Digital challenges for competition policy

Hal Varian
Maurits Dolmans
Gavin Baird
Sept 2018

We welcome the Commission's invitation to discuss "Shaping competition policy in the era of digitisation" as a basis for the ongoing debate and dialogue between industry regulators and academics. This paper comments on each of the three themes identified in the Commission's call for contributions.

1. Competition, data, privacy and AI

In a world of ubiquitous data, thanks to, for example, 5G, the Internet of Things and connected cars, where would we have data bottlenecks – or, conversely, data access, data sharing or data pooling – causing competition issues? In which ways should privacy concerns serve as an element of the competition assessment? Since data is the raw material of artificial intelligence, how do we ensure that AI technology is as competitive as possible?

Companies have been acquiring and utilizing data for centuries, but recent technological developments have led to widespread availability of sensors, mobile devices, networking, cloud computing and databases that have made it easier than ever to acquire, organize and analyze data. Nowadays raw data is plentiful. The difficulty is finding analysts with the experience and expertise to extract value from this data, and even that problem is being addressed by colleges, universities, and online education.

Data as a barrier to entry

One question that has arisen in competition policy is whether data creates a barrier to entry. An incumbent who is currently producing and selling a product to its customers will typically have more information, knowledge, and experience than a new entrant. Should we consider these to be barriers to entry? If not, why would one consider data to be a barrier to entry?

It is true that a competitive asset can be a barrier to entry, if the fact that a first mover has created or acquired it makes it more difficult or more costly for a potential entrant to create or acquire a similar asset. This is generally not the case with data, and not even with the information, knowledge and experience gained from the data. Information can often be acquired or duplicated, while knowledge and experience are typically embedded in human beings, who can be enticed or inspired to move from one employer to another. In fact, the large internet companies have spawned hundreds of startups. In many ways these companies serve as providers of post-graduate education, similar to the role played by IBM, HP, McKinsey, and

other iconic companies. It is almost impossible to hoard knowledge in a free labor market. If you have the information, knowledge, and experience to build a new company, the data will typically become available as a byproduct of operations.

Data as a non-rivalrous good

We hear that “data is the new oil”. Data and oil have one thing in common: they both need to be refined before they become useful. Data by itself is essentially useless. If it is never turned into information or knowledge it has no economic value. It is only by investing in understanding and utilizing data that it can become valuable.

Unlike oil, data is non-rival. Two people can’t consume the same barrel of oil, but two people can easily consume the same data. Data is more like sunshine than it is like oil in this respect.

The non-rival nature of data means that economic terms like “ownership” are not adequate to describe the variety of economic transactions involving data. There are many possible contractual arrangements involving data such as access restrictions, use restrictions, licenses, sharing, pooling, merging, and mining data. Debating “ownership of data” in this context is pointless since data can have many owners. One needs a much richer set of concepts to understand and appreciate the economics of data.

Diminishing returns to data

Artificial intelligence often uses large amounts of data. However, like other factors of production data tends to be subject to diminishing marginal returns: the first million observations are much more valuable in improving predictions than the last million. In fact, according to a study by Hestness et al [2018], accuracy as a function of sample size fits a “power law” very closely for machine translation, language modeling, image processing, and speech recognition.
Other studies using the ImageNet data and the Stanford Dogs Project have exhibited similar diminishing marginal returns.

Applying AI used to be very costly due to the large amount of computation required. But now everyone has access to data centers that make these computations feasible to all. See, for example, this example of how to estimate clickthrough rates for millions of online advertisers, or browse the list of competitions and freely available datasets available from Kaggle.

Data as labor

Another issue that has arisen lately is that of “data as labor”. If products can be improved by collecting and analyzing data, should not those who generate that data be paid for providing it? First, in the case of advertising-supported businesses, such as radio, TV, and magazines consumers offer attention to ads in exchange for free or inexpensive content. Second, when

firms use customer data to improve products, the compensation is the product improvement themselves that consumers enjoy.

