Fordham Urban Law Journal

Volume 46
Number 2 *Artificial Intelligence and Predictive Algorithms: Why Big Data Can Lead To Big Problems*
Article 3

2019

Coding Over the Cracks: Predictive Analytics and Child Protection

Stephanie K. Glaberson

CODING OVER THE CRACKS: PREDICTIVE ANALYTICS AND CHILD PROTECTION

Stephanie K. Glaberson[*]

ABSTRACT

Across the nation, child protective authorities are turning to machines to assist them in their work, developing predictive analytic tools to forecast risk to children and families. While there is clear evidence that current child welfare decision-making processes are flawed and in need of change, the advent of predictive analytics carries with it numerous risks to children and families that cannot be ignored. This Article explains the fundamentally human processes that go into the creation of predictive analytic tools and highlights some of the risks that these tools pose. It argues that the choices made in developing predictive tools implicate some of the most fundamental and as-yet unanswered questions in our child welfare system. As a result, the advent of predictive analytics in child welfare presents a moment for systemic reflection. Without careful attention to the issues that predictive analytics raise, communities risk simply coding over the cracks in the foundation of a flawed system, burying problems of bias, transparency, and accountability deeper, and imbuing the status quo with an undue patina of inevitability. Instead, communities should use this moment to demand more of their child welfare systems and see these tools as opportunities to build better,

---

[*] Clinical Teaching Fellow, Georgetown University Law Center; J.D., New York University School of Law. I am deeply indebted to all those who helped me shape the ideas included in this piece and encouraged me along the way, including Eduardo Ortiz, David C. Vladeck, Paul Ohm, Deborah Epstein, Nila Natarajan, Molly Ryan, Kelly Capatosto, Lucas Merrill Brown, Froilan Irizarry, Sean Hill, Aubrey Rose, Michele Gilman, Shanta Trivedi, the participants in the Family Law: Children's Welfare group at the 2018 Clinical Law Review Writer's Workshop at New York University School of Law, my colleagues in the 2017–19 class of clinical teaching fellows at GULC, and the editors at the *Fordham Urban Law Journal*. All errors are my own.

more humane systems that focus more on support and prevention and less on too-little, too-late crisis response.

TABLE OF CONTENTS

INTRODUCTION

Child welfare authorities across the nation are engaged in a high-stakes project every day: they must predict which children might be at risk of harm in their homes, and whether and when authorities can, or should, intervene. Erroneous predictions carry grave consequences. Agency failures to intervene that result in harm to children at the hands of their caregivers are highly publicized.[1] Less publicized, but equally grave and much more common, are harms that result from unnecessary agency interventions, both those that separate children from their families and those that do not, each of which risks inflicting lifelong trauma.[2] This risk-prediction project relies heavily on human decision-makers, who often handle crushing caseloads under high stress, with little training, limited time, and imperfect information. Bias inevitably creeps into these vital human-powered decisions, resulting disproportionately in the breakup of poor families and families of color.[3]

In this age of automation and artificial intelligence, a tempting new prospect has emerged: using the "magic" of predictive analytics to forecast whether and when children need the government to intervene. Algorithmic prediction systems already are impacting many areas of life, from employment, to college admissions, to tax audits, to the criminal justice system.[4] State and local governments are using algorithmic models to attempt to predict future dangerousness of alleged offenders in setting terms of release or bail, to forecast likely recidivism through sentencing prediction tools, and to project likely probation violations.[5] Additionally, law enforcement agencies nationwide are attempting to use "predictive policing" to

---

1. *See infra* Section I.B.3.

2. The recent uproar over the Trump Administration's separation of immigrant families has brought with it a rare but welcome spotlight on the harms of unnecessary family separation. The traumatic effects of even short stays in foster care have been well documented. *See* Vivek S. Sankaran & Christopher Church, *Easy Come, Easy Go: The Plight of Children Who Spend Less than Thirty Days in Foster Care*, 19 U. PA. J.L. & SOC. CHANGE 207, 210–13 (2017). Even unnecessary investigations that do not result in family separation can be traumatic. *See infra* notes 178–83 and accompanying text.

3. *See infra* Section I.B.2.

4. *See* Mary Madden et al., *Privacy, Poverty and Big Data: A Matrix of Vulnerabilities for Poor Americans*, 95 WASH. U. L. REV. 53 (2017).

5. Andrew Guthrie Ferguson, *Policing Predictive Policing*, 94 WASH. U. L. REV. 1109, 1113–14 (2017).

"stop[] crime before it happens," modeling not only where and how to allocate resources, but increasingly who to police.[6]

Child protective authorities are now dipping their toes into these waters as well. Agencies are building and deploying tools that pull together vast quantities of data stored by various government entities, and return a "risk score" purporting to predict the likelihood of child maltreatment.[7] When constructing these predictive tools, developers must make myriad complex, value-laden, and ultimately human decisions that touch on some of the most fundamental — and unanswered — questions in child welfare policy.[8] These questions range from how to define child maltreatment, to how much risk we are willing to tolerate, to how we value different types of error, to how we address bias and disproportionality in child welfare practices. As Cathy O'Neil, author of *Weapons of Math Destruction*, wrote, "models are opinions embedded in mathematics."[9]

This Article argues that the advent of predictive analytics risks simply coding over the cracks in the foundation of our child welfare system. Unless careful attention is paid to the assumptions, biases, and realities of our child welfare system at this critical juncture, algorithmic decision-making risks perpetuating and magnifying existing problems. The Article proceeds in four parts. Part I describes the "decision points" that a child welfare agency encounters as a child or family moves through the system. It considers the evidence showing that this human-based decision-making system is deeply flawed. Part II defines predictive analytics and explains the fundamentally human process of developing a machine learning algorithm. This Part also identifies those jurisdictions that are implementing or considering instituting predictive risk tools, highlighting those developed by Allegheny County, Pennsylvania and the Florida non-profit, Eckerd Kids. Part III analyzes the risks posed by predictive analytics as they are introduced into child protective decision-making. It addresses the ways in which predictive analytics can introduce error into decisions, further embed old prejudices and biases into new systems, and generate new risks for children and families. Finally, acknowledging that predictive analytics already are

6. *Id.* at 1112.

7. *See infra* Part II.

8. *See infra* Section II.A.

9. CATHY O'NEIL, WEAPONS OF MATH DESTRUCTION: HOW BIG DATA INCREASES INEQUALITY AND THREATENS DEMOCRACY (Reprint ed. 2017). Cathy O'Neil is a mathematician and former Wall Street "quant." She is now the founder of ORCAA, an algorithmic auditing company.

taking hold, Part IV provides some preliminary recommendations for ways in which advocates, scholars, and officials can attempt to ensure that any predictive tools employed where they work and live are developed responsibly and in line with community values.

I. CHILD PROTECTIVE DECISION-MAKING

To assess the role predictive analytics may play in child welfare, and the risks and benefits such tools offer, it is important to review the decisions that child welfare authorities are called upon to make in the course of a child protective case. This Part first will provide an overview of the various "decision points" that occur throughout a child protective case. It then will discuss the methods agencies generally use to arrive at decisions and conclude by reviewing the evidence that the current system of decision-making is flawed and must be improved.[10]

A. Decision Points in a Child Protective Case

As children and their families move through the child protective system, moments arise when individuals, agencies, and courts are called upon to make vital decisions. Oft-discussed examples include whether to report suspected maltreatment or whether a child should be removed from or returned to his or her home, but there are a multitude of other decisions made along the way. This Article refers to each of these moments as "decision points."[11] At each of these points, the quality of the decision may have grave consequences for the life of the child or family.

In most instances, child welfare authorities begin their involvement in the life of a family when someone makes a phone call to report suspected abuse or neglect to state authorities.[12] The decision whether to report, therefore, is the first decision point in the process.

---

10. *See* Anupam Chander, *The Racist Algorithm?*, 115 MICH. L. REV. 1023, 1045 (2017) ("When we appraise emerging technologies, we must be careful not to romanticize a pretechnological past. New technologies must be examined both in comparison to their less-technological alternatives and in the context of the world that we now inhabit.").

11. *See, e.g.*, Marian S. Harris & Wanda Hackett, *Decision Points in Child Welfare: An Action Research Model to Address Disproportionality*, 30 CHILD. & YOUTH SERV. REV. 199 (2008).

12. *See, e.g.*, 42 U.S.C. § 5106(a)(b)(2)(i); CONN. GEN. STAT. ANN. §17a-103a (2016); FL. ST. ANN. § 39.201(2)(b) (West 2016); IND. C. 31-33-7-3; IOWA C. ANN. § 232.70 (2016); VERNON'S ANN. MISS. STAT. §210.145 (2018); N.Y. SOC. SERV. L. § 422(2)(a) (2018); 10A OK. STAT. T. 10A § 1-2-101(A)(1) (2018).

Many states permit community reporters to remain anonymous.[13] State law often mandates that professionals who work closely with children and families, like teachers, social workers, and doctors, must report child abuse or neglect if they become aware of it.[14] In many places, the laws obligating mandated reporters to take action are vague and quite broad.[15]

Regardless of whether a mandated reporter or some other individual places the call, it goes to the same place: a centralized call center,[16] where a hotline worker screens the call.[17] This brings us to the second decision point: the call screener must decide whether the allegations of abuse or neglect warrant an investigation. Call screeners usually have latitude to "screen in" the call, meaning to forward it to a local child protective office for investigation, to simply take note of the call and add it to a central file kept on the child or family, or to reject the call altogether.[18]

---

13. *See* Dale Margolin Cecka, *Abolish Anonymous Reporting to Child Abuse Hotlines*, 64 CATHOLIC U. L. REV. 51, 54–55 (2015) [hereinafter Cecka, *Abolish Anonymous Reporting*]. For years, advocates have called for an end to anonymous reporting, arguing that it allows bad actors to "weaponize" child protective authorities and does not improve children's safety. *See, e.g.*, *id.*; *see also* Dale Margolin Cecka, *How Child Abuse Hotlines Hurt the Very Children They're Trying to Protect*, WASH. POST (May 6, 2015), https://www.washingtonpost.com/posteverything/wp/2015/05/06/how-child-abuse-hotlines-hurt-the-very-children-theyre-trying-to-protect/?utm_term=.62170e567bbe [https://perma.cc/H9YF-Q82J]; Richard Wexler, *Why It's So Easy to Troll Someone with Anonymous Child Abuse Report*, YOUTH TODAY (Aug. 6, 2018), https://youthtoday.org/2018/08/why-its-so-easy-to-troll-someone-with-anonymous-child-abuse-report/ [https://perma.cc/HNU2-GG8E].

14. *See, e.g.*, 23 PA. STAT. AND CONS. STAT. ANN. § 6311 (West 2015); FLA. STAT. ANN. § 39.201 (West 2016); IOWA CODE ANN. § 232.69 (West 2018); LA. CHILD. CODE ANN. art. 603(17) (2018); MASS. GEN. LAWS ANN. ch. 119, § 21 (West 2018); MO. ANN. STAT. § 210.115 (West 2018); N.Y. SOC. SERV. LAW § 413 (McKinney 2018); 40 R.I. GEN. LAWS ANN. § 40-11-3 (West 2018); R.I. GEN. LAWS ANN. § 40-11-6.1 (West 2018).

15. In recent years, some jurisdictions have attempted to strengthen punishments for mandated reporters who fail to make reports. *See, e.g.*, S.B. 135, 2017 Gen. Assemb., 437th Sess. (Md. 2017) (proposing criminal penalties for mandated reporters who "knowingly fail" to make a report of abuse or neglect).

16. *See* Cecka, *Abolish Anonymous Reporting*, *supra* note 13, at 56; 42 U.S.C. § 5106a(b)(2)(B)(i),(iv) (requiring states to have "provisions or procedures for an individual to report known and suspected instances of child abuse and neglect" and "procedures for the immediate screening, risk and safey assessment, and prompt investigation of such reports").

17. *See, e.g.*, LA. CHILD. CODE ANN. art. 612 (2018); FLA. STAT. ANN. §39.201(4) (West 2018); N.Y. SOC. SERV. LAW § 422 (McKinney 2016); 23 PA. STAT. AND CONST. STAT. ANN. §§ 6332–33.

18. *See, e.g.*, 42 U.S.C. § 5106a(b)(2)(B)(iv)–(v) (requiring states to screen and triage reports of child abuse or neglect); PA. STAT. AND CONS. STAT. ANN. § 6334

If the worker "screens in" the call, he or she refers the family to a local agency office for an investigation.[19] Generally, the agency has a specific amount of time in which to complete its investigation and make a determination as to whether the allegations of abuse or neglect are supported or warrant further intervention.[20] This is the third decision point. Different jurisdictions use different terms for the determinations that can result at this stage, some marking cases "indicated" or "unfounded," others marking them "substantiated" or "unsubstantiated."[21] Regardless of the terms used, the legal standard of proof to justify government intervention generally is low.[22] In many states, if the investigation uncovers "some credible evidence" that the child is at risk of harm, the case worker may "substantiate" the report.[23] If the investigating team determines that the report is "unfounded," the case is closed, and — at least in theory — the allegations are put to rest.[24]

If a report is "substantiated," the agency has some discretion in how it chooses to move forward.[25] Federal law requires that state agencies provide "reasonable efforts" to "prevent or eliminate the need for removing the child from the child's home."[26] States put this

---

(West 2014); FLA. STAT. ANN. §39.301 (West 2017); N.Y. SOC. SERV. LAW §§ 422, 24 (McKinney 2016 & 2017).

19. *See* WALTER R. MCDONALD, ASSESSING THE FEASIBILITY OF CREATING AND MAINTAINING A NATIONAL REGISTRY OF CHILD MALTREATMENT PERPETRATORS, U.S. DEP'T OF HEALTH & HUMAN SERV. 13 (2012), https://aspe.hhs.gov/system/files/pdf/177131/ResearchReport.pdf [https://perma.cc/9TY7-6QCX].

20. CHILDREN'S BUREAU, CHILD MALTREATMENT 8 (2011), https://www.acf.hhs.gov/sites/default/files/cb/cm11.pdf [https://perma.cc/7LN5-FSYD] [hereinafter CHILDREN'S BUREAU, CHILD MALTREATMENT]; *see, e.g.*, N.Y. SOC. SERV. LAW § 424(5-b) (McKinney 2017).

21. *See* CHILDREN'S BUREAU, CHILD MALTREATMENT, *supra* note 20, at 15.

22. *See* MCDONALD, *supra* note 19, at 13.

23. As of 2009, twenty states used a standard defined as "some credible evidence," "probable cause," or similar. Twenty-nine states required a "preponderance of the evidence," while only three required a more stringent "clear and convincing evidence" standard. *Id.*

24. In reality, evidence of the report may remain on file with the state for many years. *See* CHILD WELFARE INFO. GATEWAY, ESTABLISHMENT AND MAINTENANCE OF CENTRAL REGISTRIES FOR CHILD ABUSE REPORTS 2 (2018), https://www.childwelfare.gov/pubPDFs/centreg.pdf [https://perma.cc/NV5L-XRX3]. Although federal law requires states to have a procedure for removing or expunging "unsubstantiated or false" allegations from state registries that are open to the public or used for employment-related decisions, it permits child protective agencies to keep "information on unsubstantiated reports in their casework files to assist in future risk and safety assessment." 42 U.S.C. § 5106a(b)(2)(B)(xii) (2017).

25. *See* Cecka, *Abolish Anonymous Reporting*, *supra* note 13, at 56.

26. 42 U.S.C. § 671(a) (15) (2018).

requirement into action in a variety of ways, including providing "preventive services" to families, where appropriate. For some families, the investigating worker might meet with the parents, offer support such as home cleaning, counseling, or other services, and conclude the interaction. For others, the agency will remain involved in the family's life for months or even years through an informal service relationship.[27] In either case, the availability of preventive services that meet the needs of the family will be important to ensuring positive outcomes. In many cities and states, however, preventive services that meet the needs of the community may not be readily available, or there may be long waitlists for enrollment.[28] In these services-only cases, the agency will encounter one more decision point: whether and when to close the case and cease its involvement with the family.

There are circumstances, however, where the agency may decide to take further action. The most drastic avenue is to remove the child prior to going to court — an action referred to as an "emergency removal."[29] Short of removing the child on an emergency basis, the

---

27. *See* CHILDREN'S BUREAU, CHILD MALTREATMENT, *supra* note 20, at 81–83.

