AN FDA FOR ALGORITHMS

Andrew Tutt[1]

CONTENTS

Contact Author Before Citation or Circulation

[1] The views expressed in this essay are the author's only and do not necessarily reflect the views of the Department of Justice or the Office of Legal Counsel.

INRODUCTION

May 6, 2010. It was an otherwise sleepy afternoon for the Dow Jones Industrial Average. Stocks were down on bad news out of Europe, but not unusually so. Then at 2:32 pm, prices began falling market wide. Fifteen minutes later, the market was down 7 percent. Between 2:40 and 2:45, $500 billion dollars vanished from the economy.[2] The stocks involved were anomalous; the price swings massive and irrational. CNBC reported that the blue-chip stock Proctor & Gamble—owner of brands like Charmin, Tide, and Crest—was down 24 percent, to $47 a share. Jim Cramer, CNBC's stock market guru immediately exclaimed on air, "Well, if that's true, if that stock is there, you just go and buy it. It can't be there. That is not a real price. Just go buy Proctor."[3] P&G dropped 36% in less than four minutes.[4] Another blue chip, 3M, the maker of Scotch Tape, fell 19 percent.[5] Shares of Accenture, a multi-billion dollar consulting company, at one point fell from $40 to one cent.[6] Meanwhile, nearly a thousand shares of Apple stock were bought for $100,000 a share—a trade so outlandish the exchange refused to honor it.[7] Then, just as suddenly as it began, it was over. By 3 o'clock the market had recovered most of its losses.[8] The event would come to be known as the "Flash Crash."

Fast forward a few months. February 15, 2011. It was the second day of the Jeopardy IBM Challenge, and everything was coming up Watson. IBM's supercomputer cruised into Final Jeopardy with an insurmountable lead. Then, on the final question, Watson made an inexplicable stumble. The category was "U.S. Cities" and the clue: "Its largest airport is named for a World War II hero; its second largest for a World War II battle." The humans both answered correctly: "Chicago." Watson answered "Toronto."[9]

The stories have something in common, and isn't just that they feature algorithms making mistakes. Algorithmic mistakes are old news.[10] Google's image recognition algorithm once labeled photos of people as "Gorillas."[11] Toyota recalled 1.9 million Prius cars because a

[2] *See* Charles R. Korsmo, *High-Frequency Trading: A Regulatory Strategy*, 48 U. RICH. L. REV. 523, 524-25 (2014); *see also* Edward E. Kaufman Jr. & Carl M. Levin, *Preventing the Next Flash Crash*, N.Y. TIMES, May 6, 2011, at A27

[3] Graham Bowley, *U.S. Markets Plunge, Then Stage a Rebound*, N.Y. TIMES, May 6, 2010, at A1; Korsmo, *supra* note 8, at 524-25; *see also* WALLACH, *supra* note 9, at 41.

[4] *See* Staffs of the CFTC & SEC, Findings Regarding the Market Events of May 6, 2010, at 1, 84 (2010), available at https://www.sec.gov/news/studies/2010/marketevents-report.pdf

[5] *Id.* at 85.

[6] *Id.* at 83.

[7] *Id.* at 65, 86.

[8] Korsmo, *supra* note 2, at 524-25.

[9] *See* WENDELL WALLACH, A DANGEROUS MASTER 226-27 (2015); Betsy Cooper, *Judges in Jeopardy!: Could IBM's Watson Beat Courts at Their Own Game?*, 121 YALE L.J. ONLINE 87, 98 (2011); Steve Hamm, *Watson on Jeopardy! Day Two: The Confusion over an Airport Clue*, BUILDING A SMARTER PLANET (Feb. 15, 2011), http://asmarterplanet.com/blog/2011/02/watson-on-jeopardy-day-two-the-confusion-over-an-airport-clue.html

[10] *See, e.g.*, Ed Finn, A*lgorithms Aren't Like Spock*, SLATE, Feb. 26, 2016, http://www.slate.com/articles/technology/future_tense/2016/02/algorithms_are_like_kirk_not_spock.html (describing many examples of algorithmic errors). Some additional examples of possible algorithmic errors can be found in Andrew Tutt, *The New Speech*, 41 HASTINGS CONST. L.Q. 235, 275-78 (2014).

[11] *See* Jessica Guynn, *Google Photos Labeled Black People "Gorillas*," USA TODAY, July 1, 2015, http://www.usatoday.com/story/tech/2015/07/01/google-apologizes-after-photos-identify-black-people-as-

software glitch risked causing their gas-electric hybrid systems to shut down.[12] A "rogue algorithm" once lost Knight Capital $440 million dollars in about 45 minutes.[13]

What Watson's blunder and the Flash Crash have in common is that they both show that when and why complex algorithms fail is increasingly difficult to predict and explain. IBM, on the one hand, and the CFTC and SEC, on the other, both tried to explain the causes of their respective algorithmic lapses. But those explanations were basically guesses.

Watson's fathers at IBM thought Watson might have malfunctioned because Watson doesn't approach problems like humans do. Indeed, in creating Watson, "[t]he IBM team paid little attention to the human brain . . . . Any parallels to the brain are superficial, and only the result of chance."[14] Watson might have said Toronto—when the category was "U.S. Cities"—because Watson knows that "categories only weakly suggest the kind of answer that is expected" and "downgrades their significance."[15] Watson may have been confused because "there are cities named Toronto in the United States" and "the Toronto in Canada has an American League baseball team." Watson may have had trouble linking the names of Chicago's airports to World War II.[16] Maybe it was any of these explanations—or all of them. The truth is Watson's programmers did not really know; nor did they have a ready-made way to "teach" Watson not to make the same mistake again.[17]

Similar guesswork plagued the CFTC and SEC's investigation into the cause of the Flash Crash. At first, the CFTC and SEC were confident they knew its cause. In the months following the crash they explained in a joint report that "an effort by Waddell & Reed to sell some E-mini futures with an inept algorithm" caused it.[18] Five years after the crash, however, the CFTC and SEC revised their theory. It was a rogue trader in London who caused the crash, successfully fooling trading algorithms by placing big fake orders to sell E-mini futures. The trader priced and cancelled orders in such a way as to ensure that he made it look like a lot of people wanted to sell E-mini futures, without ever selling any himself. The practice—known as "spoofing"[19]—is

gorillas/29567465; Alistair Barr, *Google Mistakenly Tags Black People as 'Gorillas,' Showing Limits of Algorithms*, WALL ST. J.: DIGITS (Jul. 1, 2015).

[12] Hiroko Tabuchi & Jaclyn Trop, *Toyota Recalls Newest Priuses Over Software*, N.Y. TIMES, Feb. 12, 2014, at B1.

[13] *See* WALLACH, *supra* note 9, at 24-25; Christopher Steiner, *Knight Capital's Algorithmic Fiasco Won't Be The Last of its Kind*, FORBES: ENTREPRENEURS (Aug. 2, 2012), http://www.forbes.com/sites/christophersteiner/2012/08/02/knight-capitals-algorithmic-fiasco-wont-be-the-last-of-its-kind/#3815c51a462f; Caroline Valetkevitch & Chuck Mikolajczak, *Error By Knight Capital Rips Through Stock Market*, REUTERS, Aug. 1, 2012, http://www.reuters.com/article/us-usa-nyse-tradinghalts-idUSBRE8701BN20120801.

[14] ERIK BRYNJOLFSSON & ANDREW MCAFFE, THE SECOND MACHINE AGE 255 (2014) (quoting Gareth Cook, *Watson, the Computer Jeopardy! Champion, and the Future of Artificial Intelligence*, SCI. AM., Mar. 1, 2011, http://www.scientificamerican.com/article/watson-the-computer-jeopa).

[15] Steve Hamm, *Watson on Jeopardy! Day Two: The Confusion Over an Airport Clue* (Feb. 15, 2011), http://asmarterplanet.com/blog/2011/02/watson-on-jeopardy-day-two-the-confusion-over-an-airport-clue.html.

[16] *Id.*

[17] *Id.*

[18] Matt Levine, *Guy Trading at Home Caused the Flash Crash*, BLOOMBERG VIEW (Apr. 21, 2015), http://www.bloombergview.com/articles/2015-04-21/guy-trading-at-home-caused-the-flash-crash; *see Flash-Crash Story Looks More Like a Fairy Tale*, BLOOMBERG VIEW (May 7, 2012), http://www.bloombergview.com/articles/2012-05-07/flash-crash-story-looks-more-like-a-fairy-tale; *see also* CFTC & SEC REPORT, *supra* note 4,

[19] *See* FRANK PASQUALE, THE BLACK BOX SOCIETY 132 (2015).

illegal, but human traders are rarely fooled. Apparently, algorithms are more vulnerable, because the fake sell orders tipped the algorithms into a trading frenzy that quickly spiraled out of control.[20] The CFTC and SEC's change in theories was a surprise to some. One observer remarked, "I'm dumbfounded that they missed this until now."[21]

* * *

My purpose in this paper is to explain why our inability to understand, explain, or predict algorithmic errors is not only unsurprising but destined to become commonplace. My other aim is to show that dealing with that challenge in the future will require a federal regulatory agency. Making those points will require cutting through diverse legal and technological fields, ranging from the cutting edge of algorithm design, to the legal-policy literature that analyzes the merits of centralized federal regulation, to the history of the Food and Drug Administration. The goal is to show, once the foundation is laid, that a dedicated agency charged with the mission of supervising the development, deployment, and use of algorithms will soon be highly desirable, if not necessary.

