RESPONSE

A Mild Defense of Our New Machine Overlords

*Andrew D. Selbst**

INTRODUCTION: A RESPONSE BRIEF ON BEHALF OF THE MACHINES

In *Plausible Cause*,[1] Kiel Brennan-Marquez takes on a pair of vexing questions in Fourth Amendment jurisprudence: Why does "probable cause" seem to have so little to do with probability? And if not probability, what does it stand for? Brennan-Marquez demonstrates convincingly that despite the name, probable cause was never about simple probability. He shows that the Supreme Court has time and again refused to assign a number to the probability required by probable cause,[2] and that something else must be driving the results.

* Visiting Fellow, Yale Information Society Project; Visiting Researcher, Georgetown University Law Center. J.D. 2011, University of Michigan. Thanks to Jane Bambauer, Solon Barocas, Kiel Brennan-Marquez, and Kathy Strandburg for help with earlier drafts, and to the editors of Vanderbilt Law Review En Banc for their editing assistance.

1. Kiel Brennan-Marquez, *Plausible Cause*, 70 VAND. L. REV. 1249 (2017).
2. *Id.* at 1265–73.

In probability's place, Brennan-Marquez offers "plausibility," an "epistemologically distinct" concept that has to do with "explanatory power."[3] He argues that probable cause is a determination that, given "an observer's understanding of the world,"[4] guilt is a better explanation of the evidence in total than innocence. "[T]he police must be able to explain why observed facts give rise to the inference."[5] He traces Supreme Court cases, arguing that the Justices seek to determine whether explanations that imply guilt are more likely than explanations that imply innocence.[6]

As a matter of Fourth Amendment analysis, he makes a convincing case. There really is something other than probability going on, and explanations are an important part of it. But the Fourth Amendment is not his ultimate goal, and beyond the basic Fourth Amendment analysis, he makes "a more general argument about explanatory standards and judicial review in the age of powerful machines."[7] He argues that explanations are required to provide the legal system with the authority that liberal democracies imbue it with, and further, that machines, which ultimately produce numerical outputs, are incapable of such explanations. The authority, he argues, can only come from humans interacting with other humans' narratives, with machines involved only occasionally, as tools.[8] The bulk of *Plausible Cause* is not actually about the Fourth Amendment at all, but rather about explanation as a value and a uniquely human concept.

This discussion is an important one. Among others, I have emphatically warned against the seemingly unbridled enthusiasm that business and government have for automated decisionmaking.[9] Machines simply cannot be trusted on blind faith. But neither can they be rejected as easily as Brennan-Marquez argues. In making his argument, Brennan-Marquez inadvertently sets up a false dichotomy

3. *Id.* at 1258. Brennan-Marquez expressly analogizes plausibility in this context to that of pleading standards after *Twombly* and *Iqbal*. *Id.* at 1258 n.24 (citing Kiel Brennan-Marquez, *The Epistemology of* Twombly *and* Iqbal, 26 REGENT U. L. REV. 167, 191 (2013)).

4. *Id.* at 1259.

5. *Id.* at 1253.

6. *Id.* at 1265–66.

7. *Id.* at 1258.

8. *Id.* at 1280–97.

9. Solon Barocas & Andrew D. Selbst, *Big Data's Disparate Impact*, 104 CALIF. L. REV. 101 (2016); *see also* Solon Barocas, Edward Felten, Joanna Huey, Joshua Kroll, Joel Reidenberg, David Robinson, & Harlan Yu, *Accountable Algorithms*, 165 U. PA. L. REV. 633 (2017); Danielle Keats Citron & Frank Pasquale, *The Scored Society: Due Process for Automated Predictions*, 89 WASH. L. REV. 1 (2014); Kate Crawford & Jason Schultz, *Big Data and Due Process: Toward a Framework to Redress Predictive Privacy Harms*, 55 B.C. L. REV. 93 (2014); Pauline T. Kim, *Data-Driven Discrimination at Work*, 58 WM. & MARY L. REV. 857 (2017); Andrew D. Selbst, *Disparate Impact in Big Data Policing*, 49 GA. L. REV. (forthcoming 2017).

between human reason and machines as quasi-magical objects. But machines are designed and can be deconstructed. Even if humans cannot understand machines in the same way we understand each other, that is not to say we cannot understand them at all. Brennan-Marquez offers a vision of explanation that is too monolithic and machine prediction that is too atomistic. Neither is quite correct, and as a result, he paints an unrealistic picture of what the Fourth Amendment and broader legal system actually require of humans and machines. In hopes of furthering this important discussion, I offer this response brief in defense of the machines.

I. MACHINES ARE NOT MAGICAL

If *Plausible Cause* were something other than legal scholarship, it could have been titled "The Parable of the Contraband Detector." The Contraband Detector is a hypothetical piece of police detection equipment that serves as the primary vehicle for Brennan-Marquez's arguments. Understanding the limitations of the Contraband Detector will therefore help to demonstrate the limitations of the arguments in *Plausible Cause* regarding machines and humans.