28. *See, e.g.*, CHILDREN'S BUREAU, CHILDREN OF COLOR IN THE CHILD WELFARE SYSTEM: PERSPECTIVES FROM THE CHILD WELFARE COMMUNITY 20 (2003), https://www.childwelfare.gov/pubs/otherpubs/children/findings/ [https://perma.cc/9P4A-7T6P] (describing the problem that poor and minority communities lack supportive services, and quoting a direct services worker reporting that "[w]e have waiting lists forever to get any kind of services, [including] substance abuse, domestic violence, [and] parenting classes. When you go into different neighborhoods, Caucasian neighborhoods, we make a referral . . . within days, they have the services they need. My clients wait months. If we put in the referral or the case is in court but the client hasn't gotten services yet, they'll pull those kids."); Melissa Russo, *I-Team: A Proven Tool for At-Risk Families in NYC Child Abuse Cases Is in Crisis, Experts Say*, NBC NEW YORK (Feb. 17, 2017), https://www.nbcnewyork.com/news/local/ACS-Abuse-Prevention-Programs-Strapped-Overloaded-NYC-de-Blasio-414105273.html [https://perma.cc/X7J9-752Q]; ABIGAIL KRAMER, NEW SCH. CTR. FOR N.Y.C. AFFAIRS, ACS IN OVERDRIVE: SINCE THE DEATH OF A HARLEM 6-YEAR-OLD, ARE FEWER FAMILIES GETTING THE HELP THEY NEED? 5–7 (2017), https://static1.squarespace.com/static/53ee4f0be4b015b9c3690d84/t/598217239f7456c3d58f5b3d/1501697829473/ACS+In+Overdrive+%E2%80%94+Center+for+New+York+City+Affairs.pdf [https://perma.cc/3G2Y-WAKU].

29. The Supreme Court has continually reaffirmed the fundamental nature of parents' liberty interest in the upbringing of their children. *See, e.g.*, Troxel v. Granville, 530 U.S. 57, 65 (2000) ("[T]he interest of parents in the care, custody, and control of their children — is perhaps the oldest of the fundamental liberty interests recognized by this Court."). As a result, courts generally hold that parents are entitled to a pre-deprivation hearing before an agency can intervene in the care and custody of their children. *See, e.g.*, Tenenbaum v. Williams, 193 F.3d 581, 593 (2d Cir. 1999) ("As a general rule, therefore, before parents may be deprived of the care, custody or management of their children without their consent, due process —

agency can file a petition in the relevant court seeking an order placing the family under the supervision of the agency, or mandating that the parents comply with certain services or conditions to keep their child at home.[30] The agency also may seek the court's approval to remove the child or children from the home at that time.[31]

If the agency goes to court to seek removal, the court will require the agency to show that removal is necessary in order to safeguard the life or health of the child.[32] This is the next — and perhaps most consequential — decision point in a child welfare case: whether the child will remain at home. This decision is made by the agency, but ultimately must be approved by a court.[33]

If the child is removed, federal law mandates that agencies engage in reasonable efforts to reunify the child with his or her family.[34] The agency will develop a service plan for the family, and the court will monitor the agency's efforts through semi-annual "permanency hearings."[35] In these hearings, the agency must demonstrate that it has made satisfactory efforts in the preceding six months, and secure court approval for its goal for the family for the next six months.[36]

---

ordinarily a court proceeding resulting in an order permitting removal — must be accorded to them." (citing Stanley v. Illinois, 405 U.S. 645, 651 (1972))). State laws provide for an emergency-based carve-out to this general rule, allowing for removal of children without a hearing in an emergency. In these instances, parents generally have the right to a post-deprivation hearing within a proscribed time frame. *See, e.g.*, N.Y. FAM. CT. ACT § 1028 (McKinney 2010). Most states provide that parents facing a disruption in the care of their children also have the right to other protections at this stage, such as a right to counsel. *See, e.g.*, N.Y. FAM. CT. ACT § 262 (McKinney 2012); 705 ILL. COMP. STAT. ANN. 405/1-5(1) (West 2014); 42 PA. CONS. STAT. ANN. §6337 (2012) (detailing right to counsel in any abuse proceeding, including a child abuse and/or neglect proceeding). *But see* MINN. STAT. ANN. § 260C.163, subd. 3(b) (West 2017) (stating that the appointment of counsel is discretionary). The Supreme Court has yet to hold that access to counsel is constitutionally required. *See* Lassiter v. Dep't of Soc. Servs. of Durham Cty., N.C., 452 U.S. 18, 32 (1981) (holding that even at termination, whether counsel is required is a case-by-case determination).

30. *See, e.g.*, N.Y. FAM. CT. ACT § 1027 (McKinney 2016).

31. *See supra* note 28. This entire process can take months, or it can happen in a day. For some families, the agency initially decides to provide services to the family without removing the children or turning to the courts, only to change that decision later.

32. *See, e.g.*, N.Y. FAM. CT. ACT § 1028(b) (McKinney 2010).

33. *See generally* KRAMER, ACS IN OVERDRIVE, *supra* note 28, at 1 (describing how ACS has "drastically increased the number of families it brings into the system, filing more cases in Family Court and placing more children in foster care").

34. *See* Adoption Assistance and Child Welfare Act of 1980, Pub. L. No. 96–272, 1980; Adoption and Safe Families Act (ASFA) of 1997, Pub. L. No. 105–89 §101, 1997.

35. *See, e.g.*, 42 U.S.C. § 675 (2018); N.Y. FAM. CT. ACT § 1089 (McKinney 2018).

36. *See supra* note 35.

Here the agency encounters another decision point: whether and when to seek to reunify the child with his or her parents.

Up to and including this point, the court in many states will not yet have ruled as to whether the agency has shown that the parents did the alleged acts on which the case is based — whether the child actually was abused or neglected. This decision is made by the court at a "fact finding" hearing.[37] At this hearing, the agency has the burden to prove its allegations.[38] If the court determines that the agency failed to make its case, the petition will be dismissed. If the court finds that the agency carried its burden to show that the child was abused or neglected, however, the case proceeds to a dispositional hearing where a plan for the family is set.[39] This plan can mean that the family remains together (or is reunited), that the child enters or remains in foster care, or that the parents or family must engage in specific services to remain together (or be reunited).[40] If the child is placed in foster care, the case will proceed to or be scheduled for a permanency hearing.[41]

If the child does not return home, the final decision point is whether the agency should move to terminate the parent's rights. Under the Adoption and Safe Families Act (ASFA), agencies may be required to move for termination if a child remains in foster care for fifteen out of the most recent twenty-two months.[42] There are, however, exceptions to this requirement that the agency may invoke to avoid its obligation to move for termination.[43] Ultimately, the court will make the termination decision, but it is the agency that makes the initial choice whether to move forward towards termination or to invoke an exception to ASFA.[44]

Throughout this process, it should be noted that the child protective agencies are tasked with playing a conflicting, dual role:

---

37. States may refer to this hearing by different names, including "fact finding" or "adjudication." *See, e.g.*, Child and Family Services Review Information Portal, Adjudication Hearing, https://training.cfsrportal.acf.hhs.gov/section-2-understanding-child-welfare-system/3022 [https://perma.cc/6764-NMGT].

38. *See, e.g.*, *In re* Negus T., 996 N.Y.S.2d 544, 545 (2d Dep't 2014); N.Y. FAM. CT. ACT § 1046 (McKinney 2009).

39. *See, e.g.*, N.Y. FAM. CT. ACT §§ 1051–52 (McKinney 2016).

40. *See id.* § 1052.

41. *See, e.g.*, *id.* § 1055.

42. 42 U.S.C. § 675(5)(E).

43. These include situations where the child is being cared for by a relative, where the agency documents that there is a "compelling reason" why filing for termination "would not be in the best interests of the child," or the State has failed to provide the necessary services to accomplish the goal of reunification. *Id.*

44. *See id.*

they are both support and service provider, as well as investigator, prosecutor, and enforcer.[45]

B. Flaws in Human Decision-Making

At each of the decision points discussed above, a community member, the child welfare agency, or the court is called upon to make a vital, highly consequential decision. Of these actors, it is the agencies that have the most control, deciding which cases to investigate, which to bring to court, when and how to provide services or remove children from their homes, and how to proceed through the court process. It also is the agencies that are implementing predictive analytics. For this reason, this section focuses on the agency and asks: what do we know about the quality of the agency's decision-making?

There is ample evidence that the way child welfare agencies make decisions is deeply flawed.[46] Social science shows that workers tasked with making important child protective decisions often fall victim to the effects of heuristics and bias.[47] As a systemic issue, there are also trends that indicate that child welfare decision-making can and should be improved. Among those, we can count the racial disproportionality evident in child welfare systems nationwide, as well as the repeated occurrence of "foster care panics" — when agencies ramp up investigations and removals in the wake of tragedies.

---

45. *See* DOROTHY ROBERTS, SHATTERED BONDS: THE COLOR OF CHILD WELFARE 148 (Reprint ed. 2002); *see also* Wendy A. Bach, *The Hyperregulatory State: Women, Race, Poverty, and Support*, 25 YALE J.L. & FEMINISM 317, 353–54 (2014).

46. Anecdotally, anyone who has worked in a family court can rattle off a handful of stories on command that detail problems ranging from a lack of attention or information on the part of a caseworker, to the disconnected override of a fact-based, on-the-ground decision by a removed, faceless superior, to at times seemingly targeted, malicious harassment. For example, it was not an infrequent occurrence for lawyers in my office to meet parents on the first day of their family court case who had just come from a meeting with the agency, called a "child safety conference." Sometimes, the decision in the room at that conference — arrived at after a lengthy discussion among the case worker, facilitator, parent, and sometimes the child or other person intimately involved with the family — was that the child would be safe remaining at home with certain supports in place. But when the caseworker stepped out of the room to run this conclusion by his or her supervisor, who often had never met the family, that decision was changed unilaterally.

47. *See infra* notes 48–52 and accompanying text.

1. The Social Science of Agency Decision-Making

Human decision-makers are known to be fallible. Researchers across many academic disciplines have studied human decisions and identified a number of "common errors" to which we fall victim.[48] Perhaps one of the most documented logical errors to which human decision-makers succumb is confirmation bias — the tendency to readily assimilate or recall evidence that supports a previously-held belief, and to discount evidence that contradicts that belief.[49] People also are swayed by "framing effects."[50] This means that people make different decisions based on how choices are presented to them: their decision changes based on the reference point against which it is evaluated.[51] And human decision-makers are affected by the availability heuristic: the tendency to overestimate the importance or likelihood of an event that is "cognitively available" or, as Cass Sunstein put it, "vivid and salient."[52]

Expertise is not corrective and can even have an adverse effect. Indeed, experts have been found to be "particularly bad at prediction."[53] As noted psychologist Daniel Kahneman has shown, experts across fields, "from stock brokers to job recruiters," make poor decisions and are inconsistent in their choices.[54] Only where experts receive "frequent and immediate" feedback on the outcomes of their decisions do they tend to improve.[55]

When it comes to making risk predictions, particular errors emerge. Among these are findings that human decision-makers have difficulty accurately assessing the likelihood that a particular event will occur.[56] Decision-makers tend to rely insufficiently on the base

---

48. *See, e.g.*, Amy D'Andrade et al., *Risk and Safety Assessment in Child Welfare*, 5 1/2 J. EVIDENCE-BASED SOC. WORK 31, 3–4 (2008).

49. *See, e.g.*, Raymond S. Nickerson, *Confirmation Bias: A Ubiquitous Phenomenon in Many Guises*, 2 REV. GEN. PSYCHOL. 175, 175 (1998).

50. Daniel Kahneman & Amos Tversky, *The Psychology of Preferences*, 246 SCI. AM. 160, 166 (1982).

51. *Id.*

52. Cass Sunstein, *Precautions Against What? The Availability Heuristic and Cross-Cultural Risk Perception*, 57 ALA. L. REV. 75, 75 (2005) (noting that the availability heuristic "can make some risks stand out as particularly salient, regardless of their actual magnitude").

53. Marsha Garrison, *Taking the Risks out of Child Protection Risk Analysis*, 21 J.L. & POL'Y 5, 19 (2012).

54. *Id.* (citing DANIEL KAHNEMAN, THINKING FAST AND SLOW 209–44 (2011)).

55. *Id.*

56. *See generally* ANGELA WHITE & PETER WALSH, CTR. FOR PARENTING & RESEARCH, RISK ASSESSMENT IN CHILD WELFARE (2006),

rate of an event's occurrence when attempting predictions about "uncommon" events.[57] They often overestimate their own ability to predict a future event and have a hard time assigning appropriate weight to factors related to a given decision.[58] And "illusory correlations," or the "tendency to see two events as related when they are not," can affect the accuracy of human predictions.[59]

Researchers have studied child welfare workers directly, and the evidence suggests that each of these errors is present in their decisions. A 1999 study, for example, found great variability among case workers in assessing the need for removal of the same child.[60] Case workers were found to suffer from a version of the availability heuristic, basing their decisions on a "limited range of data" by focusing more on more memorable experiences and failing to access information about less salient occurrences.[61] Another study reviewed forty-five child abuse inquiry reports published in Britain between 1973 and 1994, and found that social workers in these cases made "three general types of errors."[62] These errors included being "slow to revise their judgments," interpreting and weighing new evidence in light of their initial assessment of the family.[63] The case workers also exhibited confirmation bias, approaching information that conflicted with their initial assessment of a family with skepticism, but failing to apply a critical eye to new information that comported with their preexisting view.[64] Finally, the case workers gave more weight to information coming from certain sources over others. For example, doctors carried greater influence than neighbors, regardless of whether those sources truly had more or better information.[65] At each of the decision points described above, the case workers are tasked with making decisions that can have life-changing consequences for families. But scientific research shows that case workers, being human, fall victim to bias in their thinking and are not

---

http://www.community.nsw.gov.au/__data/assets/pdf_file/0005/321647/research_riskassessment.pdf [https://perma.cc/H6U4-C9UY].

57. *Id.*

58. Garrison, *supra* note 53, at 32–34.

59. WHITE & WALSH, *supra* note 56, at 4.

60. *Id.*

61. *Id.* (discussing Eileen Munro, *Common Errors of Reasoning in Child Protection Work*, 23 CHILD ABUSE & NEGLECT 745 (1999)).

62. *Id.*

63. *Id.*

64. *Id.*

65. *Id.* at 4–5.

particularly good at weighing evidence to accurately predict events (especially infrequent events such as child fatalities).

Notably, the one identified corrective for the failures of expert decision-making — frequent and immediate feedback — generally is unavailable to child protective decision-makers. The outcomes of their decisions likely will not be known for years, perhaps decades, as the child matures. The intervening effects of myriad other decisions, made by numerous actors like supervisors, other workers, and the court, make determining the actual impact of any one decision nearly impossible. And it generally is not possible to assess any particular decision in reference to the counter-factual. Where an agency determines that removal is necessary, for example, it is not possible to measure the outcome that would have occurred if that child had remained with his or her parent.

2. Disproportionality

The prior section has shown that human decision-making, and in particular human risk-prediction, is vulnerable to error. The fact that decision-makers fall victim to the heuristics described above also means that their decisions become entry-points for bias — both explicit and implicit.[66] The disproportionate number of poor children and children of color represented at every stage in the child welfare process is strong evidence of the insidious effect of bias.[67] It is well-documented that Black children are overrepresented nationally in the child welfare system.[68] Hispanic children also are overrepresented in a substantial number of state systems, and this overrepresentation is

66. As used here, explicit bias refers to the intentional act of favoring one group over another. Implicit bias is "the automatically activated evaluations or stereotypes that affect [an] individual's understanding, actions, and decisions in an unconscious manner . . . . All humans exhibit implicit bias, and having these biases does not reflect the intent to cause harm." KELLY CAPATOSTO, KIRWAN INST., FORETELLING THE FUTURE: A CRITICAL PERSPECTIVE ON THE USE OF PREDICTIVE ANALYTICS IN CHILD WELFARE 2 (2017), http://kirwaninstitute.osu.edu/wp-content/uploads/2017/05/ki-predictive-analytics.pdf [https://perma.cc/E9J3-DVN6].

67. Although disproportionality is well documented, there exists a debate in the child welfare literature around its cause. For a detailed discussion of this debate, see Tanya A. Cooper, *Racial Bias in American Foster Care: The National Debate*, 97 MARQ. L. REV. 215, 229–39 (2013). This Article does not intend to resolve this debate, but only to point out that this disproportionality exists, calling into question the legitimacy of child welfare decisions. As discussed further *infra*, without careful attention to both the underlying problem and the design of new risk assessment tools, predictive analytics will only serve to compound these problems because they are based on historical data that is, itself, infected by bias. *See infra* Part III.

