# Artificial Intelligence, Data-Driven Learning, and the Decentralized Structure of Platform Ecosystems

**David R. Clough**
Organizational Behaviour and Human Resources Division
Entrepreneurship and Innovation Group
Sauder School of Business
University of British Columbia
ORCID: 0000-0002-2556-7033
Email: david.clough@sauder.ubc.ca

**Andy Wu***
Strategy Unit
Harvard Business School
Harvard University
ORCID: 0000-0002-9107-5731
Email: awu@hbs.edu

**ABSTRACT**

Gregory, Henfridsson, Kaganer, and Kyriakou (2020) highlight the important role of data and AI as strategic resources that platforms may use to enhance user value. However, their article overlooks a significant conceptual distinction: the installed base of decentralized users who connect with a platform lie outside the boundaries of the platform-owning firm, whereas the accumulated data derived from that installed base exists internal to the boundaries of the firm and under firm control. Accounting for this distinction brings forth two key departures from their theory. First, the decentralized structure of a platform ecosystem makes value capture by the platform an essential consideration when analyzing the implications of data-driven learning for users. Because AI and data allow a platform to increase the share of value the platform owner captures from the users, the value perceived by users can often decline. Second, as an internal asset of the platform firm, data from users and complementors exhibits different dynamics compared with the dynamics that govern the installed base itself. As a result, the quantity and quality of the platform's stock of data are only loosely coupled with the size of the platform's installed base. We highlight the strategic implications of this distinction for a manager launching a new multi-sided platform.

**Acknowledgements.** We thank Associate Editor Allah Afuah and two anonymous reviewers for their valuable feedback. Jeff Strabone provided copyediting support. Andy Wu gratefully acknowledges financial support form the HBS Division of Research and Faculty Development.

* Corresponding Author

We read with great interest the article, “The Role of Artificial Intelligence and Data Network Effects for Creating User Value” by Gregory, Henfridsson, Kaganer, and Kyriakou (2020). We agree with the authors that data is a strategic resource that enhances platform value and, as such, data-driven learning deserves greater attention within platform strategy research. Their article touches on many of the important strategic considerations that affect how platform owners can use AI capabilities and data to create value. However, their article overlooks a key conceptual difference that distinguishes the installed base of users and complementors on a platform from the data derived from that installed base. The users and complementors who connect with a platform lie outside the boundaries of the platform-owning firm, and the platform ecosystem has a decentralized governance structure. In contrast, the data accumulated by the firm is internal to the boundaries of the firm and centrally controlled by the firm.

We highlight two consequences of this essential but overlooked distinction. First, the decentralized structure of a platform ecosystem makes value capture—not just value creation—an essential consideration when analyzing an innovation ecosystem (Adner & Kapoor, 2010; Brandenburger & Nalebuff, 1996). We discuss how AI and data allow a platform to increase the share of value the platform owner captures from the ecosystem, such that the value *perceived* by users declines. Second, as an internal asset of the firm, data from users and complementors exhibits different dynamics compared with the dynamics that govern the installed base itself. As a result, the quantity and quality of the platform’s stock of data are only loosely coupled with the size of the platform’s installed base. Collectively, our arguments constitute an important refinement of the theory put forward by Gregory et al. (2020).

## DATA-DRIVEN VALUE CAPTURE

Gregory et al. (2020) focus their theory on how data-driven learning increases value creation for users. While we agree that value creation deserves attention, we believe this focus misses a key strategic implication of digitization: the impact of data-driven learning on value capture is as large as, or larger than, its impact on value creation. The outcome variable in the Gregory et al. (2020) framework is "perceived user value," and their framework predicts that AI capabilities and data increase perceived user value. But any analysis of the value perceived by a user, we suggest, must inherently take value capture into account. Decentralized users and complementors lie outside the boundaries of the platform-owning firm, creating tensions between the value they capture and the value the firm captures. First, if a platform captures a larger share of value, say by increasing price, the user will perceive less of the total value created. Second, value capture by the platform can also have a negative effect on total value created if the mechanism of value capture reduces the intrinsic quality or installed base of the platform. In this section, we show that AI capabilities and data allow a platform owner to increase the share of value they capture from interactions with platform users.

