DATA NETWORK EFFECTS: KEY CONDITIONS, SHARED DATA, AND THE DATA VALUE DUALITY

(post-print version of:

Gregory, R., Henfridsson, O., Kaganer, E., Kyriakou, H. "Data Network Effects: Key Conditions, Shared Data, and the Data Value Duality," *Academy of Management Review,* forthcoming)

Robert Wayne Gregory
University of Virginia, rg7cv@virginia.edu

Ola Henfridsson
University of Miami, ohenfridsson@miami.edu

Evgeny Kaganer
Moscow School of Management, SKOLKOVO, evgeny_kaganer@skolkovo.ru

Harris Kyriakou
ESSEC Business School, kyriakou@essec.edu

Clough and Wu (2020) provide an interesting and thought-provoking response to our article (Gregory, Henfridsson, Kaganer, & Kyriakou, 2020) on the role of Artificial Intelligence (AI) and data network effects for the creation of user value. We welcome the debate around data network effects as a new category of network effects. In this response note, we build upon the points raised by Clough and Wu (2020) to outline three clarifications to our theory of data network effects concerning: (1) conditions when data network effects accrue, (2) the importance of theorizing shared data, and (3) the model's ability to explain the cumulative effect of data-driven learning on value creation and value capture.

KEY CONDITIONS FOR DATA NETWORK EFFECTS

Clough and Wu's (2020) response calls for further clarification about the conditions under which we can refer to data network effects. As recently explained by Cennamo (2020), building on the ideas presented by Hagiu and Wright (2020), there are two key conditions for data network effects. The first condition is that learning from one user should translate into a better product or experience for other users, not just that single user. In other words, as more users use the product, the product must improve the experience for all users. The second condition is that the product experience enhancement from learning should happen fast enough to affect the current value of the product. It should benefit its current users, not the next product generation's users. In other words, the product improves over the consumption lifetime with more users adopting it.

Clough and Wu (2020) suggest that the term "data network effects" is misleading, assuming that the two conditions stated above are typically not met. This view, however, overlooks the role of AI which is central to the activation of data network effects. As stated in our original article (Gregory et al., 2020), the widespread use of AI on modern platforms makes a significant contribution to fulfilling the two conditions for data network effects stated above. First, a core characteristic of AI is learning from individual cases to identify and translate patterns into predictive models that feed into the iterative improvement of products and experiences for other users, not just one single user from which data is collected and analyzed. Second, another core characteristic of AI is the ability to efficiently scale data-driven learning and instantly release the resulting improvements to the product experience to affect the current value of the product for each user. While we explained these characteristics of AI in our original article (Gregory et al., 2020), we had not explicitly linked them to the key conditions for network effects, and we thank

Clough and Wu (2020), Hagiu and Wright (2020), and Cennamo (2020) for triggering us to add this clarification to the debate.

DATA: FROM FIRM RESOURCE TO SHARED

Another key argument presented by Clough and Wu (2020) as a foundation for their ideas about value creation and capture is that the accumulated data derived from the installed base of decentralized users exists internal to the boundaries of the firm within centralized data structures. This argument resonates well within the realm of the resource-based view of the firm in which resources under control of the firm including intangible resources such as data bases serve as potential sources of competitive advantage (Hall, 1993). However, data should not be treated as any other type of resource or production factor such as capital, labor, or oil (Parra-Moyano, Schmedders, & Pentland, 2020). We surmise that Clough and Wu's (2020) argument potentially breaks down when you consider at least three emerging characteristics of data (omitted in Clough and Wu (2020) as well as our own original article). To understand data network effects and their role for value creation and capture, we suggest it is critical to pay closer attention to the following characteristics of data.

First, data are seldom solely strategic resources that exist internal to the boundaries of any individual firm. In a great deal of contexts, they are also a medium of signification and carrier of facts and meanings that serve as a basis for learning and discovery (increasingly via machine learning or combinations of human and machine intelligence, that is, metahuman systems (Lyytinen, Nickerson, & King, 2020)) from which new insight and knowledge can emerge. As Alaimo, Kallinikos, and Aaltonen (2020) explain, viewing data as medium of signification and representation, alongside as resource, helps identifying core qualities of data that are closely associated with value creation processes, helping us further understand the role of AI and data network effects for the creation of user value. Core qualities of data include editability (e.g., through aggregation, filtering, reordering and expansion of data), portability (e.g., through the adoption and diffusion of common standards for structuring and sharing data), and re-contextualizability (e.g., through combination of ground truth data with local domain expertise and knowledge) (Alaimo et al., 2020). By focusing on these inherent qualities of data, we should expect the focus of firms competing on big data, AI, and data network effects to shift away from data control and towards data sharing (Wixom, Sebastian, & Gregory, 2020), supported by AI alignment to manage diverse stakeholder interests (Wixom, Someh, & Gregory, 2020). This leads us to our next point about data ownership versus access.

