Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy

By Cathy O'Neill. Crown Publishing Group. $ 26.00. Hardcover. Pp 259.

Apurv Jain, Microsoft Research.

Cathy O'Neil is a quant, and her experience comes across in her entertaining book "Weapons of Math Destruction." It is an insider's disillusioned tale about the dark side of large scale predictive analytics. These data and models affect many areas of our life from the financial markets where we invest our savings, to selecting educational institutions, the labor market, our consumer behavior such as obtaining insurance and credit, our civic life, and even the prison system. The author cautions us against the religious fervor with which we, as a society, are embracing quantitative models in our quest for efficiency. She shows how models are biased, often built on flimsy understanding, lack a self-correcting feedback loop and accountability, and tend to create their own reality at the expense of the poor and the vulnerable. The author's demands for more transparency in algorithms and the sacrificing of mathematical efficiency in the interest of fairness as a society should be a catalyst for important policy conversations. The title of the book might be an homage to Warren Buffet's 2002 Berkshire annual report where citing flawed models and counterparty concentration risk (among others) he famously called derivatives, financial weapons of mass destruction.

After a brief discussion of her childhood love of mathematics that eventually led her to a tenure track mathematics professor position, the author starts her journey of disillusionment with her experiences at D.E. Shaw, a prominent hedge fund during the financial market crash of 2008. Models failed spectacularly and trades that had seemed like cleverly scouted "free lunches" (arbitrages) turned out to be expensive, with a sudden appearance of risk premia such as previously unseen high borrowing costs. From believing that quants were helping markets be efficient, she goes on to seeing them as contributing to market fragility via bad models. What turns a model from merely a bad representation of reality to a weapon of math destruction (in her terminology) is the opaqueness of inputs to the subjects of the model, vast scale of application, and the absence of a self-correcting feedback loop, which results in models' outcomes only justifying their assumptions rather than improving them.

A good example of the many stories in the book is how the *U.S. News* algorithm that begun as an effort to quantitively capture what made Harvard, Stanford and other established schools elite in the 1980s, resulted in a standardized formula that changed the behavior of the entire university community. The author argues that the *U.S. News'* largely quantitative model encouraged every school to behave more like the incumbent elite schools in very specific, and measurable ways. For example, one of the metrics of being an elite school is student selectivity or the high percentage of applicants rejected. To achieve this goal universities started using expensive consultants to "manage their student populations" and engineer algorithms to reject academically good students who were less likely to matriculate at their institution. These universities also got football teams, expensive student centers, luxury dorms and whirlpool baths. She concedes that some of the imitation probably produced beneficial outcomes by forcing the schools to improve "reasonable" metrics but points out that as an example of unintended consequences, all schools became more expensive at rates four times inflation. The criticism naturally being that the average student had to pay for the whirlpool and the consultants with debt that did not improve her education outcomes. The author suggests that even in this seemingly unbiased system driven by algorithms, wealth creates a significant advantage since it helps afford the experts needed to game the system. Thus, the system is socially inefficient. One could easily and endlessly argue with the author, for instance suggesting that herding around the expensive solutions suggested by the algorithm may not be the only alternative for all colleges, and the causality of the creation of the quantitative model resulting in the entire community changing its behavior would be

Electronic copy available at: https://ssrn.com/abstract=3187660

hard to prove. However, the intuitive appeal of the story is hard to deny. Her point may be made even more compelling by explicitly evaluating the current system against the previous alternative and demonstrating that the previous system achieved better social outcomes.

In sum, the story fulfills its role by making us take a harder look at our society and the data and algorithms that seems to be guiding it. O'Neil recommends transparency- releasing data as well as algorithmic audits, as an antidote to bad algorithms. In the college rankings example, the Education Department released data and software enabling students to select the variables that matter to them, and create their own personal ranking to choose more effectively. Such actions may even create a positive social feedback loop, thus helping individuals and societies become more aware.