This paper is divided into four parts. Part I is a basic primer on algorithms. The primer is meant to bring the reader up to speed on the current trajectory of algorithmic development. It shows that the future of sophisticated algorithms is a future of "taught" rather than "programmed" computer programs. Trained algorithms will increasingly rely on the probabilistically determined interaction between less complex elements to produce emergent behaviors that rival or surpass human cognition in a number of applications. The upshot is that in the future no one will have programmed an algorithm to do anything—rather they will have taken an efficient general-purpose algorithm that can be trained to do a number of tasks and trained it to do something. Importantly, because algorithms do not "think" like humans do, it will soon become surpassingly complicated to deduce how trained algorithms take what they have been taught and use it to produce the outcomes that they do. As a corollary, it will become hard, if not impossible,\ to know when algorithms will fail and what will cause them to do so.

Part II builds on the explanation in Part I to offer some suggestions of the kinds of benefits a regulatory agency could offer. It suggests that an agency could provide a comprehensive means of organizing and classifying algorithms into regulatory categories by their design, complexity, and potential for harm (in both ordinary use and through misuse). It also suggests that such an agency could prevent the introduction of certain algorithms onto the market until their safety and efficacy have been proven through evidence-based pre-market trials. Finally, it suggests that such an agency could impose disclosure requirements and usage restrictions to prevent certain algorithms' harmful misuse.

Part III addresses the legal-policy arguments for regulating algorithms through a centralized

[20] Nonetheless, Watson's programmers were delighted with Watson's answer. Why? Because Watson knew it didn't know. Watson was aware it was unsure of the answer and guessed with very little confidence. Its programmers saw that as a victory. Levine, *supra* note 18; Nathanial Popper & Jenny Anderson, *Trader Arrested in Manipulation That Contributed to 2010 "Flash Crash,"* N.Y. TIMES, Apr. 21, 2015, at A1

[21] Popper & Anderson, *supra* note 20.

federal regulatory agency, rather than leaving such regulation to the States or other federal agencies. Ultimately, the argument is that centralized federal regulation is likelier to be responsive and nuanced. The result would be a regulatory system that strikes a more acceptable balance between regulation and innovation than other approaches offer.

Finally, Part IV turns from algorithms to pharmaceuticals to highlight the analogy between complex algorithms and wonder drugs. Anticipating some objections, it offers a brief history of the FDA, to show that objections to that agency—registered throughout its century-long life—have long been overcome by the public's desire to prevent major public health crises. Analogous concerns are likely to create pressure to regulate algorithms as well.

The paper concludes that, regardless of the path we take, there is now a need to think seriously about the future of algorithms and the unique threats they pose. A piecemeal approach may be incapable of addressing the problems presented by trained algorithms. The paper also includes an appendix with the first cut at a statute creating a National Algorithmic Technology Safety Administration ("NATSA").

I. What "Algorithms" Are and Soon Will Be

A. The Basics

At their most elemental, algorithms are just instructions that can be executed by a computer.[22] Software programs are algorithms running atop algorithms. The computers we interact with each day have a set of extremely basic algorithms known as the BIOS (the "basic input/output system") that carry out the gnomish task of telling the mechanical parts in the computer what to do. Atop those algorithms runs the OS (the "Operating System") that can start other software programs and shut them down. And all the programs we use, from web browsers to word processors, are simply algorithms bundled together to accomplish specific tasks.

Most algorithms are extremely straightforward. The instructions are relatively basic and the outcomes relatively deterministic.[23] The algorithm responds to specific inputs with specific outputs that the programmer anticipated in advance. If something goes wrong, the programmer can go back through the program's instructions to find out why the error occurred and easily correct it.

Many extremely impressive algorithms are basically not much more complicated than that. Take Google's "PageRank Algorithm," the algorithm that made Google the company to beat in search.[24] The algorithm is conceptually quite simple: it determines the rank of a page by

[22] *See* PEDRO DOMINGOS, THE MASTER ALGORITHM 2-6 (2015) ("An algorithm is a sequence of instructions telling a computer what to do."); 1 DONALD E. KNUTH, THE ART OF COMPUTER PROGRAMMING 1-9 (2d ed., 1973). In the 1930s, Alan Turing, Kurt Gödel, and Alonzo Church formalized what it means for a problem to be computable by an algorithm. *See id.* For more on the basics of algorithms, see Jennifer Golbeck, *How to Teach Yourself About Algorithms*, SLATE, Feb. 9, 2016, http://www.slate.com/articles/technology/future_tense/2016/02/how_to_teach_yourself_about_algorithms.single.html & Jacob Brogan, *What's the Deal With Algorithms?*, SLATE, Feb. 2, 2016, http://www.slate.com/articles/technology/future_tense/2016/02/what_is_an_algorithm_an_explainer.html.

[23] *See, e.g.*, DOMINGOS, *supra* note __, at 9.

[24] *See* U.S. Patent No. 6,285,999 (filed Jan. 9, 1998) ("Method for Node Ranking in a Linked Database").

determining how many other webpages link to that page, and then it determines how much to value those links by determining how many pages link to those pages. The revolutionary thing about the PageRank algorithm was not necessarily or even primarily the idea that webpages should be ranked that way, but that Larry Page and Sergey Brin figured out how to write an algorithm that could rank the whole web, at the time 26 million web pages, "in a few hours on a medium size workstation" using a "simple iterative algorithm."[25] PageRank, brilliant as it is, is fairly easy to understand.

Or consider another famous algorithm: Deep Blue, the supercomputer-driven software program that defeated Gary Kasparov in 1997.[26] Deep Blue is conceptually rather simple. On its turn the computer tried its best to make the move that would maximize its chances of winning. To do that, it would hypothesize each of the moves it could make, each of the moves that could be made in response, and so on, out to as many as six to eight moves ahead, and then it would choose the next move based on what would give it the best position several moves down the road. The tough part about programming Deep Blue was figuring out how to know how good a particular board arrangement was. To do that, Deep Blue's programmers came up with over eight thousand different parameters (known as "features") that might be used to determine whether a particular board position was good or bad.[27] Yet, remarkably, "the large majority of the features and weights in the Deep Blue evaluation function were created/tuned by hand."[28] Deep Blue was kind of like a Swiss watch. It ran extremely well, but to make it tell the time its designers had to decide that they were building a watch and then handcraft all the components.[29]

Increasingly, however, algorithms are not "programmed" in the way that PageRank and Deep Blue were programmed. Rather, it would be more apt to say that they are "trained." Put simply, rather than building an algorithm that plays chess very well, programmers are now developing algorithms that can learn to play chess well. That difference will reshape the world.

B. Trained Algorithms

The future of algorithms is algorithms that learn. Such algorithms go by many names, but the most common are "Machine Learning,"[30] "Predictive Analytics,"[31] and "Artificial

[25] *See* Sergey Brin & Lawrence Page, *The Anatomy of a Large-Scale Hypertextual Web Search Engine*, 30 COMPUTER NETWORKS & ISDN SYS. 107 (1998), *available at* http://ilpubs.stanford.edu:8090/361/1/1998-8.pdf.

[26] *See generally* FENG-HSIUNG HSU, BEHIND DEEP BLUE: BUILDING THE COMPUTER THAT DEFEATED THE WORLD CHESS CHAMPION (2002).

[27] MURRAY CAMPBELL, A. JOSEPH HOANE JR., & FENG-HSIUNG HSU, DEEP BLUE 19 (2001), *available at* http://sjeng.org/ftp/deepblue.pdf.

[28] *Id.*

[29] *See* Kunihito Hoki & Tomoyuki Kaneko, *Large-Scale Optimization for Evaluation Functions with Minimax Search*, 49 J. Artificial Intelligence Res. 527, 527 (2014) ("[F]ully automated learning of the heuristic evaluation functions remains a challenging goal in chess variants. For example, developers have reported that the majority of the features and weights in Deep Blue were created/tuned by hand.").

[30] *See* PETER FLACH, MACHINE LEARNING: THE ART AND SCIENCE OF ALGORITHMS THAT MAKE SENSE OF DATA 3 (2012); Harry Surden, *Machine Learning and Law*, 89 WASH. L. REV. 87, 88 (2014).

[31] *See* ERIC SIEGEL, PREDICTIVE ANALYTICS at 3-4, 9 (2013) ("Each of the preceding accomplishments is powered by prediction, which is in turn a product of machine learning").

Intelligence,"[32] although the use of "Intelligent" and its variants can be misleading because it is more important to distinguish between algorithms that learn and algorithms that do not, than it is to distinguish between algorithms that appear intelligent and those that do not. Learning algorithms can be almost impossibly complex, while non-learning algorithms are rarely as difficult to understand. As one author put it, "as of today people can write many programs that computers can't learn," but "more surprisingly, computers can learn programs that people can't write."[33]

Basic machine-learning algorithms are already ubiquitous. How does Google guess whether a search query has been misspelled? Machine learning. How do Amazon and Netflix choose which new products or videos a customer might want to watch? Machine learning.[34] How does Pandora pick songs? Machine learning.[35] How do Twitter and Facebook curate their feeds? Machine learning. How did Obama win reelection in 2012? Machine learning.[36] Even online dating is guided by machine learning.[37] The list goes on and on.[38]

Algorithms that engage in Machine Learning differ fundamentally from other algorithms.[39] Machine-learning algorithms require the programmer to answer a question conceptually different from the question a programmer confronts when building other kinds of algorithms. A programmer designing a typical algorithm for use in a particular task confronts the question: "How can I make this algorithm good at performing this task?" A programmer designing a machine-learning algorithm confronts the question: "How can I make this algorithm good at learning to perform this task?"

