For Good Measure

Letting Go of the Steering Wheel

DAN GEER

Dan Geer is the CISO for In-Q-Tel and a security researcher with a quantitative bent. He has a long history with the USENIX Association, including officer positions, program committees, etc. dan@geer.org

This issue I will talk about Data, not data, that is to say about capital-D Data as the raw material for the world we are now creating. We'll return to working with a small-d data set next time.

I was trained first as an electrical engineer and then as a biostatistician. From engineering, beyond all else the fundamental lesson is that getting the problem statement right is what determines the future, that if you don't get it right then you end up solving a problem you don't have. From biostatistics, beyond all else the fundamental lesson is that all data has bias: the question is whether you can correct for it, and that correcting for data bias in an imperfect world will itself be imperfect. Combining the two, engineering and biostatistics, one is left with two steering questions: where do you actually want to go and what failure modes can you tolerate?

For some time, security training has been both necessary and widely available. The curve of its sophistication and value has been generally upward. We have better tools, we have better understood practices, and we have more and better colleagues. That's the plus side. But I'm interested in the ratio of skill to challenge, and as far as I can estimate, we are expanding the society-wide attack surface faster than we are expanding our collection of tools, practices, and colleagues. If your country is growing more and more food, that's great. If your population is growing faster than your improvements in food production can keep up, that's bad. As with most decision-making under uncertainty, statistics have a role, particularly ratio statistics that magnify trends so that the latency of feedback from policy changes is more quickly clear. Yet statistics require data.

That cybersecurity is hard will come as no surprise, and it has been four years now since the U.S. National Academy of Sciences concluded that cybersecurity should be seen as an *occupation* and not a *profession* because the rate of change is too great to enable professionalization [8]. That rate of change is why cybersecurity is perhaps the most intellectually demanding occupation on the planet, and it may well be that the hybrid vigor of retreading other professions, other skill sets for cybersecurity practice has been and remains crucial to cybersecurity outcomes rather than a random bit of historical trivia.

Winston Churchill said, "The further back I look, the further forward I can see." Churchill was arguing for looking back centuries so as to discern the patterns of human affairs, to find commonalities within the dynamics of competition at whatever scale fit the age in which those competitions occurred so as to see forward and win the then current competition. But should we measure time in constant units—a day, a week, a month, a year—or should it be something akin to a log scale denoted not by the rate at which the seconds pass on a constant clock but by the number of events that have passed? Does a rapid rate of change mean we only have to look back a littler bit in chronologic time but further back in ever-denser event logs? Or must we look back further still both in time and event counts if we are to damp out the noise of the present?

When I look back to earlier stages of my own career, the principal difficulty of any particular stage has oscillated between getting the problem statement right and picking the failure mode that is tolerable given what data was available on which to make a decision. I don't think that has changed. As of today, data acquisition wouldn't seem to be the problem insofar as instrumentation is cheap and mostly reliable. But data has to be collected with an hypothesis in mind, or, as Charles Darwin said, "All observation must be for or against some view if it is to be of any service." That brings us back to the problem statement, that is to say what problem are you trying to solve and, therefore, what data do you need to collect to have it be of any service?

To repeat, you need to know something about what problem you are trying to solve and what data would help you make the decisions that solve that problem. Over a small number of years, the term "data science" has become commonplace. It seems first to have been used over 50 years ago, but the current usage stems most directly from a 1997 lecture by Jeff Wu with the title "[Does] Statistics=Data Science?" [5]. Wu characterized statistical work as a trilogy of data collection, data modeling, and decision-making. In his lecture's conclusion, he initiated the modern usage of the term "data science" and advocated that statistics be renamed data science and statisticians be renamed data scientists. Those semantics seem to add little clarity to what the collection, modeling, and use of data provide, but argument over terminology is a hallmark of how a science develops.

But Darwin's remark that all observation must be for or against some view is not quite right, at least not quite right for us here. It is not so simple as Wu's data collection, data modeling, and decision-making, either. When you collect data and with it build a model, your goal, your problem statement, matters. If your purpose in building a model is to come to a definitive conclusion about causality, about how nature works, then you are saying that the inputs to your model and the coefficients that calibrate their influence within your model are what matters in the final analysis. Parsimony in the sense of Occam's Razor is your judge, or, as Antoine de Saint-Exupéry put it, "You know you have achieved perfection in design, not when you have nothing more to add, but when you have nothing more to take away." So it is when you are chasing causality.

By contrast, when your purpose in building a model is to enable control of some process or other, then you will not mind if your input variables are correlated or redundant—their correlation and their redundancy are not an issue if your goal is to direct action rather than to explain causality.