The Article opens as follows:

> Suppose, in the near future, that police start using an algorithmic tool—the Contraband Detector—to locate residences likely to contain illegal weapons. . . . [I]ts accuracy rate hovers around eighty percent, and data scientists . . . report that the tool's performance will only continue to improve. When the tool locates a suspicious residence, it does not explain why; it simply displays an address. And because of the tool's complexity—it draws on more than 100 input variables—officers have no idea which variables are determinative in a given case.
>
> Here is the puzzle. Imagine the Contraband Detector, deployed in New York City, turns up "285 Court St., Apt. 2L," prompting the NYPD to seek a search warrant. When the judge asks about probable cause, the officers point to one, and only one, fact: the tool's performance rate. Should the judge sign the warrant? . . .
>
> There is a powerful and widespread intuition that the answer to th[is] question[] is no. Performance aside, blind reliance on an algorithmic tool feels uncomfortable. It misses the point of particularized suspicion. But why? On its face, probable cause would seem to depend on the *probability* that a "person[ ], house[ ], paper[ ] or effect[ ]" is linked to wrongdoing. In the example, it is eighty percent probable that 285 Court St., Apt. 2L contains an illegal weapon. So probable cause, literally construed, should be satisfied.[10]

Immediately, Brennan-Marquez sets up a dichotomy. On one side is explanation, evidence, reason, and lawfulness. On the other is "one, and only one, fact:" what he later terms "statistical accuracy."[11]

---

10. Brennan-Marquez, *supra* note 1, at 1251–53.

11. Arguably, "statistical accuracy" can have more meaning than he attributes to it, once one takes into account context through things like Bayesian analysis. *See* Part II, *infra*. But Brennan-

Some background will help flesh out the intuition on which Brennan-Marquez relies. Data mining systems operate by finding relationships between input features and known outcomes in a set of training data.[12] In the criminal context, this might mean that they examine places, dates, weather patterns, and assorted other information about past crimes, and use it to predict future crimes.[13] What makes them different from human decisionmaking models is that the machines themselves, without direct human interference, develop the rules that best predict the known outcomes based on the input data.[14] This set of rules is known as the "model."[15] Then the model is applied to future, unobservable cases of interest and predicts the results.[16] Brennan-Marquez envisions the Contraband Detector as one of these systems that builds a model, and then in the future, when given a list of addresses, it would simply assign them each a probability of containing drugs, and no one would be able to explain why.

Many commentators have lamented the "black box" nature of machine learning-based data mining algorithms, arguing that if we cannot see the code or interact with it, we cannot regulate it.[17] And in many cases, due to trade secrecy or other reasons for lack of access, such access might prove impossible.[18] While these concerns are legitimate, they are not the whole story. What also matters is how much we understand the system. And machine learning systems often defy

---

Marquez is clear that he uses it to mean acontextual accuracy. Brennan-Marquez, *supra* note 1, at 1251 n.4.

12. See generally Usama Fayyad, *The Digital Physics of Data Mining*, 44 COMM. ACM, Mar. 2001, at 62.

13. Selbst, *supra* note 9, at *13–14.

14. Cary Coglianese & David Lehr, *Regulating by Robot: Administrative Decision Making in the Machine-Learning Era*, 106 GEO. L.J. at *18 (forthcoming 2017), *available at* http://scholarship.law.upenn.edu/cgi/viewcontent.cgi?article=2736&context=faculty_scholarship [https://perma.cc/DV45-QJGV].

15. Barocas & Selbst, *supra* note 9, at 677.

16. Something can be unobservable and thus need to be predicted for many reasons. The prediction could be about a future event, or trying to measure something that is not easily captured, like a personality trait, or just trying to capture a truth about which information is lacking. Brennan-Marquez's Contraband Detector predicts present or past crime, and is an example of the latter.

17. *See, e.g.*, FRANK PASQUALE, The BLACK BOX SOCIETY 3–4 (2015); Brenda Reddix-Smalls, *Credit Scoring and Trade Secrecy: An Algorithmic Quagmire or How the Lack of Transparency in Complex Financial Models Scuttled the Finance Market*, 12 U.C. DAVIS BUS. L. J. 87 (2011); Frank Pasquale, *Restoring Transparency to Automated Authority*, 9 J. ON TELECOMM. & HIGH TECH. L. 235, 237 (2011).

18. Mikella Hurley & Julius Adebayo, *Credit Scoring in the Age of Big Data*, 18 YALE J.L. & TECH. 148, 196–98 (2016); Citron & Pasquale, *supra* note 9, at 5.

complete understanding even when completely transparent.[19] In the simplest forms of machine learning systems, the machine finds simple correlations between inputs and outputs. But at least with models that rely on large numbers of input features, operating on that many variables at the same time is just too complex an operation for a human to comprehend. More advanced versions of machine learning can key in on variables that have no semiotic value to humans, and thus we cannot truly comprehend them even as approximation.[20]

Brennan-Marquez recognizes this, and he is not among the scholars calling for transparency, but rather for understanding.[21] Yet his treatment of the Contraband Detector actually assumes not only that it is a black box, but that no information about the device can be given to the judge other than the output. This is an extremely limited hypothetical. Black boxes can generally be tested, and the relationship between inputs and outputs is often knowable,[22] even if one cannot describe succinctly how all potential inputs map to outputs. To say that something is a black box is not to say we can understand nothing about it.[23] And there is no reason in principle why his Contraband Detector has to be a black box at all. In the absence of other secrecy-promoting concerns, it should be possible to speak with the engineers that designed it to see what assumptions they made or where they got their training data. Indeed, I have argued elsewhere that police should be required to test their models and disclose exactly this sort of information before adopting predictive policing technology.[24] Even if no one does any testing internally, there are mechanisms available to differentially perturb inputs and observe the effect on outputs, allowing mapping of the whole or part of the system.[25] Intelligibility and transparency are different concerns, yet Brennan-Marquez argues for

---

19. *See generally* Mike Ananny & Kate Crawford, *Seeing Without Knowing: Limitations of the Transparency Ideal and Its Application to Algorithmic Accountability*, NEW MEDIA & SOC'Y 1 (2016).