68. *See* Cooper, *supra* note 67, at 224–25.

only rising.[69] The same is true for Native American children in every jurisdiction in which they reside.[70] This overrepresentation is present throughout all phases of the child welfare process: Black and Native American children are twice as likely to undergo investigation, it is twice as likely that such investigations will be found credible, and it is two or three times more likely that Black or Native American children will be placed in foster care.[71] Note that this means that these levels of disproportionality only worsen as families move through the system.[72] Poor children also are vastly overrepresented in the system.[73] As Professor Dorothy Roberts put it, "the public child welfare system in America is populated almost exclusively by poor children."[74] Because families of color are disproportionately poor,[75] these two trends are inextricably intertwined and cannot be understood in isolation.[76]

---

69. ROBERT B. HILL, CASEY-CSSP ALL. FOR RACIAL EQUITY IN CHILD WELFARE, AN ANALYSIS OF RACIAL/ETHNIC DISPROPORTIONALITY AND DISPARITY AT THE NATIONAL, STATE, AND COUNTY LEVELS (2007), https://www.aecf.org/m/resourcedoc/aecf-AnalysisofRacialEthnicDisproportionality-2007.pdf [https://perma.cc/TV4B-PM33].

70. *Id.*

71. *Id.* at 10.

72. *Id.* at 9. *See also* Harris & Hackett, *supra* note 11, at 211 (conducting study of racial disproportionality across decision points in King County, Washington child welfare system, and finding that "[a]t each successive decision point, the gap in outcomes widens according to race"). In Allegheny County, Pennsylvania, the jurisdiction at the forefront of developing its own predictive analytic tool for child welfare, disproportionality is evident throughout a family's involvement in the system. *See* ALLEGHENY CTY. DEP'T OF HUMAN SERVS., DATA BRIEF: RACIAL DISPROPORTIONALITY IN ALLEGHENY COUNTY'S CHILD WELFARE SYSTEM 2 (Sept. 2017). Notably, the largest share of Allegheny County's disproportionality problems arises at the referral stage: in 2015, Black children were 3.5 times more likely, and bi- or multi-racial children were 2.8 times more likely, than white children to be referred to the county's child protective service for an investigation. *Id.*

73. *See, e.g.*, Harris & Hackett, *supra* note 11, at 201.

74. Dorothy E. Roberts, *Poverty, Race, and New Directions in Child Welfare Policy Access to Justice: The Social Responsibility of Lawyers*, 1 WASH. U. J.L. & POL'Y 63, 64 (1999).

75. *See* Cooper, *supra* note 67, at 229.

76. *See, e.g.*, Schuette v. Coal. to Defend Affirmative Action, 134 S.Ct. 1623, 1676 (2014) (Sotomayor, J., dissenting) ("Race also matters because of persistent racial inequality in society — inequality that cannot be ignored and that has produced stark socioeconomic disparities."); Gratz v. Bollinger, 539 U.S. 244, 298–99 (2003) (Ginsburg, J., dissenting) ("[W]e are not far distant from an overtly discriminatory past, and the effects of centuries of law-sanctioned inequality remain painfully evident in our communities and schools. In the wake of a system of racial caste only recently ended, large disparities endure" in areas such as unemployment, poverty, and access to health care (internal quotations and citations omitted)); *see also* Michele Estrin Gilman, *The Class Differential in Privacy Law*, 77 BROOK. L. REV.

3. Foster Care "Panics"

Further evidence of the failures of human decision-making comes in the form of "foster care panics" — the tendency of child welfare systems to engage in a frenzied push toward removal of children from their homes following the highly publicized death of a child.[77] New York City has recently seen a version of this phenomenon following the deaths of six-year-old Zymere Perkins and three-year-old Jaden Jordan, which happened just months apart.[78] New York had done admirable work to reduce its foster care population over the years preceding these deaths, but after they occurred, abuse and neglect reports, investigations, and case filings skyrocketed.[79] In the months and years following these tragedies, New York's child protective authority, the Administration for Children's Services (ACS), investigated drastically more cases than before.[80] Once families were under investigation, ACS was far more likely to pursue action in court, and the number of emergency removals increased by thirty percent.[81] Advocates for families on the ground during this panic reported that ACS seemed to be making unjustified removals more frequently.[82] This anecdotal reporting is borne out by the evidence.

1389, 1394 (2012) ("Because poor Americans are disproportionately minority and female, it is impossible to talk about class without taking into account how subordination is linked to race and gender.").

77. *See, e.g.*, Petula Dvorak, *Child Deaths Led to Excessive Foster Care Placements, Critics Say*, WASH. POST (Jan. 8, 2009), https://cdn.ymaws.com/www.law.udc.edu/resource/resmgr/fraidin/fraidin_washpost_010809.pdf [https://perma.cc/C38Q-L2QX].

78. Yoav Gonen, *Child Abuse Reports Surge After High-Profile Tragedies*, N.Y. POST (July 24, 2018), https://nypost.com/2018/07/24/child-abuse-reports-surge-after-high-profile-tragedies/ [https://perma.cc/23LS-JURF].

79. KRAMER, ACS IN OVERDRIVE, *supra* note 28, at 2. According to a later report by this same author, referrals to New York's child maltreatment hotline increased by nearly eleven percent in the twenty months following Perkins's death when compared with the twenty-month period that preceded it. ABIGAIL KRAMER, NEW SCH. CTR. FOR N.Y.C., CHILD WELFARE SURGE CONTINUES: FAMILY COURT CASES, EMERGENCY CHILD REMOVALS REMAIN UP 2, 4 (2018), https://static1.squarespace.com/static/53ee4f0be4b015b9c3690d84/t/5b50c5c38a922d3211b6606d/1532020164221/Child+Welfare+Surge+Continues.pdf [https://perma.cc/AU56-PFSB].

80. KRAMER, CHILD WELFARE SURGE CONTINUES, *supra* note 79, at 1.

81. *Id.* ("Between October 2016 and May 2018, ACS reported approximately 2,300 emergency removals — a 28 percent jump from a corresponding 20-month span before the crisis."). These statistics do not capture the true numbers of children separated from their families during this "panic." According to Kramer, "ACS does not publicly report emergency removal cases in which a child is reunited with his or her family at the initial court hearing." *Id.*

82. *Id.* at 3.

Although ACS's use of its emergency removal power jumped considerably, the foster care population ultimately did not see an equally drastic increase.[83] The number of children placed in foster care in the ten months following Perkins's death rose by about four percent,[84] indicating that the court system may have acted as a somewhat effective check on the agency's panic. The trauma of these emergency removals, however, had already been inflicted.[85]

This type of foster care panic is not new, not even to New York City itself. New York experienced a similar panic in 1996 following the death of Elisa Izquierdo, and again in 2006 following the death of Nixmary Brown.[86] Illinois experienced its own version of the phenomenon in the 1990s, following the 1993 death of three-year-old Joseph Wallace.[87] Countless other cities and states have cycled through their own foster care panics.[88]

Although these societal reactions to high-profile child deaths may be understandable, they are not logical. There is no evidence that levels of child abuse or neglect in the population increase during these periods,[89] but there is ample evidence that foster care panics actually make children less safe. In Illinois, for example, child welfare authorities abruptly abandoned the state's developing family preservation program, Families First, in the wake of Wallace's death in 1993.[90] Following this move, deaths of children attributed to abuse

83. *Id.* at 2.

84. *Id.*

85. *See* Sankaran & Church, *supra* note 2, at 211–12.

86. *Once Again, New York City Children Pay the Price of Foster Care Panic*, NCCPR CHILD WELFARE BLOG (Dec. 15, 2016), https://www.nccprblog.org/2016/12/once-again-new-york-city-children-pay.html [https://perma.cc/CS54-F73T] [hereinafter *Once Again, New York City Children Pay the Price*].

87. ROBERTS, *supra* note 45, at 147–48.

88. *See, e.g.*, NAT'L COAL. FOR CHILD PROT. REFORM, FOSTER CARE PANICS: ISSUE PAPER 2, https://drive.google.com/file/d/0B291mw_hLAJsMGdWQ2NhYTM2Ulk/view [https://perma.cc/ES6G-J7WJ] (documenting Illinois's and New York's experience as well as a panic set off in 1999 in Florida following the death of a child "known to the system") [hereinafter NCCPR, *Foster Care Panics*]; Dvorak, *supra* note 77 (describing Washington, D.C. panic).

89. In fact, children in New York may have been safer in the months surrounding Zymere Perkins's death than at any time in recent years. *Once Again, New York City Children Pay the Price*, *supra* note 86. ("The percentage of children reabused in 2016 was the lowest since 2009. The rate of foster-care recidivism was the lowest since 2006.").

90. ROBERTS, *supra* note 45, at 147; *see also* NCCPR, *Foster Care Panics*, *supra* note 88, at 1.

rose from seventy-eight prior to the cancellation of Families First to ninety-one in 1997.[91] Sharp increases in investigations increase worker case-loads, resulting in higher worker attrition and diverting attention from those children and families that may actually need support.[92] This means that less experienced workers are making decisions about children's safety with less time and under increased pressure.[93] More children are needlessly exposed to the trauma of removal. Case workers operating during panics admit that decisions made during these times are not based on sound assessments of children's safety, but instead on dangerous, self-protective mentalities. As one caseworker put it:

> At this point everybody's so afraid, they'd rather cover their ass . . . . Take the case to court and let the judge say no. Then we can document we tried. Nobody wants to end up with their face in the Daily News. They don't want to face criminal charges.[94]

It is clear that responsible agency decision-making breaks down during foster care panics.

C. Decision-Making Models Employed by Child Welfare Systems

In part because of concerns regarding inconsistent, biased, and otherwise flawed decision-making, child protective agencies have shifted from relying on clinical decision-making models to ever more formalized and structured models. Predictive analytics represent the latest and perhaps most powerful step in this shift, but the advent of these tools is an extension of a process that has been ongoing for many years.[95] Inspired by public health methodologies, child welfare agencies across the country have already incorporated a micro-level version of an algorithm into their decision-making processes, in the form of "structured decision-making" models, or "SDM."[96]

In contrast to older, clinical models that require experienced, trained clinicians to make individualized judgments in each case,

---

91. ROBERTS, *supra* note 45, at 148.

92. KRAMER, ACS IN OVERDRIVE, *supra* note 28, at 4–5; KRAMER, CHILD WELFARE SURGE CONTINUES, *supra* note 79, at 3.

93. KRAMER, ACS IN OVERDRIVE, *supra* note 28, at 4–5; KRAMER, CHILD WELFARE SURGE CONTINUES, *supra* note 79, at 3.

94. KRAMER, CHILD WELFARE SURGE CONTINUES, *supra* note 79, at 3.

95. *See* Garrison, *supra* note 53, at 17.

96. *Id.* at 25 ("[A]ctuarial risk assessment has, over time, come to dominate the field."); D'Andrade et al., *supra* note 48, at 32 ("Most states in the US formalize the process of assessing risk by using some type of structured decision-making process or tool.").

SDM models are themselves a version of an algorithm.[97] They provide case workers with a structured guide for arriving at decisions. SDM tools often take the form of a questionnaire, which the case worker or team will fill in with answers according to their on-the-ground knowledge of the family.[98]

These tools come in two forms: consensus-based and actuarial. Consensus-based questionnaires can be lengthy, and include questions about a broad range of topics.[99] Many of the questions require highly subjective judgments on the part of the case worker, and grant the worker a fair amount of flexibility.[100] Actuarial questionnaires "employ the methods of epidemiology" by using samples of actual substantiated cases, through which researchers "determine which case characteristics are significant predictors of filing and substantiation recurrence."[101] In other words, these actuarial risk assessments are precursors to the predictive analytic tools being developed today. They require case workers and teams to fill out a "score sheet" by answering questions geared to elicit specific information.[102] For example, the score sheet may ask for the type of allegation (e.g., abuse or neglect), demographic information about the family (e.g., the number of children or adults in the home and their ages), or the number of prior cases the family has had.[103] But such actuarial tools also may demand subjective judgments, such as the worker's perception of the primary caretaker's level of motivation to improve parenting skills, whether the parent "viewed the situation less seriously than the investigator," or "failed to cooperate satisfactorily."[104]

Actuarial structured decision-making tools have their advantages and have been found to lead to more consistent decision-making in child welfare cases relative to consensus-based decision-making models.[105] However, actuarial tools also present risks. There is a

---

97. *See* Garrison, *supra* note 53, at 22–23.

98. *Id.* at 22.

99. *See id.*

100. *Id.* One example of a consensus-based tool is the Washington Risk Assessment Matrix (WRAM), which "requires inquiry into thirty-seven different issues, some of which (e.g., caregiver-child relationship) are difficult to measure objectively." *Id.*

101. *Id.* at 23.

102. *Id.*

103. *Id.*

104. *Id.* at 24 fig.2, 28–29.

105. *Id.* at 23. Studies conducted by the developer of one of the most widely-used SDM tools also found that SDM did a "significantly better job in predicting

fundamental flaw in attempting to graft public health methods of epidemiological prediction onto child protective work.[106] Public health as a field concentrates on populations, not on individuals, and seeks to offer prevention programs, not treatment. Imposing public health tools on child protective models thus poses significant risks, as child welfare generally aims to make more difficult, individual-level predictions about each child.[107] To date, child protection has oriented itself as a "blame and cure system," which focuses resources on investigating and, when needed, prosecuting individual reports, rather than seeking to implement broader, population-level prevention programs.[108]

The actuarial questionnaires themselves also present serious problems. First, actuarial models have been developed without a standard definition of abuse, neglect, or maltreatment.[109] Without such standardized definitions, it is impossible to be sure what is being measured.[110] Second, these tools often incorporate substantial subjectivity, and thereby "reintroduce all the problems with intuitive judgments that decision-making algorithms were designed to avoid."[111] Third, actuarial SDM models tend to focus exclusively on "negative risk factors," and are static, meaning they do not incorporate changes in the family's situation easily.[112] Lastly, researchers have expressed concern that actuarial tools are not used

---

subsequent investigation and substantiation of maltreatment" than consensus-based models. *Id.* at 23–25. These studies also found that SDM tools made it more likely that an agency would close a low-risk case and involve high-risk families (as defined by the tools) in services and that high-risk families tended to have fewer subsequent substantiated reports and lower rates of foster care placement. *Id.*; *see also* WASH. STATE INST. FOR PUB. POLICY, STRUCTURED DECISION MAKING® RISK ASSESSMENT: DOES IT REDUCE RACIAL DISPROPORTIONALITY IN WASHINGTON'S CHILD WELFARE SYSTEM? 2 (2011), http://www.wsipp.wa.gov/ReportFile/1086/Wsipp_Structured-Decision-Making-Risk-Assessment-Does-it-Reduce-Racial-Disproportionality-in-Washingtons-Child-Welfare-System_Full-Report.pdf [https://perma.cc/VC35-DN6C] (describing finding that the consensus-based WRM "was not a good predictor of future abuse or neglect").

106. Garrison, *supra* note 53, at 7–17.

107. *Id.* at 8. For example, we may be confident that smoking is highly correlated with lung cancer, and successfully target prevention efforts at all smokers. However, we cannot predict that any one particular smoker will develop lung cancer. *Id.* at 16–17.

108. *Id.* at 5, 7.

109. *Id.* at 25–26.

110. *Id.* at 26.

111. *Id.* at 29.

112. *Id.* at 29–30.

properly.[113] For instance, research into the ways these tools are used in practice has revealed that case workers or other users may enter information only after a decision has already been reached, changing their answers to support their conclusion.[114]

II. PREDICTIVE ANALYTICS IN CHILD WELFARE

This Article has thus far demonstrated that child welfare decision-making must be improved. As the system stands today, even experts tasked with making vital child welfare decisions fall victim to the influence of heuristics and bias. Children of color and children from poor households are vastly over-represented in the system, and the system itself is vulnerable to panics that unjustifiably pull children away from their homes in record numbers. Tools developed to combat some of these problems have thus far failed to do so successfully, and present their own slew of risks, including undefined standards, subjectivity, focus on static, negative risk factors, and misuse.

A tempting prospect has emerged in response to this reality — using the growing field of "predictive analytics" to improve child welfare decision-making. But can machines actually help us in this complex and fundamentally human endeavor? To begin to answer this question, we must first understand what predictive analytics are and how they function. This Part provides an overview of the processes that produce predictive algorithms. It shows that these efforts are not "purely scientific," but instead are the result of myriad, fundamentally human choices that "packag[e] policy as data science."[115] It then goes on to describe some of the predictive analytics tools currently in use in child protective efforts.