Sometimes the two questions are essentially the same. Consider one of the most basic machine-learning algorithms: the Spam Filter. Unwanted emails containing malicious software programs, links to dangerous websites, and advertisements for Viagra are sent to inboxes by the thousands each day. A challenging task is to figure out how to distinguish "spammy" emails from good ones.[40] Even if we think we know what makes an email likely to be spammy—say, it includes an executable attachment or the words "Nigerian Prince"—it would be extremely challenging for a human to figure out precisely how much the inclusion of those things should matter when trying to distinguish spam from legitimate emails. A machine-learning algorithm can automate that task by seeing which emails the humans consider spam, and being told what information in an email might be relevant to deciding on its spamminess, and then calculating for itself the optimal weights to place on each factor that together most accurately determine how to separate spam from other email. At its most abstract, a spam filter could simply be given all of

[32] DOMINGOS, *supra* note __, at xix, 8.

[33] *See* DOMINGOS, *supra* note __, at 6.

[34] *See* SIEGEL, *supra* note __, at 5-9, 142-43; DOMINGOS, *supra* note __, at xi-xxv.

[35] *See* SIEGEL, *supra* note __, at 5-9; DOMINGOS, *supra* note __, at xi-xxv.

[36] *See* SIEGEL, *supra* note __, at 6, 213-217; DOMINGOS, *supra* note __, at 16-17 ("Machine learning was the kingmaker in the 2012 presidential election.").

[37] *See* SIEGEL, *supra* note __, at 5-9; DOMINGOS, *supra* note __, at xi-xxv.

[38] *See* SIEGEL, *supra* note __, at 5-9 (listing dozens of examples of the real-world use of machine learning from predicting mortality and injury rates to decoding from MRI scans what people are thinking, to engaging in automated essay grading).

[39] *See* DOMINGOS, *supra* note __, at xi, 9.

[40] *See* FLACH, *supra* note __, at 1-6.

the information in tens of millions of emails and be told at the outset which are spam and which are not. The algorithm could then decide not only how much *weight* to put on the information in an email, but also *which* information in an email is relevant in the first place. That's how, for example, a machine-learning algorithm can intuit that the inclusion of the word "via6a" is likely to mean an email is spammy without a human needing to tell it so.[41]

That last bit is, in a nutshell, both the promise and peril of the future of machine-learning algorithms. Knowing what in a dataset might be relevant to solving a problem is known to the AI community as extracting a "feature."[42] "In essence, *features* define a language in which we describe the relevant objects in our domain, be they emails or molecules."[43] And traditionally, "[t]he *feature construction* process [has proven] absolutely crucial for the success of a machine-learning algorithm."[44] For example, a programmer might select as the features of an email: the words in the body of the email, the names of the attachments, and the words in the subject line, and leave it to the algorithm to figure out which words are spammy, and how spammy they are. As noted earlier, Deep Blue had more than 8,000 features in its evaluation function, most of them handpicked and hand-weighted.

Algorithms with even basic features can be hugely complex and powerful. Consider what Google was able to do with the H1N1 virus using an algorithm about as complex as a spam filter in combination with the massive Google search database.[45] In 2009, H1N1 was spreading rapidly but because it took a while for ill patients to consult their doctors after infection, the Center for Disease Control (CDC) was only able to track the spread of the disease with a two-week delay.[46] Google unleashed an algorithm that used search terms as the feature, simply looking for correlations between search terms and H1N1 infection rates.[47] The algorithm struck gold, discovering 45 search terms that could be used to predict where H1N1 was in real time, without a two-week lag.[48]

Even Watson has fairly straightforward features. Watson uses the outputs of lots of other algorithms as its features.[49] When Watson is asked a question, it uses natural language processing algorithms to extract keywords, categories, and concepts from the question.[50] Watson combines the outputs of its natural language processing algorithms with information retrieval

[41] *See* VIKTOR MAYER-SCHONBERGER & KENNETH CUKIER, BIG DATA 11 (2013).

[42] *See* FLACH, *supra* note __, at 13, 50; *see also* Solon Barocas & Andrew Selbst, *Big Data's Disparate Impact*, 104 CAL. L. REV. __ (2016) ("Data miners refer to the process of settling on the specific string of input variables as 'feature selection.'").

[43] FLACH, *supra* note __, at 13.

[44] FLACH, *supra* note __, at 41.

[45] The example in the text is adapted from MAYER-SCHONBERGER & CUKIER, *supra* note __, at 1-3.

[46] *Id.*

[47] *Id.*

[48] *See id.* Google published its results in the Journal *Nature*. *See* Jeremy Ginsburg et al., *Detecting Influenza Epidemics Using Search Engine Query Data*, 457 NATURE 1012, 1012-14 (2009).

[49] *See* SIEGEL, *supra* note __, at 165 ("Watson merges a massive amalgam of methodologies. It succeeds by fusing technologies.").

[50] Those natural language algorithms have a certain hard-coded edge to them. For example, Watson had a dedicated algorithm designed extract puns from questions by relying on at least a few *Jeopardy!*-specific quirks (such as the fact that on *Jeopardy!* puns are often set off by quotation marks). As Eric Brown, one of Watson's programmers revealed in one Q&A session about the puns algorithm, "it was probably not done in as general a way as you would like." *See* https://www.youtube.com/watch?v=gcmhXOR7LJQ.

algorithms—similar to the algorithms that power search engines—applied to a massive database (the database included, for example, the entire contents of Wikipedia).[51] Watson was then fed thousands of *Jeopardy!* questions, along with the correct answers, and told to figure out which natural language algorithms, combined with which information retrieval algorithms, maximized the likelihood that Watson would guess a correct answer.[52] The more questions Watson saw, the better Watson got at predicting which combinations of search results were likeliest to be right answers.[53]

But one can easily see that it is miserably difficult, if not impossible, to figure out what Watson will do with a question it has never been asked before. And in the grand scheme, Watson is the algorithmic equivalent of a single-celled organism. It will one day be regarded as little more than a curio. In the future, ultra-sophisticated algorithms unleashed on huge amounts of data with only the vaguest of goals will decide for themselves, based on the data, both what in the data is relevant and how relevant it is.[54] They will be "algorithms that make other algorithms."[55] That is, they will determine the features and weight them. Indeed, it is that development—the development of algorithms that can "extract high-level features from raw sensory data"—that has led "to breakthroughs in computer vision and speech recognition."[56]

To see the difference between old-school machine-learning algorithms and the new-school algorithms, consider "Giraffe," an algorithm that uses deep reinforcement learning to play chess.[57] Deep Blue, like nearly every other chess engine ever made, relied on human chess experts to determine how to evaluate the relative strength or weakness of a particular board arrangement by tweaking the features of the evaluation function. And as any computer programmer would tell you, "almost all improvements in playing strength among the top engines nowadays come from improvements in their respective evaluation functions"[58]—often improvements made by hand.

Giraffe improves its evaluation function by going "beyond weight tuning with hand-designed features," instead using "a learned system [to] perform feature extraction" through "a powerful and highly non-linear universal function approximator which can be tuned to approximate complex functions like the evaluation function in chess."[59] Giraffe's algorithm makes it capable

[51] *See* SIEGEL, *supra* note __, at 151-185

[52] *See, e.g.,* URVESH BHOWAN & D.J. MCCLOSKEY, GENETIC PROGRAMMING FOR FEATURE SELECTION AND QUESTION-ANSWER RANKING IN IBM WATSON 1 (2015) (explaining that Watson "uses ML [machine learning] to rank candidate answers generated by the system in response to an input question using a large extremely heterogeneous feature set derived from many distinct and independently developed NLP [natural language processing] and IR [information retrieval] algorithms"), https://www.researchgate.net/publication/281842058_Genetic_Programming_for_Feature_Selection_and_Question-Answer_Ranking_in_IBM_Watson.

[53] *See* SIEGEL, *supra* note __, at 175.

[54] *See* Quoc V. Le at al., Building High-level Features Using Large Scale Unsupervised Learning, arXiv:1112.6209, at 1 (July 12, 2012), http://arxiv.org/pdf/1112.6209v5.pdf.

[55] DOMINGOS, *supra* note __, at 6; SIEGEL, *supra* note __, at 115 (a "computer [that] is literally programming itself")

[56] Volodymyr Mnih et al., *Playing Atari with Deep Reinforcement Learning*, arXiv:1312.5602, at 1 (Dec. 19, 2013), http://arxiv.org/pdf/1312.5602v1.pdf

[57] *See* Matthew Lai, *Giraffe: Using Deep Reinforcement Learning to Play Chess*, arXiv:1509.01549, at 2, 8-9, 12-13 (Sept. 14, 2015), http://arxiv.org/pdf/1509.01549v2.pdf.

[58] *Id.* at 12.