**Capturing Value with User-Level Price Discrimination**

A basic result in economics is that a seller can capture more value from its customers when the seller is able to price discriminate. If a seller can only charge the same price to all customers, the customers with higher willingness to pay gain a greater consumer surplus from the transaction. Two factors limit a firm's ability to price discriminate: how well it can categorize specific consumers as members of a certain segment and how well it ascertains the willingness to pay of each segment (e.g., Varian, 1989). For example, cinemas price discriminate by offering lower prices to students only when a customer's status as a student can be verified with a piece of ID.

By using accumulated data, a platform can overcome both factors that limit price discrimination. Platforms identify and track specific users through several methods, such as user account logins, device identifiers, web browser cookies, and IP addresses (Athey, Calvano, & Gans, 2016; D'Annunzio & Russo, 2019). Platforms create detailed databases that allow for the granular categorization of a specific customer along numerous dimensions. Platforms can then estimate the customer's willingness to pay for a good or service based on the platform's accumulated data on their general activity (e.g., web-browsing history), past transactions, and A/B tests conducted on similar customers (Shiller, 2020; Thomke, 2020). This allows the platform to customize the prices displayed to a given user in ways that extract value from the user, i.e., individual-level price discrimination, a practice that is well established in the pricing of flights (Sengupta & Wiggins, 2014) and e-commerce retail (Garbarino & Lee, 2003).

Data also allows the platform to optimize its value capture based on the user's present psychological state. Facebook told its advertisers that its data and algorithms can identify "moments when young people need a confidence boost" and target ads to teenagers when they feel "worthless," "insecure," "defeated," "anxious," "useless," "stupid," "overwhelmed," "stressed," and "a failure" (Reilly, 2017). Perhaps the optimal (in the purely economic sense) time to capture value from users is when they are at their lowest, which some might interpret as opportunistic or predatory.

**Value Capture at the Expense of Value Creation**

It is common practice to design platforms in ways that capture value for the platform owner at the expense of the total value being created (Cusumano, Gawer, & Yoffie, 2019; Parker, Van Alstyne, & Choudary, 2016; Zhu & Liu, 2018); this design choice impacts how platforms use the data they collect. Consider that many platforms deliver services to users at a price of zero and generate revenue by showing users advertisements (Ghose & Yang, 2009). Here, the platform owner has an incentive to create value for users in order to attract them to the platform and retain them as recurring users of the platform. But the platform's goal should not be misconstrued: the user is valuable to the platform because they pay attention to the advertisements posted on it. A decision that may reduce the value a user captures, while providing more value to advertisers, is something the platform would rationally enact, even if some users leave because of the advertisements thereby reducing the total value created by the platform ecosystem.

A platform has an incentive to manipulate the psychological state of its own users to keep them addicted to the platform. App developers write and share playbooks of tips and techniques for retaining users by taking advantage of psychological reward mechanisms (Zichermann & Cunningham, 2011). For example, some video games with freemium business models surreptitiously let players win more often immediately after the player makes a monetary in-game purchase, known as a microtransaction; this conditions the user to make repeated purchases in order to maintain the flow of dopamine-inducing victories (Drummond & Sauer, 2018; King & Delfabbro, 2018). Recently, there has been controversy around in-game purchases of "lootboxes" that go further and introduce a gambling-like element of chance to the experience, facilitating gambling addiction (Zendle & Cairns, 2018) and drawing the attention of regulators (Griffiths, 2018).

As the platform collects more data on how users respond to attempts at manipulation, the platform learns to manipulate its users more effectively. This data-driven learning may induce the user to spend more time and/or money on the platform, even if the user derives less overall enjoyment from their experience using it. Sometimes, users sense when they are being exploited. In 2017, Electronic Arts (EA) released its game Star Wars Battlefront II. To "unlock" popular characters such as Darth Vader, users either had to play the game for an overwhelming number of hours or pay to purchase the character in-game. EA utilized data on user behavior to dynamically make the hours-of-play requirement unachievable for most users. Many expressed disdain for EA's monetization strategy: "The truth is [EA] know[s] very few people are going to sink a full work week into this game and you're hoping that somebody is desperate enough to buy credits to unlock the character. It has nothing to do with providing a 'sense of pride and accomplishment.’ This is a flat-out lie and [EA] know[s] it.” (bookem_danno, 2017). In this case, the situation turned out poorly for EA: user complaints snowballed, earning EA the Guinness World Record for the most “downvoted” comment in the history of Reddit, a popular internet forum (Leskin, 2019). While one lesson is that you should not get between Star Wars fans and Darth Vader, another is that user or complementor backlash against platforms using data against them increasingly occurs in high-stakes settings (Zhu & Liu, 2018), such as drivers suing Uber for access to their personal data because they worry Uber uses it to manipulate them into driving longer hours or to cut their pay to a bare sustenance level (Holder, 2019).