---

113. Philip Gillingham, *Predictive Risk Modelling to Prevent Child Maltreatment and Other Adverse Outcomes for Service Users: Inside the 'Black Box' of Machine Learning*, 46 BRIT. J. SOC. WORK 1044, 1045 (2016) ("[A] criticism has been that even the best risk-assessment tools are 'operator-driven' as they need to be applied by humans. Research about how practitioners actually use risk-assessment tools has demonstrated that there is little certainty that they use them as intended by their designers." (internal citations omitted)).

114. *Id.*

115. Julie Ciccolini & Cynthi Conti-Cook, *Rationing Justice: Risk Assessment Instruments in the American Criminal Justice System*, EUROPENOW (Nov. 8, 2018), https://www.europenowjournal.org/2018/11/07/rationing-justice-risk-assessment-instruments-in-the-american-criminal-justice-system/ [https://perma.cc/NYR2-PET5].

A. Predictive Analytics: An Overview

Predictive analytics is an overarching term that broadly describes the creation of algorithms designed to predict specific outcomes based on historical data. In its most basic form, an "algorithm" is simply a set of instructions for solving a problem.[116] The creation of an algorithm may be the result of traditional statistical modeling or machine learning processes, which are automated processes geared toward "discovering correlations . . . between variables in a dataset."[117]

By using these statistical or machine learning approaches, developers mine vast amounts of "administrative data" — data gathered for operational purposes and held by various government entities — to attempt to discover correlations: tendencies for certain factors to occur together.[118] The data mined by these systems in the child welfare context often includes data arising directly out of the child welfare agency's involvement with the family, such as past substantiated reports.[119] Depending on the jurisdiction, however, the

116. Jennifer Golbeck, *How to Teach Yourself About Algorithms*, SLATE (Feb. 9, 2016), https://slate.com/technology/2016/02/how-to-teach-yourself-about-algorithms.html [https://perma.cc/R7TP-GYXF]. *See also* Solon Barocas & Andrew D. Selbst, *Big Data's Disparate Impact*, 104 CALIF. L. REV. 671, 674 n.10 (2016) ("An 'algorithm' is a formally specified sequence of logical operations that provides step-by-step instructions for computers to act on data and thus automate decisions."); Diamond v. Diehr, 450 U.S. 175, 186 (1981) ("We defined 'algorithm' as a 'procedure for solving a given type of mathematical problem.'" (quoting Gottschalk v. Benson, 409 U.S. 63, 65 (1972))). In the context of machine learning, an algorithm specifically is "a set of mathematical steps for achieving . . . learning . . . when exposed to historical or example data." David Lehr & Paul Ohm, *Playing with the Data: What Legal Scholars Should Learn About Machine Learning*, 51 U.C. DAVIS L. REV. 653, 671 (2017).

117. Lehr & Ohm, *supra* note 116, at 671. In his book *Predictive Analytics*, Eric Siegel describes "machine learning" as the "academic term" for predictive analytics, stating that, in "commercial, industrial, and government applications," this process is "called something else": predictive analytics. ERIC SIEGEL, PREDICTIVE ANALYTICS: THE POWER TO PREDICT WHO WILL CLICK, BUY, LIE, OR DIE 33 (2016). Siegel goes on to define predictive analytics as "[t]echnology that learns from experience (data) to predict the future behavior of individuals in order to drive better decisions." *Id.* While the conclusion of this definition, that predictive analytics drive "better decisions," involves a complicated value judgment implicated by the discussion in Part III, *infra*, the description of predictive analytics as "technology that learns from experience" in the form of data in order to "predict future behavior" serves us well.

118. *See* David C. Vladeck, *Consumer Protection in an Era of Big Data Analytics*, 42 OHIO N.U. L. REV. 493, 494 (2016) ("Decisions that matter will now be based on correlations, not hard facts.").

119. *See* RHEMA VAITHIANATHAN ET AL., DEVELOPING PREDICTIVE RISK MODELS TO SUPPORT CHILD MALTREATMENT HOTLINE SCREENING DECISIONS: ALLEGHEY COUNTY METHODOLOGY AND IMPLEMENTATION (Apr. 2017),

data may have been gathered by the government from any number of public touchpoints, such as criminal justice, public benefits, public hospitals, birth records, education records, and more.[120]

Where machine learning approaches are used, predictive algorithms are developed by a multi-stage process.[121] Each step requires developers to make a series of choices that implicate their knowledge of the system in which they operate, their own biases or value judgments, and the biases and value judgments of the community directing their work.

The effort of designing any predictive risk model begins with the problem definition, in that a predictive model "must predict or estimate *something*, and the first step of any analysis is to define what that something should be."[122] The prediction the tool makes can be called an "outcome of interest" or a "target variable."[123] Selecting a "target variable" is an inherently subjective and open-ended design process, requiring data scientists to translate the amorphous "outcome of interest" identified into a question "that computers can parse."[124]

The question of what outcome a model is predicting is fundamental to the validity of any predictive tool.[125] In the "learning" process, the "outcome variable acts as a teacher."[126] For computers to predict a particular outcome successfully, that outcome must be well-defined and must relate well to on-the-ground reality, or "ground truth."[127]

---

https://www.alleghenycountyanalytics.us/wp-content/uploads/2017/04/Developing-Predictive-Risk-Models-package-with-cover-1-to-post-1.pdf [https://perma.cc/JJV4-VRBS].

120. *See id.*

121. *Id.* These stages are not necessarily linear; the process of developing any given model can and should dance back and forth across and between these stages, from focusing on design back to data, and back and forth again. Lehr & Ohm, *supra* note 116, at 672–73. Decisions made at each stage affect and are affected by decisions made at earlier and later stages. *Id.* at 671.

122. *Id.* at 672–73.

123. Barocas & Selbst, *supra* note 116, at 678.

124. *Id.*

125. Gillingham, *supra* note 113, at 1052 (describing the "particular challenge in applying predictive machine learning techniques" to child welfare, of "finding valid and reliable outcome variables within data about service activity").

126. *Id.* at 1052.

127. CHRISTOPHER KINGSLEY & STEFANIA DI MAURO-NAVA, METROLAB NETWORK, FIRST DO NO HARM: ETHICAL GUIDELINES FOR APPLYING PREDICTIVE TOOLS WITHIN HUMAN SERVICES (2017), https://metrolabnetwork.org/wp-content/uploads/2017/09/Ethical-Guidelines-for-Applying-Predictive-Tools-within-Human-Services_Sept-2017.pdf [https://perma.cc/KY3W-GDDB].

Otherwise, the predictive tool will learn unhelpful rules and make less accurate predictions.[128]

Once the problem is defined, the data that will make up the fibers of the model must be collected. As part of this stage, developers determine what sources of data are available to them and how they will be used.[129] Once data is collected, developers must go through a series of processes to prepare it for use in the model. Developers may be called upon to make normative decisions while preparing the data that affect the later functioning of the model.[130] For instance, they may need to decide where to populate missing data with average terms, or where to exclude data when some elements are missing.

Once the data is in good shape, the development team can move to designing the algorithm itself. Although algorithms often are discussed as "black boxes"[131] that spit out predictions based on hidden, impenetrable processes, they actually are the product of intentional design choices by their creators. Data scientists have developed various categories or types of algorithmic models from which developers can choose. None is inherently better than all others; each has particular strengths and drawbacks.[132] For example, one type of algorithm may be designed more simply and therefore be easier to explain to a lay audience, but its predictions may be less accurate. Others may generate more accurate predictions, but the processes by which they arrive at those predictions may be vastly more difficult (or impossible) to explain.[133]

Once selected, a machine learning algorithmic model must be "trained." During this process, the "algorithm is run on the training data set and, in the process, learns rules for predicting the outcome."[134] As the algorithm runs, sometimes many times over,

---

128. *See* Lehr & Ohm, *supra* note 116, at 676.

129. *See id.* at 679.

130. *See* Jessica M. Eaglin, *Constructing Recidivism Risk*, 67 EMORY L.J. 59, 80 (2017).

131. *See, e.g.*, FRANK PASQUALE, THE BLACK BOX SOCIETY 3 (2015) (describing an algorithm as "a system whose workings are mysterious; we can observe its inputs and outputs, but we cannot tell how one becomes the other").

132. *See* Emily Berman, *A Government of Laws and Not of Machines*, 98 B.U. L. REV. 1277, 1288 (2018).

133. Lehr & Ohm, *supra* note 116, at 688–90.

134. *Id.* at 695–96. In the case of the predictive risk model employed by the Allegheny County research team, the "algorithm 'learns' by calculating the correlation between each predictor, or independent, variable (a piece of information about the child, parent or parent's partner) and the outcome, or dependent, variable (a substantiation or not of maltreatment by age five) across all the individual cases in the training data set." Gillingham, *supra* note 113, at 1048.

data scientists assess and adjust it in various ways. Each adjustment requires the developers to make subjective and value-laden judgments. One way in which developers adjust models during training is to engage in "feature selection." Of the data points available to input into the model, the developers must "prune" away those that are not required in the final model, leaving only those that are necessary for the model to run effectively.[135]

Finally, an algorithm that has been designed, trained, and tested becomes a "running model," graduating from its training and test data, to begin making predictions based on live data in the real world.[136] After an algorithm has been deployed, this process of adjustment and re-adjustment can and should continue, as flaws, biases, and other issues with the algorithm become apparent. Additionally, real-world conditions may change, affecting the accuracy of even the best-trained model.[137]

B. The Current State of Affairs

Numerous jurisdictions around the country already are experimenting with and using predictive analytic models in their child protective efforts, in a variety of ways and likely to varied effect. As of 2018, there was publicly-available evidence that child protective authorities in more than a dozen states were using or developing predictive analytic tools.[138] Many other jurisdictions are interested in

---

135. Lehr & Ohm, *supra* note 116, at 700–01.

136. *Id.* at 670.

137. Cecelia Klingele, *The Promises and Perils of Evidence-Based Corrections*, 91 NOTRE DAME L. REV. 537, 576 (2015) ("[T]o retain their accuracy, risk instruments must be constantly re-normed for changing populations and subpopulations.").

138. New York City, for example, is developing its own tools in-house. According to agency officials, New York's Administration for Children's Services (ACS) is using its model primarily as a "quality assurance" vehicle. Child Welfare Organizing Project & New School Center for New York City Affairs, *Automating Inequality in Child Welfare Using Predictive Analytics Panel*, FACEBOOK (July 17, 2018), https://www.facebook.com/166753923391170/videos/1798471443552735/ [https://perma.cc/X95L-MARD] [hereinafter *Automating Inequality Panel*]. As described by ACS Deputy Commissioner Andrew White in 2018, ACS's "analytic team" created a predictive tool that "does a pretty excellent job predicting the likelihood that a child will experience severe physical harm or sexual abuse at some point in the two years following the start of an investigation." *Id.* New York's ACS has yet to publish details or studies of this tool and its performance. Similarly, Wisconsin undertook an effort to develop a model to help the agency allocate post-reunification services. *Id.* The Wisconsin model sought to assess the likelihood of a child's reentry into foster care after reunification, which the agency would then use to decide whether families would receive services upon reunification. Families with higher scores were eligible for services, and those with lower scores were not. *Id.*

or considering implementation of these tools.[139] The two most well-documented predictive tools currently in use by child welfare agencies are those developed by and used in Allegheny County, Pennsylvania, called the Allegheny Family Screening Tool, or AFST,[140] and a tool developed by a Florida non-profit, the Eckerd Rapid Safety Feedback tool.

1. The Allegheny Family Screening Tool

Allegheny County, Pennsylvania, is at the forefront of using predictive analytics in its child protective efforts. In 2014, Allegheny County's Department of Human Services (DHS) put out a request for proposals, seeking partners with whom to develop a tool based on the data available to the county.[141] The county ultimately selected a team of researchers from Auckland, New Zealand, and California, who developed the Allegheny Family Screening Tool (AFST).[142] The AFST went into use in August 2016.[143]

Allegheny County's DHS uses the AFST at the screening stage, when call center workers must determine whether to "screen in" a call.[144] The research team stated that its objective was "to develop a decision aid to support hotline screeners in determining whether a maltreatment referral is of sufficient concern to warrant an in-person investigation," and not to "replace human decision-making."[145]

The AFST pulls together data from a number of county systems that document families' relationships with public services in nearly every area of life.[146] In addition to data documenting a family's prior involvement with child protective authorities, the AFST pulls data from the County's jail, juvenile probation, public welfare, behavioral

---

139. Interview with Marina Martin, Tech. & Democracy Fellow, New Am. Found. (Aug. 2, 2018).

140. *See* VAITHIANATHAN ET AL., *supra* note 119, at 1.

141. *Id.*

142. *Id.*

143. *Id.*

144. *Id.* at 4.

145. *Id.* The research team has been relatively open with its process, publishing a report detailing much of its work. *Id.*; *see also* Robert Brauneis & Ellen P. Goodman, *Algorithmic Transparency for the Smart City*, 20 YALE J.L. & TECH. 103, 146 (2018).

146. Allegheny County started this process in a slightly different place than many jurisdictions around the country. Before designing and implementing its predictive risk-modeling tool, Allegheny County had already created an "integrated client service record and data management system" that brought together data from various government service touchpoints. VAITHIANATHAN ET AL., *supra* note 119, at 5.

health, and census systems.[147] This data includes, among other things, dates of past bookings into the Allegheny County Jail or past involvement with the Allegheny County Juvenile Probation Office, whether and when a family received public benefits such as Temporary Assistance for Needy Families (TANF) or the Supplemental Nutrition Assistance Program (SNAP, formerly known as food stamps), whether and when an individual received behavioral health services, including diagnoses, as well as information such as a family's zip code, linked with Census information on the poverty status of each zip code area.[148]

The AFST algorithm analyzes 112 variables drawn from these data sets, and pulls together these data points to return a "risk score" for each specific child with one of two outcomes in mind: re-referral and placement.[149] The re-referral model predicts the likelihood that a child who is "screened out" during the initial call will be re-referred to the agency by someone in the community within two years of that screen-out.[150] The placement model is designed to predict the likelihood that a child who is "screened in" will be removed from his or her home within two years.[151] For each model, the risk score is reported on a scale from 1 to 20.[152] Each number along the scoring continuum represents the strata of risk in which the child's score sits, with 20 representing that "the child is in the top 5% of risk scores" from the model.[153]

As a practical matter, the AFST is incorporated into the county's process at the screening stage. When a referral is made, call center workers must evaluate the report and provide their own recommendations; once these recommendations are logged into the system, they press a "big blue button" on their computer screen, and

147. *Id.* at 12 ("Allegheny County has additional data sets such as birth records, homeless services and educational outcomes from local school districts that were not tested in the first iteration of the model for various reasons. Birth records, for example, were not regularly being integrated into Allegheny County's data warehouse at the time the model was developed. Education data were not included since Allegheny County does not have full coverage of the county; it only partners with a subset of local school districts. The research team will consider adding additional data sets to future iterations of the model but does not expect that they will lead to significant increases in the accuracy of the model.").

148. *Id.*

149. *Id.* at 9–10.

150. *Id.*

151. *Id.*

152. *Id.*

153. *Id.* at 27.

the child's AFST risk score pops up.[154] If a risk score is above a certain level, the family is automatically "screened in," a result that can only be overridden by a supervisor.[155]

Since the AFST's deployment, Allegheny County officials have announced their intention to move forward with plans for an algorithm that will assess every child born in the county at the time of their birth.[156] This proposal is controversial, to say the least.[157]

154. VIRGINIA EUBANKS, AUTOMATING INEQUALITY: HOW HIGH-TECH TOOLS PROFILE, POLICE, AND PUNISH THE POOR 128, 140 (2018) [hereinafter EUBANKS, AUTOMATING INEQUALITY].

155. Virginia Eubanks reported that "[i]f a family's AFST risk score is high enough, the system automatically triggers an investigation." Virginia Eubanks, *A Child Abuse Prediction Model Fails Poor Families*, WIRED (Jan. 15, 2018), https://www.wired.com/story/excerpt-from-automating-inequality/ [https://perma.cc/YKX4-HYC2] [hereinafter Eubanks, *A Child Abuse Prediction Model Fails Poor Families*]. The county's Department of Human Services publicly responded, claiming that "Ms. Eubanks is incorrect in her statement" and that "[i]n fact, it is Allegheny County policy that children classified by the AFST as being at the very highest risk level should be screened in for investigation unless a supervisor deems this unnecessary." Press Release, Marc Cherna, Dir., Dep't of Human Servs., DHS Response to Automated Inequality by Virginia Eubanks (Jan. 31, 2018), http://www.alleghenycounty.us/Human-Services/News-Events/Accomplishments/Allegheny-Family-Screening-Tool.aspx [https://perma.cc/L35X-H24U]. As Eubanks pointed out in reply, the distinction drawn by the county is a hollow one: when a sufficiently high-risk score is generated, a "system *has been automatically triggered* if it will move forward without being overridden by a supervisor." Virginia Eubanks, *A Response to Allegheny County DHS* (Feb. 16, 2018) (emphasis in original), https://virginia-eubanks.com/2018/02/16/a-response-to-allegheny-county-dhs/ [https://perma.cc/5HJM-ZZC4]. Allegheny County has yet to release the threshold that triggers this chain of events. Brauneis & Goodman, *supra* note 145, at 145.

