

February 2021

Data-enabled learning: policy implications

(10000)] ===



Contact Matthew Johnson Partner

Despite superficial similarities, dataenabled learning does not necessarily create network effects, and when it does, data network effects are usually weaker and less conducive to lockin than standard network effects. In this article, Andrei Hagiu, Associate Professor of Information Systems at Boston University, and Julian Wright, Oxera Associate and Professor of Economics at the National University of Singapore, consider how policies that aim to correct market inefficiencies associated with data-enabled learning (such as mandatory data-sharing by incumbents or data privacy restrictions) have unintended consequences that may end up hurting customers overall

This is the second of two Agenda in focus articles on the topic of data-enabled learning. The first article discusses factors that determine whether data-enabled learning creates a sustainable competitive advantage, https://bit.ly/3aMVzUX.

In our previous Agenda in focus article,¹ we summarised some insights concerning firms' competitive advantage that came out of our work on data-enabled learning—the academic research article 'Data-enabled learning, network effects and competitive advantage',² and the Harvard Business Review article 'When Data Creates Competitive Advantage'.³ Our research also contains some important implications for policymakers, which are the focus of this article.

Data network effects and lock-in

There is a widespread tendency to assume



that data-enabled learning necessarily creates data network effects, and to equate the latter with standard network effects. This leads to exaggerated claims of competitive advantage and lock-in based on dataenabled learning.

Recall that a product or service exhibits regular network effects when the value to a user directly increases with the number of other users that buy the same product or use the same service because users want to interact or transact with one another. This covers both 'same-side network effects', where all users are essentially the same from the provider's perspective (e.g. users on social networks), and 'cross-side network effects', where there are two or more distinct customer groups that interact/ transact with one another (e.g. buyers and suppliers on marketplaces; users and app developers on iOS or Android).

Part of the confusion between data network effects and regular network effects, in our view, arises because many large tech companies that are built around platforms with regular network effects (such as Airbnb, Alibaba, Amazon, Facebook and Google) also have access to valuable data that enhances their existing network effects. In such cases, the strong competitive positions these companies already enjoy from network effects may be incorrectly (or excessively) attributed to their data.

Another source of confusion is that in some cases, what might seem like dataenabled learning leading to network effects is actually just regular network effects at play. For example, data-enabled learning is not present when users directly share information with each other (e.g. user reviews on TripAdvisor and Yelp, usergenerated playlists on Spotify, and user questions and answers on Quora and Stack Overflow). In such cases, the more such users there are, the more information will be shared and so the more valuable the service becomes. Such network effects are potentially quite powerful, but need not enhance a company's ability to learn from its customers' data.

Focusing on situations where true dataenabled learning is at play, data network effects only arise to the extent that the learning is across users. In other words, what is learned from some users must translate to the firm being able to offer a better experience for other users. This is in contrast to within-user learning, which means learning from any given user's history is relevant only to that user. For example, smart connected devices (e.g. thermostats) rely mostly on within-user learning. Such within-user learning may create switching costs, but it does not create anything like a network effect.

Even when they exist, data network effects are usually not as long-lasting and secure as regular network effects. There are a number of reasons for this.

First, there are usually many more ways around data network effects than regular network effects. Buying data is generally easier than acquiring customers. With data network effects, customers do not care about the presence of other customers per se, so it is usually possible to at least partially compensate for a smaller customer base by acquiring alternative sources of data and/or developing better algorithms. This means the cold-start problem is less severe and challengers have an easier time catching up to incumbents than in the case with regular network effects, where the only way to catch up is to get more customers.

Second, in many cases, data network effects run out of steam after attracting a relatively small number of customers. This may be because, with improvements in algorithms, it does not take a lot of data to extract most of the valuable learning. Even when large amounts of customer data are necessary, sometimes that data can be obtained from just a few large business customers (e.g. a few big farms, hospitals or law firms, depending on the application). This can be a lot easier to achieve than attracting the large number of customers that would typically be required under network effects. In some applications (e.g. speech recognition), dramatic improvements in AI and the emergence of publicly available datasets have reduced the need for unique customer data to the point where the value of data-enabled learning has largely disappeared. Regular network effects, on the other hand, often extend further and are more resilient: an additional customer still typically enhances value for existing customers (who can interact or transact with that newcomer), even when the number of existing customers is already large.

Third, self-reinforcing user expectations are less likely to play a role with data

network effects. With regular network effects, when choosing which firm to buy from, customers have a direct reason to care about what other customers will do in the future, because that will directly affect the benefit they get. For instance, such expectations play an important role in the adoption choices for new video game consoles. Every user wants to buy the console that they expect will attract a larger number of high-quality games and that will be adopted by a larger number of their friends. Such expectations usually favour incumbents and can make it challenging for

a new entrant to break into the market.

In contrast, with data network effects, the mechanism by which a product gets better when more customers adopt it is less direct and therefore less likely to be well understood by users, particularly in B2C contexts. Thus, users are more likely to behave myopically with data network effects, basing their decision only on a comparison of the value offered by the firms' current offerings. And by being myopic, present users looking to buy a product and use it for some time will end up underestimating how much better an incumbent's product would get relative to a challenger's if all future users were to join the incumbent. This means, other things being equal, it is easier for a new entrant to attract users and compete with data network effects than with regular network effects.