[59] *Id.* at 15.

of learning from self-play, and the result of training for "72 hours on a machine with 2x10-core Intel Xeon E5-2660 v2 CPU" was the development of an algorithm capable of playing chess "at least comparably to the best expert-designed counterparts in existence today, many of which have been fine-tuned over the course of decades."[60]

The outputs of machine-learning algorithms that engage in their own feature extraction are sometimes almost indistinguishable from magic.[61] A team of researchers was able to use deep reinforcement learning to create a single super-algorithm that could be taught to play more than a half-dozen Atari games using information "it learned from nothing but the video input, the reward and terminal signals, and the set of possible actions—just as a human player would."[62] The trained algorithm surpassed the performance of previous game-specific AIs on six of the seven games and exceeded human expert performance on three of them.[63] Video of the expert algorithm playing the Atari games is stunning.[64]

The development of ever-more-abstract and sophisticated learning algorithms is happening at an accelerating pace. Only a few years ago it was thought that problems like accurate speech recognition, image recognition, machine translation, and self-driving cars, were many years from satisfactory algorithmic solutions. But it is now apparent that learning algorithms can apply extraordinary processing power to immense datasets to achieve results that come close to human-level performance.[65] And as "remarkable" as the growth in machine-learning algorithms is, "it's only a foretaste of what's to come."[66] When "algorithms in the lab make it to the front lines, Bill Gates's remark that a breakthrough in machine-learning would be worth ten Microsofts will seem conservative."[67]

Game-changing breakthroughs will involve combining learning algorithms with other learning algorithms and unimaginable amounts of data to create systems that meet or exceed human performance.[68] Self-driving cars, for example, will combine algorithms that can learn to distinguish objects based on sensory input with algorithms that can use that information to learn how to drive a car.[69] Better-than-human machine translation will come from scaling up the

[60] *Id.* at 25,

[61] *See, e.g.*, DOMINGOS, *supra* note __, at xv (calling them "seemingly magical technologies); *see also* Andrej Karpathy, *The Unreasonable Effectiveness of Recurrent Neural Networks*, ANDREJ KARPATHY BLOG (May 21, 2015), http://karpathy.github.io/2015/05/21/rnn-effectiveness ("There's something magical about Recurrent Neural Networks (RNNs).").

[62] *See* Mnih, et al., *supra* note __, at 2.

[63] *Id*. at 2.

[64] *See* https://www.youtube.com/watch?v=EfGD2qveGdQ.

[65] *See* BRYNJOLFSSON & MCAFFE, *supra* note __, at 14-37 (describing the accelerating sophistication of algorithms in a number of areas once thought to be intractable for computers, predicting that "we're at an inflection point").

[66] DOMINGOS, *supra* note __, at 22.

[67] DOMINGOS, *supra* note __, at 22.

[68] *See* DOMINGOS, *supra* note __, at 7, 15; *see also* Kate Allen, *How a Toronto Professor's Research Revolutionized Artificial Intelligence*, THE STAR, Apr. 17, 2015, http://www.thestar.com/news/world/2015/04/17/how-a-toronto-professors-research-revolutionized-artificial-intelligence.html ("The holy grail is a system that incorporates all these actions equally well: a generally intelligent algorithm. Such a system could understand what we are saying, what we mean by what we say, and then get what we want.").

[69] *See* Alexis C. Madrigal, *The Trick That Makes Google's Self-Driving Cars Work*, THE ATLANTIC, May 15, 2014, http://www.theatlantic.com/technology/archive/2014/05/all-the-world-a-track-the-trick-that-makes-googles-self-driving-

number of sentences used to teach the algorithm from millions to hundreds of billions,[70] relying, for example, on complementary algorithms that can discern similarities between languages to greatly increase the available training data.[71]

The upshot is algorithms are becoming increasingly self-reliant or semi-autonomous. We will soon no longer need (or wish) to provide algorithms with hard-coded hints about how to solve problems. Instead, algorithms will be provided with some basic tools for solving problems and then left to construct for themselves tools to solve intermediate problems, on the way to achieving abstract goals.[72]

Looking twenty to forty years ahead, a fear of many futurists is that we may develop an algorithm capable of recursive self-improvement, i.e. producing learning algorithms more efficient and effective than itself.[73] That development is popularly known in the Artificial Intelligence community as the "singularity."[74] A learning algorithm capable of developing better learning algorithms could rapidly and exponentially improve itself beyond humanity's power to comprehend through methods humans could never hope to understand.[75] Again, however, that development is probably a long way off.

C. Predictability and Explainability

Looking to the more immediate future, the difficulties we confront, as learning algorithms become more sophisticated, are the problems of "predictability" and "explainability."[76] An

cars-work/370871.[70] *See* MAYER-SCHONBERGER & CUKIER, *supra* note __, at 37-39 (explaining how Google's decision to use "95 billion English sentences, albeit of dubious quality" to train its translation algorithm resulted in the most accurate and rich machine-translation algorithm available).

[70] *See* MAYER-SCHONBERGER & CUKIER, *supra* note __, at 37-39 (explaining how Google's decision to use "95 billion English sentences, albeit of dubious quality" to train its translation algorithm resulted in the most accurate and rich machine-translation algorithm available).

[71] *See, e.g.*, Tomas Mikolov, *Exploiting Similarities Among Languages for Machine Translation*, arXiv:1309.4168, at 1 (Sept. 17, 2013), http://arxiv.org/pdf/1309.4168v1.pdf.

[72] *See, e.g.*, MAYER-SCHONBERGER & CUKIER, *supra* note __, at 55-56 (describing how algorithms do not need to be designed with a theory about how they are supposed to make predictions); DOMINGOS, *supra* note __, at 23-26, 40-45.

[73] *See* Eliezer Yudkowsky, *Artifical Intelligence as a Positive and Negative Factor in Global Risk*, *in* CATASTROPHIC GLOBAL RISKS 308, 308-345 (Nick Bostrom & Milan M. Ćirković eds., 2008).

[74] This footnote acknowledges the fact that what type of algorithm would be considered the singularity is disputed. *See Singularity*, LESSWRONG WIKI (Feb. 10, 2014), https://wiki.lesswrong.com/wiki/Singularity. The majority view, however, seems to be that the singularity would be the result of the development of an algorithm that could make itself smarter or otherwise engage in "recursive self-improvement." Initially, however, the concept of the "Singularity" was coined to describe the achievement of "intelligences greater than our own." *See* BRYNJOLFSSON & MCAFFE, *supra* note __, at 254-55 (quoting Vernor Vinge).

[75] Yudkowsky, *supra* note __, at 313-14, 323-28.

[76] *See* Yu Zhang et al., *Plan Explainability and Predictability for Cobots*, arXiv: 1511.08158, at 1 (Nov. 25, 2015), http://arxiv.org/pdf/1511.08158v1.pdf; Ryan Turner, *A Model Explanation System* (2015) http://www.blackboxworkshop.org/pdf/Turner2015_MES.pdf (conference paper); David Barbella et al., *Understanding Support Vector Machine Classifications via a Recommender System-Like Approach* (2009), http://bret-jackson.com/papers/dmin09-svmzen.pdf; Mark G. Core et al., *Building Explainable Artificial Intelligence Systems*, American Association for Artificial Intelligence (2006), https://www.aaai.org/Papers/AAAI/2006/AAAI06-293.pdf; *see also* MAYER-SCHONBERGER & CUKIER, *supra* note __, at 179 ("'Explainability,' as it is called in artificial intelligence circles, is important for us mortals, who tend to want to know why, not just what.").

algorithm's predictability is a measure of how difficult its outputs are to predict; its explainability a measure of how difficult its outputs are to explain.[77] Those problems are familiar to the robotics community, which has long sought to grapple with the concern that robots might misinterpret commands by taking them too literally (i.e. instructed to darken a room, the robot destroys the lightbulbs).[78] Abstract learning algorithms run headlong into that difficulty. Even if we can fully describe what makes them work, the actual mechanisms by which they implement their solutions are likely to remain opaque: difficult to predict and sometimes difficult to explain.[79] And as they become more complex and more autonomous, that difficulty will increase.

Explainability and predictability are not new problems. Technologies that operate on extremely complex systems have long confronted them. Consider pharmaceutical drugs. When companies begin developing those drugs, their hypotheses about why they might prove effective are little better than smart guesses. And even if the drug proves effective for its intended use, it is almost impossible to predict its side effects because the body's biochemistry is so complex. Pfizer was developing Viagra as a treatment for heart disease when it discovered that the drug is actually a far more effective treatment for erectile dysfunction.[80] Rogaine first came to market as Loniten, a drug used to treat high blood pressure before it was discovered that it could regrow hair.[81] Sometimes, once a drug is discovered, its mechanisms (including the reasons for its side effects) can be easily explained; sometimes not. But efficacy and side effects can be very difficult to predict in advance.

Humans are another example of an often unpredictable and inexplicable system. Legal rules, incentives, entitlements, and rights are all designed to change human behavior, but it is sometimes difficult to know in advance whether a given social intervention will be effective, and even if it is effective, whether it will produce unintended consequences.[82] An important difference between machine-learning algorithms and humans, however, is that humans have a built-in advantage when trying to predict and explain human behavior: millions of years of co-evolution. Humans are social creatures whose brains have evolved the capacity to develop theories of mind about other human brains. There is no similar natural edge to intuiting how algorithms will behave.[83]

Determining that an algorithm is sufficiently predictable and explainable to be "safe" is

[77] *See* Zhang et al., *supra* note __, at 1.

[78] *See* Zhang et al., *supra* note __, at 1.

[79] Barbella et al., *supra* note __, at 1 ("Because support vector machines are 'black-box' classifiers, the decisions they make are not always easily explainable. By this we mean that the model produced does not naturally provide any useful intuitive reasons about why a particular point is classified in one class rather than another.").

[80] *See* Naveen Kashyap, *Why Pfizer Won in the United States but Lost in Canada, and the Challenges of Pharmaceutical Industry*, 16 T.M. COOLEY J. PRAC. & CLINICAL L. 189, 202-03 (2014).