Second, data do not have to be owned (yet accessed) to learn and improve through the use of AI. With the proliferation of data sharing agreements and widespread adoption and diffusion of standardized interfaces for data exchange, so-called Application Programming Interfaces (APIs), firms are increasingly able to leverage the portability of data to access and create value with data by training unique machine learning models, without necessarily owning and controlling the training data. Once a machine learning model has been trained, it is able to function independently from the training data with which it was fed and developed. Growing availability of open datasets and emerging markets for data with built-in appropriation regimes and guardrails for quality control and data provenance also contribute to the shift away from a focus on exclusive ownership of data which creates significant challenges for privacy protection (Thomas, Leiponen, & Koutroumpis, 2020). In fact, the emergence of data exchanges promises to serve as "platforms that gather data from many different sources and that allow third parties to run algorithms on these data" (Parra-Moyano et al., 2020).

Third, a significant overarching development that will predictably shift the focus further away from internally managing and controlling the data collected from the installed base of users (e.g., Facebook platform) is the ongoing transition from platform to token economy, and the associated shift from data monopolies to data sovereignty (Voshmgir, 2020). With few exceptions (e.g., Cennamo, Marchesi, & Meyer, 2020), the vast majority of existing studies about network effects in the last two decades have been carried out in platform-based settings. As a result, these studies have assumed a centralized, if only virtually, cloud and data storage model as inherent in the business models of platforms such as Facebook, Google, Amazon, and Youtube. The gradual transition from client-server Internet facilitating interactions to the decentralized Internet facilitating agreements and value exchange, enabled by blockchain networks, first started to manifest in the form of peer-to-peer money without banks. The use of blockchain technology, however, is not limited to storage and transfer of financial value, and examples such as Filecoin that seek to incentivize distributed and decentralized data storage highlight that tokenization (tokenization is a special form of digitization and has been described as a method that converts rights to an asset into digital tokens that can be securely bought, sold, and traded on blockchains (Sazandrishvili, 2020)), when applied more broadly to transform economies and markets beyond trying to revolutionize money, can potentially reduce further the data monopolies of the contemporary platform economy that Clough and Wu (2020) assume.

VALUE CREATION AND CAPTURE AS A DUALITY

A significant point made by Clough and Wu (2020) is that data-driven learning (i.e., platform AI capability) enhances the firm's capacity not only to create but also to capture value and, accordingly, "firms (must) strategically decide whether more creation or capture is needed." We concur that the focus on value capture is a crucial addition to the conversation about how firms compete in a data-rich world. However, rather than viewing data-driven value creation and capture in terms of a trade-off, we argue that in this context the strategic balancing act between favoring value creation or value capture needs to be considered holistically in terms of a duality. The proposed model of data network effects in our original article (Gregory et al., 2020) offers a solid foundation to develop such an inquiry.

Clough and Wu (2020) provide several convincing examples wherein digital platforms choose to leverage AI capability to maximize value capture rather than value creation. The authors allude to the role of one of the moderating variable in the data network effects model, *viz.,* data stewardship (i.e., *data quantity and quality*), in enhancing the platform's ability to "manipulate" its users through predatory price discrimination and/or malicious user experience. The example of Electronic Arts (EA) employing data about user behavior to make the unlocking of popular game characters unachievable is particularly telling.

Such behaviors on the part of the platform, as Clough and Wu aptly demonstrate, are often viewed with disdain by users who, at the extreme, may choose to leave the platform. To theorize these unintended outcomes and to better understand the compound effect of data-driven learning on value creation and value capture, the other two moderating factors of the data-network effects model need to be brought into the fold. Incorporating *performance expectancy* (a part of user-friendly design) into our explanatory account, for instance, would suggest that making the unlocking of the game characters unachievable undermines the users' belief that the desired outcomes are attainable, reduces their engagement and, ultimately, may diminish the firm's capacity to capture value. Similarly, considering the effect of *personal data use* (an element of platform legitimation) could help us predict whether maximizing value capture by manipulating psychological state of the users is a good long-term strategy. If deemed "the wrong thing to do" in the backdrop of prevalent moral beliefs, such strategy puts the firm at risk of losing access to vital resources provided by investors, regulators and partners.

