JOURNAL OF BUSINESS LOGISTICS
STRATEGIC SUPPLY CHAIN RESEARCH

Journal of Business Logistics, 2013, 34(4): 249–252

Click Here for a Data Scientist: Big Data, Predictive Analytics, and Theory Development in the Era of a Maker Movement Supply Chain

Matthew A. Waller[1] and Stanley E. Fawcett[2]

[1]*University of Arkansas*
[2]*Weber State University*

Predictive analytics is impacting many diverse areas, ranging from baseball and epidemiology to forecasting and customer relationship management. Manufacturers, retailers, software companies, and consultants are creatively discovering new applications of big data using predictive analytics in supply chain management and logistics. In practice, predictive analytics is generally atheoretical; however, we develop a 2 × 2 model to explain the role of predictive analytics in the theory development process. This 2 × 2 model shows that in our discipline we have traditionally taken one path to theory development, but that predictive analytics can be a salient component of a comprehensive theory development process. The model points to a number of research questions that need to be addressed by our research community. These questions are not just highly relevant to the academic community but also in urgent need of answers to help practitioners execute the right strategies with greater precision and efficiency. We also discuss how one disruptive trend, the maker movement, changes the nature of who the producers are in the supply chain, making big data even more valuable. As we engage in higher levels of dialogue we will be able to make meaningful progress addressing these vital research topics.

Keywords: data science; predictive analytics; big data; data scientist; maker movement; 3D printing; additive manufacturing; theory development; supply chain management; logistics; education; research

He gets on base a lot. Do I care if it's a walk or a hit?—Billy Beane

INTRODUCTION[1]

The epigraph, a quote by Brad Pitt's character Billy Beane in the 2011 movie *Moneyball*, sums up the Oakland A's unconventional approach to building a winning baseball team. As told in the book *Moneyball: The Art of Winning an Unfair Game* by Michael Lewis (2003), the Oakland A's used predictive analytics to win more games with less money. The idea is increasingly familiar. Long-held approaches to doing something may not be the best way to get the job carried out. For the Oakland A's, this meant moving away from traditional metrics like RBIs, batting average, stolen bases, and defense metrics because they are not predictive of success. Out of desperation, the A's used predictive analytics to identify which metrics really made a difference. With this knowledge in hand, the A's build a competitive team—with about a third of the budget of the Yankees in 2002. The A's made the transition from losers to winners and reached the playoffs, dramatically increasing the team's value. The rest of major league baseball soon followed, hiring data scientists to build more competitive teams. The key learning point: The A's did not pick the independent variables ex ante based on theory. They picked the vital decision criteria based on statistical associations among the variables.

Others have shown that the enabling power of leveraging big data with predictive analytics is not a fluke. Consider the following two examples.

- Google sought to predict the timing of flu outbreaks geographically based on search term frequency (Mayer-Schönberger and Cukier 2013). Google's model was completely empirical, not based on theory or even logic. You may be thinking, "Google got lucky. A more theoretical approach in the hands of experts would have worked even better." The Centers for Disease Control and Prevention pursued the more traditional approach, developing models based on biology and epidemiological theory. But, the Google model outperformed the CDC's models, more accurate at predicting H1N1 outbreaks in 2009 (Mayer-Schönberger and Cukier 2013).
- Walmart contributed to folklore among disaster relief experts by studying point-of-sale data from stores that were in areas where hurricanes were imminent. Logic would suggest that facing disaster, people would buy saws, shovels, and safety equipment. But, the data showed that people buy Pop-Tarts, among other things, in unusually large volumes. This was not deduced with the use of a theory but through correlation analysis. This unusual insight turns out to be useful for forecasting and inventory management prior to a hurricane. That is, data-driven insight reduces the cost of lost sales and other logistics costs, in addition to supplying hurricane-stricken areas with sufficient Pop-Tarts and water.

[1]Our earlier editorial, "Data Science, Predictive Analytics, and Big Data: A Revolution that will Transform Supply Chain Design and Management" (Waller and Fawcett 2013), defined data science, predictive analytics, and big data, within the context of logistics and supply chain management (SCM).