20. Jenna Burrell, *How The Machine "Thinks": Understanding Opacity In Machine Learning Algorithms*, 3 BIG DATA & SOC'Y 1, 7–8 (2016). Burrell uses spam filtering as an example. There is no clear definition for spam, and filtering programs are triggered on a "big bag of words." *Id.* "There is no posited semiotic relationship between the words and no meaning in the messages is extracted, nor is there an attempt in the algorithm at narrative analysis." *Id.* at 8.

21. Brennan-Marquez, *supra* note 1, at 1280. ("The idea here is simple: we cannot effectively regulate what we do not understand.")

22. *See, e.g.*, Kroll, et al., *supra* note 9, at 650–51 (referring to "black-box testing" as "consider[ing] . . . the inputs and outputs of a system or component").

23. *Id.*

24. Selbst, *supra* note 9, at *48–55.

25. JULIUS ADEBAYO, FAIRML: AUDITING BLACK-BOX PREDICTIVE MODELS (2017), http://blog.fastforwardlabs.com/2017/03/09/fairml-auditing-black-box-predictive-models.html [https://perma.cc/Q9SW-GRVH].

intelligibility based on a completely opaque system. To test his argument, he must allow the black box to be opened.

The Contraband Detector seems to have another conceptual hole that distinguishes it from real systems in ways that complicate the argument. Brennan-Marquez states at one point that the Contraband Detector operates without context.[26] At another point, Brennan-Marquez notes that the Contraband Detector "draws on more than 100 input variables."[27] But these ideas are in conflict.

When a data miner creates a model, he takes into account a huge amount of information about the world. A system that would decide to pull a car over for drunk driving, for example, would include in the model past information about time of day, model and color of the car, and traffic patterns of the place the arrest is to be made, among other data. Then, the model would need much of the same information about the potential arrestee in order to classify him—that is, determine whether he was driving drunk or not. The system would not be able to give a useful answer without a host of background data about the person, place, and time. Simply entering a name or any other sole piece of information would not accomplish anything, unless the system has some way to connect that information to the context it needs.

So which is it? Is context provided or not? At no point in the Article do the actual input variables ever enter Brennan-Marquez's analysis, and whenever he discusses a result of the Contraband Detector, it is always a single instance, independent of context and inputs. Imagine a Contraband Detector where the eventual model depends on whether an address is an apartment or house. Now, suppose the police found an 80% probability of drugs at an address, and though the model depends on the apartment/house information, it has not been offered. As soon as the police supply more context by re-processing the request with the apartment number, the 80% likelihood will change, in one direction or another. That is, by definition, what it means for the model to depend on the factor. This type of change never enters his analysis. Thus, his model does appear to operate without context, with probabilities at the output operating independently of the input data. But by setting it up this way, Brennan-Marquez actually fashions an impossible hypothetical. It cannot be a single Contraband Detector; it is instead a cross section of an infinite set of detectors, all tuned differently. It is a magical device.

---

26. *Id.* at 27 ("In fact, the whole point of a tool like the Contraband Detector is to make predictions from correlative variables *out* of context—a process that, by its nature, frustrates inquiry into the tool's case-by-case performance, as plausibility analysis requires.")

27. *Id.* at 2.

Brennan-Marquez then argues that certain probabilistic determinations will not be tailored enough to satisfy the Fourth Amendment. But that concern is *also* explained by failure to consider the effect of input data. He writes:

> Imagine, for example, that . . . [elevated] electricity usage . . . —say, ten times the average amount—has correlated eighty percent of the time, in the past, with drug manufacturing. Furthermore, suppose . . . drug manufacturers will be unable to avoid outsized electricity usage. . . . On these facts, would the observation that a given residence uses ten times the average amount of electricity be sufficient, by itself, to warrant a search? [No.] Here, the problem is not that the explanatory theory behind the prediction is unknown. On the contrary, the theory of wrongdoing—that drug manufacturing led to high electricity usage—is plain, and certainly plausible. The problem is that heightened electricity usage has many innocuous explanations. From the fact of heightened usage alone, it is impossible to assess the relative plausibility of criminality in any given case by comparison to innocent explanations.[28]

Machine learning systems do take background data into account. Thus, if a robust enough machine learning system discovered that electricity usage was such a dominant factor that it made up 80% of the determination, it is precisely because other hypotheses are less likely. If there were other factors present that help to distinguish innocent explanations from criminal ones, they would express themselves in the data, as long as the machine learning system has been exposed to them. That is, while humans may catch some things that machines do not,[29] machines will also factor in many competing explanations that Brennan-Marquez argues they cannot.

Ultimately, the difficulties with the argument Brennan-Marquez presents stem from the limitations of this hypothetical technology. Had he instead considered a more realistic technology, his arguments would have had to change considerably.

II. EXPLANATIONS MUST BE TIED TO A QUESTION

Explanation is central to the very concept of law. Judges write opinions to explain their rulings. Litigators offer theories of the case to try to better explain the facts than the other side. Administrative agencies write long explanations to accompany new rules and regulations. Explanations are the difference between authority vested in a decision and authority vested in a person or system.[30] Giving reasons is at the very core of what lawyers do.