156. *See, e.g.*, Accenture, *Predictive Analytics in Action - Marc Cherna, Allegheny County, PA*, YOUTUBE (Sept. 5, 2017), https://www.youtube.com/watch?v=Gpb3G31l3lM [https://perma.cc/P3BX-WYU5] ("It's only the first step of our predictive analytics. Its controversial, predictive analytics, so we wanted to start with something that is really pretty straightforward, and then we're going into things that are a little bit more difficult, like preventing child abuse, because we can tell at birth that there's a good likelihood that somebody's going to be in the system, but how do you do preventive services in a way that's not stigmatizing, that doesn't violate people's rights, that is helpful, and that's the challenge that we're doing a lot of community work around, figuring out how to do that piece."); Marc Cherna, *We Will Use All Resources to Keep Children Safe*, PITTSBURGH POST-GAZETTE (Mar. 23, 2018), http://www.post-gazette.com/opinion/letters/2018/03/23/We-will-use-all-resources-to-keep-children-safe/stories/201803230094 [https://perma.cc/D7R8-36WK].

157. *See* Richard Wexler, *Pittsburgh Misuses Big Data to Target Poor Children for Abuse Investigations*, YOUTH TODAY (Mar. 28, 2018), https://youthtoday.org/2018/03/pittsburgh-misuses-big-data-to-target-poor-children-for-foster-care/ [https://perma.cc/CCP5-BGXA] (arguing that a model that brands "every infant with a scarlet number at birth, a number that will even affect the

2. *Eckerd Rapid Safety Feedback*

A Florida non-profit, Eckerd Kids, markets a tool called the "Rapid Safety Feedback" (RSF) process, which it says combines "a baseline of serious risk factors with real-time quality assurance."[158] For agencies that contract to use RSF, Eckerd Kids supplies the "coaching, training and technical assistance, and fidelity reviews," while its "tech partner, Mindshare, is building in the 'prediction piece.'"[159] As of early 2017, at least ten states were at some stage of development with Eckerd's RSF, including Florida, Alaska, Connecticut, Indiana, Maine, Oklahoma, Louisiana, Tennessee, Ohio, and Illinois.[160]

By December 2017, however, Illinois's Department of Children and Family Services (DCFS) ended its trial of the RSF, citing the tool's unreliability.[161] DCFS Director Beverly Walker told the Chicago Tribune: "We are not doing the predictive analytics because it didn't seem to be predicting much."[162] Reportedly, case workers using the tool were "alarmed and overwhelmed by alerts as thousands of children were rated as needing urgent protection."[163] According to press accounts, the RSF predicted a ninety percent or greater likelihood of death or injury for more than 4,100 children.[164] For more than 350 children, case workers received alerts saying the child had a "100 percent chance of death or serious injury in the next two years."[165] At the same time, Illinois's DCFS felt that the tool failed to flag some of the most serious cases. In 2017, two young children, Semaj Crosby and Itachi Boyle, died within a month of one another.[166] Neither was rated as "high risk" by the model.[167] Upon

---

number assigned to their children and grandchildren," may be "inherently unethical").

158. John Kelly, *Rapid Safety Feedback's Rapid Ascent*, CHRON. SOC. CHANGE (Feb. 28, 2017), https://chronicleofsocialchange.org/child-welfare-2/rapid-safety-feedbacks-rapid-ascent [https://perma.cc/Y467-CUE2].

159. *Id.*

160. *Id.*

161. David Jackson & Gary Marx, *Data Mining Program Designed to Predict Child Abuse Proves Unreliable, DCFS Says*, CHI. TRIB. (Dec. 6, 2017), http://www.chicagotribune.com/news/watchdog/ct-dcfs-eckerd-met-20171206-story.html [https://perma.cc/VY4B-QL8J].

162. *Id.*

163. *Id.*

164. *Id.*

165. *Id.* Responding to inquiries in the wake of Illinois's discontinuance of its trial, an Eckerd spokesperson stated: "We all agree that we could have done a better job with that language. I admit it is confusing." *Id.*

166. *Id.*

review, Illinois officials found that the data fed into the model from its case management system for each of these children was "riddled with error."[168]

III. ASSESSING THE MODELS

As child protective agencies push ahead with efforts to implement predictive analytics, the question arises: what role should these tools play in child protective work? Proponents of incorporating predictive analytic tools into child welfare decision-making argue that these tools will help agencies make better decisions more consistently, and reduce bias.[169] But predictive algorithms are built on data that reflects the existing problems in the child welfare system. They are the result of myriad human choices, many of which implicate important value judgments about the way the child welfare system should work. Unless careful attention is paid at every stage of development and use, predictive analytics risk not only papering over existing problems in the child welfare system, but also introducing new risks for families.

Some of the most important risks predictive analytics present can be categorized along three axes: accuracy, fairness, and misuse.[170] First, these tools incorporate data that itself is riddled with error, and may introduce new errors through the way the models are designed. Second, algorithmic prediction presents serious fairness concerns. Algorithms can perpetuate or magnify historical patterns of bias, their use of individual data in these models may implicate data privacy, and they can work to erode necessary transparency. Finally, predictive models are subject to misuse. Users may overestimate the capabilities

---

167. *Id.*

168. *Id.*

169. For example, Walter Smith Jr., a deputy director of Allegheny County's Office of Children, Youth, and Families has argued that "[w]e know there are racially biased decisions made" in current child welfare practice; he continued:

> There are all kinds of biases. If I'm a screener and I grew up in an alcoholic family, I might weigh a parent using alcohol more heavily. If I had a parent who was violent, I might care more about that. What predictive analytics provides is an opportunity to more uniformly and evenly look at all those variables.

Dan Hurley, *Can an Algorithm Tell When Kids Are in Danger?*, N.Y. TIMES (Jan. 2, 2018), https://www.nytimes.com/2018/01/02/magazine/can-an-algorithm-tell-when-kids-are-in-danger.html [https://perma.cc/42XL-F7GD].

170. This Article does not intend to provide an exhaustive list of all the issues presented by the advent of predictive analytics.

of these tools and allow their predictions to take on outsize importance or be used in ways for which they were not intended. The following Part describes these risks.

A. Accuracy

The first area of concern raised by the advent of predictive analytics is whether and to what degree they provide accurate predictions. The introduction of predictive analytic tools is only justified if they in fact serve to improve decision-making. The existence of error in the underlying data and the difficulty of identifying appropriate outcomes to measure lead to serious questions about these tools' accuracy. Even with perfect data, probabilistic predictive tools will make classification errors, identifying some families as high-risk when they are not, and some as low-risk when the opposite may be true. Developers may be able to prioritize one type of error over another, meaning that fundamental value judgments about the way child welfare systems should work may be baked in to the models. And these models may discount or ignore changes to real-world conditions, leaving them making "zombie predictions" that do not track reality. For all these reasons, careful attention must be paid to the accuracy of these models and how the errors they make implicate fundamental systemic concerns. The following section details some of these accuracy concerns.

1. Flaws in the Data: Garbage In, Garbage Out

The foundation of any predictive model is the data on which the model is trained and is meant to analyze. If there are flaws in the underlying data, those flaws will be translated into the output of the model. Algorithms are only as good as their data — simply put: "garbage in, garbage out."[171]

In the case of child welfare predictive tools, there is reason to be extremely concerned about the level of pure error present in the data being fed into the algorithm. Data in administrative systems is entered, initially, by humans. Names, addresses, or other vital information may be wrong, information from one individual may be erroneously tied to another, old and outdated information may persist, or information may be missing altogether.[172] Error that

---

171. Tal Z. Zarsky, *An Analytic Challenge: Discrimination Theory in the Age of Predictive Analytics*, 14 I/S: J.L. & POL'Y INFO. SOC'Y 11, 24 (2017).

172. I can easily think of a handful of examples from my own practice. One client, who was incarcerated in federal prison for a number of years, was listed on all his

changes an important input into a risk prediction model for a particular child makes the tool useless to decision-makers.[173] The Illinois DCFS's short-lived experiment with Eckerd's predictive tool provides a stark example of this problem. The 2017 deaths of Semaj Crosby and Itachi Boyle provided a window into what is likely widespread error in the underlying data.[174] When the agency looked at their records, they found that the tool had not predicted that these children were at risk, in part because the Department's "automated case-tracking system was riddled with data entry errors" in both cases.[175] The reverse problem is easy to imagine as well. Error in the data that artificially inflates a child's score could lead the agency down a destructive path.[176]

2. *Error in the Model: Not All Errors Are Created Equal*

Even if perfect data were available, any predictive tool would still produce errors because models are simply statistical predictions — not perfect visions of the future. The inevitability of error raises two issues. First is the question of how the model prioritizes the various types of errors. Second is the question of what the stakes of a potential error tell us about the model itself. Each concern is considered in turn.

---

prison forms with his name misspelled and his first and last names reversed; this made even simple things, like setting up a telephone conference with him, inordinately difficult. Another client lost out on months of visitation with her young son, derailing her reunification efforts at a critical time, because a referee located an old order of protection in an online system kept by the courts, which appeared there as active. In reality, that order of protection had been vacated more than a year earlier. Resolving this issue took weeks, multiple trips to the courthouse, and reams of paper — all because of a data error. Similarly, the Allegheny County team described how their efforts to validate their model against hospital data were hampered when they encountered missing data in records they received from the Children's Hospital of Pittsburgh. VAITHIANATHAN ET AL., *supra* note 119, at 20 n.14.

173. *See* Julia Angwin & Jeff Larson, *How We Analyzed the COMPAS Recidivism Algorithm*, PROPUBLICA (May 23, 2016), https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm [https://perma.cc/Z9L9-HZ2H] (finding, in a study of the criminal justice COMPAS tool, that "sometimes people's names or dates of birth were incorrectly entered in some records — which led to incorrect matches between an individual's COMPAS score and his or her criminal records. We attempted to determine how many records were affected. In a random sample of 400 cases, we found an error rate of 3.75 percent (CI: +/- 1.8 percent).").

174. *See* Jackson & Marx, *supra* note 161.

175. *Id.*

176. *See* Kahenman & Tversky, *supra* note 50, at 166 (discussing how framing impacts decision-making); WHITE & WALSH, *supra* note 56, at 4–5 (describing how bias impacts decision-making in child protection risk assessment).

First, we may all agree that prediction errors are of concern, but how we value them may vary. Not all errors are created equal. In the case of models that purport to predict risk to a particular child, these errors will consist of false positives (e.g., classifying some individuals as high risk when they are not), and false negatives (e.g., classifying others as at low risk when the reverse is true). When models that purport to predict individual-level risk are used, these two types of error lead to different potential harms: a false positive may lead to unwarranted state intervention, possibly even to family separation that otherwise would not have occurred. A false negative, on the other hand, could lead the agency to fail to intervene when it should have.

Understanding that different types of error may implicate different harms, developers may have latitude to optimize the tool for one type of error over another.[177] This might appear at first blush to be a technical decision for model-developers to make, but it in fact implicates one of the most fundamental questions in this field. It is a new battleground for the same debate scholars have engaged in throughout the development of the modern "child welfare" system: how to calibrate the fine balance of rights at stake when the state intervenes in family life.[178]

Often, the risks of false negatives are over-emphasized while the risks of false positives are underappreciated. Take, for example, the Second Circuit's description of the choice protective services workers face when investigating a family: "If they err in interrupting parental custody, they may be accused of infringing the parents' constitutional rights. If they err in not removing the child, they risk injury to the child and may be accused of infringing the child's rights."[179] Or this description of the possible risks of error, which appears in a 2008 article: "In child welfare, these classification errors are of concern because false negatives can be dangerous to the child, while false

177. In fact, scholars have suggested that it would be impossible to "be fair in both ways at the same time." Sam Corbett-Davies et al., *A Computer Program Used for Bail and Sentencing Decisions Was Labeled Biased Against Blacks. It's Actually Not That Clear.*, WASH. POST: MONKEY CAGE (Oct. 17, 2016), https://www.washingtonpost.com/news/monkey-cage/wp/2016/10/17/can-an-algorithm-be-racist-our-analysis-is-more-cautious-than-propublicas/?noredirect=on&utm_term=.3ff410f6c794 [https://perma.cc/8N7D-QK64]. *See also infra* notes 213–16 and accompanying text.

178. *See generally* MARTIN GUGGENHEIM, WHAT'S WRONG WITH CHILDREN'S RIGHTS 181–212 (2005) (discussing the shift in child welfare towards termination of parental rights and adoption as goals).

179. Tenenbaum v. Williams, 193 F.3d 581, 596 (2d Cir. 1999).

positives can result in poor targeting of agency resources."[180] While not untrue, these statements of the potential risks set up a false dichotomy, placing parents' and children's interests on opposite sides when they are much more intertwined than these descriptions imply. The conception of parents as benefiting exclusively from family integrity and children benefiting exclusively from separation, discounts the danger posed to the child and family by the false positive. We must be concerned not only with "poor targeting of agency resources," but also with the serious harms that families experience from unnecessary or unwarranted system involvement. This is obvious in the case of an unnecessary child removal,[181] but we cannot disregard the harms that an unnecessary visit by a child protective worker can cause. The experience of being visited by a child protective worker can be destabilizing and disempowering. Case workers enter private family spaces, bringing with them accusations and judgments about the parents and their abilities or worth. Families often are forced apart to speak with the case worker separately, and are asked to open cabinets and drawers to prove that there is food and clothing in the home — regardless of the substance of the allegation.[182] Neighbors often are interviewed or questioned regarding the habits of the parents or children, destroying a family's sense of autonomy and privacy.[183] This experience can leave parents and children with the feeling that their home is no longer a safe space, that their family is at imminent risk of being torn apart, and that their worst fears have been confirmed: that, as many parents often secretly

---

180. D'Andrade, *supra* note 48, at 37.

181. *See* Shanta Trivedi, *The Harm of Child Removal*, N.Y.U. REV. LAW & SOC. CHANGE (forthcoming 2019). In the throes of the family separation crisis created by the Trump Administration, more than 12,600 mental health professionals signed a petition summarizing the serious harms that even brief disruptions in families can cause:

> From decades of research and direct clinical experience, we know that the impact of disrupted attachment manifests not only in overwhelming fear and panic at the time of the separation, but that there is a strong likelihood that these children's behavioral, psychological, interpersonal, and cognitive trajectories will also be affected.

Petition from Mental Health Professionals: Stop Border Separation of Children from Parents!, CHILD'S WORLD AMERICA, https://childsworldamerica.org/stop-border-separation/stop-border-separation-text-preview/ [https://perma.cc/XU53-UQTB].

182. *See, e.g.*, Doriane Lambelet Coleman, *Storming the Castle to Save the Children: The Ironic Costs of a Child Welfare Exception to the Fourth Amendment*, 47 WM. & MARY L. REV. 413, 436–37, 441 (2005).

183. *Id.* at 431 n.38.

worry, they are not good enough, are not doing a good job as parents.[184]

Throughout the various stages of predictive model design and implementation, and especially during the model training stage, developers may choose to optimize predictive tools to allow for more of one type of error than another.[185] This is a value judgment embedded deep within each model. For this reason, it is important that communities in which predictive analytics are being developed understand the models' error rates and can assess the ways that the developers have weighted different types of error.

Complicating this prospect is the difficulty of assessing whether a model is making truly accurate predictions. There are a number of problems in assessing the accuracy of models that purport to make individual-level predictions of future risk to a particular child. First, as discussed above, algorithmic tools must be designed to predict well-defined "outcome variables."[186] The child welfare system has never successfully defined child maltreatment, making outcome selection extremely difficult. Definitions of "neglect" are notoriously vague.[187] They can range from a child missing school to severe deprivations of basic life necessities, and everything in between.[188]

184. *See, e.g.*, Brief for the N.Y.U. Sch. of Law Family Def. Clinic et al. as Amici Curiae Supporting Respondents at 26–27, Camreta v. Greene, 563 U.S. 692 (2011) (No. 09–1454 and 09–1478), http://www.familydefensecenter.net/images/stories/NYU_School_of_Law_Family_Defense_Clinic.pdf [https://perma.cc/SK3T-3BRF] ("The interview may undermine family integrity and trust and suggest that it is not safe to trust parents."). These concerns may be particularly acute in situations involving domestic violence. *See* Brief of Battered Women's Res. Ctr. et al. in Support of Respondents at 13–14, Camreta v. Greene, 563 U.S. 692 (2011) (No. 09–1454 and 09–1478), http://www.familydefensecenter.net/images/stories/Battered_Womens_Resouce_Center.pdf [https://perma.cc/2RFB-ZX39]; *see also* Gillingham, *supra* note 113, at 1052 (arguing that labelling people in particular ways has consequences for their construction of identity and the ensuing subject positions offered to them by such constructions).