Nowadays many companies can continually improve their products, even after they've been sold. For example, smartphones provide operating system upgrades, as do TVs and even automobiles. This data enables continuous improvement---"computer kaizen"---that is immensely valuable to users as it gives them enhanced products at zero cost. This is due to the non-rival nature of data: multiple parties access the same data with minimal additional cost to the user or to the company that sells the product in question.

Furthermore some users *are* paid for their data. All the large internet companies subscribe to services like Nielsen, Comcast, and the like that collect data on user behavior. These services have hundreds of thousands of users who are compensated to provide data concerning their online and offline behaviors, which is primarily for marketing studies.

Most companies do extensive testing of their products by recruiting people to evaluate the product performance. Google, for example, pays more than 10,000 search quality raters to evaluate various sorts of changes to its search ranking system. The bible for these raters is the Search Quality Raters Handbook, a 160-page book that describes the criteria Google uses to evaluate websites.

Finally we must point out that the value of data on a per user basis is tiny. Facebook had earnings in 2017 of about 10 billion dollars and claims to have 2 billion users. That amounts to $5 per user per year, or 1.6 cents per day. This applies to most large internet companies: they make a small profit per user, but they can have billions of users.

Data access

Some large internet companies make it easy for a user to download their data and have done so for years. The Google Takeout project enables users to download their data to their own computers or to the online storage service of their choice. More recently the Data Transfer Project allows users to directly transfer their data from one platform to another. The initial participants in this project are Google, Facebook, Microsoft, Twitter, Flickr, and Instagram, but it is expected that more firms will adopt the open-source APIs that facilitate data transfers.

Privacy

Mandatory data transfer requirements should be carefully considered because any government measure mandating the sharing of assets reduces incentives to invest in their creation. Proportionate alternatives should be considered first, like the data portability policy mentioned above.

In addition, mandatory data transfer will likely raise privacy concerns since much data is managed by several parties. For example, data about a household's electricity usage could plausibly be considered to be controlled by both the utility service that generates and distributes that electricity, and by the household that consumes the electricity. If a regulator wants to enhance competition by requiring the utility to share that data, does this require approval by the household? This problem arises even with a simple case like email. If A sends email to B, do they each have equal rights about how that message is disseminated or where it is stored?

This issue is especially important in the case of healthcare. Everyone values health and there are likely significant benefits to machine learning analysis of medical records. Should people be encouraged to share healthcare data? Should this be opt-in or opt-out? It is likely that different countries will make different choices. Think of the different privacy regimes in China, UK, US, and others. This will make data sharing across countries more difficult, but it may be helpful in understanding the tradeoffs involved in different policy regimes.

2. Market power

The interests of platforms are not always aligned with the interests of their users, which can, as a result of platforms' market power, give rise in particular to: a) leveraging concerns (digital platforms leveraging their positions from one market to another); and b) lock-in concerns (network externalities, switching costs, better service due to accessibility of data make it difficult for users to migrate to other platforms, and allow platforms to "exploit" their user bases). What should/can competition policy do to address these concerns and how?

Online competition

The industry structure for online tech companies is unusual. Each company has its historical core competencies: Amazon in shopping, Apple in mobile devices, Facebook in social networks, Google in search, and Microsoft in operating systems and office applications, Netflix in streaming video access. It is often claimed that these firms are monopolists in their core area. But at the same time, each of these companies competes intensely with the others outside their core area, through pricing, product design, and innovative services. For example, Apple, Amazon, Google, and Microsoft all provide operating systems. Apple, Google, and Microsoft also provide competing office applications and so on.

There are in fact dozens of areas where these companies compete. Amazon's historical core was book sales, which evolved into general retailing which in turn spawned cloud computing services. Google once specialized in search, but now is also heavily involved with driverless cars, cloud computing, and entertainment.

This competition keeps these firms on their toes. Even if they continue to be the main player in their core area, that is not a sign of lack of competition: it's often a sign that they continue to invest and innovate, in the knowledge that if they slacken even briefly, they will be pushed out.