[81] *See* John N. Joseph et. al., *Enforcement Related to Off-Label Marketing and Use of Drugs and Devices: Where Have We Been and Where Are We Going?*, 2 J. HEALTH & LIFE SCI. L. 73, 100-01 (2009); *see also* W. Nicholson Price II, *Making Do in Making Drugs: Innovation Policy and Pharmaceutical Manufacturing*, 55 B.C. L. REV. 491, 525 n.229 (2014) (noting that Rogaine's first patented use was as a treatment for high blood pressure).

[82] *See, e.g.*, Samuel Issacharoff & George Loewenstein, *Unintended Consequences of Mandatory Disclosure*, 73 TEX. L. REV. 753, 785-86 (1995) (explaining that changing a certain procedural rule to respond to a problem that emerged in a limited minority of actual cases was likely to have harmful unintended consequences).

[83] *See* Yudkowsky, *supra* note __, at 308-14.

difficult, both from a technical perspective and a public policy perspective. If an algorithm is insufficiently predictable, it could be more dangerous than we know. If an algorithm is insufficiently explainable, it might be difficult to know how to correct its problematic outputs. Indeed, it may be extremely difficult even to know what kinds of outputs are "errors." For example, a few recent articles have made the point that self-driving cars will need to be programmed to intentionally kill people (pedestrians or their occupants) in some situations to minimize overall harm and thereby implement utilitarian ethics.[84] Crash investigators deconstructing a future accident may want to know whether the accident was a result of the "ethics" function or a critical algorithmic error. The self-driving car algorithm's explainability will be crucial to that investigation.

What we know—and what can be known—about how an algorithm works will play vital roles in determining whether it is dangerous or discriminatory.[85] Algorithmic predictability and explainability are hard problems. And they are as much public policy and public safety problems as technical problems. At the moment, however, there is no centralized standards-setting body that decides how much testing should be done, or what other minimum standards machine-learning algorithms should meet, before they are introduced into the broader world. Not only are the methods by which many algorithms operate non-transparent, many are trade secrets.[86]

II. THINGS AN AGENCY COULD SORT OUT

The rising complexity and varied uses of machine-learning algorithms promise to raise a host of challenges when those algorithms harm people. Consider three: (1) algorithmic responsibility will be difficult to measure, (2) algorithmic responsibility will be difficult to trace, and (3) human responsibility will be difficult to assign.[87]

Consider the difficulty of measuring algorithmic responsibility. The problem is many faceted. Algorithms are likely to make decisions that no human would have made in a variety of circumstances no human has confronted or even could confront. Those decisions might be a "bug" or a "feature." Often it will be difficult to know which.[88] A self-driving car might intentionally cause an accident to prevent an even-more-catastrophic collision. A stock trading

[84] *See, e.g.*, Jean-François Bonnefon et al., *Autonomous Vehicles Need Experimental Ethics: Are We Ready for Utilitarian Cars?*, arXiv:1510.03346 (Oct. 12, 2015), http://arxiv.org/pdf/1510.03346v1.pdf; *Why Self-Driving Cars Must Be Programmed to Kill*, MIT TECH. REV., Oct. 22, 2015, https://www.technologyreview.com/s/542626/why-self-driving-cars-must-be-programmed-to-kill.

[85] *See* Barocas & Selbst, *supra* note __.

[86] *See* Frank Pasquale, *Beyond Innovation and Competition: The Need for Qualified Transparency in Internet Intermediaries*, 104 NW. U. L. REV. 105 (2010).

[87] For scholarly articles explaining and addressing some of the issues raised in the paragraphs that follow, see, for example, Jack Boeglin, *The Costs of Self-Driving Cars: Reconciling Freedom and Privacy with Tort Liability in Autonomous Vehicle Regulation*, 17 YALE J. L. & TECH. 171, 186 (2015); F. Patrick Hubbard, *"Sophisticated Robots": Balancing Liability, Regulation, and Innovation*, 66 FLA. L. REV. 1803 (2014); and David C. Vladeck, *Machines Without Principals: Liability Rules and Artificial Intelligence*, 89 WASH. L. REV. 117 (2014). Ryan Calo has also written about this issue with respect to robots and reached a similar conclusion that a federal agency is warranted. *See* RYAN CALO, THE CASE FOR A FEDERAL ROBOTICS COMMISSION (2014). Frank Pasquale has suggested the creation of

[88] *See* Calo, *supra* note __, at 7.

algorithm may make a bad bet on the good faith belief (whatever that means to an algorithm) that a particular security should be bought or sold. The point is, we have a generally workable view of what it means for a person to act negligently or otherwise act in a legally culpable manner, but we have no similarly well-defined conception of what it means for an algorithm to do so.[89]

Next, consider the difficulty of tracing algorithmic harms. Even if algorithms were programmed with specific attention to well-defined legal norms, it could be extremely difficult to know whether the algorithm behaved according to the legal standard or not in any given circumstance. The stock trading algorithm that made the bad bet might have made its decision based solely upon the “signal” in its training data—i.e. the algorithm was right about the circumstance it was confronting, but the event it predicted did not come to pass. Or it might have made its decision based on “noise” in the training data—i.e. the algorithm looked for the wrong thing, in the wrong place. Algorithms that engage in discrimination offer a good example. Suppose a company used a machine-learning algorithm to screen for promising job candidates. That algorithm could end up discriminating on the basis of race, gender, or sexual orientation—but tracing the discrimination to a problem with the algorithm could be nearly impossible. To be sure, the discrimination could be a result of a bug in the design of the training algorithm, or a typo by the programmer, but it could also be because of a problem with the training data, a byproduct of latent society-wide discrimination accidentally channeled into the algorithm, or even no discrimination at all but instead a low-probability event that just happened to be observed.[90]

Finally, consider the difficulty in fixing human responsibility. Algorithms can be sliced-and-diced in a number of ways that many other products are not. A company can sell only an algorithm’s code or even give it away. The algorithm could then be copied, modified, customized, and reused or put to use in a variety of applications its initial author never could have imagined. Figuring out how much responsibility the original developer bears when any particular harm arises down the road will be a difficult question. Or consider a second company that sells training data for use in developing one’s own learning algorithms, but does not sell any algorithms itself. Depending on the algorithm the customer trains, and the use to which the purchaser wishes to put the data, the data’s efficacy could be highly variable, and the responsibility of the data seller could be as well. Or imagine a third company that sells algorithmic services as a package, but the algorithm it offers relies partially or extensively on human interaction when determining its final decisions and outputs (e.g., a stock trading algorithm where a human must confirm all of the proposed trades). Divvying up responsibility between the algorithm and the human is likely to prove complicated.

With those challenges in mind, the following subsections suggest the kinds of issues a federal agency could sort out.

[89] *See* WALLACH, *supra* note __, at 239-43.

[90] *See* Barocas & Selbst, *supra* note __.

A. Acting as a Standards-Setting Body

At its most basic, a federal agency could act as a standards-setting body that coordinates and develops classifications, design standards, and best practices.[91]

Classification. An agency could develop categories for classifying algorithms, varying the level of regulatory scrutiny on the basis of the algorithm's complexity. Under a sufficiently nuanced rubric, the vast majority of algorithms could escape federal scrutiny altogether. For example, the agency could classify algorithms into types based on their predictability, explainability, and general intelligence, but only subject the most opaque, complex, and dangerous types to regulatory scrutiny—thereby leaving untouched the vast majority of algorithms with relatively deterministic and predictable outputs.

Table 1. A Possible Qualitative Scale of Algorithmic Complexity

Algorithm Type	Nickname	Description
Type 0	"White Box"	Algorithm is entirely deterministic (i.e. the algorithm is merely a pre-determined set of instructions)
Type 1	"Grey Box"	Algorithm is non-deterministic but its non-deterministic characteristics are easily predicted and explained.
Type 2	"Black Box"	Algorithm exhibits emergent proprieties making it difficult or impossible to predict or explain its characteristics
Type 3	"Sentient"	Algorithm can pass a Turing Test (i.e. has reached or exceeded human intelligence)
Type 4	"Singularity"	Algorithm is capable of recursive self-improvement (i.e. the algorithm has reached the "singularity")

Performance Standards. An agency could also establish guidance for design, testing, and performance to ensure that algorithms are developed with adequate margins of safety. That guidance, in turn, could be based on knowledge of an algorithm's expected use, types of critical versus acceptable errors it might make, and the suggested predicted legal standard to apply to accidents involving that algorithm.

Table 2. Sample Possible Performance Standards

Algorithm	Performance Standard	Based On
Self-Driving	With 95% statistical confidence, the algorithm	Risk of death and injury

[91] *See* Calo, *supra* note __, at 3-5, 11-12 ("[A]gencies, states, courts, and others are not in conversation with one another. Even the same government entities fail to draw links across similar technologies; drones come up little in discussions of driverless cars despite presenting similar issues of safety, privacy, and psychological unease. Much is lost in this patchwork approach.").

Car (Autonomous)	must be involved in fewer than 1.13 fatal accidents per 100 million vehicle miles, and there must be fewer than 80 injuries per 100 million miles traveled.	per 100 million miles driven in 2012.[92]
Stock Trading Algorithm (Autonomous)	An algorithm's average return volatility must be predicted with 95% confidence based on historical data, and that volatility must be reported to investors	Typical measure of the risk of a security (price volatility)
Job Applicant Screening Algorithm (Autonomous)	With 95% confidence, the pool of favored applicants drawn from a set of applicants must not underrepresent any protected class (based on EEOC guidance) by more than 20%.	The "80% rule" in the Uniform Guidelines on Employee Selection Procedures[93]

<u>Design Standards.</u> An agency could also look into the knotty problem of establishing satisfactory measures of predictability and explainability and promulgate guidance for developing algorithms that meet those standards. Especially with respect to explainability, there is reason to believe that algorithm designers could design machine-learning algorithms with attention to ensuring explainability. For example, through testing they might develop more transparent algorithms that match the performance of black-box algorithms by discovering the hidden features that make the black-box algorithm effective.[94] If explainability can be built into algorithmic design, the presence of a federal standard could nudge companies developing machine-learning algorithms into incorporating explainability from the outset.