In rejecting the sole output of his Contraband Detector, Brennan-Marquez can claim all the virtues of non-arbitrary

28. Brennan-Marquez, *supra* note 1, at 1261–62.
29. *See* Part III, *infra*.
30. Frederick Schauer, *Giving Reasons*, 47 STAN. L. REV. 633, 636–37 (1995).

decisionmaking as arguments in his favor. And there are many. But Brennan-Marquez's view of what counts as an explanation is underspecified. Because the Contraband Detector gave so little information to the hypothetical judge, all he had to say was that *some* explanation was required. Once the machinery starts looking like a real system, the question instead becomes "what kind of explanation?"

The explanation Brennan-Marquez would find most troubling is also the first one some data scientists might have been tempted to offer: "the data says so." They would say that the system simply captures the state of the world, and in the past the same hundred or so factors that act as inputs here have combined to indicate a certain result. At first blush, this is pretty similar to the non-explanation that the Contraband Detector offers, though it has more content. People who design machine learning systems usually reserve about 10% of the training data to use in validity testing.[31] If the machine can predict the known results, the model is working as it is supposed to. This *is* a form of explanation that responds to questions of validity, but it is also precisely the explanation that motivates the Contraband Detector hypothetical.

"The data says so" is indeed an unsatisfactory explanation. But Brennan-Marquez chooses the wrong reason to reject it. At times he echoes a common refrain that the detector is simply a correlation of past data and past outcomes, not an explanation.[32] This rationale recalls the old adage that "correlation is not causation," and for good reason. Simplifying immensely, typical legal explanations of facts—the narrative explanations Brennan-Marquez seeks—are causal in nature with limited factors.[33] So it is "A caused B," or "C and D caused E," not "these 100 factors predict F, in some combination." In the Fourth Amendment context, this could be represented by a police officer explaining that she saw a suspect enter a known drug house and come out with a rectangular package and therefore it is more likely than not

---

31. *See, e.g.*, MOHSSEN MOHAMMED, MACHINE LEARNING: ALGORITHMS AND APPLICATIONS § 1.2 (2016) (discussing "holdout data").

32. Brennan-Marquez, *supra* note 1, at 1291–92, 1296–97; *see also* Kim, *supra* note 9, at 13–14 (making a similar argument). This mindset has driven a lot of the worry in the media about Big Data as well, beginning with Chris Anderson's famous exclamation that Big Data heralds "the end of theory." Chris Anderson, *The End of Theory: The Data Deluge Makes the Scientific Method Obsolete*, WIRED (June 23, 2008), https://www.wired.com/2008/06/pb-theory/ [https://perma.cc/NY5Y-UEKS].

33. *See* VIKTOR MAYER-SCHÖNBERGER & KENNETH CUKIER, BIG DATA 63–64 (2013). The number-of-factors limitation is a practical one about complexity of concepts. The analysis needed to fully justify this limitation is beyond the scope of this Essay, but it relates to the work on interpretability in machine learning. *Cf.* Zachary P. Lipton, *The Mythos of Model Interpretability*, at 98 (2016), *available at* https://arxiv.org/abs/1606.03490 [https://perma.cc/TX6F-ZXWB], (suggesting that even simple decision procedures such as trees are not intrinsically interpretable when they become too complex).

he is buying drugs. This is what Brennan-Marquez refers to as "narrative" explanation.[34]

At least with models that rely on large numbers of input features, it may not be possible to provide a simple narrative, because the outcome depends on the interaction of all the variables.[35] They all contribute to the outcome to some degree, or they would not be included in the model at all. There may be cases where a few variables are dominant and make up the bulk of the decision, and for those special cases, it might be appropriate to approximate those variables as the "primary" predictors, but such models are not the general case. In the general case, the models will depend on more complex relationships than humans can easily comprehend.

But this is neither inherently a problem in Fourth Amendment jurisprudence specifically, nor in law generally. Causation and correlation are not as easily disentangled as people typically think them to be.[36] The rectangular package and particular building the suspect was spotted leaving did not explain why the officer concluded that there was a drug deal. Rather, in the past, the police have found that people in similar circumstances were buying drugs, so these facts predicted it, and human reasoning relies on causal intuitions to make sense of such predictions.[37] And those two pieces of evidence were not the only facts the officer would have relied on in this hypothetical – perhaps it was night when she saw the suspect, perhaps the suspect looked around suspiciously, his clothing indicated a certain socioeconomic class, or his skin color or gender factored in. Whether consciously or unconsciously, these factors combined in the police officer's mind to indicate a drug deal. And as Jane Bambauer has observed, the more factors that go into a warrant application, the more likely it is ultimately approved.[38]

There is therefore nothing inherently wrong with a long, correlative list of factors that add up to wrongdoing via inductive reasoning.[39] Indeed, though Brennan-Marquez holds up *United States v. Sokolow*[40] as an example of his theory, it amounts to nothing more

34. *Id.* at 17, 27.

35. This is a not a feature unique to machines, either. It was understood at least by the late 1970s, with the advent of credit scoring. *See* Winnie F. Taylor, *Meeting the Equal Credit Opportunity Act's Specificity Requirement: Judgmental and Statistical Scoring Systems*, 29 BUFF. L. REV. 73, 105–07 (1980). Moreover, it is likely a decent description of what police are *actually* doing inside the black boxes of their brains, but not what they submit to the court.

