Volume Title: The Economics of Artificial Intelligence: An Agenda

Volume Authors/Editors: Ajay Agrawal, Joshua Gans, and Avi Goldfarb, editors

Volume Publisher: University of Chicago Press

Volume ISBNs: 978-0-226-61333-8 (cloth); 978-0-226-61347-5 (electronic)

Volume URL: http://www.nber.org/books/agra-1

Conference Date: September 13–14, 2017

Publication Date: May 2019

Chapter Title: Comment on "Prediction, Judgment, and Complexity: A Theory of Decision-Making and Artificial Intelligence"

Chapter Author(s): Andrea Prat

Chapter URL: http://www.nber.org/chapters/c14022

Chapter pages in book: (p. 110 – 114)

Dimensional Methods and Inference on Structural and Treatment Effects." *Journal of Economic Perspectives* 28 (2): 29–50.
Benzell, Seth G., Laurence J. Kotlikoff, Guillermo LaGarda, and Jeffrey D. Sachs. 2015. "Robots Are Us: Some Economics of Human Replacement." NBER Working Paper no. 20941, Cambridge, MA.
Bolton, P., and A. Faure-Grimaud. 2009. "Thinking Ahead: The Decision Problem." *Review of Economic Studies* 76:1205–38.
Brynjolfsson, Erik, and Andrew McAfee. 2014. *The Second Machine Age*. New York: W. W. Norton.
Dogan, M., and P. Yildirim. 2017. "Man vs. Machine: When Is Automation Inferior to Human Labor?" Unpublished manuscript, The Wharton School of the University of Pennsylvania.
Domingos, Pedro. 2015. *The Master Algorithm.* New York: Basic Books.
Forbes, Silke, and Mara Lederman. 2009. "Adaptation and Vertical Integration in the Airline Industry." *American Economic Review* 99 (5): 1831–49.
Hastie, Trevor, Robert Tibshirani, and Jerome Friedman. 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed. New York: Springer.
Hawkins, Jeff. 2004. *On Intelligence*. New York: Times Books.
Jha, S., and E. J. Topol. 2016. "Adapting to Artificial Intelligence: Radiologists and Pathologists as Information Specialists." *Journal of the American Medical Association* 316 (22): 2353–54.
Lusted, L. B. 1960. "Logical Analysis in Roentgen Diagnosis." *Radiology* 74:178–93.
Markov, John. 2015. *Machines of Loving Grace*. New York: HarperCollins Publishers.
Ng, Andrew. 2016. "What Artificial Intelligence Can and Can't Do Right Now." *Harvard Business Review Online*. Accessed Dec. 8, 2016. https://hbr.org/2016/11/what-artificial-intelligence-can-and-cant-do-right-now.
Simon, H. A. 1951. "A Formal Theory of the Employment Relationship." *Econometrica* 19 (3): 293–305.
Tadelis, S. 2002. "Complexity, Flexibility and the Make-or-Buy Decision." *American Economic Review* 92 (2): 433–37.
Tirole, J. 2009. "Cognition and Incomplete Contracts." *American Economic Review* 99 (1): 265–94.
Varian, Hal R. 2014. "Big Data: New Tricks for Econometrics." *Journal of Economic Perspectives* 28 (2): 3–28.

Comment

Andrea Prat

One of the key activities of organizations is to collect, process, combine, and utilize information (Arrow 1974). A modern corporation exploits the vast amounts of data that it accumulates from marketing, operations, human resources, finance, and other functions to grow faster and be more

Andrea Prat is the Richard Paul Richman Professor of Business at Columbia Business School and professor of economics at Columbia University.
For acknowledgments, sources of research support, and disclosure of the author's material financial relationships, if any, please see http://www.nber.org/chapters/c14022.ack.

productive. This exploitation process depends on the kind of information technology (IT) that is available to the firm. If IT undergoes a revolution, we should expect deep structural changes in the way firms are organized (Milgrom and Roberts 1990).