The reason for this multiproduct competition is that digital technology is a General Purpose Technology---it is highly flexible and can be used in many different ways. The computer engineering skills used to develop a word processor can also be used to develop an ebook. A platform is, quite literally, a structure on which others can build. Companies that create production platforms involving hardware, software, and wetware (human expertise) can utilize those capabilities in a variety of applications. Or, for that matter, provide those digital tools to third parties via cloud computing capabilities, which allows external developers to have access to the same computer infrastructure as the larger firms.

Product	AMZN	AAPL	GOOG	FB	MSFT
advertising platforms	✓		✓	✓	✓
artificial intelligence	✓	✓	✓	✓	✓
browsers	✓	✓	✓		✓
cloud services	✓		✓		✓
digital assistants	✓	✓	✓	✓	✓
ebooks	✓	✓	✓		
email and messaging		✓	✓	✓	✓
games	✓	✓	✓	✓	✓
general purpose search engines			✓		✓
home delivery services	✓		✓		
maps		✓	✓		✓
office tools		✓	✓		✓
operating systems	✓	✓	✓		✓
smartphones	✓	✓	✓		✓
social networks			✓	✓	
special purpose search engines	✓	✓	✓	✓	✓
streaming video	✓		✓	✓	
video and music distribution	✓	✓	✓		
video conferencing		✓	✓	✓	✓

From a static perspective, it may look as though these firms are invincible in their core areas. But a static view is not useful in high tech, since the very nature of the industry is based on creation and dissemination of new innovations.

Advertising and competition

Advertising is a common business model for online firms with 4 out of the 5 large online firms actively engaged in this industry. But large competitors are only the tip of the iceberg. There are hundreds of special purpose firms that also compete for advertising dollars. Facebook is just the tip of the iceberg in social networks; other smaller specialist companies such as LinkedIn, Twitter and Pinterest also compete in this area. In search, there are also hundreds of special purpose search engines such as Amazon, eBay, Travelocity, TripAdvisor, Hotels.com, Trivago, Idealo, eBay, Criteo, Shopify, MoneySupermarket, Pinterest and a host of other services that help users find products and services and are compensated by the providers of

those products and services. Individually and in the aggregate they are a formidable competitive force because they focus on the most monetizable aspects of search.

If we look at general purpose search engines like Google and Bing, only 6% of the clicks are ad clicks; the other 94% are organic clicks. New entrants to the search business quite naturally focus on commercial search, since that's where the money is. In the last several years, it has become very easy to create comparative shopping sites, since merchants can utilize standardized data formats to upload their product listings to comparative shopping sites. Of course as the entry costs have declined this segment has become highly competitive. Investing in building a general purpose search engine is costly since you have to offer capabilities and capacity to handle the 94% of clicks that make no money as well as the 6% of the clicks that do make money. So it is not surprising that much of the investment in new services focuses on the commercial side of search.

Entry costs

One competitive advantage of large internet companies is that they have designed and built massive computing facilities that can manipulate vast amounts of data to produce useful information. This required huge investments in software and hardware design. Ironically, all that investment has now been packaged up and is being sold to all comers (including competitors) via cloud computing. Entrants can now avail themselves at a moment's notice of systems that took years of effort and billions of dollars to develop. Intense competition is pushing prices down so that these services cost far less via the cloud that they would cost if they were managed internally and companies like Netflix, Apple, Snap, SAP and others are now taking advantage of these low prices by using the cloud to host their services. Users of cloud services can easily multihome and container technologies such as Docker and Kubernetes provide software and data portability among cloud providers.

Switching costs

In the early days of the automobile, different manufacturers used different user interfaces. Cars could be steered using pedals, levers, joysticks, wheels, and even reins. This made user switching costs high. This user interface fragmentation was beneficial to some manufactures but was costly to other providers and to the users themselves. As time went by, manufactures standardized on the interface we know today.

Similarly, the "desktop" interface of a PC operating system and the "touch" interface on smartphones have converged to a basic set of standard components. The result is that switching costs are quite low. The same is true in basic digital applications. You may be a power Excel user, but that doesn't impede you from creating and sharing a Google spreadsheet with a colleague on the other side of the globe.