36. MAYER-SCHÖNBERGER & CUKIER, *supra* note 33, at 85.

37. DANIEL KAHNEMAN, THINKING FAST AND SLOW 185–86 (2011).

38. Jane Bambauer, *Hassle*, 113 MICH. L. REV. 461, 496 (2014).

39. Technically, abductive reasoning. *See infra* notes 60–62, and accompanying text.

40. 490 U.S. 1 (1989).

than this. In *Sokolow*, the Supreme Court approved police reliance on six factors to stop Andrew Sokolow on suspicion of drug trafficking:

> (1) he paid $2,100 for two airplane tickets from a roll of $20 bills; (2) he traveled under a name that did not match the name under which his telephone number was listed; (3) his original destination was Miami, a source city for illicit drugs; (4) he stayed in Miami for only 48 hours, even though a round-trip flight from Honolulu to Miami takes 20 hours; (5) he appeared nervous during his trip; and (6) he checked none of his luggage.[41]

The Court reasoned that while taken alone, these observations were easily explained by innocent conduct, in concert, the factors combined to predict that Sokolow was on the trip for drug-related purposes.[42] The Court did not require the officers to explain why those six factors work—they just did. Brennan-Marquez offers this as an example of " 'plausibilistic' reasoning,"[43] but it is unclear that the Court did anything other than consider extra factors, and with each factor added, decide that the overall likelihood of criminality was higher. The Court did not need to consider the factors in any specific order, so a given narrative was not important, except as assembled post hoc by the Court. And the Court specifically took issue with the lower court's division of facts into those indicating "ongoing criminal activity" and those that were mere "personal characteristics."[44] In reality, those differences are a matter of degree, not kind, and all the facts were probative. This sounds a lot like the inferences of machine learning.

There are indeed good reasons to be suspicious of claims that "the data said so," even if its reliance on correlations is not one. No matter what some data scientists say, data is not neutral, and design decisions that modelers make will affect the ultimate outcome from both an accuracy and fairness perspective.[45] Thus, the real reasons to be suspicious of the models are the myriad ways that uncritically relying on the data can be misleading.

Explanation is therefore still important, but one cannot just ask for an explanation absent a purpose for the explanation.[46] As Brennan-Marquez noted, certain explanations are not legally valid: "An official cannot, for example, rely on the explanation that he strongly wished to perform Act X as authority to perform Act X. Nor can he rely on the explanation that God told him to."[47] These explanations may well be the

---

41. *Id.* at 3.
42. *Id.* at 9.
43. Brennan-Marquez, *supra* note 1, at 1270.
44. *Sokolow,* 490 U.S. at 6.
45. *See* Barocas & Selbst, *supra* note 9, at 677–92.
46. W. Bradley Wendel, *Explanation in Legal Scholarship: The Inferential Structure of Doctrinal Analysis*, 96 CORNELL L. REV. 1035, 1059 (2011).
47. Brennan-Marquez, *supra* note 1, at 1288.

most salient to the people offering them; they are just not legally relevant one way or another.[48]

What does it mean to be legally justified? It means to give explanations that correspond to a host of legal concerns and to satisfy them all. Instead of asking for an explanation in general, then, what the law does is ask for specific types of explanation to vindicate different normative concerns.[49] If the concern is discrimination,[50] then an explanation of how a model treats protected classes is important. This can be tested by varying the protected class variables while holding others constant[51] and seeing how the output is affected. If the concern is how much we rely on association and other First Amendment values,[52] then a similar operation can be done on those variables.[53] But a desire for an "explanation" in general is underspecified and only makes sense when offered in opposition to "no explanation"—the scenario set up by the Contraband Detector.

The one type of explanation that is tough for some machine learning systems to offer is a full descriptive account detailing how they get from input to output in all cases. It may be possible to look at how some specific sensitive input variables affect the outputs individually, but as discussed above, the model will usually be too complex to describe in total and have a person understand it intuitively. But it is not clear that the Fourth Amendment ever asks for this specific type of explanation.

As an illustration, consider the line of cases involving anonymous tips. The anonymous tip here acts as the single point of information, without any explanation, like the Contraband Detector. In

---

48. Offering these explanations may be seen as evidence of absence of a legitimate legal justification, but the content of explanations themselves is legally irrelevant.

49. Wendel, *supra* note 46, at 1049 (describing legal explanation as "inherently contrastive").

50. *See* Brennan-Marquez, *supra* note 1, at 1281.

51. Technically, this is an oversimplification. One actually needs to account for the interdependency of different variables like race and class for example, but that can be done as well. *See, e.g.*, ADEBAYO, *supra* note 25. The results might not ever be perfectly explanatory, but it is possible to get a picture of whether protected class correlates a lot or a little with outcome.