Agrawal, Gans, and Goldfarb explore the effects that an IT revolution centered on artificial intelligence could have on organizations. Their analysis highlights an insightful distinction between *prediction*, the process of forecasting a state of the world θ given observable information, and *judgment*, the assessment of the effects of the state of the world and the possible action x the organization can take in response to it, namely, the value of the payoff function $u(\theta,x)$.

This is an important point of departure from existing work. Almost all economists—as well as computer scientists and decision scientists—assume that the payoff function $u(\theta,x)$ is known: the decision maker is presumed to have a good sense of how actions and states combine to create outcomes. This assumption, however, is highly unrealistic. The credit card fraud example supplied by the authors is convincing. What is the long-term cost to a bank of approving a fraudulent transaction or labeling a legitimate transaction a suspected fraud?

Organizations can spend resources to improve both their prediction precision and their judgment quality. Agrawal, Gans, and Goldfarb characterize the solution to this optimization problem. Their main result is that, under reasonable assumption, investment in prediction and investment in judgment are complementary (Proposition 2). Investing in prediction makes investment in judgment more beneficial in expected value.

This complementarity suggests that moving from a situation where prediction is prohibitively expensive to one where it is economical should increase the returns to judgment. In this perspective, the AI revolution will lead to an increase in the demand for judgment. However, judgment is an intrinsically different problem—one that cannot be solved through the analysis of big data.

Let me suggest an example. Admissions offices of many universities are turning to AI to choose which applicants to make offers to. Algorithms can be trained on past admissions data. We observe the characteristics of applicants and the grades of past and present students. Leaving aside the censored observations problem arising from the fact that we only see the grades of successful applicants who decide to enroll, we can hope that AI can provide a fairly accurate prediction of an applicant's future grades given his or her observable characteristics. The obvious problem is that we do not know how admitting someone who is likely to get high grades is going to affect the long-term payoff of our university. The latter is a highly complex object that depends on whether our alums become the kind of inspiring, successful, and ethical people that will add to the academic reputation and financial sustainability of our university. There is likely to be a connection between grades and this long-term goal, but we are not sure what it is. In this setting, Agrawal, Gans, and Goldfarb teach us an important lesson. Progress in AI should induce our university leaders to ask deeper questions about the relationship between student quality and the long-term goals of our higher-learning institutions. These questions cannot be answered within AI, but rather with more theory-driven retrospective approaches or perhaps more qualitative methodologies.

As an organizational economist, I am particularly interested in the implications of Agrawal, Gans, and Goldfarb's model for the study of organizations. First, this chapter highlights the importance of the dynamics of decision-making—a seriously underresearched topic. In a complex world, organizations are not going to immediately collect all the information they could possibly need about all possible contingencies they may face. Bolton and Faure-Grimaud (2009), a source of inspiration for Agrawal, Gans, and Goldfarb, model a decision maker who can "think ahead" about future states of the world in yet unrealized states of nature. They show that the typical decision maker does not want to think through a complete action plan, but rather focus on key short- and medium-term decisions. Agrawal, Gans, and Goldfarb show that Bolton and Faure-Grimaud's ideas are highly relevant for understanding how organizations are likely to respond to changes in information technology.

Second, Agrawal, Gans, and Goldfarb also speak to the organizational economics literature on mission. Dewatripont, Jewitt, and Tirole (1999) develop a model where organizational leaders are agents whose type is unknown, as in Holmstrom's (1999) career concerns paradigm. Each agent is assigned a mission, a set of measured variables that are used to evaluate and reward the agent. Dewatripont, Jewitt, and Tirole identify a tension between selecting a simple one-dimensional mission that will provide the agent with a strong incentive to perform well or a "fuzzy" multidimensional mission that will dampen the agent's incentive to work hard but will more closely mirror the true objective of the organization.