185. Lehr & Ohm, *supra* note 116, at 698 ("Real-world stakeholders rarely view different kinds of errors as holding the same normative valence. Therefore, analysts frequently rely on tuning parameters to implement these asymmetries.").

186. *See supra* notes 116–19 and accompanying text.

187. *See, e.g.*, Lassiter v. Dep't of Soc. Servs., 452 U.S. 18, 45 (1980) (Blackmun, J., dissenting) ("The legal issues posed by the State's petition [for removal on the basis of abuse or neglect] are neither simple nor easily defined. The standard is imprecise and open to the subjective values of the judge."); Jane C. Murphy, *Legal Images of Motherhood: Conflicting Definitions from Welfare "Reform," Family, and Criminal Law*, 83 CORNELL L. REV. 688, 706 (1998).

188. *See, e.g.*, N.Y. FAM. CT. ACT § 1012 (McKinney 2018).

For this reason, neglect often is confused with "parenting while poor."[189]

Even were we to agree on some well-defined notion of "child maltreatment," however, it is unlikely that developers could identify a data point or a constellation of data points that accurately reflect when and where maltreatment takes place in the real world.

Accordingly, predictive analytics developers have been forced to identify proxies for child maltreatment — values that will stand in for the real thing. But good proxies are hard to find. "Substantiation,"[190] for instance, is not a good measure.[191] Substantiation is influenced by bias, determined based on an insufficient legal standard, and does not distinguish well between outcomes for individual children. The population whose cases are substantiated is necessarily dependent on the population for whom referrals for investigation are made.[192] The legal standard for "substantiation" — generally something similar to "some credible evidence" — is too low to be meaningfully tied to "truth." Children whose siblings may have experienced abuse or neglect, but who themselves did not, may be included in substantiated reports, and research has shown that myriad factors unrelated to actual child maltreatment may influence substantiation decisions.[193] Such a measure, therefore, is replete with inaccuracies and bias, and does not serve as a useful proxy for real-world maltreatment.

In the case of the Allegheny County algorithm, for example, the developers chose two flawed proxies for the tool to predict: re-referral and placement.[194] Re-referral, subject as it is to the

189. *See* Eubanks, *A Child Abuse Prediction Model Fails Poor Families*, *supra* note 155.

190. Substantiation refers to the agency's initial finding that there is some basis to the report of child maltreatment. *See supra* notes 21–25 and accompanying text.

191. *See* Gillingham, *supra* note 113, at 1049–52.

192. *See supra* notes 11–23 and accompanying text. As Erin Dalton, director of Allegheny County's Office of Data Analysis, Research and Evaluation stated, "[w]ho we investigate is not a function of who abuses. It's a function of who gets reported." Hurley, *supra* note 169.

193. Gillingham, *supra* note 113, at 1049–50 ("The term 'substantiation' may be applied to cases in more than one way, as stipulated by legislation and departmental procedures. It might be applied in cases not only where there is evidence of maltreatment, but also where children are assessed as being 'in need of protection' or 'at risk.' Substantiation in some jurisdictions may be an important factor in the determination of eligibility for services and so concerns about a child or family's need for support may underpin a decision to substantiate rather than evidence of maltreatment. Practitioners may also be unclear about what they are required to substantiate, either the risk of maltreatment or actual maltreatment, or perhaps both." (internal citations omitted)).

194. *See supra* notes 142–44 and accompanying text.

idiosyncratic conception of what constitutes "maltreatment" on the part of the community-member who makes the call, cannot be tied directly to true harm. What's more, community referral is one of the decision points in the system where the most disproportionality is evident, indicating that biases play a role in these decisions.[195] Similarly, placement is a manifestation not necessarily of risk to the child, but of the myriad decisions of the existing system, most made on the basis of vague standards and under the influence of bias. For these reasons, it may be difficult to assess whether individual-prediction models have weighted error properly.

This discussion leads to the question: if models will inevitably produce errors, is it responsible to use them to make individual-level predictions? The debate about optimizing for false positives or false negatives is a vital one, but the more important question is whether these models can make predictions on the individual level that are accurate enough to justify the stakes of any error. Where predictions are geared toward individuals' future behavior, the stakes are inexorably higher than they would be for population-level prediction. Consider, for example, the risk posed by a false positive prediction that an individual child is at high risk of harm. This false prediction may set off a series of events with dire consequences: if the model being used is like Allegheny County's AFST, the child and his or her family may become the subjects of a child protective investigation in which they otherwise might not be involved. This could lead to trauma and even removal. Now compare this individual false positive with a population-level false positive. If a model were built to assess where families are likely to need preventive services, for example, the result of any given family being erroneously included in the "high risk" set results in only a slightly higher likelihood that an agency will direct preventive services toward their community, services that will act to better support that community.

3. Zombie Predictions

Predictive analytic tools also may fail to account for resiliency and change over time. Just like SDM tools, predictive analytic algorithms tend to rely on negative risk factors, to the exclusion of protective,

195. Allegheny County, for instance, has stated that most of the disproportionality present in its system can be attributed to disproportionate rates of referral of black families. *See* ALLEGHENY CTY. DEP'T OF HUMAN SERVS., *supra* note 72, at 2–3. In this way, substituting re-referral for maltreatment as the outcome variable serves to entrench and to magnify the biases that lead to disproportionality. *See* Eubanks, *supra* note 154, at 155; *see also infra* Section III.B.1.

positive factors.[196] Of particular concern in child welfare are tools that rely on a parent's own history of placement in foster care as a child.[197] These tools tend to project static conditions from a family's past into the future, without considering rehabilitative actions taken or resiliency-supporting factors that may be present.

These tools pose particular risks when on-the-ground conditions have changed. Scholars John Logan Koepke and David G. Robinson warn of the risk of "zombie predictions."[198] These are predictions that revive old data and outcomes for a system that has undergone reforms, which results in tools that systematically overestimate risk and potentially undercut reform efforts.[199] In child welfare, there has been much discussion about a renewed focus on prevention nationwide, surrounding the recent enactment of the Family First Prevention Services Act in early 2018.[200] As reforms get underway in line with this renewed focus, we must be particularly cognizant of the risks of "zombie predictions."

B. Fairness

The second axis along which concerns arise with predictive analytics relates to fairness. First and most importantly, algorithmic prediction can perpetuate, magnify, and even shroud from public view the pernicious effects of bias. Second, the use of individuals' information in these big data systems may raise privacy concerns. And third, algorithmic decision-making risks reducing necessary transparency on the workings of government systems, which in turn makes oversight more difficult and degrades public confidence in the legitimacy of such systems. The following section discusses each of these concerns.

---

196. *See, e.g.*, VAITHIANATHAN ET AL., *supra* note 119, app. at 37–43.

197. *See, e.g.*, Lindsay Williams & Kristine Monroe, Ohio Dep't of Job & Fam. Servs., *Using Predictive Analytics to Identify High Risk Child Welfare Cases*, OHIO CASA 2017 CELEBRATE KIDS! CONFERENCE (Sept. 20, 2017), https://www.ohiocasa.org/wp-content/uploads/2017/09/T18.pdf [https://perma.cc/C7JR-MHQ6].

198. John Logan Koepke & David G. Robinson, *Danger Ahead: Risk Assessment and the Future of Bail Reform*, 93 WASH. L. REV. 1725, 1725 (2018).

199. *Id.* at 1730.

200. Bipartisan Budget Act of 2018, Pub. L. No. 115-123, 132 Stat. 64 (codified as amended in scattered sections of 42 U.S.C.).

1. Bias in, Bias out

Just as errors in data invalidate the predictions that an algorithm provides based on that data, so too does bias in data infect the results of predictive algorithms. In other words, "bias in, bias out."[201] Algorithms trained on data representing historical biases will inherit those biases.[202] Where the data fed into the algorithm represents the "prejudicial or biased behavior of prior decision makers," the inclusion of that data in the model will cause the algorithm to "learn from the bad example" of the past and import those same biases into the future.[203] More than simply inheriting existing bias, these algorithms may also "launder" that bias, giving an "aura of logical inevitability" to historical patterns or to the status quo in the form of the predictive model's authoritative, "clean, mathematical apparatus."[204]

In the child welfare context, a long history of over-surveillance and over-policing of poor communities and communities of color means that those communities are disproportionately represented in any child welfare or criminal justice data set. This means that, without careful attention to the details of tool development and training, algorithms trained on these biased data sets are likely to rate individuals coming from these communities as higher risk, while systematically discounting the risk of wealthier, white families.[205]

---

201. KINGSLEY & DI MAURO-NAVA, *supra* note 127, at 8.

202. Barocas & Selbst, *supra* note 116, at 687 ("The efficacy of data mining is fundamentally dependent on the quality of the data from which it attempts to draw useful lessons. If these data capture the prejudicial or biased behavior of prior decision makers, data mining will learn from the bad example that these decisions set."). Anupam Chander refers to this as the "problem of *viral discrimination*," writing that "algorithms simply compound the errors of the past . . . . Algorithms trained or operated on a real-world data set that necessarily reflects existing discrimination may well replicate that discrimination." Chander, *supra* note 10, at 1036.

203. Barocas & Selbst, *supra* note 116, at 687.

204. Dave Gershgorn, *Companies Are on the Hook if Their Hiring Algorithms Are Biased*, QUARTZ (Oct. 22, 2018), https://qz.com/1427621/companies-are-on-the-hook-if-their-hiring-algorithms-are-biased/ [https://perma.cc/HHL4-XE87].

205. *See* Barocas & Selbst, *supra* note 116, at 684–86 ("If a sample includes a disproportionate representation of a particular class (more or less than its actual incidence in the overall population), the results of an analysis of that sample may skew in favor of or against the over- or underrepresented class."). In critiquing machine learning mainly in finance and employment, scholars have also warned of the reverse risk — that bias can result from a community's relative absence from the data set. *Cf.* Kate Crawford, *Think Again: Big Data*, FOREIGN POL'Y (May 10, 2013), https://foreignpolicy.com/2013/05/10/think-again-big-data/ [https://perma.cc/M67J-CY2V] ("Because not all data is created or even collected equally, there are 'signal problems' in big-data sets — dark zones or shadows where some citizens and

Even proponents of extant predictive analytic tools acknowledge that they may replicate existing inequities. Erin Dalton, director of Allegheny County's Office of Data Analysis, Research and Evaluation has acknowledged that the AFST "definitely oversample[s] the poor."[206] Ms. Dalton stated even more explicitly when speaking to the New York Times, "It's a conundrum . . . . All of the data on which the algorithm is based is biased. Black children are, relatively speaking, over-surveilled in our systems, and white children are under-surveilled."[207]

This problem is compounded by the fact that predictive algorithms incorporate and learn from multiple data sets, each of which is infected by bias. For example, predictive tools might learn from data in the criminal and juvenile justice systems; but this means that the biases and disproportionality present in those systems are imported into child welfare risk assessments.[208] Similarly, historical patterns of housing discrimination and poverty concentration mean that even facially neutral variables, such as zip code, can serve as stand-ins for decades of bias. By relying on such variables, we risk magnifying these problems.[209]

The inclusion of even facially neutral inputs that import protected variables like race into the algorithm's predictions risks not only building historical patterns of bias into the algorithms' predictions about the future, but also obscuring the true impact those protected variables have.[210] One way this problem manifests is in "omitted variable bias."[211] This term refers to the problem presented when an important, influential variable — for instance, race or socioeconomic

communities are overlooked or underrepresented."). There is a similar corollary present in child welfare data, as the relative absence of white and wealthy families in historical data means that children from those communities risk being systematically under-counted and underserved, while black and poor families continue to be victimized by the system's harsh spotlight.

206. Eubanks, *A Child Abuse Prediction Model Fails Poor Families*, *supra* note 155.

207. Hurley, *supra* note 169.

208. *See* KINGSLEY & DI MAURO-NAVA, *supra* note 127, at 10. ("[R]isk models are extremely likely to 'rediscover' protected variables through their close correlation with other factors. (For example, in many communities, a client's race will be highly correlated with her zip code of residence.)"); *see, e.g.*, Sonja B. Starr, *Evidence-Based Sentencing and the Scientific Rationalization of Discrimination*, 66 STAN. L. REV. 803, 837–38 (2014).

209. *See* Starr, *supra* note 208, at 838.

210. *See* Chander, *supra* note 10, at 1037–39.

211. *See, e.g.*, GARY KING ET AL., DESIGNING SOCIAL INQUIRY SCIENTIFIC INFERENCE IN QUALITATIVE RESEARCH 168–82 (1994).

status — is absent during the training stage. The absence of that variable, despite its outsize influence on the likelihood of the outcome, will cause other, correlated variables to take on weight that they do not warrant, acting as a sort of algorithmic Trojan Horse by importing the effect of the missing variable while obscuring its role.[212] For this reason, the exclusion of vital data, like information identifying individuals as members of protected classes, could itself lead to discriminatory predictions.[213]

The Allegheny development team's description of the role of race in the AFST raises questions about whether race has a larger impact than acknowledged.[214] The accompanying ethical review, in fact, concedes that "potentially reinforcing" racial disparities in the model's underlying data may lead the model to "overstate the actual risk status of a child or family."[215] This review does little to assuage concern over racial bias in the algorithm, instead finding that such bias is not problematic in the child welfare context because — according to the authors — child welfare system involvement, unlike criminal justice system involvement, is designed to be rehabilitative.[216] As this Article has shown, however, the

---

212. *See, e.g.*, *Omitted Variable Bias: A Comprehensive Econometrics Review*, ALBERT (Sept. 20, 2016), https://www.albert.io/blog/omitted-variable-bias-econometrics-review/ [https://perma.cc/Q79M-W8S6].

213. *See* Sam Corbett-Davies & Sharad Goel, *The Measure and Mismeasure of Fairness: A Critical Review of Fair Machine Learning* 5 (Stanford Univ., Working Paper, Sept. 11, 2018), https://5harad.com/papers/fair-ml.pdf [https://perma.cc/4HLP-2Y7U].

214. VAITHIANATHAN ET AL., *supra* note 119, at 15, 29–30 ("Of course, it is important to note that not including race is not to imply that race does not feature into the model because there are other predictors that are highly correlated with race due to potentially institutionalized racial bias (e.g., criminal justice history) that would imply that race is still a factor.").

215. TIM DARE & EILEEN GAMBRILL, ALLEGHENY CTY. DEP'T. OF HUMAN SERVS., ETHICAL ANALYSIS: PREDICTIVE RISK MODELS AT CALL SCREENING FOR ALLEGHENY COUNTY 6 (Mar. 2017), https://www.alleghenycountyanalytics.us/wp-content/uploads/2017/04/Developing-Predictive-Risk-Models-package-with-cover-1-to-post-1.pdf [https://perma.cc/7MXK-GY77].

216. *See id.* ("We think it matters in the AFST case that while a history of engagement with child protection services may lead the AFST to overstate the actual risk status of a child or family, the intervention which flows from that classification is designed and intended precisely a) to identify that family or individual's actual risk status through home visits and professional judgement, and b) to address in so far as possible any risk factors which are found to exist. It matters, ethically, this is to say, that a high risk score will trigger further investigation and positive intervention rather than merely more intervention and greater vulnerability to punitive response. We believe, that is, that the fact that the AFST will prompt further detailed inquiry into a family's situation and that any intervention is designed to assist gives grounds to

rehabilitative goal of child welfare systems often falls short in practice, and even brief interactions with system actors may be destabilizing.[217] We must be vigilant any time the gaze of the state is focused on an individual, even for purportedly "beneficial" purposes, but especially so when that gaze is focused as a result of the person's race.[218]

2. *Privacy Concerns*

Individual data is being tracked constantly. Our locations are monitored through our phones, our purchases logged, and our interests catalogued and traded among corporations.[219] Generally, we expect people to "police their own data disclosures," relying on a privacy model that prioritizes individual choice and is based on consent.[220] But when the government obtains and uses our data in ways we may not have expected — such as to influence decisions about the future integrity of our families — questions arise about whether we as a community are comfortable with this new use.