The second big theme of the book is the increase in inequality and the threat to democracy due to the exploitation and targeting of the poor and vulnerable using big data and models. Examples cover Payday loans, law enforcement using crime prediction software that evokes the movie Minority Report, and for profit universities such as Corinthian College that target "isolated", "impatient" individuals with "low self-esteem" who were "unable to plan well for the future." O'Neil points out using math only to efficiently exploit the poor may not be the social objective we had in mind. She suggests leaving out some data from the models, thus sacrificing some efficiency in the interest of fairness and equality. Tongue barely in cheek, she provides an example of a feedback loop that could be created by tossing the wealthy citizens of the district of Gold Coast in the city of Chicago, into jail for minor offences, giving the esteemed citizens a taste of the experience of the poor citizens.

The well narrated vignettes in the book combine to form a cautionary tale against blindly trusting models, bring forth the various unintended consequences that large scale applications of models, and ask us if we really believe we are a meritocratic society. People with more resources are the ones who can afford the education to make the models, and the models made by them reflect and shape a reality in which they win.

Recently there has been renewed questioning of our quest of immutable laws in social sciences like the ones in physical sciences. These challenges are not restricted to mathematical ones such as "noisy data" or poorly applied statistics in the form of- say an omitted variable bias that might have an econometric solution, but philosophical issues raised by Karl Popper. In his book "The Poverty of Historicism," Popper points out the theoretical misconceptions in extending the methods of natural sciences to social sciences, as well as the social dangers such as totalitarianism and authoritarianism that receive legitimacy from a belief in the existence of historical laws of social development. Another example of a philosophical challenge is the one raised by the social theory of Reflexivity about the circular relationship between cause and effect. Reflexivity suggests that the very act of observation (and modeling) might change the social system, in contrast to a physical system, where for instance, the laws of motion remain the same whether we observe them or not. Investor George Soros (2008) is a big proponent of Reflexivity in economics, social scientists like Campbell (1976), and Goodhart (1975), have also engaged with some version of the concept. More recently, Andrew Lo and Mark Mueller (2010) also tackle a related idea in their paper on Physics Envy where they extend Knight's (1921) differentiation between risk (randomness fully captured by probability and statistics) and uncertainty (all other types of randomness) to a more nuanced scale with fully reducible uncertainty, partially reducible uncertainty and irreducible uncertainty. They place religion at the end of the spectrum of fully reducible uncertainty and physics and mathematics at the reducible uncertainty. O'Neil's concerns about models creating their own truth seem to echo the scholars above, and an explicit discussion of how their ideas relate to hers would be interesting. On the policy side, providing an example of a system that improved outcomes upon embracing transparency would make this book (or the next) even more powerful.

Electronic copy available at: https://ssrn.com/abstract=3187660

These minor issues aside, the author has succeeded in writing a cautionary tale about the increasing role of models and data in our life, that is informative, understandable and worth our attention. All resource allocation systems have flaws since they are designed by humans, and models using big data are not exempt. The enormous power that these flawed models wield under the guise of science is the danger. So, rather than trying to put the data and model genie back into the bottle, I prefer to pay more attention to the author's calls for algorithmic audits, transparency, and privacy. Let us, as a society, better answer the moral question about efficiency vs. fairness she has confronted us with.

References

Berkshire Hathaway Inc., 2002, *Annual Report*, page 15

Popper, Karl, 1961, "*The Poverty of Historicism*", 2nd edition, Routledge, London.

Thornton, Stephen, "Karl Popper", *The Stanford Encyclopedia of Philosophy* (Winter 2016 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/win2016/entries/popper/>.

Soros, George, 2009, "*The Crash of 2008 and What it Means*," Pereseus Book Group, New York.

Goodhart, Charles A.E., 1975, "Problems of Monetary Management: The U.K. Experience," *Papers in Monetary Economics, vol. 1.*, Reserve Bank of Australia, Sydney.

Campbell, Donald T., "Assessing the Impact of Planned Social Change," Evaluation and Program Planning 2 (1), 67-90.

Lo, Andrew, and Mueller, Mark, 2010, *"Warning!" Physics Envy May Be Hazardous To Your Wealth,"* Journal of Investment Management 8, 13-63.

Knight, F., 1921, *Risk, Uncertainty, and Profit,* Boston, MA, Houghton Mifflin.