This tension is also present in Agrawal, Gans, and Goldfarb's world. Should we give the organization a mission that is close to a pure prediction problem, like admitting students who will get high grades? The pro is that it will be relatively easy to assess the leader's performance. The con is that the outcome may be weakly related to the organization's ultimate objective. Or should we give the organization a mission that also comprises the judgment problem, like furthering the long-term academic reputation of our university? This mission would be more representative of the organization's ultimate objective, but may make it hard to assess our leaders and give them a weak incentive to adopt new prediction technologies. One possible lesson from Agrawal, Gans, and Goldfarb is that, as the cost of adopting AI goes down, the moral hazard problem connected with judgment becomes relatively more important, thus militating in favor of incentive schemes that reward judgment rather than prediction.

Third, Agrawal, Gans, and Goldfarb's section on reliability touches on an important topic. Is it better to have a technology that returns accurate predictions with a low probability or less accurate predictions with a higher probability? The answer to this question depends on the available judgment technology. Better judgment technology increases the marginal benefit of prediction accuracy rather than prediction frequency. More broadly, this type of analysis can guide the design of AI algorithms. Given the mapping between states, actions, and outcomes, and given the cost of various prediction technologies, what prediction technology should the organization select? A general analysis of this question may require using information theoretical concepts, introduced to economics by Sims (2003).

Fourth, Agrawal, Gans, and Goldfarb show that economic theory can make important contributions to the debate over how AI will affect optimal organization. There is a related area where the interaction between economists and computer scientists can be beneficial. Artificial intelligence typically assumes a stable flow of instances. When a bank develops an AI-based system to detect fraud, it assumes that the available data, which is used to build and test the detection algorithm, comes from the same data-generating process as future data on which the algorithm will be applied. However, the underlying data-generating process is not an exogenously given natural phenomenon: it is the output of a set of human beings who are pursuing their own goals, like maximizing the chance of getting their nonfraudulent application accepted or maximizing their chance of defrauding the bank. These sentient creatures will in the long term respond to the fraud-detection algorithm by modifying their application strategy, for instance, by providing different information or by exerting effort to modify the reported variables. This means that the data-generating process will be subject to a structural change and that this change will be endogenous to the fraud-detection algorithm chosen by the bank. A similar phenomenon occurs in the university admission example discussed above: a whole consulting industry is devoted to understanding admissions criteria and advising applicants on how to maximize their success chances. A change in admissions practices is likely to be reflected in the choices that high school students make.

If the data-generating process is endogenous and depends on the prediction technology adopted by the organization, the judgment problem identified by Agrawal, Gans, and Goldfarb becomes even more complex. The organization must evaluate how other agents will respond to changes in the prediction technology. As, by definition, no data is available about not yet realized data-generating processes, the only way to approach this problem is by estimating a structural model that allows other agents to respond to changes in our prediction technology.

In conclusion, Agrawal, Gans, and Goldfarb make a convincing case that the AI revolution should increase the benefit of improving our judgment ability. They also provide us with a tractable yet powerful framework to understand the interaction between prediction and judgment. Future research should focus on further understanding the implications of improvements in prediction technology on the optimal structure of organizations.

References

Arrow, Kenneth. J. 1974. *The Limits of Organization*. New York: W. W. Norton.
Bolton, P., and A. Faure-Grimaud. 2009. "Thinking Ahead: The Decision Problem." *Review of Economic Studies* 76: 1205–38.
Dewatripont, Mathias, Ian Jewitt, and Jean Tirole. 1999. "The Economics of Career Concerns, Part II: Application to Missions and Accountability of Government Agencies." *Review of Economic Studies* 66 (1): 199–21.
Holmstrom, Bengt. 1999. "Managerial Incentive Problems: A Dynamic Perspective." *Review of Economic Studies* 66 (1): 169–82.
Milgrom, Paul, and John Roberts. 1990. "The Economics of Modern Manufacturing: Technology, Strategy, and Organization." *American Economic Review* June: 511–28.
Sims, Christopher. 2003. "Implications of Rational Inattention." *Journal of Monetary Economics* 50 (3): 665–90.