In some circumstances you can do both, that is you can both explain causality and enable control. In those situations, it is your model's ability to predict that both satisfies the reader that you have captured a causal relationship and that operationalizes the model's predictions irrespective of any underlying truths [7]. A goal of understanding causality in its full elegance leads to $F=ma$ or $E=mc^2$. A goal of control leads to econometric models with thousands of input variables each of whose individual contribution is neither clear nor relevant.

Consider anomaly detection and its role in current cybersecurity products. Anomaly detection presumes something about distributions of detectible events, namely that within a selected interval anything outside some bounding box is worth investigation. It is not concerned with causality; it is concerned with control irrespective of causality. This is a coherent strategy, though with side effects.

Or consider "Big Data" and deep learning. Even if Moore's Law remains forever valid, there will never be enough computing, and hence data-driven algorithms must favor efficiency above all else as data volume grows. Yet the more efficient the algorithm, the less interrogatable it is, that is to say that the more optimized the algorithm is, the harder it is to know what the algorithm is really doing. That was the exact theme of a workshop held in New York by Morgan Stanley and the Santa Fe Institute three Octobers ago titled, "Are Optimality and Efficiency the Enemies of Robustness and Resilience?"

The more desirable some particular automation is judged to be, the more data it is given. The more data it is given, the more its data utilization efficiency matters. The more its data utilization efficiency matters, the more its algorithms will evolve to opaque operation. Above some threshold of dependence on such an algorithm in practice, there can be no going back. As such, preserving algorithm interrogatability despite efficiency-seeking, self-driven evolution is the pinnacle research-grade problem that is now on the table, and I mean for all of cybersecurity. If science does not pick up this challenge, then Larry Lessig's characterization of code as law is fulfilled. A couple of other law professors have seized on that very idea and suggested that price-fixing collusion among robots will be harder to detect than collusion among people [12].

The point is this: if we choose control as the purpose of our efforts, then we will have to let causality become harder to see because our models will submerge any causal relationships in a thicket of confounding. If, instead, we focus on causality, the very things that we need to measure become harder to get if, for no other reason, our sentient opponents will make it so. I'm for measurement as decision support, i.e., I am in the control camp, not the causality camp. At the same time, I very much do demand that I be able to ask some algorithm, "Why did you do that?" and get a meaningful answer. Overall, having both control and interrogatability is a difficult problem to say the least.

And that may be the most important thing I have to say here, that the real problem statement is not about cybersecurity per se but about the side effects of our pursuit of it. Some years ago, in a lecture at Harvard's Kennedy School of Government, the speaker listed the Four Verities of Government as:

- Most exciting issues are not important.
- Most important issues are not exciting.
- Not every problem has a good solution.
- Every solution has side effects.

I think that those are the verities of cybersecurity, too. The din of press coverage of cybersecurity is only about the exciting failures, not the important successes nor that even more important trendline for the ratio of skill to challenge. Perhaps this simply reinforces Donald Knuth's remark that "Premature optimization is the root of all evil." Perhaps it is simply that evolution in our digital world follows the same patterns as evolution in the natural world.

If that is so, then what we see in Nature is what we should expect to see in cybersecurity. Well, in Nature there are two alternative games for survival, r-selection and K-selection [2]. R-selected species produce many offspring, each of whom has a relatively low probability of surviving to adulthood, while K-selected species are strong competitors in crowded niches. K-selected species invest more heavily in much fewer offspring, each of whom has a relatively high probability of surviving to adulthood. If we change the term from "produce many offspring" to "re-image frequently" you now have precisely the world of VMs. Or, to be more current still, you have the kind of components in a DevOps setting where it is arguable whether moving target defense or minimizing new product introduction latency is the paramount goal or value.

Stephen Jay Gould's idea of punctuated equilibrium [4] as the fundamental cadence of evolution has a hold on me. In his formulation, long periods of stasis are the norm. In computing, we would call that "legacy." I trace the birth of the cybersecurity industry to Microsoft's introduction of a TCP/IP stack as a freebie in the Windows 95 platform, thereby taking an operating system designed for a single owner/operator on a private net, if any, and connecting it to a world where every sociopath is your next door neighbor. That event was the birth of our industry, though the fact was unnoticed at the time.

The second of these punctuations occurred around a decade ago when our principal opponents changed over from adventurers and braggarts to professionals. From there on, mercenaries, some armed with zero-days, dominated. The defense response has been varied, but the rise of bug-bounty programs and software analysis companies are the most obvious. An Internet of Things (IoT) with a compound annual growth rate of 35% will be like anabolic steroids for at least those two.

In August 2016, we passed a third such punctuation. The DARPA Cyber Grand Challenge [1] showed that what has heretofore required human experts will shortly come within the ken of fully automatic programs, or, shall we say, algorithms that are today at the upper level of skill, with intelligence, per se, soon to follow. As with both of the previous two punctuations, the effects of the third will reverberate for the indefinite future. I have long argued that all security technologies are dual use, and the day after the Cyber Grand Challenge, Mike Walker, its DARPA program manager, said as much: "I cannot change the reality that all security tools are dual-use."