52. Brennan-Marquez, *supra* note 1, at 1282.

53. Brennan-Marquez offers a few other concerns, such as vagueness. *Id.* at 1288–94. Vagueness as he posits it is misleading, though. He seeks to "guarantee[ ] that the reasons for intrusion are, at least to some extent, predictable, because it ties intrusion to activity that *appears more plausibly guilty than innocent*." *Id.* at 1293. But that is not what vagueness doctrine is. Vagueness doctrine is about giving people fair notice of what acts are criminal, not what acts *appear* to be criminal. *See* Johnson v. United States, 135 S. Ct. 2551, 2556 (2015). The intrusions that concern Brennan-Marquez are actually to those of specifically innocent people, the police officer's false positives. When he suggests that we should know what innocent activities *appear* criminal so we can avoid them, he ironically suggests that law create chilling effects on innocent conduct. Therefore, it might achieve greater liberty to work to reduce false positives, but in the meantime, be publically vague about what specific conduct leads to them.

these cases, the Court has been concerned that the tip might not be reliable because the tipster might have questionable motivations. Therefore the explanations they seek are "indicia of reliability."[54] In *Florida v. J.L.*, the Supreme Court held that an anonymous tip, standing alone, is not sufficient to generate reasonable suspicion for a *Terry* stop.[55] The Court contrasted the informant in the case with "a known informant whose reputation can be assessed and who can be held responsible if her allegations turn out to be fabricated."[56] The Court did not state that it needed a full explanation of who the tipster was, how he came by the information, what his motivation was to offer the tip, and everything else about the person. In *Alabama v. White*, the Court held that an anonymous tip corroborated by further police work has sufficient indicia of reliability, even though not every aspect of the tip was corroborated.[57] In neither case was a full explanation of how the tip occurred necessary.[58]

Explanations are certainly required by the Fourth Amendment, as they are in virtually all areas of law.[59] But those explanations have to answer a specific concern. The explanations the Court sought in the tipster cases were ones that addressed the particular concern the Court had. The analogous case for a machine learning system is whether it, too, made reliable predictions in the past, or can be held accountable for intentionally false information. The former has to do with validity testing, and the latter suggests that the system should be auditable, rather than inaccessible under something like trade secret doctrine. Both are legitimate validity concerns that should be addressed before trusting a system. But importantly, neither of these concerns have to do with how the system works internally, and both directly respond to a specific extant concern in Fourth Amendment law. Once other questions about a machine learning system's validity are raised,

54. Florida v. J.L., 529 U.S. 266, 270 (2000).

55. *Id.*

56. *Id.* (citing Adams v. Williams, 407 U.S. 143 (1972)).

57. Alabama v. White, 496 U.S. 325, 331 (1990).

58. The "internal explanation" analogue for a tipster would be to ask the tipster how he connected the information to an ongoing crime. But no one ever asks that.

59. One notable exception is jury deliberations, which are famously designed as black boxes. Fed. R. Evid. 606(b); Tanner v. United States, 483 U. S. 107 (holding that marijuana and cocaine use during trial could not be used to impeach a jury verdict). But in a recent case, the Supreme Court held that jury verdicts could be impeached if the jurors relied on "racial stereotypes or animus to convict a criminal defendant." Peña-Rodriguez v. Colorado, No. 15–606, 2017 WL 855760, at *17 (Mar. 6, 2017). That is, the normative requirement that criminal trials not be marred by intentional racism is so great that it demands explanation of the jury deliberations enough to determine whether that occurred. But other normative concerns here do not. Similarly, the Fourth Amendment will demand explanation for some normative concerns and not others, but explanation in general is not meaningful.

whether on equal protection, First Amendment, or some other grounds, explanations to satisfy those concerns should be investigated, and predictions that do not satisfy those concerns could be barred. But it is hard to see what Fourth Amendment question is implicated by a need to explain the entire model. In general, for explanation to be useful as a requirement, there has to be a question in mind that needs answering.

III. HUMANS AND MACHINES HAVE DIFFERENT STRENGTHS

Thus far, I have tried to present a more realistic view of the role of machines and explanations in the Fourth Amendment warrant process. But machines are not being evaluated against a blank slate. Rather, they must be compared to what was until recently the only option: human decisionmaking. This human-machine comparison is the ultimate goal of both Brennan-Marquez's Article and this Essay. A call for "narrative" explanations seems to be a call specifically for those kinds of explanations that accord easily with human reasoning and experience. But does that actually flow from plausibility analysis?

Plausibility analysis is an instance of abductive reasoning.[60] Abductive reasoning is a flavor of induction by which people infer the best explanation given a set of facts and competing hypotheses.[61] An abductive inference is the statement that a given hypothesis better explains the available data than any other hypothesis. The most familiar instance of abductive reasoning is the causal inference; we cannot know *for sure* that the glass shattered because it hit the ground, but it's the best hypothesis available.[62]

Brennan-Marquez claims that machines cannot test for relative plausibility.[63] But that is not correct for two reasons. First, though he mentions input variables, he ignores the effect of changing them. When factors that would implicate alternative hypotheses are added to a model, if they are relevant, they will change the probability at the output; if not, they do not implicate viable alternative hypotheses. There is a difference between a percentage likelihood offered by one factor model (e.g., a person leaving a drug house) and one with many more factors (e.g., a person leaving a drug house, at night, with his face hidden, holding a package under a big coat). Second, he treats his notion of "statistical accuracy" as equivalent to all forms of probability. But

60. *See* Brennan-Marquez, *supra* note 1, at 1253 n.9.

61. Ronald J. Allen & Michael Pardo, *Juridical Proof and the Best Explanation*, 27 L. & PHIL. 223, 229–33 (2008); John Josephson, *On the Proof Dynamics of Inference to the Best Explanation*, 22 CARDOZO L. REV. 1621, 1622–24 (2001).

