data analytics and machine learning capabilities continue to be well-placed to extract innovation out of data. Hence, they may see giving data access to smaller competitors as less of a ruinous threat, especially compared with some of the regulatory alternatives – like breakups – being discussed.

Moreover, mandating data sharing isn't an entirely novel policy measure. Its principle of enabling data access is embedded in a number of regulatory measures that have been enacted around the world. In the US and the European Union for instance, the legal right of phone subscribers to keep their phone numbers as they switch operators essentially disappropriated phone companies from valuable assets in the name of lowering switching cost.[32] It resulted in an increase in competition in the phone markets, and improved consumer welfare. Perhaps emboldened by this success, the European Union later passed legislation to let bank customers get access to their bank account data in machine-readable form.[33] This was intended not only to lower switching cost (and enhance competition in the banking sector), but also to create a wide stream of informational raw material that innovative fintechs can avail themselves of to enter the market. The goal is a diverse and innovative ecosystem of financial insight driven by drastically enhanced access

[31] *Data Is Giving Rise to a New Economy*, The Economist, May 6, 2017.

[32] In the US, mobile phone number portability was enacted as part of the Telecommunications Act 1996 and codified as 47 U.S.C. § 251(b)(2); in the European Union it is Article 30 of the Universal Service Directive (Directive 2002/22/EC of the European Parliament and of the Council of 7 March 2002 on universal service and users' rights relating to electronic communications networks and services, OJ L 108, 24.4.2002, 51–77); similar legislation exists in many other nations around world.

[33] Directive (EU) 2015/2366 of the European Parliament and of the Council of 25 November 2015 on payment services in the internal market, amending Directives 2002/65/EC, 2009/110/EC and 2013/36/EU and Regulation (EU) No 1093/2010, and repealing Directive 2007/64/EC, OJ L 337, 23.12.2015, 35–127, also referred to as the Payment Services Directive 2 (PSD2).

to data. And finally and most dramatically, the EU's General Data Protection Regulation[34], which came into force in 2018, explicitly mandates "data portability" – the right of individuals to get all personal data from a data processor in machine-readable form.[35] It's phone number portability and bank account portability spread across the board and applied to all personal data.

There is an important difference, however, between data portability and the progressive data-sharing mandate. Data portability's immediate aim is a rebalancing of informational power away from large data processors and towards individuals. Only if individuals then make their "portable data" accessible to other processors can the market concentration process be halted. The health of market competition thus hinges on the behavior of individuals, who do not all have strong incentives to share their data.[36] Moreover, because their data contains personal identifiers – it's the unaltered personal dataset – individuals have to trust data processors each time they share their data with them. This puts an unfair burden on individuals and results in an unfortunate disincentive for data sharing.

Data portability is simply not predictable and sustainable enough a policy measure to ensure competition and diversity in markets. But it is, no doubt, a powerful case in point that even legislative mandates that constrain the power of large data processors can get enacted – and in this case even on a pan-European level. This bodes well for the chances of a progressive data-sharing mandate.

CONCLUSIONS

In the past, markets have remained competitive in significant part because scale and network effects have been counterbalanced by innovation. Competition law could thus be focused on uncompetitive behavior, and not on market concentration in general. As the source of innovation shifts from human ingenuity to data-driven machine learning, behavioral constraints are no longer sufficient to protect competition.

The situation is exacerbated when the raw material of innovation in digital decision assistance is feedback data, collected and used by the providers of both

[34] 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), OJ L 119, 4.5.2016, 1–88.

[35] Art 20 GDPR: "The data subject shall have the right to receive the personal data concerning him or her, which he or she has provided to a controller, in a structured, commonly used and machine-readable format and have the right to transmit those data to another controller without hindrance from the controller to which the personal data have been provided, where: the processing is based on consent pursuant to point (a) of Article 6(1) or point (a) of Article 9(2) or on a contract pursuant to point (b) of Article 6(1); and the processing is carried out by automated means."

[36] This is essentially a "collective action problem", as described by Mancur Olson (The Logic of Collective Action: Public Goods and the Theory of Groups (Harvard University Press 1971)) and highlights negative externalities of such an individuals-based approach.

markets and digital assistants. Then, in addition to worries about a concentration of innovation, we may also face a single point of failure, exposing the market itself to a structural vulnerability.

The progressive data-sharing mandate is the policy measure I propose to address this unique situation. It is narrowly tailored to spread access to the raw material of innovation, with incentives for data utilization and renewed competition based on the ability to tease valuable insights from the raw data. While novel as a competition measure, it is based on principles of lowering switching cost and enhancing competition that are well-rooted in existing policy practices. If enacted, the progressive data-sharing mandate will act as a powerful antidote to market concentration, foster broad innovation, and prevent systemic vulnerabilities of online markets.[37]

[37] The idea of a data-sharing mandate has cropped up in a number of legislative and regulatory proposals, especially in Europe. For instance, in May 2019, the European Commissioner for Competition, Margrethe Vestager called for data-sharing (see John Detrixhe, *Instead of Breaking Up Facebook, the EU May Force It to Share Its Data*, Quartz, May 17, 2019, https://qz.com/1622036/break-up-facebook-eus-margrethe-vestager-favors-making-it-share-data-instead/); similar suggestions were put forward on a national level: in May 2019 the Dutch government sent a letter to the Dutch parliament on amending competition policy and opined: "The responsible competition authority must be given the power to take ex ante action if a platform risks gaining a position where it becomes impossible for businesses or consumers to avoid it. For example, it should be possible to impose obligations on a platform to share data with other companies", https://www.government.nl/latest/news/2019/05/27/dutch-government-change-competition-policy-and-merger-thresholds-for-better-digital-economy; in March 2019 the UK Digital Competition Expert Panel wrote "Active efforts . . . to make data available for competitors, offering benefits to consumers and also facilitating the entry of new businesses", https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/785547/unlocking_digital_competition_furman_review_web.pdf; in 2018 the Swiss Expert Group on the Future of Data Use and Data Security suggested that the Swiss government should consider a data-sharing mandate for non-personalized data, https://www.newsd.admin.ch/newsd/message/attachments/53591.pdf.