And everywhere the talk is about "Big Data" and how much better an instrumented society will be. The cumulative sum of the curves for computing, storage, and bandwidth is this: in 1986 you could fill the world's total storage using the world's total bandwidth in two days. Today, it would probably take nine months of the world's total bandwidth to fill the world's total storage [6], but because of replication, synchronization, and sensor-driven autonomy, it is no longer really possible to know how much data there is. Decision-making that depends or depended on knowing how much data there is is over.

In other words, whatever the future holds, it is clear that it will be data rich and that the tools acting on it will be dual use. The classic triad of cybersecurity has long been confidentiality, integrity, and availability, and we have heretofore prioritized confidentiality, especially in the military sector. That will not be the case going forward, and not just because the rising generations have a relaxed complacency about the tradeoffs in information sharing between what it enables and what it disables. In the civilian sector, integrity will supplant confidentiality as the pinnacle goal of cybersecurity. In the military sector, weapons against data integrity already far surpass weapons against data confidentiality.

This trend—the eclipse of confidentiality by integrity and availability—is solidly entrenched now. Already algorithms learn rather than being taught. What they learn depends on how their learning is scored. This is behavioral reinforcement of a form that would be entirely familiar to B. F. Skinner—you don't teach the subject the desired behavior, you reward the subject for exhibiting the desired behavior. You don't look into the mind of the human subject nor into the structure of the self-modifying algorithm, you just look at the objective reality of behavior itself. This is not so much our creation of an intelligence but an unforced assumption that an intelligence will appear if given enough training sets.

But because of how that kind of learning works, it can be fooled by data as easily as it can be enriched by it. In a 2013 paper, Szegedy et al. found that

> [D]eep neural networks learn input-output mappings that are fairly discontinuous to a significant extent. We can cause the network to misclassify an image by applying a certain imperceptible perturbation, which is found by maximizing the network's prediction error. In addition, the specific nature of these perturbations is not a random artifact of learning: the same perturbation can cause a different network, that was trained on a different subset of the dataset, to misclassify the same input [10].

Not even a year ago, researchers in France and Switzerland found that

> Given a state-of-the-art deep neural network classifier, we show the existence of a *universal* (image-agnostic) and very small perturbation vector that causes natural images to be misclassified with high probability. We propose a systematic algorithm for computing universal perturbations, and show that state-of-the-art deep neural networks are highly vulnerable to such perturbations, albeit being quasi-imperceptible to the human eye. We further empirically analyze these universal perturbations and show, in particular, that they generalize very well across neural networks. The surprising existence of universal perturbations reveals important geometric correlations among the high-dimensional decision boundary of classifiers. It further outlines potential security breaches with the existence of single directions in the input space that adversaries can possibly exploit to break a classifier on most natural images [11].

Note to reader: look up "adversarial perturbations."

So why am I making a point about image classification by deep neural networks? Because it raises the fundamental question: given data richness and self-modifying algorithms becoming ever more prevalent in the cybersecurity regime, is keeping a human in the loop a liability or a failsafe? I've already written that the central requirement for security is keeping a human in the loop, that of interrogatability. But there are others, not the least of which is reaction time.

Nevertheless, as data volume grows it creates a challenge far beyond the parlor exercise of how long would it take to fill all the world's storage with all the world's bandwidth, yet it is bandwidth itself that is a limiting coefficient. It is safe to predict that the F-35 will be the last manned fighter plane; drone fleets make more sense going forward. Those drone fleets require ever more massive compute power handling, ever more massive data flows. Lt. Colonel Rhett Hierlmeier heads up the training center for the F-35. He believes that what is today a training simulator will tomorrow be a control point, not a simulator. *Popular Science*'s interview with him includes this telling snippet: "Standing outside the cockpit, he peers into the darkened dome, and says he believes we will one day fight our enemies from inside one of these things. When I ask what that will take, he says flatly, 'Bandwidth'" [9]. Just that point about bandwidth is why "engineers are focused on things like improving artificial intelligence so planes can act with more autonomy, thus cutting down on communication bandwidth."

And the same thing will apply in our field. My estimate is that the Internet of Things has a 35% compound annual growth rate. If I am approximately correct, then IoT growth is already outdistancing the growth rate for installed bandwidth, and for us as much as for fighter pilots the pressure for autonomy is and will be driven by the data-sensing capacity of a rapidly increasing installed base.

Let me therefore suggest that when sentience is available, automation will increase risk, whereas when sentience is not available, automation can reduce risk. Note that parsing, that replacing available sentience with something that is not sentient *will* increase risk but that substituting automation for whatever you have absent sentience *can* make things better. It won't do so necessarily, but it can.