62. Allen & Pardo, *supra* note 61, at 228.

63. Brennan-Marquez, *supra* note 1, at 1262.

abductive reasoning *can* involve probability. When determining the best explanation of the facts, the best hypothesis is likeliest to explain all the facts. Brennan-Marquez recognizes this when he praises a mathematical model of relative plausibility, but it does not factor into his analysis.[64] The multi-factor model above will predict a higher probability of criminality than the one-factor model precisely because of the extra facts that narrow the universe of likely explanations. And a model that did not take into account the variables that could implicate alternative explanations simply would not be accurate enough to use. Thus, while Brennan-Marquez offers plausibility-as-narrative-explanation as the only way out of pure probability, he never addresses the impact of the number or relative importance of the different factors considered in a data mining model. The model's reaction to input data is exactly the consideration of alternative hypotheses he seeks.

We can also ask what value "narrative" adds to the formulation. There are two possibilities. One is that a narrative is just a series of additional facts that enter the evaluation. The other is that the narrative somehow triggers some innate human quality of explanation. But in all his examples, the only work that "narrative" does is to add contextual facts. Consider *United States v. Sokolow* again. Brennan-Marquez holds it up as an example of his theory, but the Court merely said that more factors in combination were more persuasive.[65] Brennan-Marquez then suggests that perhaps if it had been Super Bowl weekend, "the innocent explanation might have prevailed."[66] Therefore an additional fact would have changed the narrative. But there is nothing special about the *narrative* there. The only question for a machine learning model is whether the additional fact was or was not taken into account.

So, if not narratives, what are humans actually contributing over machines? There are two differences between *Sokolow* and a more realistic Contraband Detector.[67] First, the input data is known in *Sokolow* but hidden in the Contraband Detector. Second, the members of the *Sokolow* majority had an intuition that the data they saw correlated to drug activity. Now imagine a case like *Sokolow*, except that it involves a machine learning system where the input data *is* known. With this new "Kontraband Detektor," the court is shown the

---

64. *Id.* at 1259 n.26.

65. 490 U.S. 1, 9 (1989).

66. Brennan-Marquez, *supra* note 1, at 1270.

67. Here I'm specifying "realistic" because the factors that *Sokolow* takes into account each change the probability of wrongdoing, as opposed to the "infinite-cross section" version of the Contraband Detector, where the output probability is independent of input. *See* Part I, *supra*. We should assume we're comparing to a hypothetical, but physically possible machine.

inputs as well as the resulting probability. In this case, if the inputs were the same six variables as in *Sokolow*, we know the outcome, because the Justices could have put it together on their own, and would just ignore the machine.[68] Now alter the hypothetical, to a different six variables with the same ultimate probability of criminality. If the Justices come out differently, it is not because of anything different about the technology. It is exactly the same Kontraband Detektor, relying on six inputs to get to a probability of criminality. Rather, the only remaining difference is the intuition they had about whether the six factors added up to drug running, independently of the machine. That intuition is the same type of calculation that a data mining system would do—finding correlations that lead to a predictive outcome—but done in a human mind. That *intuition* is therefore the core of the work that plausibility analysis is doing for Brennan-Marquez.

Intuition is not necessary to providing relative plausibility of different hypotheses, because that would have been expressed in a more realistic data model. Human intuition instead serves as a sanity check. If there is an additional fact to be taken into account—if Sokolow had been traveling Super Bowl weekend, for example—a machine learning system might not have modeled that.[69] If it *had* been included in the model, then it would have reduced the outcome probability, just as if a human had considered it. To then rely on the additional fact in addition would actually be incorrect; it would be double counting. The human's role, therefore, is to detect when input information might be missing from the model, and then supply it.

Human intuition, sometimes called "common sense,"[70] is one major point of difference between humans and machines. Machines are myopic. They can only understand facts about the world that they are exposed to, and as a result, it is possible they will miss something that a human will see with the human's broader knowledge base. Judges must remain involved in the process in case the machines fail to take into account certain contextual facts that are important, in case the results implicate values unconsidered by the data miners, or in case the results just make no sense.[71]

68. Here I assume the outcome of *Sokolow* was correct, and those variables in combination *were* a good indicator of drug activity.

69. Michael L. Rich, *Machine Learning, Automated Suspicion Algorithms, and The Fourth Amendment*, 164 U. PA. L. REV 871, 871–75, 897 (2016).

70. Curtis E.A. Karnow, *The Application of Traditional Tort Theory to Embodied Machine Intelligence*, *in* ROBOT LAW 51, 75 (Ryan Calo, A. Michael Froomkin & Ian Kerr eds., 2016).

71. Rich, *supra* note 69, at 897–901; Orin Kerr, *Why Courts Should Not Quantify Probable Cause*, *in* THE POLITICAL HEART OF CRIMINAL PROCEDURE 131, 131 (Michael Klarman et al. eds., 2012).

This is, not coincidentally, a big part of the role Brennan-Marquez sees judges occupying. "[J]udges must have an opportunity to scrutinize" the technology,[72] "to consider the plurality of values implicated by the exercise of state power," and "to resolve conflicts between those values in a context-sensitive way."[73] And he notes that prediction tools are useful and can permissibly aid police.[74] He just systemically overestimates the innate humanness of explanation and underestimates the auditability of machines, and therefore argues that judges should oversee machines' overall intelligibility as well,[75] which is both unnecessary and often impossible.