Fourth and finally, to produce lasting data-enabled network effects, a firm has to keep putting in work to learn from customer data. This work can include the ongoing tasks of gathering, cleaning, securing and processing data, as well as improving algorithms to remain competitive. In contrast, as one of us remembers Intuit cofounder Scott Cook saying, 'products that benefit from [regular] network effects get better while I sleep.' Indeed, with regular network effects, interactions between customers can continue to create value even if the firm stops innovating. For example, even if a new online site for classifieds offered buyers and sellers objectively better features than Craigslist does, it would still have to contend with Craigslist's network effects—buyers prefer going to the online classifieds site where most sellers are, and vice versa. Of course, even regular network effects require some maintenance work (e.g. minimising fraud or abuse by ill-intentioned participants). We would certainly not recommend that firms fall asleep—rather, they should continue to make investments in improving the quality of the interactions enabled. Still, once in motion, the self-reinforcing mechanism associated with regular network effects usually requires less fuel provided by the firm than the one associated with data network effects.

For all these reasons, in our view, policymakers should be less concerned about lock-in and dominant competitive positions with data-enabled learning than with traditional network effects. In fact, in our academic research paper,⁴ we have examined the conditions for the competitive outcome to be distorted away from the efficient outcome when firms benefit from data-enabled learning. We do this in a setting where an incumbent that has more data competes with an entrant that has less. Regardless of which firm has more data to start with, we find that the firm that creates the most overall value for users (taking into account how it learns from future customer data) will end up winning. This shows that despite the presence of a virtuous cycle for the winning firm, which will keep winning as it accumulates more data, the outcome is efficient. The result reflects that a forwardlooking entrant that has less data but has more scope to benefit from learning will be willing to sacrifice more profits in the current period than the incumbent-by offering low/ negative prices and a high-quality productin order to win customers (so it can benefit from their data).

We find the only distortion from this efficient outcome arises when consumers face some cost of switching between firms, and so form expectations about how much each firm's product will improve in the future when deciding which firm to join. This creates a coordination game very similar to that at play with regular network effects. In this case, if consumers hold expectations that favour the incumbent, it is possible that they all keep buying from the incumbent because they expect other consumers to keep buying from it in the future (and so its product to improve more), even though buying from the entrant would be more efficient. Note that consumers' switching costs can arise endogenously when the firm they use gets better for them the more they use it (i.e. within-user learning). Without switching costs, or if users do not take into account how future improvements in the firms' products will arise from other users' decisions, there will be no coordination problem, and our results imply that there is no reason to worry about the efficiency of the market outcome.

Unintended consequences of policy interventions

Although we find the firm that is most efficient and should win ends up attracting consumers, this does not mean that consumers always benefit from dataenabled learning. As a firm pulls ahead, attracting ever more customers and data, the ability of its rivals to exercise a competitive constraint on the leading firm will be weakened. This means that the leading firm no longer has to offer as much surplus to consumers, as it accumulates more data. It is therefore theoretically possible that data-enabled learning can actually hurt consumers (even though the outcome maximises *total* welfare)—and we show this possibility formally in our paper.

This might suggest that helping a firm that is behind to catch up by forcing the incumbent to share (some of) its data would put more competitive pressure on the incumbent's pricing, and always benefit consumers. But that is not necessarily the case. Such a data-sharing policy may reduce an entrant's incentive to compete to attract customers in the first place, because it anticipates the possibility of obtaining data via data-sharing instead. This is a form of free-riding, in which firms naturally invest less in accumulating data if they expect to be able to get it for free. In the case of data-enabled learning, this reduction in investment takes the form of competing less aggressively to attract customers, given that it is these customers that provide the source of data. Taking into account the negative ex ante effects and positive ex post effects of a data-sharing policy on consumers, the net effect is ambiguous. In our research paper, we provide conditions that determine which of these effects dominates.

Data privacy policies can also have unintended consequences once one takes into account how they interact with data-enabled learning. If such policies make it easier for consumers to keep their data private from firms, they will slow down the rate at which new customers generate additional useful data for firms, and so the rate at which firms learn as they attract more customers. This will have a disproportionate effect on firms that are behind, because such firms are more reliant on learning from new customers to catch up with the market leader. As a result, data privacy policies can have the unintended effect of strengthening the competitive advantage of the firms with more data-i.e. incumbents.

One can see this most clearly in the extreme case of an incumbent that has already learned everything it needs from data and which competes with an entrant that still needs to learn. The incumbent would be unaffected by a data privacy policy, whereas the entrant would be disadvantaged. And as a result, consumers could end up worse off, because the entrant would exert less competitive pressure on the incumbent if the data privacy policy were implemented. This is of course true for policies that affect incumbents and entrants symmetrically; other data privacy policies that target incumbents only might help entrants.

So what have we learned about data-enabled learning?

There has been a tendency in recent years for policymakers and commentators to assume that large platforms enjoy almost unbounded advantages from the large amounts of data they have access to, and the ability to leverage that data to dominate an ever-expanding array of markets. We suspect that this view often overstates the power of data and may result from conflating the barriers to entry created by traditional network effects and economies of scale with those created by data-enabled network effects.

The ability to leverage existing data to new applications is very often much more

limited once one takes into account the subtle (but important) nuances in the way that data is used across different applications in training algorithms. And as our research shows, implementing data policies can lead to unintended consequences for consumers, even though these policies may be well intentioned.

Professor Andrei Hagiu

Professor Julian Wright

Contact

matthew.johnson@oxera.com

Matthew Johnson

¹ Oxera (2021), 'Is knowledge power? Data-enabled learning and competitive advantage', *Agenda in focus*, January, https://bit. ly/3uBeGt9.

² Hagiu, A. and Wright, J. (2020), 'Data-enabled learning, network effects and competitive advantage', June, https://bit.ly/3pPLUkD.

³ Hagiu, A. and Wright, J. (2020), 'When Data Creates Competitive Advantage', *Harvard Business Review*, January–February, https:// bit.ly/3khEohr.

⁴ Hagiu, A. and Wright, J. (2020), 'Data-enabled learning, network effects and competitive advantage', June, https://bit.ly/3aOVG2r.