As a child of the hillbilly South, I have nothing against automating away drudgery; a 110-year-old woman interviewed for the book *Supercentenarians* was asked what was the most important invention during her lifetime. Her answer was the washing machine. But with the spread of computers, we have tended to use automation as soon as it is cheaper than human labor. No single replacement of labor by automation matters, but the sum of it does. Yet as we sit here today, the equation of automation is not that of eliminating drudgery but eliminating the need for sentience. Is there enough available sentience to indict cybersecurity automation as risk creating or, alternately, is there far too little sentience that is up to the task at hand and therefore automation is essential and risk reducing?

The embedded systems space has long since made the attack surface of the non-embedded space trivial by comparison. It was two years ago when the count of networked devices exceeded the count of human beings [3]. Qualcomm's Swarm Lab at UC Berkeley predicts 1000 radios per human by 2025, while Pete Diamandis' *Abundance* calls for $45x10^{12}$ networked sensors by 2035. These kinds of scale cannot be supervised, they can only be deployed and left to free-run. If any of this free-running is self-modifying, the concept of attack surface is just plain over as is the concept of trustworthy computing, at least as those are presently understood. This will echo John McAfee's April 2017 interview in *Newsweek*: "Any logical structure that humans can conceive will be susceptible to hacking, and the more complex the structure, the more certain that it can be hacked."

So the situation with data in cybersecurity is richly complex. We need ever more of it if we are to capture increasingly subtle attack vectors, and especially so if we want autonomous, learning-capable algorithms that need to be faster than we are or which don't have the bandwidth to tell us what they are seeing. Yet the more important the decision to be made, the more vital it is to keep a human in the loop.

That is a tall problem statement, and to go with it we need to carefully consider what the tolerable failure modes are. Do we want to trust no sensor data that can't be corroborated? Do we want to accept algorithms as better managers than we are even when we can't tell how it is that they do what they do? Do we want to keep humans in the loop and, if so, how do we protect their legal culpability when it can be shown that some algorithm would not have made mistakes as costly as the ones the human made?

I urge you to take in data that you have some feel for, that is to say for which you have at least some calibrated understanding such that your presence in the loop is prima facie meaningful. If you are designing algorithms, work hard on making them interrogatable. If you are of necessity relying on self-modifying algorithms, let Santayana remind you that "Skepticism is the chastity of the intellect." Remember that all data has bias and that that, too, is in the equation for what failure modes you can tolerate.

References

[1] DARPA Cyber Grand Challenge, August 4, 2016: http://archive.darpa.mil/cybergrandchallenge/.

[2] E. Pianka, "On r and K Selection," *American Naturalist,* vol. 102 (1970), pp. 592–597: http://bit.ly/2fmrZf8.

[3] D. Geer, "For Good Measure: Implications of the IoT," *;login:,* vol. 41, no. 4 (December 2016): geer.tinho.net/fgm/fgm.geer.1612.pdf.

[4] N. Eldredge and S. J. Gould, "Punctuated Equilibria: An Alternative to Phyletic Gradualism," in *Models in Paleobiology* (Freeman Cooper, 1972), pp. 82-115: www.blackwellpublishing.com/ridley/classictexts/eldredge.asp.

[5] www2.isye.gatech.edu/~jeffwu/presentations/datascience.pdf as drawn from Wikipedia, "Data Science": en.wikipedia.org/wiki/Data_science.

[6] M. Hilbert, "World's Information Capacity PPTS": www.martinhilbert.net/WorldInfoCapacityPPT.html (reflecting M. Hilbert & P. Lopez, *Science*, vol. 332, no. 6025 (2011), pp. 60–65) extrapolated with concurrence of its author; see also http://bit.ly/2hmHArN.

[7] "You see, there is only one constant, one universal. It is the only real truth. Causality."—Merovingian in *The Matrix Reloaded*.

[8] National Research Council, "Professionalizing the Nation's Cybersecurity Workforce? Criteria for Decision-Making": www.nap.edu/openbook.php?record_id=18446.

[9] K. Gray, "The Last Fighter Pilot," *Popular Science*, December 22, 2015: www.popsci.com/last-fighter-pilot.

[10] C. Szegedy, W. Zaremba, I. Sutskever, J. Bruna, D. Erhan, I. Goodfellow, R. Fergus, "Intriguing Properties of Neural Networks," February 2014: arxiv.org/pdf/1312.6199.pdf.

[11] S.-M. Moosavi-Dezfooli, A. Fawzi, O. Fawzi, P. Frossard, "Universal Adversarial Perturbations," October 2016: arxiv.org/pdf/1610.08401v1.pdf.

[12] S. Farro, "When Robots Collude: Computers Are Adopting a Legally Questionable Means to Crush the Competition" (algorithms can learn to do so), *Business Insider*, April 28, 2015: http://read.bi/2feNmuW.