<u>Liability Standards.</u> An agency could also make progress toward developing standards for distributing responsibility for harms among coders, implementers, distributors, and end-users. The development of such standards will prove complex and require careful consideration of many factors, including impacts on innovation, compensation for victims, and problems of justice and fairness. An agency could bring together diverse stakeholders—from the open source community, to commercial firms, to customers, to potential victims—to develop flexible guidelines that do not unduly stifle innovation.

B. Acting as a Soft-Touch Regulator

A federal agency could also nudge algorithm designers through soft-touch regulations. That is, it could impose regulations that are low enough cost that they "preserve freedom of choice" and do not substantively limit the kinds of algorithms that can be developed or when or how they

[92] NHTSA, TRAFFIC SAFETY FACTS 2012 (2012), http://www-nrd.nhtsa.dot.gov/Pubs/812032.pdf

[93] *See Uniform Guidelines on Employee Selection Procedures*, 29 C.F.R. § 1607.4D (1997); *The Four-Fifths or Eighty Percent Rule*, 5 Emp. Coord. Employment Practices § 23:28 ("Under the Uniform Guidelines, a test or other selection procedure is generally regarded as having an adverse impact where its selection rate for any race, sex, or ethnic group is less than four-fifths (or 80%) of the rate for the identifiable group with the highest rate.").

[94] *See, e.g.*, sources cited *supra* note __.

can be released.[95]

Transparency. Among the most meaningful soft-touch regulations an agency could impose would be requirements of openness, disclosure, and transparency.[96] There appears to be a growing consensus among scholars that the ability to require transparency should be one of the first tools used to regulate algorithmic safety.[97] Transparency can take many forms and can range from feather light to brick heavy.

Table 3. A Spectrum of Disclosure (reproduced from Frank Pasquale, *The Black Box Society*)[98]

	Depth of Disclosure	Scope of Disclosure	Timing of Disclosure
Preserving Secrecy	Shallow and cursory	To small group of outside experts	Delayed for years or decades
Providing Transparency	Deep and thorough	To the public generally	Immediate

On the lighter end, an agency could require that certain aspects of certain machine-learning algorithms (their code or training data) be certified by third-party organizations, helping to preserve the trade secrecy of those algorithms and their training data. Intermediately, an agency could require that companies using certain machine-learning algorithms provide qualitative disclosures (analogous to SEC disclosures) that do not reveal trade secrets or other technical details about how their algorithms work but nonetheless provide meaningful notice about how the algorithm functions, how effective it is, and what errors it is most likely to make.

On the heavier end, in appropriate circumstances, the agency could require that technical details be disclosed, potentially preempting state-level trade secret protections in the name of public safety. Frank Pasquale has discussed the pros and cons of requiring various kinds of transparency in depth in his book *The Black Box Society*.[99] Without addressing the benefits and drawbacks of striking any particular balance, it is worth emphasizing that the complex tradeoffs between innovation and safety will demand extensive and careful study. An agency could strike that difficult balance in a granular way, by drawing together many stakeholders and mandating only those disclosures that are most appropriate to certain kinds of algorithms used in specific contexts.

[95] *Cf.* Cass R. Sunstein, *Nudges vs. Shoves*, 127 HARV. L. REV. F. 210 (2014) (terming low-cost choice-preserving regulations "nudges"); Cass R. Sunstein, *The Storrs Lectures: Behavioral Economics and Paternalism*, 122 YALE L.J. 1826, 1830-31 (2013) (same); Cass R. Sunstein, *The Ethics of Nudging*, 32 YALE J. ON REG. 413, 414 (2015) (same). For a book-length treatment, see RICHARD H. THALER & CASS R. SUNSTEIN, NUDGE (2008).

[96] *But see* Eric A. Posner, E. Glen Weyl, *An FDA for Financial Innovation: Applying the Insurable Interest Doctrine to Twenty-First-Century Financial Markets*, 107 NW. U. L. REV. 1307, 1355 (2013) (explaining that although disclosure requirements are a "less heavy-handed form[] of regulation" they are "notoriously weak").

[97] *See* MAYER-SCHONBERGER & CUKIER, *supra* note __, at 176-184 (describing recommended accountability mechanisms as disclosure and certifications); PASQUALE, *supra* note __, at 140-88 (offering a detailed account of the types of transparency that could be required and the public policy motivations that might drive particular disclosure solutions).

[98] PASQUALE, *supra* note __, at 142.

[99] *See* PASQUALE, *supra* note __, at 140-88.

C. Acting as a Hard-Edged Regulator

Finally, a federal agency could act as a hard-edged regulator that imposes substantive restrictions on the use of certain kinds of machine-learning algorithms, or even with sufficiently complex and mission-critical algorithms, act as a regulator that requires pre-market approval before algorithms can be deployed.

Pre-Market Approval. Among the most aggressive positions an agency could take would be to require that certain algorithms slated for use in certain applications receive approval from the agency before deployment. That pre-market approval process could provide an opportunity for the agency to require that companies substantiate the safety performance of their algorithms. For example, a self-driving car algorithm could be required to replicate the safety-per-mile of a typical vehicle driven in 2012. The agency could work with an applicant to develop studies that would prove to the agency's satisfaction that the algorithm meets that performance standard. Algorithms could also be conditionally approved subject to usage restrictions—for example, a self-driving car algorithm for cruise control could be approved subject to the condition that it is only approved for highway use. Off-label use of an algorithm, or marketing an unapproved algorithm, could then be subject to legal sanctions.

III. OTHER REGULATORY OPTIONS AND THEIR INADEQUACY

Although the regulation of complex algorithms is inevitable, there are at least two competing alternative regulatory paths that might be pursued other than the creation of a centralized federal agency. One alternative to a federal agency would be regulation State-by-State.[100] In that scenario, most algorithmic regulation could be left to the tort and criminal law systems of the several States, or regulation could be performed by a combination of State-level agency, statutory, tort, and criminal regulation. A second alternative to a single federal agency would be regulation across several agencies regulating algorithms incident to their primary jurisdiction. In that scenario the National High Transportation Safety Administration (NHTSA) would regulate self-driving cars, the Federal Trade Commission (FTC) might regulate the internet, the Federal Aviation Administration (FAA) would regulate drones, etc. Both of those alternatives seem, on balance, to be inferior to regulation through a centralized agency.

A. The Case for State Regulation

A weak case could be made that algorithm regulation should be left to the States to develop. The tort regulatory system has effectively, if imperfectly, dealt with transformational technological change in the past, adapting common-law tort precepts to the problems posed by modern industrial society, perhaps most notably the development of the automobile.[101] The

[100] The suggestion in the text that the two alternatives are state-level regulation or regulation by a federal agency assumes that Congress will not attempt to regulate algorithms through detailed and responsive legislation nor through a kind of federal tort regulatory system. In light of the long practice of Congress in these matters, that seems like a safe assumption.

[101] *See* JOHN FABIAN WITT, THE ACCIDENTAL REPUBLIC (2006); G. Edward White, *The Emergence and Doctrinal Development*

States are famously laboratories of legal innovation,[102] and competition between the States can sometimes produce a race to the top that tends toward optimal legal rules.[103] One could argue that, for those reasons, State-level regulation might prove agile, responsive, and effective.

Moreover, even if one were inclined to think that State-by-State regulation would not be particularly effective, one might nonetheless prefer it because it would be more effective than federal regulation. Federal agencies have been criticized for tending toward three forms of failure: (1) "tunnel vision," in which they do not engage in cost-justified regulation because they are unduly focused on carrying out their narrow mission without attention to broader side effects of regulatory choices,[104] (2) "random agenda selection," in which they tend to focus on high-salience political issues rather than the issues that pose the greatest threat to public safety,[105] and (3) "inconsistency," in which they treat similarly situated risks differently.[106] It might be argued that State-level regulation could better grapple with those sources of failure than a federal agency could, because, for example, State legislatures are more attuned to competing priorities and stakeholders, and so will not as readily fall prey to tunnel vision and inconsistency.

But the case for State-level regulation is really rather weak. An appropriately structured federal agency is as capable of solving the tunnel vision, random agenda selection, and inconsistency problems as the States are. Indeed, the solution offered by those who levy those criticisms of federal regulation is that federal agencies should place a greater premium on expertise and should be more politically insulated.[107] Moreover, the efforts of generalist State judges to adapt common law principles to rapidly evolving technological developments are likely to be fitful, imperfect, and slow. And State legislatures are as susceptible to political capture as federal agencies—sometimes even more so.

Moreover, algorithms pose national problems, and such problems generally call for national solutions. The mobile nature of algorithms makes their regulation a national problem. Most of the technologies in which algorithms are embedded or extensively used are likely to be involved in national commerce—be it because they provide their services through the internet, or because they are embedded in technologies like cars, planes, and drones. Absent a compelling case that algorithmic regulation would lead to a rapid race-to-the-top regulatory effort by the States, the most likely outcome of State-level regulation will be a checkerboard of regulatory efforts, with different standards of safety applicable in different geographic regions. That outcome is likelier to stifle innovation than to promote it. Algorithm designers would probably prefer meeting a single national standard than attempting to figure out how to comply with the State-level

of Tort Law, 1870-1930, 11 U. ST. THOMAS L.J. 463, 465-66 (2014) (describing the emergence of transportation accidents as central to the development of early twenty-first century tort law).