A firm that switches suppliers, or a consumer that switches supermarkets, likely incurs more switching costs than one that changes map software or search engines. Of course, there may be quality differences in products, or aesthetic preferences that are important to users, but by and large, software and hardware tools have converged to similar user interfaces today. It is easy to try out a smartphone app and to switch back if you don't like it.

It is therefore ironic that critics express concerns about switching costs and lock-in in digital industries. Yes, being a popular service confers some advantages, but consumers find it very easy to switch if a better alternative comes along.

In 2014 Firefox made Yahoo its default search engine and Yahoo usage initially increased. But as users realized what had happened, they switched back to their preferred search engine and Yahoo search share on Firefox declined. In 2017 Mozilla terminated its agreement with Yahoo two years ahead of time, saying in a lawsuit that "*Yahoo Search consistently failed to retain users and search volume over time, reducing the potential revenue* [for Mozilla] *under the Strategic Agreement.*"

Switching from one mobile operating system to another, or one application to another is not hugely difficult. The problematic case arises when there are direct or indirect network effects involved. In such instances, there may be coordination problems that give rise to *collective* switching costs. Yes, it is easy for me to switch to a different social network...but how do I get all of my friends to switch with me at the same time? This is a challenge, but not an insurmountable one. After all, users did switch from MySpace to Facebook. A combination of measures that enable data portability and reduce barriers to multihoming will maintain the ability to switch, and thus maintain competitive pressure.

3. Preserving digital innovation using competition policy

Do network effects, economies of scale and 'copycat' products impede innovation? In digital merger cases, is there scope to apply theories of harm based on a loss of innovation and/or loss of "potential competition" more often? Would a focus on innovation require updating our analytical tools?

Network effects

In economics, positive feedback can give rise to economies of scale. There are two sorts of economies of scale.

> **Demand side economy of scale.** The more users that adopt a product or service the more valuable it becomes to prospective users.

Supply side economies of scale. The more output that is produced by a firm, the lower the average cost of production.

Demand side economies of scale are also known as network effects. They come in two flavors: direct network effects and indirect network effects. An example of the first case is a social network, where a larger network is more attractive than a smaller one since a potential adopter's friends are more likely to be in the larger network. An example of the second case is an operating system, where the larger operating system attracts more developers, and hence has more applications, making it more valuable to potential users.

These terms all have a reasonably precise definition among economists but are still commonly misused outside that community. "Network effects" are frequently invoked as a force that inevitably leads to industry concentration. However, when closely examined such claims lose credibility.

For example, some people refer to a "data network effect". The idea is that the more data a firm collects the better product its products become. And a better product attracts more customers, which in turn leads to more data being collected.

The trouble with this definition is that it applies to any factor of production. The more skilled workers a firm has the better the product it can build. And a better product attracts more customers which in turn leads to more workers being hired.

The problem with the concept of "data network effects" is twofold. First, the hypothesized increase in quality doesn't "just happen". The data has to be collected, organized, and analyzed to be useful, just as the workers have to be hired, trained, and managed effectively. Data that sits around is about as useful as workers that just sit around.

Second, there are costs to acquiring and managing more data just as there are costs to hiring and managing more workers. These costs cannot simply be ignored.

Here is a passage from an industry observer trying to explain why there is a "network effect" for search engines that stems "from the link between the attractiveness of the online search advertising side of a general search platform and the revenue of that platform.

1. The higher the number of advertisers using an online search advertising service, the higher the revenue of the general search platform;
2. revenue which can be reinvested in the maintenance and improvement of the general search service so as to attract more users."

It is likely that a search engine with a larger number of advertisers may have a higher revenue. The same is true of a TV station, a radio station, a magazine, a newspaper or any other ad-supported medium. But that revenue must first be used to pay the cost of operations. It is not *revenue* that can be reinvested, but rather *profits*. Reinvesting profits to maintain and

improve a product may well be an attractive strategy, but of course it is necessary to have the profit to reinvest.

So the alleged “network effect” described above simply says “First make a profit. Then reinvest your profit in the improvement of your product which will attract more customers which (perhaps) will generate more profits”. The hard part is “first, make a profit.” Plowing some of that profit back into your business may be good advice, but that has nothing to do with network effects.