Information science scholar Helen Nissenbaum has put forward a theory to explain the great variability among situations and communities in concepts of data privacy, termed "contextual integrity."[221] According to Nissenbaum's theory, what causes anxiety and resistance — what "raise[s] privacy hackles" — is not the sharing of data in and of itself, but moments when personal information flows in ways we may not expect, flouting established "informational norms."[222] With predictive analytics, individuals may have expected their data to reside with or be used by health systems or social support systems, but may not have expected it to end up in a powerful algorithm that scores their family for "risk" and leads to the knock of

think the model is not vulnerable to the legitimate concerns generated by the existence of disparities in data used in punitive contexts."). *Id.*

217. *See, e.g.*, *supra* notes 178–83.

218. *See* Duchesne v. Sugarman, 566 F.2d 817, 828 n.24 (2d Cir. 1977) ("[O]f all tyrannies a tyranny sincerely exercised for the good of its victims may be the most oppressive . . . . (T)hose who torment us for our own good will torment us without end for they do so with the approval of their own conscience.") (quotation marks and citations omitted).

219. *See, e.g.*, Michele Gilman & Rebecca Green, *The Surveillance Gap*, 42 N.Y.U. REV. L. & SOC. CHANGE 254, 292 (2018).

220. *Id.* at 291–92.

221. Helen Nissenbaum, PRIVACY IN CONTEXT: TECHNOLOGY, POLICY, AND INTEGRITY OF SOCIAL LIFE 127 (2010).

222. *Id.* at 3, 127, 148; *see also* Gilman & Green, *supra* note 219, at 295 (describing Nissenbaum's theory).

a child protective worker at their door. This use may violate notions of what an "appropriate information flow" looks like and thereby raise a "red flag."[223]

For some of the information included in predictive analytic tools, we may believe that informed consent is required before data can be used in new and different ways.[224] Child protective agencies put forward various arguments as to why there are no problems with consent when they deploy predictive tools, most focusing on the idea that the agency is using data that already is available to it or to its associated governmental entities.[225] There is no guarantee, however, that agencies will not reach beyond government boundaries into the wider world of data being collected and mined every day about everyone.[226]

Even assuming that child welfare agencies confine themselves to data already available to government systems, there is a strong argument that extracting individual data, in many cases, violates the notion of informed consent. In her recent book, scholar Khiara Bridges argues that "poor mothers have been deprived of privacy rights."[227] Bridges outlines two ways of understanding this argument, a "moderate" claim and a "strong" claim.[228] The moderate claim is that although poor women may nominally have a right a to privacy, the practical effect of our system results in "no *effective* privacy rights"[229] The strong claim, by contrast, argues that poor mothers have no privacy rights at all.[230]

In line with Bridges' distinction between the moderate and strong claims, there are two ways to view consent, or lack thereof, when individual data ends up in public systems. In accordance with Bridges' moderate claim, we can conceive of a family's (in particular, a mother's) data as having been handed over as payment in a barter for public goods or services.[231] When a mother receives public goods

---

223. Nissenbaum, *supra* note 221, at 148–50; Gilman & Green, *supra* note 219, at 295.

224. Nissenbaum, *supra* note 221, at 145–47 (discussing "transmission principles").

225. *See* DARE & GAMBRILL, *supra* note 215, at 3.

226. David C. Vladeck, *Consumer Protection in an Era of Big Data Analytics*, 42 OHIO N.U. L. REV. 493, 498 (2016) ("Indeed, government agencies routinely buy information from data brokers to round out their own profiles.").

227. KHIARA M. BRIDGES, THE POVERTY OF PRIVACY RIGHTS 11 (2017).

228. *Id.*

229. *Id.*

230. *Id.*

231. *See also* Dorothy E. Roberts, *The Only Good Poor Woman: Unconstitutional Conditions and Welfare*, 72 DENV. U. L. REV. 931, 939 (1995) (noting that public

or services, the argument goes, she implicitly consents to her data being shared with the government as part of the cost of those services. However, when viewed in the context of the state's coercive child welfare authority and removal power, that barter begins to look precipitously inequitable. Poor parents are not free and willing participants in this market — the specter of their child's removal hangs overhead. A poor parent who refuses to access available public services as a way of "preserving a sphere of individual privacy from government interference"[232] risks opening her home to a different, involuntary state invasion: a child protective investigator knocking on the door.[233] No matter what "choice" this parent makes her data ends up in the system. As to Bridges' strong claim, if poor parents are thought to lack privacy rights altogether, voluntary consent simply cannot exist. Regardless of which lens one chooses to view the gathering of this data, then, there are serious questions as to whether the state's receipt and control of the data is the result of any kind of satisfactory consent.

3. Transparency

The fact that many child welfare predictive analytic tools are being developed by private companies for public systems leads to another significant risk — lack of transparency. As is already evident in the criminal justice sphere, private companies regularly and successfully assert trade secret protections to shield predictive algorithms from public scrutiny.[234] Courts have not required that individuals

assistance recipients "barter away their rights in exchange for benefits"); Gilman, *supra* note 76, at 1396–1400 (describing the litany of privacy invasions the poor must endure in exchange for their receipt of public benefits).

232. Roberts, *supra* note 231, at 939.

233. BRIDGES, *supra* note 227, at 9–10, 66, 84–86. Even progressive state laws that recognize that poverty alone should not be considered neglect provide for a finding of neglect where a parent willingly foregoes available public services. *See, e.g.*, N.Y. FAM. CT. ACT § 1012(f)(i)(A) (McKinney 2018) ("'Neglected child' means a child less than eighteen years of age whose physical, mental or emotional condition has been impaired or is in imminent danger of becoming impaired as a result of the failure of his parent or other person legally responsible for his care to exercise a minimum degree of care in supplying the child with adequate food, clothing, shelter or education . . . though [the parent or other person legally responsible is] financially able to do so *or offered financial or other reasonable means to do so*.") (emphasis added).

234. *See* Natalie Ram, *Innovating Criminal Justice*, 112 NW. U. L. REV. 659, 683–86 (2018); Jeff Larson & Julia Angwin, *Machine Bias*, PROPUBLICA (May 23, 2016), https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing [https://perma.cc/HT9G-FC3E] ("Northpointe's software is among the most widely used assessment tools in the country. The company does not publicly

negatively affected by risk assessment tools be provided with information about how those tools function, even when fundamental constitutional interests are at stake.[235]

In-house development by child welfare authorities is no promise of transparency either. Although the Allegheny County team has been a leader in openness about its project, the team already publicly disclaimed any true ability to explain the inner workings of its predictive model, stating that "even researchers cannot interpret the final list of variables and their corresponding weights."[236] This despite the fact that the team commendably, if perhaps nominally, prioritized "explainability" during the model-selection phase.[237] Ultimately, the team also decided to add a less transparent, but (according to the researchers) slightly more accurate, algorithmic model to their tool.[238] Similarly, although New York City's ACS boasts of its in-house development and commitment to "using [predictive analytics] carefully, ethically, and thoughtfully,"[239] the agency has already begun employing a predictive analytics tool, yet little information has been made publicly-available.

Transparency — not just about the inputs and outputs, but that provides visibility on the human decisions being made at every stage of the development process — is vital for a number of reasons. First

---

disclose the calculations used to arrive at defendants' risk scores, so it is not possible for either defendants or the public to see what might be driving the disparity. (On Sunday, Northpointe gave ProPublica the basics of its future-crime formula — which includes factors such as education levels, and whether a defendant has a job. It did not share the specific calculations, which it said are proprietary.)").

235. In *State v. Loomis*, a 2016 case decided by Wisconsin's Supreme Court, defendant Eric Loomis challenged his sentencing court's reliance on a risk assessment tool, COMPAS, on due process grounds. 881 N.W.2d 749, 761 (Wis. 2016). Information about how COMPAS produced its risk score was withheld from Loomis because its developer "considers COMPAS a proprietary instrument and a trade secret" and therefore "does not disclose how the risk scores are determined or how the factors are weighed." *Id.* at 760. The Wisconsin court sanctioned this secrecy, finding that Loomis was not denied due process by the sentencing court's consideration of his risk score. *Id.* The court distinguished Loomis's case from others in which "the court relied on information the defendant did not have any opportunity to refute, supplement or explain." *Id.* at 761. The Court was satisfied by the fact that Loomis could "at least review and challenge the resulting risk scores set forth in the report attached to the PSI." *Id.*

236. VAITHIANATHAN ET AL., *supra* note 119, at 14.

237. *Id.* (discussing alternative methods considered for constructing their algorithm, and citing as one weakness of these alternate approaches their tendency "to be [a] 'black box' in the sense that it is more difficult to understand why a family received a high score").

238. *See id.* at 35 (discussing addition of a "random forest model" to the tool).

239. *See Automating Inequality Panel*, *supra* note 138.

is accuracy, as open-source tools generally are more accurate and contain fewer errors.[240] Second is oversight, as without visibility on each of the "levers" developers adjusted and the decisions they made throughout the tool's creation, advocates and oversight bodies cannot fully assess whether a tool is being developed and utilized responsibly and in line with community values.[241] Third is legitimacy, as shrouding these tools in secrecy — even if they are in fact valid, reliable, and useful — serves to undermine public confidence in their fairness and quality.[242]

C. Misuse

Predictive analytic tools may have some utility in child welfare,[243] but users of these tools must understand and acknowledge their limitations and guard against assigning their predictions undue influence in order to employ them responsibly. If not carefully monitored, the production of algorithmic, individual risk scores might compromise the delicate constitutional balance necessary to protect child safety and respect the fundamental right of family integrity. The following section details some of the ways that predictive analytics are vulnerable to misuse.

1. Acknowledging Limitations

No probabilistic tool can predict whether any given child will or will not experience abuse or neglect.[244] These tools work on the level of averages, drawing upon probabilities and correlations.[245] They cannot make predictions as to the source of any risk, whether it be one caregiver or another, an outside party, or environmental, nor can they predict risk in short time frames. Both the Eckerd and Allegheny County tools, for example, predict risk to children over a period of two years.[246] Such long-term predictions are of limited

---

240. Ram, *supra* note 234, at 687.

241. *Id.*; *see also* Brauneis & Goodman, *supra* note 145, at 129.

242. Ram, *supra* note 234, at 691.

243. *See, e.g.*, YVONNE HUMENAY ROBERTS ET AL., CASEY FAMILY PROGRAMS, CONSIDERATIONS FOR IMPLEMENTING PREDICTIVE ANALYTICS IN CHILD WELFARE 13 (2018), https://caseyfamilypro-wpengine.netdna-ssl.com/media/Considerations-for-Applying-Predictive-Analytics-in-Child-Welfare.pdf [https://perma.cc/524W-BWGE] (identifying potential for predictive analytics to assist with "gap analysis, service evaluation, and service matching").

244. *See* Starr, *supra* note 208, at 842.

245. *See* Ram, *supra* note 234, at 684 (discussing how a particular risk assessment algorithm works).

246. The research team responsible for Allegheny County's AFST wrote:

utility in many child protective decisions, and are inappropriate in the context of the most important decisions, such as whether a child should remain at home with his or her parents.

Long-term, generalized risk is not sufficient to warrant state intervention in a family. Although child neglect laws are troublingly broad and vague, to pass constitutional muster courts have asserted they must permit state intervention only upon the most limited terms practicable, focusing solely on true danger to children.[247] To justify removing a child, for instance, risk to that child must be "imminent."[248] This imminence requirement is one way that courts have attempted to strike the delicate balance needed to protect the varied fundamental interests at stake in child protective cases. In New York, for example, a finding of neglect must be based upon "proof of *actual (or imminent danger of)* physical, emotional or mental impairment to the child."[249] Danger to the child "must be near or impending, not merely possible."[250]These time limitations ensure that the court, "in deciding whether to authorize state intervention, will focus on serious harm or potential harm to the child, not just on what might be deemed undesirable parental behavior."[251]

For these reasons, amorphous, long-term, probabilistic risk predictions simply do not map onto most of the decisions that child

---

> While there is not universal agreement on the degree to which the current clinical assessment at point of referral is focused on the longer-term risk of adverse events versus assessing the current crisis of alleged abuse or neglect, the research team and Allegheny County chose to design a model to predict long arc risk.

VAITHIANATHAN ET AL., *supra* note 119, at 7–8. Public reporting regarding the Eckerd system employed by Illinois and a number of other states makes clear that its algorithm also is focused on long-term risk:

> Illinois child care agencies told the Tribune they were alarmed by computer-generated alerts like the one that said: "Please note that the two youngest children, ages 1 year and 4 years have been assigned a 99% probability by the Eckerd Rapid Safety Feedback metrics of serious harm or death in the next *two years*."

Jackson & Marx, *supra* note 161.

247. *See, e.g.*, Nicholson v. Scoppetta, 3 N.Y.3d 357 (N.Y. 2004).

248. Theo Liebmann, *What's Missing from Foster Care Reform? The Need for Comprehensive, Realistic, and Compassionate Removal Standards*, 28 HAMLINE J. PUB. L. & POL'Y 141, 146 n.19 (2006) (collecting statutes requiring "imminent" or "immediate" risk to the child in order to justify removal).

249. *Id.* (emphasis added).

250. N.Y. FAM. CT. ACT §1012 (McKinney 2018).

251. *Id.*

welfare authorities and courts must make. This disconnect highlights the tension in attempting to graft public health strategies onto the framework of the existing child welfare system. To justify their approach, for example, the Allegheny County researchers attempt to compare their predictions to predictive risk modeling in health care.[252] But this comparison only underscores how poorly predictive algorithms fit into many — if not all — of the decisions that must be made in the course of a child protective case under our current legal regime. As things stand today, most child welfare systems are constructed to adhere to the "blame and cure" paradigm, wherein child welfare agencies investigate and respond to individualized reports of child maltreatment.[253] Agencies are tasked with playing conflicting dual roles — they act as both support and service provider, and as prosecutor and enforcer.[254] Resources are limited, and what resources exist generally have been poured into triage and foster care, while programs aimed at preventing the need for such crisis interventions are chronically underfunded.[255] As the child welfare system currently exists, its focus is on identifying bad actors and families already in or approaching crisis, and intervening only at that point.

---

252. VAITHIANATHAN ET AL., *supra* note 119, at 8 ("[Predictive risk modeling] is a way of supplementing clinical decision-making. By offering clinicians a risk score that stratifies that the patient is at long term risk of, for example, readmission to hospital, the clinicians could be alerted to looking at the wider context of patient's situation than simply the current medical crisis that brought the patient to the attention of the clinician. Similarly, targeting the PRM on long arc-risk complements the role of the screening staff who are focused on the information about the allegation contained in the referral.").

253. *See* Garrison, *supra* note 53, at 5.

254. *See supra* note 46 and accompanying text.

255. *See* Vivek Sankaran, *Is the Solution Really for More Children to Enter Foster Care?*, CHRON. SOC. CHANGE (Aug. 29, 2018), https://chronicleofsocialchange.org/child-welfare-2/is-solution-really-more-children-entering-foster-care [https://perma.cc/2E3B-AR7Q] ("[T]he United States has never provided adequate funds to support effective prevention programs. Recently, Dr. Jerry Milner, associate commissioner at the Children's Bureau, stated that the federal government spends approximately $6 billion on foster care, yet it only spends $83 million on prevention programs. We know these underfunded programs can safely keep children with their families. For example, studies have documented the effectiveness of programs like Families First, in which families get resources and support to keep children in their care. A study done by Dr. Sacha Klein, a social work professor at Michigan State University, found that Head Start was effective in reducing the need for foster care placements. And another study found that raising the minimum wage by just $1 reduced the number of referrals to child welfare agencies by nearly 10 percent.").

This system stands in stark contrast to public health efforts that truly strive for prevention. Public health workers "also investigate risks[, b]ut they do so with a focus on populations and conditions instead of individuals; their aim is not to categorize or prosecute, but instead to identify the circumstances associated with adverse health consequences so that those circumstances can be altered."[256] Understanding the limitations of importing predictive models developed in prevention-oriented contexts onto the child welfare system is vital to ensuring that algorithms are not misused in the child welfare sphere.

2. *The Human Role*

Predictive analytic tools can be wrong, they can be biased, and they may fit poorly within the existing decisional frameworks our system demands. And yet, despite these many shortcomings, humans tend to imbue machines with outsize authority. This phenomenon has been called "automation bias."[257] Seeing automated systems as authoritative and neutral, users tend to be "less likely to search for information that would contradict a computer-generated system" and to "rely on automated decisions even when they suspect system malfunction."[258] In criminal contexts, it is reported that judges find it "very hard . . . to go against [an automated] risk assessment program because it's couched in science."[259]

Where predictive analytic tools enter into the child welfare decision tree, we should be wary of the power of automation bias. Call center workers, social workers, case workers, and judges may "over-credit the reliability of machine-generated risk scores or, in extreme cases, begin to make decisions" about important questions such as whether to remove a child with blind reliance on the risk score.[260] This danger is especially salient in times of foster care panic,

---

256. Garrison, *supra* note 53, at 7.

257. Danielle Keats Citron, *Technological Due Process*, 85 WASH. U. L. REV. 1249, 1271–72 (2008) (defining "automation bias" as "the 'use of automation as a heuristic replacement for vigilant information seeking and processing."); *see also* Gershgorn, *supra* note 204.