[102] Robert A. Schapiro, *Toward A Theory of Interactive Federalism*, 91 IOWA L. REV. 243, 267 (2005); Michael C. Dorf, *Foreword: The Limits of Socratic Deliberation*, 112 HARV. L. REV. 4, 60-61 (1998) (describing the states-as-laboratories theory).

[103] *See* ROBERTA ROMANO, THE GENIUS OF AMERICAN CORPORATE LAW (1993) (arguing that jurisdictional competition between States for corporate charters produces efficient corporate law).

[104] *See* STEPHEN BREYER, BREAKING THE VICIOUS CIRCLE 10-19 (1993).

[105] *See* BREYER, *supra* note __, at 19-21.

[106] *See* BREYER, *supra* note __, at 21-29.

[107] *See* BREYER, *supra* note __, at 55-81.

standards of fifty jurisdictions.

B. The Case for Federal Regulation By Other Subject-Matter Agencies

It might also be argued that algorithms should not be treated as a single regulatory category but should instead be thought of as a kind of helper technology that should be regulated incident to the regulation of other technologies or fields, such as vehicles, aircraft, and the internet. The argument would be that the bureaucratic burden of imposing double or overlapping regulatory jurisdiction would outweigh the benefit of obtaining the expertise of a single central agency.

But the arguments for a central agency are rather strong. Machine-learning algorithms will pose systematic complex challenges that will transcend the particular technology with which they are associated. The same machine-learning algorithm could one day be deployed to drive a car and fly an airplane. Watson could be used to yield expert guidance in fields ranging from medicine to finance. Placing regulatory jurisdiction in multiple agencies would only make the problems of tunnel vision, random agenda selection, and inconsistency more acute. The same algorithm could be regulated two different ways depending on whether it is deployed in a car or a drone. In addition, lessons learned in developing regulatory solutions for one set of algorithms would not be readily available to other agencies developing solutions to identical or highly similar algorithms.

Moreover, even if other agencies had overlapping jurisdiction, that would not necessarily undermine the case for a single central expert agency. Often two or more agencies share regulatory jurisdiction and work jointly to develop comprehensive regulatory strategies. Thus, a new federal agency in this space could add significant value—in the form of centralized expertise—even if other agencies retained primary jurisdiction over specific technologies.

C. The Case for A Central Federal Agency

The case for regulation by a single expert agency outweighs the case for regulation by the States or jurisdiction distributed across multiple agencies because algorithms have qualities that make centralized federal regulation uniquely appealing.

There are at least three qualities intrinsic to algorithms that make a national regulatory solution warranted—and, in particular, a national regulatory solution that may include pre-market approval requirements for some algorithms.

<u>Complexity.</u> First, the kinds algorithms that are most concerning are by their nature opaque, with benefits and harms that are difficult to quantify without extensive expertise. That feature of the market for algorithms contrasts sharply with the market for most products where individuals are able easily to assess the benefits and safety risks posed. Highly opaque and complex products benefit more from expert evaluation by a regulator than other products.

<u>Opacity.</u> Second, the difficulties with assigning and tracing responsibility for harms to algorithms, and then associating that responsibility with human actors, further distinguish algorithms from other products. Algorithms could commit small but severe long-term harms or may commit grievous errors with low probability. Therefore, unlike many other products for which a combination of tort regulation and reputation will correct for accidents at an acceptable

pace, the market and tort regulatory system are likely to prove too slow to respond to algorithmic harms.

Dangerousness. Third, at least in some circumstances, algorithms are likely to be capable of inflicting unusually grave harm. Whether a machine-learning algorithm is responsible for keeping the power grid operational, assisting in a surgery, or driving a car, an algorithm can pose an immediate and severe threat to human health and welfare in a way many other products simply do not and cannot.

A central regulatory agency with pre-market review would be better able to contend with those problems than the States or an amalgam of subject-matter agencies working independently. Take expertise: To the degree significant expertise is required to understand the possible dangers algorithms pose, a single central regulatory agency is more likely to be able to pool top talent together than are fifty jurisdictions or ten agencies seeking to hire experts to help them make sense of the problem. Take agility and nuance: A single federal regulator could grapple with the dangers algorithms pose holistically rather than piecemeal—effectively distinguishing between algorithms on the basis of stakeholder feedback and expert judgment. A single national agency would be able to maximize the centralized expertise that can be brought to bear on the issue while offering the most agility and flexibility in responding to technological change and developing granular solutions.

D. But What Kind of Agency?

An agency with all of the regulatory powers set out above may be warranted. But many structural and institutional questions remain. For example, whether the agency should have a commission structure (like the SEC and the FTC) or a Director (like NHTSA and the CFPB). Whether the agency should be independent, quasi-independent, or politically accountable, whether the agency's enforcement powers should be internal (by ALJs) or external (through the courts), and whether the agency should be authorized to litigate on its own behalf or rather required to rely on the Department of Justice to implement its enforcement authority.

At this juncture, the Consumer Finance Protection Bureau (CFPB) appears to be the state of the art when it comes to consumer protection agency design. It combines the effectiveness of a single Director with the insulation traditionally afforded a commission-structured agency. It has the full complement of conventional agency powers: rulemaking, enforcement, and adjudication. It can litigate on its own behalf and choose between prosecuting enforcement actions before its own ALJs or in the courts. The CFPB's design has made the agency remarkably nimble, powerful, scalable, and effective. At least at this early stage, the CFPB archetype seems like a good fit for an agency designed to make difficult tradeoffs between innovation and safety in a fast-paced industry.

The Appendix to this Article offers the essentials of a potential organic statute for a possible agency modeled on the FDA and CFPB tentatively dubbed the "National Algorithmic Technology Safety Administration" or "NATSA."

IV. THE FDA MODEL: THE ANALOGY BETWEEN DRUGS AND ALGORITHMS

Many will be skeptical that a new federal agency is warranted. Several arguments could be made. It might be argued that it is too soon to develop a regulator because algorithmic technology is still in its infancy. It might be contended that algorithms are not a species of technology that calls for extensive regulation and oversight. It might be offered that regulation is harmful in principle and that the public benefits most when there are fewer regulations and fewer obstacles to private-sector innovation.

One could try to answer those arguments purely with logic. One might counter, for example, that the exponential pace at which algorithms develop means that we will likely progress from "too soon" to regulate algorithms to "too late" in the blink of an eye without much of a Goldilocks period in between. One might say that algorithms are precisely the kind of technology that calls for federal regulation: opaque, complex, and occasionally dangerous. To the argument that less regulation is always better, it might be pointed out that regulation in one form or another is inevitable, and the true choice is between regulation that is piecemeal, reactive, and slow or regulation that is comprehensive, anticipatory, and technically savvy.

But a better answer than a volume of logic is a page of history. Those objections have long been registered against what is perhaps the world's most popular, effective, and widely emulated regulatory agency: the Food and Drug Administration (FDA).[108] The products the FDA regulates, and particularly the complex pharmaceutical drugs it vets for safety and efficacy, are similar to black-box algorithms. And the crises the FDA has confronted throughout its more than one hundred years in existence are comparable to the kinds of crises one can easily imagine occurring because of dangerous algorithms. The FDA has faced steep resistance at every stage, but its capacity to respond to, and prevent, major health crises has resulted in the agency becoming a fixture of the American institutional landscape.[109] We could draw on the FDA's history, and use it as an opportunity to avoid repeating it.

The FDA was born against a backdrop of public health crisis. Adulterated and misbranded foods and drugs were being sold nationwide, and people were becoming seriously ill.[110] In a political environment heavily resistant to federal regulation of any kind, overwhelming popular sentiment forced the issue to a vote, and the result was the creation of the Food and Drug Act, signed into law in 1906.[111] The law established minimum purity requirements and labelling requirements.[112]

But the law was very limited. All non-narcotic drugs could still be sold by anyone to

[108] *See* PHILIP J. HILTS, PROTECTING AMERICA'S HEALTH xix (2003) ("The FDA . . . is the most known, watched, and imitated of regulatory bodies. Because of its influence outside the United States, it has also been described as the most important regulatory agency in the word."); DANIEL CARPENTER, REPUTATION AND POWER 1-32 (2010) (noting, inter alia, that "[i]n a nation as purportedly anti-bureaucratic as the United States, the FDA's power in the national health system, in the scientific world and in the therapeutic marketplace is odd and telling").

[109] *See* RONALD HAMOWY, MEDICAL DISASTERS AND THE GROWTH OF THE FDA (2010), http://www.independent.org/pdf/policy_reports/2010-02-10-fda.pdf.

[110] HILTS, *supra* note __, at 19-34.

[111] HILTS, *supra* note __, at 52-55.