Ultimately, firms make a strategic choice whether to use data to improve value creation or whether they use it for value capture. Data-driven learning implies a (monotonically) increasing relationship between the volume of data and the *capacity* for both creation and capture. Firms strategically decide whether more creation or capture is realized.

## DECOUPLING OF DATA AND INSTALLED BASE

Gregory et al. (2020) coin the term “data network effects,” which they describe as “a new category of network effects.” We agree with the authors that the phenomenon of data-driven learning by platforms deserves careful scholarly attention. However, we question whether referring to data-driven learning as a category of network effects is accurate. The network effects label brings with it an existing set of mental models and strategic frameworks which may not apply to data-driven learning. It is well-established that network effects stem from the size and structure of a platform’s installed base (Afuah, 2013; Farrell & Saloner, 1986; Katz & Shapiro, 1985), hence the label “data network effects” implies there is a tight coupling between data-driven learning and the installed base of users and complementors on a platform. We believe the coupling is weaker than this label suggests. The decentralized structure of a platform ecosystem means that the dynamics—i.e., changes in stock over time—of the installed base of users differ from the dynamics of the quantity and quality of platform-owned data. As a result, the extent of a platform's data-driven learning is only weakly coupled with its number of users. Hence, prescriptions derived from treating data-driven learning as a network effect may be misleading.

There are two fundamental differences between the dynamics of data-driven learning and the dynamics of direct and indirect network effects. First, data-driven learning is much more path-dependent than direct and indirect network effects. The data and behavioral insights that platforms derive from users and complementors accumulate over time, whereas direct and indirect network effects depend on the potential interactions that could take place between users and complementors at any given moment (Gawer, 2014; McIntyre & Srinivasan, 2017). A well-known mental model for direct network effects is the "critical mass" threshold for an installed base: a platform with a user base above that threshold comprises a self-sustaining ecosystem (Afuah & Tucci, 2003), and prospective users might join a platform on the expectation that its installed base will cross the threshold in the future (Fang, Wu, & Clough, 2020). Similar thresholds exist under indirect network effects. Research has yet to establish whether data-driven learning exhibits a similar critical threshold; if such a threshold exists, it would likely be defined by the platform's total accumulated experience with a set of users—akin to a learning curve—rather than the current or anticipated future size of the user base.

Second—ongoing legal disputes notwithstanding—data is presently treated as a proprietary and tradeable asset of a firm, whereas the installed base of users and complementors on a platform make autonomous decisions to join or leave platforms at will. In the current legal regime, platform users agree to lengthy terms and conditions (that most do not read), which in effect allow platforms a great deal of leeway in using and owning their data. In many cases, those terms and conditions allow the platform to sell that data: one of Twitter's highest-margin businesses is licensing user data in the form of subscriptions to historical and real-time data on the platform (Bary, 2018). As an asset, data can be purchased from third parties prior to establishing a user base;[1] it can also be retained even if users leave the platform.

To illustrate the strategic implications of these differences, consider the decision facing a manager launching a new multi-sided platform. If the manager believes that direct and indirect network effects are the key source of competitive advantage, they are likely to adopt a strategy to "get big fast" (e.g., Afuah, 2003; Eisenmann, 2006; Wu, Clough, & Kaletsky, 2019). On the other hand, if the manager believes that data-driven learning will be their key source of advantage, they are likely to prioritize learning over growth (Ries, 2011). Rather than rapidly achieve scale, they may instead spend effort trialing multiple—possibly incomplete—product alternatives on small subsets of users (Ghosh, 2020). They may seek to purchase an existing dataset to mine for insights on user behavior[2] or, if they work for an incumbent firm, capitalize on the firm's pre-existing data accumulated from its users and complementors.

As Gregory et al. (2020) suggest through their choice of label, data-driven learning and direct and indirect network effects are often positively correlated. Their article provides valuable insights that apply to those situations. We draw attention to instances when the dynamics of data-driven learning differ from the dynamics of direct and indirect network effects.