The data suggest that correlation can be a leading indicator that leads to both better and more-timely solutions than more traditional causation-based models. Mayer-Schönberger and Cukier (2013) note: "Society will need to shed some of its obsession for causality in exchange for simple correlations: not knowing why but only what. This overturns centuries of established practices and challenges our most basic understanding of how to make decisions and comprehend reality (p. 7). If Mayer-Schönberger and Cukier are right even part of the time—and again, the data suggest they are—our challenge is to understand when is correlation providing better route markers in our quest for knowledge discovery. We thus present a 2 × 2 model to explain the role of predictive analytics in the theory development process.

THEORY DEVELOPMENT AND PREDICTIVE ANALYTICS

Beyond exemplifying uses of predictive analytics (see Waller and Fawcett 2013), each of these stories is an example of an atheoretical approach to problem solving. Theory explains and predicts phenomena. Most academic work in the top journals addressing SCM topics takes an explanatory, theoretical approach. The theoretical framework may come from a wide variety of sources, such as organizational theories, psychological theories, microeconomics, or operations research.

Figure 1 depicts a 2 × 2 matrix with an explanation on the horizontal axis and a prediction on the vertical axis. Explanation and prediction are two key components of good theory. Most of the research that appears in the top supply chain related journals focuses on the high explanation side of theory, though these research studies may be high or low in predictive accuracy. In contrast, predictive analytics focuses on predictive accuracy and deemphasizes explanation. As Mayer-Schönberger and Cukier suggested, getting it right may at times be more important than understanding all of the nuances associated with the underlying relationships.

The ultimate goal of the *Journal of Business Logistics* is to publish high quality theoretical explication (upper right quadrant in Figure 1), including theory building and testing that includes thorough explanation and accurate prediction. Theoretical explication is essential in advancing our academic discipline. However, in order to create the highest level of value at the pace society should expect—especially as new concepts and phenomena emerge—we need to use all of the tools that are available to us, including descriptive exposition and exegetical constructions. We often see such manuscripts. However, we seldom see manuscripts in the upper left quadrant—Predictive Analytics. When we do, authors find it difficult to get their work through the review process because reviewers label the work as atheoretical. That is, these papers either offer no explanation, or, if there is explanation, the explanation is derived post hoc.[2]

Figure 1: Two by two matrix of prediction and explanation.

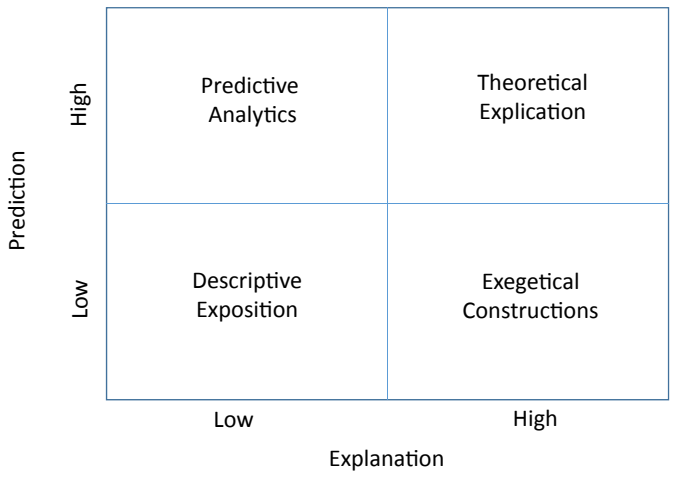

Figure 1 thus presents a conundrum. Emerging analytic tools provide power that exceeds existing explanation. As a result, some research that is cutting edge and makes a very real contribution to our ability to solve real-world problems—a vital goal in our discipline—is not and cannot be theory based—at least not yet. But, despite its lack of a priori theoretical grounding, does such research contribute to theory? Further, do we need to pursue and publish such research? We believe that it does, and we do! Let us explain. Research published about 3 years ago found that public anxiety measured through weblogs could improve predictions of stock market downturns (Gilbert and Karahalios 2010). This work falls into the predictive analytics quadrant of Figure 1. It is primarily atheoretical. Still, the work's nontheoretical findings have contributed to at least 50 other research papers, based on manuscripts that have cited it. It is leading to theory building. We believe this exemplifies how manuscripts in the predictive analytics quadrant can be seminal to theory building in our research community. Research in Quadrant 3 can identify and present new answers and new domains from which better, more relevant theory can be developed over time. Such research can drive much needed dialogue—a goal at the *Journal of Business Logistics* (Fawcett and Waller 2013).