[112] HILTS, *supra* note __, at 52-55.

anyone.[113] Homebrewed remedies were outside the Act's purview as long as they "didn't contain narcotics or one of a few listed poisons."[114] Public sentiment turned against this state of affairs when severe public health crises rocked the nation.[115]

In the summer of 1937, a prominent Tennessee pharmaceutical manufacturer developed a new medicine by mixing a foul-tasting but effective antibacterial treatment (sulfanilamide) with a somewhat sweet-tasting liquid (diethylene glycol) to make the antibacterial more palatable to children.[116] Shockingly, the company "did not bother to test for toxicity, either in humans or animals."[117] But "diethylene glycol, a chemical customarily employed as an antifreeze, was a deadly poison and known to be such by the FDA."[118] With deaths mounting, the company attempted an informal recall of the product without informing anyone that it was poison.[119] Over one hundred people, most of them children, died before the FDA was able to track down and destroy the remainder of the medicine.[120] Shortly thereafter, Congress passed the Food, Drug, and Cosmetic Act of 1938, vastly expanding the powers of the FDA, including authorizing pre-market review. In the years that followed, every "nation of the developed world would adopt its central principles."[121] "In image and in law the sulfanilamide tragedy of 1937 became an instructive moment whose essential lesson was pre-market clearance authority over new drugs."[122]

Even with greater powers, however, the agency discovered that even pre-market controls were not always enough. Another prominent health crisis led to further refinement of the FDA's regulatory mandate. In the fall of 1960, an American drug manufacturer applied for permission to market Thalidomide in the United States.[123] Thalidomide had been introduced internationally in the 1950s as a non-barbiturate sedative with supposedly few side effects and low toxicity.[124] Testing for new drugs was "still a matter of some discretion for companies, as the law did not specify what was needed,"[125] to prove that a drug was safe and effective. In the case of Thalidomide, clinical testing had been almost comically slipshod, involving no controlled clinical trials or other systematic investigation into the drug's efficacy or side effects.[126]

Before it made its formal application, the American drug maker had already distributed tens of thousands of doses without any FDA oversight because "[u]nder the 1938 law, doctors could experiment on patients with new drugs, in any numbers and with any chemical, so long as they

[113] HAMOWY, *supra* note __, at 5.
[114] HILTS, *supra* note __, at 75.
[115] CARPENTER, *supra* note __, at 73 ("By all accounts the Federal Food, Drug, and Cosmetic Act of 1938 issued from crisis.")
[116] HAMOWY, *supra* note __, at 5-6; HILTS, *supra* note __, at 89-90; CARPENTER, *supra* note __, at 85-88.
[117] HAMOWY, *supra* note __, at 6.
[118] HAMOWY, *supra* note __, at 6.
[119] HAMOWY, *supra* note __, at 6.
[120] HAMOWY, *supra* note __, at 6-7; HILTS, *supra* note __, at 92.
[121] HILTS, *supra* note __, at 93.
[122] CARPENTER, *supra* note __, at 73-74.
[123] HILTS, *supra* note __, at 152.
[124] HAMOWY, *supra* note __, at 10-11
[125] HILTS, *supra* note __, at 150.
[126] HILTS, *supra* note __, at 144-50.

called the work an experiment."[127] The American manufacturer even began to sell the drug as a treatment for nausea for pregnant women, even though it had done zero testing to determine whether the drug was safe and effective for that use.[128]

To its credit, the FDA held up the drug's approval for a year, apparently because the outlandish claims made on its behalf—that it was effective, had low toxicity, and few side effects—did not match the outcomes of even the most cursory animal trials.[129] And during that time the drug's show-stopping side effect emerged—it caused terrible birth defects.[130] Thousands of children were born with severe disabilities worldwide because of Thalidomide, including a handful in the United States because of the drug's experimental use.[131] Shortly thereafter, Congress amended the FDA's statutes to put in place strict rules to ensure that drugs were not tested on humans without safeguards and that clinical trials were strictly controlled to ensure that they adequately determine a drug's efficacy and safety.[132]

The arc of FDA's history shows a relatively stable pattern in which public health crises have caused the American public to expand the FDA's powers to ensure that drugs are proven safe and effective before they reach the marketplace.[133] Given the close analog between complex pharmaceuticals and sophisticated algorithms, leaving algorithms unregulated could lead to the same pattern of crisis and response. That is, unless we learn from the FDA's history and decide to act before those crises occur. Some of the world's largest companies are hoping to transform the way people live and work through the power of algorithms. But the algorithms of the future will operate in ways that we cannot fully understand nor, without carefully controlled trials, reliably predict. We are poised to enter a world where algorithms can cause similarly outsized risks in similarly difficult-to-know ways as pharmaceutical drugs. Rather than wait for an algorithm to harm many people, we might take the FDA's history as a lesson and instead develop an agency now with the capacity to ensure that algorithms are safe and effective for their intended use before they are released.

CONCLUSION

The purpose of this Article was to make an early case for developing a new federal agency whose goal is to ensure that algorithms are safe and effective. Any proposal to introduce legal oversight into an uncharted domain merits careful scrutiny. There are legitimate concerns that regulation stifles innovation and impedes competition. Many who are disposed to free markets may think a federal regulatory agency is too radical and more than is necessary at this early stage. But given the pace of algorithmic progress, it may not be so early. And the unique dangers algorithms pose, coupled with their complexity, make them similar to technologies we have

[127] HILTS, *supra* note __, at 152.

[128] HILTS, *supra* note __, at 149-50.

[129] HILTS, *supra* note __, at 150-53; CARPENTER, *supra* note __, at 243.

[130] HILTS, *supra* note __, at 154-55; CARPENTER, *supra* note __, at 238-40.

[131] HILTS, *supra* note __, at 158.

[132] HILTS, *supra* note __, at 162-65.

[133] *See also* Rebecca S. Eisenberg, *The Role of the FDA in Innovation Policy*, 13 MICH. TELECOMM. & TECH. L. REV. 345, 345-46 (2007).

closely regulated in the past. It may be that the future is here and that the time to treat algorithms as a mature technology deserving of society's watchful eye is now.

APPENDIX: THE NATIONAL ALGORITHMIC TECHNOLOGY SAFETY ADMINISTRATION

Sec. 1. Establishment of the National Algorithmic Technology Safety Administration

There is established an independent Administration to be known as the "National Algorithmic Technology Safety Administration," which shall regulate the development, sale, and use of algorithms.

Sec. 2. Director; Deputy Director; Functions

(a) Director

(1) In general

There is established the position of the Director, who shall serve as the head of the Administration.

(2) Head of Administration

The Director is the head of the Administration and shall have direction, authority, and control over it.

(3) Appointment

The Director shall be appointed by the President, by and with the advice and consent of the Senate.

(4) Functions vested in Director

All functions of all officers, employees, and organizational units of the Administration are vested in the Director.

(b) Deputy Director

There is established the position of Deputy Director, who shall—

(1) be appointed by the Director; and

(2) serve as acting Director in the absence or unavailability of the Director.

(c) Powers of the Director

The Director—

(1) may delegate any of the Director's functions to any officer, employee, or organizational unit of the Administration;

(2) may make contracts, grants, and cooperative agreements and enter into agreements with other executive agencies, as may be necessary and proper to carry out the Director's responsibilities;

(3) may appoint and supervise personnel, establish rules for conducting the general business of the Administration, and direct the establishment and maintenance of divisions or other offices within the Administration;

(4) may prescribe algorithmic safety, design, and compliance standards;

(5) may prevent the introduction or delivery into interstate commerce of any new algorithm or algorithmic implementation unless it has been approved by the Director in a form and manner the Director shall prescribe;

(6) may impose conditions and limitations on the use of algorithms at the Director's discretion; and

(7) may perform such other functions as may be authorized or required by law.

Sec. 3. Algorithmic Technology Advisory Board

(a) Establishment

The Director shall establish an Algorithmic Technology Advisory Board to advise and consult with the Administration in the exercise of its functions and to provide information on emerging practices and technologies relating to algorithms and other relevant information.

(b) Membership

In appointing the members of the Algorithmic Technology Advisory Board, the Director shall seek to assemble experts in appropriate technological fields.

(c) Meetings

The Algorithmic Technology Advisory Board shall meet from time to time at the call of the Director, but, at a minimum, shall meet at least twice in each year.

Sec. 4. Coordination

The Administration shall coordinate with other Federal agencies and State regulators, as appropriate, to promote consistent regulatory treatment of algorithms.

Sec. 5. Rulemaking; Investigation; Adjudication; Enforcement; Penalties

(a) Rulemaking

The Director may prescribe rules and issue orders and guidance, as may be necessary or appropriate to enable the Administration to administer and carry out the purposes and objectives of the Administration and to prevent evasions thereof.

(b) Investigation

The Administration may engage subpoenas, issue interrogators, demand production, and otherwise investigate persons that the Administration has reason to believe have violated rules prescribed by the Director.

(c) Adjudication

The Administration is authorized to conduct hearings and adjudication proceedings with respect to any person in the manner prescribed by Chapter 5 of Title 5 in order to ensure or enforce compliance with any Rules prescribed by the Administration.

(d) Enforcement

(1) In general

If any person violates a rule or final order or condition imposed in writing by the Administration, the Director may commence a civil action against such person to impose a civil penalty or to seek all appropriate legal and equitable relief including a permanent or temporary injunction as permitted by law.

(2) Representation

The Administration may act in its own name and through its own attorneys in any action, suit, or proceeding to enforce this title or to which the Administration is a party.

(e) Penalties

(1) Relief Available

Relief for violations of a rule or final order or condition imposed in writing by the Administration may include: rescission or reformation of contracts; refund of moneys or return of real property; restitution; disgorgement or compensation; payment of damages or other monetary relief; public notification regarding the violation, including the costs of notification; and other money penalties.

(2) Civil Penalties

For any violation of a rule or final order or condition imposed in writing by the Administration, the Director may impose a civil penalty not to exceed $5,000 for each day during which such violation or failure to pay continues.