As this discussion highlights, firms are increasingly under pressure and have technologies at their disposal (e.g., open digital infrastructures, decentralized data governance, and interoperability standards) to reduce the natural tendency in platform capitalism to decide the tension of data-driven value creation versus capture in favor of the latter. In contrast to Clough and Wu (2020), our original article (Gregory et al., 2020) would suggest that firms embracing new digital technologies and societal demands do not necessarily have to strategically decide whether more creation or capture is needed—value creation and capture can potentially be combined into a duality, which views the two elements as interdependent, rather than separate and opposed. Viewing value creation and capture in terms of a duality shifts the unit of analysis and managerial emphasis from individual actors (the firm serving and exploiting its external customers) towards value exchange within the network of relationships between actors (the firm embedding itself into the network and middle of peer-to-peer transactions), consistent with the rise and gradual diffusion of blockchain technology as one of the foundations for the gradual shift from centralized platforms to decentralized networks as the bedrock of building the new economy (Pentland, 2020).

CONCLUDING REMARKS: BALANCING DIVERSE STAKEHOLDER INTERESTS

As concluding remark, we would like to comment on Clough and Wu's (2020) observation that "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." The model in our original article (Gregory et al., 2020) predicts that this practice is not sustainable and may not be that common for all too long time, not least due to new regulations to be expected. We would like to remind readers of the need and growing awareness among managers to balance diverse stakeholder interests in managing data network effects, in consistency with the view of "stakeholder governance as a process of finding ways to resolve stakeholder conflicts" (Amis, Barney, Mahoney, & Wang, 2020, p. 501). While we do think that the unfolding debate about data network effects has yet to deliver solid answers to this central problem of management in today's economy, the notion of platform legitimation was a central argument in our original article. Firms are predicted by the model in our original article (Gregory et al., 2020) to create (and capture) value only insofar as they make appropriate use of the oftentimes personal data collected on each user. In other words, firms must ensure the moral desirability of the use of personal data, which includes considerations of security and privacy, and ensure explainability of predictions made my machine learning algorithms fed with this data. Without ensuring such appropriate use of personal data, our model predicts that value creation will likely not sustain itself, in turn also taking away the basis for value capture, putting the firm's long-term success at jeopardy.

REFERENCES

Alaimo, C., Kallinikos, J., & Aaltonen, A. 2020. Data and value. In S. Nambisan, K. Lyytinen, & Y. Yoo (Eds.), ***Handbook of Digital Innovation***: 162-178: Edward Elgar Publishing.

Amis, J., Barney, J., Mahoney, J. T., & Wang, H. 2020. From the Editors—Why We Need a Theory of Stakeholder Governance—And Why This is a Hard Problem. ***Academy of Management Review***, 45(3): 499-503.

Cennamo, C. 2020. Value Preserving Platform Regulation: Network Effects, Platform Value, and Regulatory Remedies. ***Digital Markets Competition Forum***.

Cennamo, C., Marchesi, C., & Meyer, T. 2020. Two sides of the same coin? Decentralized versus proprietary blockchains and the performance of digital currencies. ***Academy of Management Discoveries***.

Clough, D. R., & Wu, A. 2020. Artificial Intelligence, Data-Driven Learning, and the Decentralized Structure of Platform Ecosystems. ***Academy of Management Review***, forthcoming.

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***, forthcoming.

Hagiu, A., & Wright, J. 2020. Data-enabled learning, network effects and competitive advantage. ***Working Paper***.

Hall, R. 1993. A Framework Linking Intangible Resources and Capabiliites to Sustainable Competitive. ***Strategic Management Journal***, 14(8): 607-618.

Lyytinen, K., Nickerson, J. V., & King, J. L. 2020. Metahuman systems = humans + machines that learn. ***Journal of Information Technology***.

Parra-Moyano, J., Schmedders, K., & Pentland, A. 2020. Shared Data: Backbone of a New Knowledge Economy. In A. Pentland, A. Lipton, & T. Hardjono (Eds.), ***Building the New Economy***: MIT Press.

Pentland, A. 2020. Building the New Economy: what we need and how to get there. In A. Pentland, A. Lipton, & T. Hardjono (Eds.), ***Building the New Economy***: MIT Press.

Sazandrishvili, G. 2020. Asset tokenization in plain English. ***Journal of Corporate Accounting & Finance***, 31(2): 68-73.

Thomas, L. D. W., Leiponen, A., & Koutroumpis, P. 2020. Markets for data. ***Industrial and Corporate Change***, 29(3): 645-660.

Voshmgir, S. 2020. ***Token Economy: How the Web3 Reinvents the Internet***. Berlin, Germany: BlockchainHub Berlin.

Wixom, B. H., Sebastian, I. M., & Gregory, R. W. 2020. Data Sharing 2.0: New Data Sharing, New Value Creation, ***Research Briefing***. Boston, MA: MIT Sloan Center for Information Systems Research.

Wixom, B. H., Someh, I. A., & Gregory, R. W. 2020. AI Alignment: A New Management Paradigm, ***Research Briefing***. Boston, MA: MIT Sloan Center for Information Systems Research.