FORWARD THINKING SUBMISSIONS

If recent experience with big data—and the anecdotal evidence—persists going forward, we will discover many new phenomena and answer surprising new questions before we can explain them. That is, we will discover with data that which we cannot explain. We should not be forced to sit on this new knowledge, waiting for an explanation. At the *Journal of Business Logistics*, we have created a place for these data-driven discoveries. Our "Forward Thinking" section of the journal should be used for the upper left quadrant, predictive analytics, based research. An impactful data-driven manuscript shared early through Forward Thinking invites the entire supply chain community to join the dialogue and collaborate in the theory-building process. Consider the benefits: new, relevant insight published on a fast cycle becomes the engine to drive dialogue-leveraged theory development. Together, we can use today's "what" to gain access and insight into tomorrow's "why."

[2] One of the reviewers of this editorial, in a private communication, suggested that predictive analytics research might be more acceptable to our reviewers if the post hoc explanations are tied to specific theories in the discussion section of the manuscripts.

In addition to findings from predictive analytics (well suited for our "Forward Thinking" section), we need theory about predictive analytics in SCM per se. Such manuscripts could be submitted to the journal as a regular submission. Research questions include the following:

- When should predictive analytics be used in SCM?
- When should predictive analytics be avoided in SCM?
- How can predictive analytics be used to advance theory in SCM?
- What makes some firms more likely to use predictive analytics in SCM?
- How does predictive analytics in SCM affect firm performance and under what circumstances?
- How can predictive analytics be used to advance the total cost concept of logistics?

DISRUPTIVE TRENDS FORCE THE ISSUE

Leveraging big data with predictive analytics have proven useful, because they get us closer to definitions and predictions of individual consumer behavior. This reality shows big data's practical, relevant value. This, however, is just the beginning. New technologies and trends are emerging that will change the rules of supply chain design and management. Understanding the uses and implications of big data and predictive analytics will be urgent as additive manufacturing (also called 3D printing) makes traditional models of production, distribution, and demand obsolete in some product areas.

Today, it is easier than ever to design and to manufacture a product at distributed locations. For many parts and situations (e.g., spares and one-of-a-kind), additive technologies are making scale economies irrelevant. Digitization means the first, and every, product out of the printer benefits from quality and efficiency. Further, companies no longer need to invest in dedicated, expensive milling machines or injection molding equipment. A single printer can produce a wide range of parts—without expensive dies and jigs. Consider, for example, the extreme example: Production in lot sizes of one in your home or garage. On the Internet you can find and download digital designs for a vast array of items.[3] And for those things you cannot find, 3D scanners may allow you to develop the digital code needed to go to print/manufacture. You can even modify the designs to meet special, one-of-a-kind needs and then use an affordable 3D printer to "manufacture" the part yourself. This is exactly what Jay Leno does to maintain his personal collection of old and rare automobiles (Koten 2013). Third-party design shops (Fowler 2013) and logistics service providers like UPS are investing in these technologies to bring them to close to consumers (Diakov 2013; UPS Store 2013). Although additive technologies are poorly suited for certain types of product and cannot match the efficiency for mass production, imagine the potential for Caterpillar to re-design its spare parts' supply chain or for the military to support deployed aircraft. The cost/availability equation cannot be ignored.

Drones provide a case in point. Drones can be acquired by the average consumer for a fraction of the cost of supplying similar models for military uses. Drones can be used for business and science applications as well as for entertainment. The consumer can buy a complete drone, buy a kit, buy various components based on an existing design, or design a new one using many easily obtained components.[4] In most of these possible options, the consumer forms an integral part of the production chain, whereas in a traditional model, the consumer is primarily an endpoint. This process in which the consumer becomes part of product design and production is occurring in many industries at once. The trend is called the maker movement (Anderson 2012). The rise of the maker movement leads to a number of research questions:

- How will the maker movement affect supply chain design and operation?
  a. Modes of transportation
  b. Warehousing design and location
  c. Forecasting demand
  d. Evaluation and selection of suppliers of components and raw materials
- Which firms will be most likely to adapt to the maker movement and what capabilities do firms need to succeed?
  a. Knowledge management
  b. Information sharing
  c. Collaboration
- How will adoption of maker movement trends by large firms affect their performance?