## CONCLUSION

We are grateful to Gregory et al. (2020) for initiating an important conversation about the role that data and artificial intelligence capabilities play in the creation of value within platform

---

[1] Even though LinkedIn does not sell its data, one can still buy it: hiQ Labs scraped LinkedIn online professional profiles so employers could have more visibility on prospective and current employees, without LinkedIn's permission (*hiQ Labs, Inc. v. LinkedIn Corp.*, 2019). Management researchers should take note that hiQ Labs won against LinkedIn in the U.S. Court of Appeals for the Ninth Circuit in 2019. Based on that precedent, for the time being it may be legal to scrape LinkedIn's publicly viewable member profiles, although it does violate LinkedIn's terms and conditions. LinkedIn intends to appeal the case to the U.S. Supreme Court (Neuburger, 2020).

[2] A platform can buy data on its competitors, such as when Uber purchased data on Lyft users from Unroll.me. Unroll.me presented itself as a service to help people clean up their email inboxes, but—unbeknownst to most users—it had been taking private email data, e.g., Lyft receipts, and selling it (Isaac & Lohr, 2017).

ecosystems. In this dialogue essay, we highlight a defining attribute of platform ecosystems: the decentralized nature of users and complementors. Recognizing this attribute changes some of the theory's predictions with respect to the value perceived by users and qualifies its assumptions on the extent to which data-driven learning is coupled with network effects.

## REFERENCES


Adner, R., & Kapoor, R. 2010. Value creation in innovation ecosystems: How the structure of technological interdependence affects firm performance in new technology generations. ***Strategic Management Journal***, 31(3): 306–333.

Afuah, A. 2003. Redefining firm boundaries in the face of the internet: Are firms really shrinking? ***Academy of Management Review***, 28(1): 34–53.

Afuah, A. 2013. Are network effects really all about size? The role of structure and conduct. ***Strategic Management Journal***, 34(3): 257–273.

Afuah, A., & Tucci, C. L. 2003. ***Internet business models and strategies: Text and cases*** (2nd ed.). New York: McGraw-Hill.

Athey, S., Calvano, E., & Gans, J. S. 2016. The impact of consumer multi-homing on advertising markets and media competition. ***Management Science***, 64(4): 1574–1590.

Bary, E. 2018. Twitter earnings growth relied on selling user data. ***MarketWatch***, April 25.

bookem_danno. 2017. Seriously? I paid 80$ to have Vader locked? ***Reddit***. https://www.reddit.com/r/StarWarsBattlefront/comments/7cff0b/seriously_i_paid_80_to_have_vader_locked/.

Brandenburger, A. M., & Nalebuff, B. J. 1996. ***Co-opetition***. New York: Currency Doubleday.

Cusumano, M. A., Gawer, A., & Yoffie, D. B. 2019. ***The business of platforms: Strategy in the age of digital competition, innovation, and power***. New York: HarperCollins.

D'Annunzio, A., & Russo, A. 2019. Ad networks and consumer tracking. ***Management Science***. https://doi.org/10.1287/mnsc.2019.3481.

Drummond, A., & Sauer, J. D. 2018. Video game loot boxes are psychologically akin to gambling. ***Nature Human Behaviour***, 2(8): 530–532.

Eisenmann, T. R. 2006. Internet companies' growth strategies: Determinants of investment intensity and long-term performance. ***Strategic Management Journal***, 27(12): 1183–1204.

Fang, T. P., Wu, A., & Clough, D. R. 2020. Platform diffusion at temporary gatherings: Social coordination and ecosystem emergence. ***Strategic Management Journal***. https://doi.org/10.1002/smj.3230.

Farrell, J., & Saloner, G. 1986. Installed base and compatibility: Innovation, product preannouncements, and predation. ***American Economic Review***, 76(5): 940–955.

Garbarino, E., & Lee, O. F. 2003. Dynamic pricing in internet retail: Effects on consumer trust. ***Psychology & Marketing***, 20(6): 495–513.

Gawer, A. 2014. Bridging differing perspectives on technological platforms: Toward an integrative framework. ***Research Policy***, 43(7): 1239–1249.

Ghose, A., & Yang, S. 2009. An empirical analysis of search engine advertising: Sponsored search in electronic markets. ***Management Science***, 55(10): 1605–1622.