While human common sense is a benefit we hold over machines, machines hold two benefits over humans: the ability to process much more information and the lack of human biases, conscious or unconscious.[76] Though Brennan-Marquez accurately argues that statistical accuracy is not the only value present, we should not disregard it entirely. Machine learning systems regularly outperform humans at inferential tasks.[77] These traits are just as important as the value of common sense to understanding when we should and should not make decisions with machines. In the case where machines are *not* missing information, they might be better at the kind of analysis we want.

Returning to *Sokolow* once again, the decision produced a dissent by Justice Marshall, joined by Justice Brennan.[78] The disagreement was essentially about whether or not the factors, all told, added up to an inference of drug activity.[79] Ultimately, because both sides were working with the complete information of only six facts, there is no easy way to figure out which side got it right, to this day. Had they been employing a machine learning system with some degree

72. Brennan-Marquez, *supra* note 1, at 1253.

73. *Id.* at 1256.

74. *Id.* at 1254.

75. *Id.* at 1253.

76. *See, e.g.*, KAHNEMAN, *supra* note 37, at 199–200 (discussing the "narrative fallacy"), 224-25 (discussing the inconsistency of human judgment); Linda Hamilton Krieger, *The Content of Our Categories: A Cognitive Bias Approach to Discrimination and Equal Employment Opportunity*, 47 STAN. L. REV. 1161, 1211–17 (1995) (discussing implicit bias in the context of disparate treatment doctrine).

77. MAYER-SCHÖNBERGER & CUKIER, *supra* note 33, at 61; KAHNEMAN, *supra* note 37, at 224 ("Several studies have shown that human decision makers are inferior to a prediction formula even when they are given the score suggested by the formula! They feel that they can overrule the formula because they have additional information about the case, but they are wrong more often than not.").

78. United States v. Sokolow, 490 U.S. 1, 11–18 (1989) (Marhsall, J., dissenting).

79. *Id.* at 17.

of auditability, we would have more confidence that once side or the other at least drew the correct inference as a statistical matter.

This *Sokolow* problem is about human processing power, but it is only made worse when human biases enter the picture. Consider two more cases: *United States v. Brignoni-Ponce*[80] and *United States v. Martinez-Fuerte*,[81] in which the Supreme Court held "perceived Mexican ancestry" could be relied upon in border searches "for brief inquiry into . . . residence status,"[82] but it could not be the only factor.[83] In a data mining system taking hundreds of factors into account, there is sense to this. Race is just one more background fact, and even if it is not included, it will likely be redundantly encoded in other data to a degree.[84] But for individual officers to take race into account it is not hard to imagine that will give it more value as a factor than they should, whether intentionally or not.

Machines and humans each have our strengths, and the optimal system will recognize and use the real strengths of both.[85]

CONCLUSION

Science fiction author Arthur C. Clarke once famously wrote that "[a]ny sufficiently advanced technology is indistinguishable from magic."[86] But lawyers and policymakers must distinguish science fiction from actual technology. Machines and humans each have different strengths and naturally process the world differently. But we cannot blind ourselves to the value of the machines that are being developed. *Plausible Cause* makes some very important points about Fourth Amendment doctrine but fails to appreciate the contributions of machines, because it draws on a hypothetical machine that cannot exist in the real world.

Humans remain important too. Brennan-Marquez asks early on whether, if broad use of machine learning is permissible under the

80. 422 U.S. 873 (1975).

81. 428 U.S. 543 (1976).

82. *Id.* at 555.

83. *Brignoni-Ponce*, 422 U.S. at 886–87.

84. Barocas & Selbst, *supra* note 9, at 691.

85. *Cf.* Hope Reese, *How One AI Security System Combines Humans and Machine Learning to Detect Cyberthreats*, TECHREPUBLIC (Mar. 23, 2016), http://www.techrepublic.com/article/how-one-ai-security-system-combines-humans-and-machine-learning-to-detect-cyberthreats/ [https://perma.cc/U7G6-BHRP] (noting that in machine learning systems for cybersecurity, "[t]here will always be security analysts involved"); Cade Metz, *In Two Moves, AlphaGo and Lee Sedol Redifned the Future*, WIRED (Mar. 16, 2016), https://www.wired.com/2016/03/two-moves-alphago-lee-sedol-redefined-future/ [https://perma.cc/74AM-6GB6] (describing how the human and machine strategies complemented each other in playing Go).

86. ARTHUR C. CLARKE, HAZARDS OF PROPHECY: THE FAILURE OF IMAGINATION (1973).

Fourth Amendment, an "Automatic Warrant Machine" is as well.[87] While he has not given machines nearly enough credit for the plausibility analysis they can perform, until and unless we achieve so-called "strong AI",[88] we will always need human oversight due to machine myopia.

Ultimately, Brennan-Marquez is correct that we cannot simply "entrust our fates to the power of computation rather than the wisdom of judgment."[89] But no one is calling for that, and he has not given the machines the benefit of their best argument. While we cannot simply abdicate human reasoning to become slaves to the machine, neither can we afford to reject them out of hand. *Plausible Cause* comes a little too close to the latter.

87. Brennan-Marquez, *supra* note 1, at 1252.

88. "Strong AI" is the idea of a computer brain that is functionally equivalent to a human's. *See, e.g.*, John R. Searle, *Is The Brain's Mind a Computer Program?* 262 SCI. AM. 26 (1990).

89. Brennan-Marquez, *supra* note 1, at 1300.