We need strong predictive analytics applications and theory because the disintermediation of the traditional supply chain channels means that consumer behavior has become an integral part of both production and demand. Thus, the ability to predict consumer behavior has implications for product innovation, production, distribution, and demand. This is not a futuristic claim, but an observation of existing trends.

CONCLUDING DISCUSSION

Predictive analytics is a hot topic in industry and a growing topic in academics. Predictive analytics can be a part of the theory-building process, even when a given study does not produce or test a specific theory. It can be criticized as being subject to false positives, but even theory testing can be subject to false positives. We would like to see research published in *Journal of Business Logistics* using predictive analytics, including predictive analytics within a theoretical framework, and explaining the role

[3] One of the reviewers of this editorial pointed out that many of the "makers" are from Generation Y and are used to a sharing economy because of their exposure to open sourcing like Linux and R.

[4] A reviewer thought surfboards are a better example than drones. A serious surfer who cannot find the desired surfboard can make a customized surfboard more easily than ever before.

of predictive analytics in the theory-building process. Here is how you can participate:

- Submit manuscripts using predictive analytics to our "Forward Thinking" section
- Submit articles using predictive analytics as the subject of theory testing
- Send us a proposal for a Special Topics Forum on predictive analytics
- Propose a Thought Leader Series topic on predictive analytics

This is the second editorial we have written this year in the *Journal of Business Logistics* (see Waller and Fawcett 2013), calling for research on big data, data science, and predictive analytics in SCM. As researchers and educators, we need to be providing curriculum to our students so that at the very least they can be informed about predictive analytics and newer subjects such as additive manufacturing and the maker movement. In addition, we need to be creating data scientists who can address SCM problems. We may not be able to make SCM data scientists as easily procurable as personal drone components can be with a click of a button, but we can improve our students' competencies in this urgent field. In order to teach these concepts, we need forward-thinking research to create the content for the curriculum. We look forward to your manuscript submissions.

ACKNOWLEDGMENTS

We thank the following individuals for their helpful comments, edits, and input on earlier drafts of this manuscript: Annibal Camara Sodero, Bryan Ashenbaum, Robert Novack, and Christopher Vincent. Their input resulted in a significant improvement to the manuscript.

REFERENCES

Anderson, C. 2012. *Makers: The New Industrial Revolution*. New York: Random House.

Diakov, D. 2013. "UPS May Have Hit Pay Dirt With 3D Printing." *Forbes*. http://www.forbes.com/sites/rakeshsharma/2013/08/19/ups-may-have-hit-pay-dirt-with-3d-printing/

Fawcett, S., and Waller, M. 2013. "Inquiry and the Practice of Theoretical Conversation: Engaging in Dialogue to Elaborate Hidden Connections." *Journal of Business Logistics* 34(1):1–5.

Fowler, G. 2013 "Build a Better Mousetrap—Fast." *Wall Street Journal*, June 10. http://online.wsj.com/news/articles/SB10001424127887324082604578487163888116022

Gilbert, E., and Karahalios, K. 2010. "Widespread Worry and the Stock Market." *ICWSM* May:59–65.

Koten, J. 2013 "Who Says Jay Leno Isn't Cutting Edge?" *Wall Street Journal*, June 10. http://online.wsj.com/news/articles/SB10001424127887324866904578517090180869864

Mayer-Schönberger, V., and Cukier, K. 2013. *Big Data: A Revolution That Will Transform How We Live, Work, and Think*. New York, NY: Eamon Dolan/Houghton Mifflin Harcourt.

UPS Store. 2013. "The UPS Store Makes 3D Printing Accessible to Start-Ups and Small Business Owners." *Wall Street Journal*, July 31. http://online.wsj.com/article/PR-CO-20130731-911842.html#

Waller, M., and Fawcett, S. 2013. "Data Science, Predictive Analytics, and Big Data: A Revolution That Will Transform Supply Chain Design and Management." *Journal of Business Logistics* 34(2):77–84.