Innovation

Hi-tech companies spend more on R&D than any other sector. Four out of the top eight spenders globally are large online companies.

It is the highly competitive environment in technology industries that drives this R&D expenditure. Not only are the major high tech firms competing intensely with each other, they are also competing with automobile companies, which are investing heavily in digital technology and robotics.

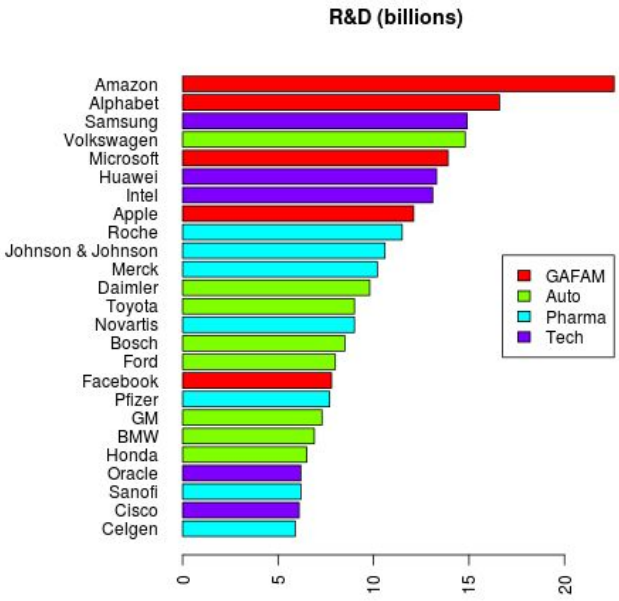

Figure 2. Global R&D spending by company. Source: Bloomberg [2018].

Venture capital

There are some studies that suggest a lack of dynamism in the US and European economies. This is particularly true of new business formation. By contrast, *venture capital* investment is robust. According to Sandhill Econometrics [2018] the total funding raised is approaching the

2000 peak and is expected to exceed it by the end of the year. There is no shortage of entrepreneurs or of capitalists who want to fund their projects.

There are two ways venture capitalists can exit from an investment: an IPO or an acquisition. Lately, firms have delayed going public, not just in high tech, but across the board. The reasons for this behavior are not clear—possible explanations may include regulatory burdens, the difficulty of accounting for R&D spending, increasing institutional ownership, and improved financial intermediation.

Regardless of the reason, if going public has become less popular the industry must rely on acquisition as a way of rewarding investors. Over the last 20 years there have been about four times as many acquisitions in the U.S. as IPOs. Half of those acquisitions had no value, implying that there were about twice as many successful acquisitions as IPOs.[1]

Critics of the large online companies accuse them of "vacuuming up" potential competitors and propose heightened restrictions on acquisitions. But if acquisitions become more difficult venture capitalists can only rely on IPOs, which could easily affect their incentives to invest in the first place. A country that imposed such a restriction would see a likely see entrepreneurs moving to a more investment-friendly environment.

Summary

We see three important principles that should govern competition law in this domain.

First, intervention should be *focused* on the fundamental objectives of competition law: the protection of competition and enhancement of consumer welfare. Competition law should not be used to pursue policy objectives that are better addressed with other rules and regulations. In particular competition law should not be used to preserve existing players' market presence and protect them from disruption.

Second, remedies should be *effective*. It makes no sense to impose a restriction or remedy on a firm if the competitive difficulties that its rivals experience are largely or entirely caused by factors other than the defendant's conduct -- for instance, by the emergence of new technologies, the rise of competitive alternatives, or changes in consumer preferences.

Third, regulation or remedies should be *necessary,* in the sense that no less restrictive and equally effective alternatives should be feasible. These include allowing firms to compete to meet consumer demand for particular features such as privacy or adequate platform management, but also technology-based solutions, and compliance by technical design.

---

[1] Authors' calculations based on data from Sand Hill Econometrics. Note that in 2018 the IPO market came back to life. See Maureen Farrell, *IPO Market Posts Blistering First Half,* Wall Street Journal, July 2, 2018.