Ghosh, S. 2020. Think before you act: The unintended consequences of inexpensive experimentation. ***Harvard Business School Working Paper***.

Gregory, R. W., Henfridsson, O., Kaganer, E., & Kyriakou, H. 2020. The role of artificial intelligence and data network effects for creating user value. ***Academy of Management Review***. https://doi.org/10.5465/amr.2019.0178.

Griffiths, M. D. 2018. Is the buying of loot boxes in video games a form of gambling or gaming? ***Gaming Law Review***, 22(1): 52–54.

hiQ Labs, Inc. v. LinkedIn Corp. 2019. ***U.S. Court of Appeals for the Ninth Circuit***, vol. 938 F.3d: 985. (9th Cir. 2019).

Holder, S. 2019. For ride-hailing drivers, data is power. ***CityLab***, August 22.

Isaac, M., & Lohr, S. 2017. Unroll.me service faces backlash over a widespread practice: Selling user data. ***New York Times***, April 24.

Katz, M. L., & Shapiro, C. 1985. Network externalities, competition, and compatibility. ***American Economic Review***, 75(3): 424–440.

King, D. L., & Delfabbro, P. H. 2018. Predatory monetization schemes in video games (e.g. “loot boxes”) and internet gaming disorder. ***Addiction***, 113(11): 1967–1969.

Leskin, P. 2019. EA’s comment on a Reddit thread about “Star Wars: Battlefront 2” set a Guinness World Record for the most downvoted comment of all time. ***Business Insider***, September 9.

McIntyre, D. P., & Srinivasan, A. 2017. Networks, platforms, and strategy: Emerging views and next steps. ***Strategic Management Journal***, 38(1): 141–160.

Neuburger, J. D. 2020. LinkedIn files petition to the Supreme Court in HiQ web scraping case. ***National Law Review***, March 12.

Parker, G. G., Van Alstyne, M. W., & Choudary, S. P. 2016. ***Platform revolution***. New York: W. W. Norton & Company.

Reilly, M. 2017. Is Facebook targeting ads at sad teens? ***MIT Technology Review***, May 1.

Ries, E. 2011. ***The lean startup: How today’s entrepreneurs use continuous innovation to create radically successful businesses***. New York: Crown Business.

Sengupta, A., & Wiggins, S. N. 2014. Airline pricing, price dispersion, and ticket characteristics on and off the internet. ***American Economic Journal: Economic Policy***, 6(1): 272–307.

Shiller, B. R. 2020. Approximating purchase propensities and reservation prices from broad consumer tracking. ***International Economic Review***, 61(2): 847–870.

Thomke, S. H. 2020. ***Experimentation works: The surprising power of business experiments***. Boston: Harvard Business Press.

Varian, H. R. 1989. Price discrimination. In R. Schmalensee & R. Willig (Eds.), ***Handbook of industrial organization***, vol. 1: 597–654. Amsterdam: North-Holland.

Wu, A., Clough, D. R., & Kaletsky, S. 2019. Nascent platform strategy: Overcoming the chicken-or-egg dilemma. ***Harvard Business School Industry and Background Note No. 719-507***.

Zendle, D., & Cairns, P. 2018. Video game loot boxes are linked to problem gambling: Results of a large-scale survey. ***PLOS One***, 13(11): e0206767.

Zhu, F., & Liu, Q. 2018. Competing with complementors: An empirical look at Amazon.com. ***Strategic Management Journal***, 39(10): 2618–2642.

Zichermann, G., & Cunningham, C. 2011. ***Gamification by design: Implementing game mechanics in web and mobile apps***. Sebastopol, CA: O’Reilly Media.

## AUTHORS

**David Clough** (david.clough@sauder.ubc.ca) is an Assistant Professor in the OBHR Division and the Entrepreneurship and Innovation Group at the University of British Columbia's Sauder School of Business. He received his PhD from INSEAD. His research interests include new venture emergence, organizational learning, and technological change in innovation ecosystems.

**Andy Wu** (awu@hbs.edu) is an Assistant Professor in the Strategy Unit at Harvard Business School. He received a PhD and MS in Applied Economics from the Wharton School of the University of Pennsylvania, where he is a Senior Fellow at the Mack Institute for Innovation Management. He researches, teaches, and advises managers on strategy for entrepreneurial technology ventures to organize for scalable innovation and mobilize resources for growth.