258. Citron, *supra* note 257, at 1271.

259. *All Things Considered: Did a Bail Reform Algorithm Contribute to This San Francisco Man's Murder?*, NAT'L PUB. RADIO (Aug. 18, 2017), https://www.npr.org/2017/08/18/543976003/did-a-bail-reform-algorithm-contribute-to-this-san-francisco-man-s-murder [https://perma.cc/W6TS-PDTF].

260. KINGSLEY & DI MAURO-NAVA, *supra* note 127, at 2.

when decision-makers are operating under the intense "cover their ass" mentalities that prevail in these crises.[261]

Early anecdotal evidence from the Allegheny County experiment shows that automation bias may already be having an effect on the way call center workers incorporate information gleaned from repeated use of the AFST. According to Eubanks's description, the developers of the AFST appear to have built a safeguard against automation bias into their system, requiring call center workers to lock in their own risk assessments based on their review of the family's information before they click on the "big blue button" to generate the AFST score.[262] After that time, only a manager may change the recommendation.[263] Despite this attempted safeguard, there is evidence that automation bias is having an effect. According to Virginia Eubanks, even experienced call center workers have asked to change their risk assessments after reviewing the score turned out by the AFST.[264] Allegheny County Intake Manager Jessie Schemm stated: "If you get a report and you do all the research, and then you run the score and your research doesn't match the score, typically, there's something you're missing. You have to back-piece the puzzle."[265] This may be automation bias in action.

Additionally, automation bias may have effects that last beyond the moment of decision. As call center workers interpret risk scores over time and compare AFST results to their own judgments in case after case, the algorithm may influence the workers, fundamentally altering the way they assess reports.[266]

IV. PRELIMINARY RECOMMENDATIONS

Predictive analytics are the result of numerous human choices, reflecting the fundamental unresolved problems in our existing child welfare system. They may be inaccurate, not only working off of flawed data but also weighting the various types of error and risk to families in ways that do not comport with community values. Predictive algorithms may be infected by, and may even magnify, the same faulty assumptions and biases that created and perpetuate a system that disproportionally affects poor families of color. They may

---

261. *See supra* Section I.B.3.

262. *See* EUBANKS, AUTOMATING INEQUALITY, *supra* note 154, at 141–42.

263. *Id.* at 142.

264. *Id.*

265. *Id.* at 141.

266. *Id.* at 142 (stating that "in practice, the algorithm seems to be training the intake workers").

use data without the consent of their subjects, even where we might believe consent is necessary. They may be misused, importing long-term, amorphous risk predictions into decisions that must be about imminent risk. And they may do all of this with an air of authority that they do not warrant.

Even so, agencies continue their efforts to implement predictive analytics. In fact, the federal government is poised to throw its weight behind the development of these tools — in June 2018, Senator Todd Young of Indiana introduced a bill in the United States Senate that, if enacted, would allocate ten million dollars for a pilot program to fund up to five tribal, local, or state governments as they develop new predictive analytics tools.[267]

Acknowledging that efforts to implement predictive algorithms into child welfare systems are likely to continue, what can be done to ensure these tools are developed responsibly and in line with community values? This Article makes four preliminary recommendations that should serve as a baseline for any responsible implementation of predictive analytics in child welfare. First, these tools must be developed in the most transparent manner possible. Second, enforceable legal constraints must be placed on their use. Third, strategies to confront and correct bias must be put in place. And fourth, communities considering implementation of predictive analytics should use this moment as an opportunity to build better, more humane, and values-driven child welfare systems, focusing more on support and prevention, and less on too-little, too-late targeting and crisis intervention.

A. Prioritize Transparency

Developing a predictive algorithm requires a multitude of parties to make myriad value-laden decisions. Agency personnel, data

267. Using Data to Help Protect Children and Families Act, S. 3039, 115th Cong. (2018). As of July 2018, the bill had been introduced by Sen. Young (R-Ind.) and referred to the Senate Health, Education, Labor, and Pensions Committee. During a committee hearing held on June 26, 2018, Senator Patty Murray (D-Wa.) stated that the committee was "getting closer to an agreement" on the bill. Press Release, Senator Patty Murray, U.S. Senate Comm. on Health, Educ., Labor & Pensions, Senator Murray Applauds Bipartisan Progress on Career and Technical Education, Maternal Mortality, and More (June 26, 2018), https://www.help.senate.gov/ranking/newsroom/press/senator-murray-applauds-bipartisan-progress-on-career-and-technical-education-maternal-mortality-and-more [https://perma.cc/A3ZV-CFUQ]; *see also* Marisol Zarate, *Congress Weighs Pilot Program for Predictive Analytics Use in Child Welfare*, CHRON. SOC. CHANGE (July 13, 2018), https://chronicleofsocialchange.org/child-welfare-2/congress-pilot-program-predictive-analytics-child-welfare/31565 [https://perma.cc/U4ZQ-PLLT].

analysts and coders may all have a hand in development and may make choices that impact the lives of children and families. For this reason, it is vital that communities have oversight of, and input into, the process of deciding whether and how to create and implement predictive tools. To accomplish this, agencies and governments embarking on a predictive analytics project must ensure that the process is transparent and open at every stage.

One level of transparency would include making the source code and details of the algorithm itself open to the public. This is not a radical thought, but is common practice in the technology world.[268] Some of the most secure digital tools are peer reviewed and open in this way — what the industry calls "open source."[269] This practice ensures that tools remain honest, reliable, and are constantly improving.[270] Where predictive tools are developed by entities outside of government, insisting on this type of transparency is likely to be particularly fraught because companies may attempt to assert trade secrets protections. Wherever possible, government agencies should insist in their contracting processes that the code be open, and that the contracting party maintains the burden of identifying any information for which protections can be asserted. States and municipalities can also take action to mandate that all algorithms be open source, through laws or regulations.[271]

---

268. *See, e.g.*, Laurie Wurster, *Open Source Software Hits a Strategic Tipping Point*, HARV. BUS. REV. (Mar. 9, 2011), https://hbr.org/2011/03/open-source-software-hits-a-st [https://perma.cc/4TRK-KSBR]; OPEN SOURCE INITIATIVE, https://opensource.org [https://perma.cc/WW7Z-RVFZ].

269. "Open source" refers to a technology for which the source code is available to users and/or the public. Wurster, *supra* note 268. The code can be downloaded, used, and modified by anyone, often leading to improvements to the code or additional innovations built on top of existing products. Wurster, *supra* note 268; *see, e.g.*, U.K. Government Digital Service, *Guidance: Be Open and Use Open Source*, GOV.UK (Nov. 6, 2017), https://www.gov.uk/guidance/be-open-and-use-open-source [https://perma.cc/6NDW-AZRY]. For example, tools developed for the National Security Administration (NSA) are open source. *Open Source @ NSA*, NAT'L SEC. ADMIN., https://code.nsa.gov/ [https://perma.cc/TTK2-K6NL]. Similarly, Signal, a secure, encrypted communication application, is open source and peer reviewed. *Signal-Android*, GITHUB, https://github.com/signalapp/Signal-Android [https://perma.cc/GM7F-MYQG].

270. *See, e.g.*, Ram, *Innovating Criminal Justice*, *supra* note 234, at 691–92.

271. *See, e.g.*, NEW YORK CITY, N.Y., Int. 1696-2017 (2018) (a bill requiring the creation of a task force that provides recommendations on how agencies may address instances where people are harmed by agency automated decision systems); Julia Powles, *New York City's Bold, Flawed Attempt to Make Algorithms Accountable*, NEW YORKER (Dec. 20, 2017), https://www.newyorker.com/tech/annals-of-technology/new-york-citys-bold-flawed-attempt-to-make-algorithms-accountable [https://perma.cc/E29N-TFKP] (detailing the significant walk-back that took a New

Although there is value in ensuring that the underlying code is open, access to the code itself is not the only way to achieve meaningful transparency; in fact, such access might not be of much help in some cases.[272] The code can be complex, and may be difficult to understand or impossible to interpret without the proper training and knowledge. In order to achieve meaningful oversight, communities will have to ensure that the choices made during development are documented with accessible, plain-language materials that explain the reasoning behind, and the effects of, the many decisions made. It is important to ensure that this information is documented in a form that is accessible to the relevant audiences such as the general public and specific, affected communities.[273]

Special attention also must be paid to the information that is provided to users of the system. In order to combat over-reliance on predictive tools, ignorance of their limitations, and the effects of automation bias, users may need to understand the pathway that the algorithm took to reach its conclusion, be trained specifically on the meaning of any score or result, and appreciate the technology's limitations.[274]

Finally, the validity, reliability, and real-world functionality of the tool must continually be studied. This includes both pre-validation and post-implementation studies, regularly performed by independent, outside groups. Further, these studies must then be made available to the public in the same plain language style as development process decisions, so that affected communities can assess any "disparate impact" the tool may be having.[275]

B. Create Enforceable Legal Constraints and Guidance

Legislatures ought to put in place clear, enforceable legal rules to limit how predictive tools are used, and to guide users in employing these tools in line with community values. This will ensure that predictive tools are developed and operated with transparency and with appropriate consideration for their limitations. Even where communities may approve of the actions taken by a certain official or

York City Council bill from its "ambitious" beginnings proposing that the source code for all city-employed algorithms be made public, to its final form instituting a fact-finding task force).

272. Brauneis & Goodman, *supra* note 145, at 175–76.

273. *Id.* at 168, 169, 171–74.

274. *Id.* at 170.

275. Chander, *supra* note 10, at 1039 (describing the concept of "algorithmic affirmative action").

feel that the agency is developing the tool responsibly — as may be the case in Allegheny County, for example — concrete rules are important in the event of mission creep or a change in leadership.[276]

In the case of child protective algorithms, it is important to set limits that will guide the many aspects of these tools that can run wild without proper protections. These include when and how these tools can be used, what they can be used to model, what information they can access, who should have access to their scores or outcomes, and what individuals can do if they feel that erroneous information is utilized, their consent was not adequately sought, or they are treated unfairly as a result of the algorithm.

It is of particular importance to set limits on when and how predictive tools can be used to inform child protective decisions. As established above, predictive algorithms cannot foresee risk on a specific, individual level, and their predictions lack the appropriate level of "imminence" to inform removal decisions. Therefore, these tools should be excluded by enforceable legal rules from removal decisions. Legislation or regulations barring agencies from employing algorithmic prediction as part of their removal decisions, given the current state of the science, are a start.

C. Confront Bias

Although predictive analytic tools are touted for their bias-reducing potential, they will not reach this potential without concerted efforts of developers and oversight from agencies and communities. As this Article has shown, without careful attention these tools run the risk of magnifying the effects of existing biases,

---

276. *See* Richard Wexler, *Pittsburgh Misuses Big Data to Target Poor Children for Abuse Investigations*, YOUTH TODAY (Mar. 28, 2019), https://youthtoday.org/2018/03/pittsburgh-misuses-big-data-to-target-poor-children-for-foster-care/ [https://perma.cc/DM57-GSQH] ("And what about [Director of the Allegheny County Department of Human Services Marc] Cherna's successor, and his successor's successor? Any system that depends for success on the benevolence of a single leader with near-absolute power is too dangerous for a free society."). The importance of this point is well illustrated by a recent change the Trump administration made to a similar tool used in immigration detention settings. Jason Tashea, *ICE Risk Assessment Tool Now Only Recommends 'Detain'*, A.B.A. J. (June 28, 2018, 7:00 AM), http://www.abajournal.com/news/article/ice_risk_assessment_tool_now_only_recommends_detain/ [https://perma.cc/AC8S-MSDG]. According to news reports, the Trump administration modified a risk assessment tool that U.S. Immigration and Customs Enforcement used in its decisions as to whether a given individual should be released or detained to remove the "release" option, such that the tool recommended only "detain." *Id.*

while simultaneously clouding visibility into their impact. When agencies and communities consider utilizing a predictive tool, therefore, they must focus efforts on reducing the effects of bias in their data sets, tool design, and implementation. This means interrogating the data set being used to feed the tool for sources of bias, paying careful attention to the choice of variables included, and closely monitoring any possible disparate impacts of the tool's predictions.

As discussed above, particular attention must be paid to the effect of protected variables such as race.[277] For this reason, some scholars have advocated that tool developers should treat protected attributes "with great care," but that they should "not eliminate them from the dataset."[278] Instead, they should, where possible, seek to "capture these variables" and study how the risk model treats them in order to monitor, and attempt to correct, any systemic biases in the data or in the tool's implementation of the relevant program.[279]

D. Focus on Prevention

As this Article has established, predictive analytic tools are not magic wands generated out of pure science. Machines are not moral. They do not implement their own value judgments, nor will they save us from our flaws. These tools are, instead, expressions of human opinion and policy.[280] Understanding this leads to a final point: agencies or communities attempting to implement data-driven predictive tools must approach these efforts with a keen eye for the values they seek to further.

The policy goals of child protective efforts are many: to prevent child maltreatment, to direct resources to where they are needed, to support families in crisis, and to avoid unnecessary family break-ups. Where agencies consider attempting to harness the power of predictive analytics, they should gear their efforts toward areas where the capabilities of predictive analytics align with the goals for which they are being employed. Instead of trying to shoehorn predictive tools into decisions machines are ill-equipped to make, agencies

---

277. *See supra* notes 206–08 and accompanying text.

278. KINGSLEY & DI MAURO-NAVA, *supra* note 127, at 10 (explaining that this does not apply only to "personal characteristics that have been vectors of historical discrimination" such as race, gender, socioeconomic status, or physical ability, but also to "behaviors the government wants to be careful not to stigmatize, such as seeking mental health services").

279. *Id.*

280. *See supra* Section II.A. and Part III.

should embrace the public health origins of these tools and use them for population-based prediction or "opportunity mapping," guided by a drive toward prevention.[281]

Child protective efforts to date have often been overly focused on triage, resulting in a system that is reactive instead of prevention-focused.[282] There is, however, evidence that the Titanic of child welfare is turning ever so slowly toward prevention. In 2018, for example, the federal government enacted the Family First Prevention Act, which in part made federal funds that previously funded foster care available for preventive services such as mental health and substance abuse prevention and treatment services.[283] And in November 2018, the Children's Bureau issued an information memorandum to "strongly encourage" child protective agencies to focus more on preventive efforts.[284]

Where agencies shift focus to improve prevention as opposed to crisis intervention, predictive tools can be used to amplify these efforts. For instance, agencies might use predictive analytics to map the communities where families are likely to be in need of services and direct additional resources to those communities, instead of attempting to target a particular family once it already is in crisis.[285]

CONCLUSION

There are clear flaws in the way our child welfare system makes decisions. These flaws must be addressed, and perhaps technology is one route to doing that. There are, however, serious risks posed by predictive analytic tools. To guard against these risks, we must be cognizant of the many value-laden human decisions that go into the development of these tools and ensure that those whose lives are affected by them have a meaningful seat at the table as they are considered or developed.

Algorithms will not fix child welfare. If introduced, they will be tools employed by and in service of a system that is itself riddled with flaws. Until we take effective action to address the underlying

---

281. *See* CAPATOSTO, *supra* note 66, at 9.

282. COMM'N TO ELIMINATE CHILD ABUSE AND NEGLECT FATALITIES, THE DISSENTING REPORT OF THE HONORABLE JUDGE PATRICIA M. MARTIN, CECANF COMM'R 7–8 (2016), https://www.cwla.org/wp-content/uploads/2016/03/Dissenting-Report-of-Judge-Patricia-Martin-Final.pdf [https://perma.cc/N8GD-64YW].

283. Bipartisan Budget Act of 2018, Pub. L. No. 115–123 § 50701, 132 Stat. 64, 212.

284. U.S. DEP'T OF HEALTH & HUMAN SERVS., ADMIN. FOR CHILDREN & FAMILIES, MEMORANDUM FROM THE CHILDREN'S BUREAU (2018).

285. *See* CAPATOSTO, *supra* note 66, at 9.

problems that truly lead to the disproportionality, unfair treatment, and backwards functioning of our system, we will simply be coding over the cracks in its foundation.