Proposing Safeguards for Governmentally-Regulated Risk Assessment Mechanisms

Sritej Attaluri
Sarah Scheffler
Law For Algorithms—Berkeley CS 294-155 / BU CAS CS 791

Abstract

In the last decade, advancement and proliferation of big data processing mechanisms have made it easier for governmental and corporate entities to collect and analyze information about citizens and non-citizens utilizing large reservoirs of data and data analysis tools. Risk assessment systems grant individuals a score which is used to determine whether rights or permissions should be withheld from the individual. We limit our attention to risk assessment systems that are governmentally-regulated, which include systems that are developed or run by the government itself, as well as those with a significant enough impact on citizens' lives that the government has chosen to regulate them.

This paper analyzes governmentally-regulated risk assessment systems by evaluating them on three axes: We examine the costs of the systems on individuals, the system holders, and society, we analyze the inputs to the system, and we describe the transparency (or lack thereof) within the systems. Using three case studies—the Unified Passenger system (UPAX), the COMPAS Risk & Need Assessment System, and the FICO score—we develop a standardized set of potential technical requirements to mitigate abuse and ensure individuals are treated fairly while remaining within the constraints levied by the system's purpose.

We propose three system requirements that apply in all contexts: that the system is *rectifying*, *reparative*, and *responsible*. The first—rectifying—is that there must be an easy-to-use mechanism for individuals to correct errors when the input data used is incorrect. The second—reparative—is that even if the input data is correct, the system should attempt to avoid perpetuating institutional injustices and contributing to "self-fulfilling prophecies." Finally, the third—responsible—is that the system must be accountable to some entity trusted by the government's constituents, be it via transparency of its methods or by frequent auditing.

1. Introduction

Risk assessment systems of various kinds are used in many different parts of American life. They are in the "national security" sphere (e.g. counterterrorism and screening systems), the "criminal" sphere (e.g. algorithms for aiding sentencing decisions or predicting recidivism), and also the "civil" sphere (e.g. credit scoring). These three use cases have very different needs and requirements for the system in order for them to function properly and fairly. Our goal is to develop broad requirements for government risk assessment systems as a whole; hence, it is important that we choose examples that differ in impact and scope. We have chosen these three carefully and justify our choices further in the cost-bearing section.

This paper uses three case studies—the Unified Passenger system (UPAX),[1] the COMPAS Risk & Need Assessment System,[2] and the FICO credit score[3]—to act as a representative sample of the much larger set of governmentally-regulated risk assessment systems. Each of these algorithms has a significant impact on the life of the individual it is judging. UPAX determines whether a person is allowed to enter or exit a country, and in some cases may lead to an individual's detention. The COMPAS recidivism tool, once limited to parole decisions in a few states, is now one of several recidivism-risk algorithms incorporated in sentencing in more than twenty states.[4] FICO scores impact where individuals are allowed to live, if they're allowed to work at a specific job, and whether they can access access to potentially life-saving credit for medical expenses.

Sections 2, 3, and 4 will examine our three case studies with regard to three different axes of analysis: their cost-bearing properties, the quality of their data inputs, and their transparency. Section 5 will analyze common trends throughout the systems. We will conclude by formulating three technical requirements—*rectifying*, *reparative*, and *responsible*—that apply to all the systems we analyzed, and that should apply broadly to all governmentally-regulated risk assessment systems.

---

[1] DHS-CBP-PIA-006(e) U.S. Department of Homeland Security, *Privacy Impact Assessment for the Automated Targeting System 2017 - Appendix Updated 2018* (2018) [*hereinafter* 2017 PIA].

[2] Case Management for Supervision, Equivant, http://www.equivant.com/solutions/case-management-for-supervision (last visited Nov. 20, 2018).

[3] FICO® Score, https://www.fico.com/en/products/fico-score (last visited Nov. 20, 2018).

[4] Joe Palazzolo, *Court Judges Can Consider Predictive Algorithms in Sentencing*, Wall St. J. (July 13, 2016, 5:04 PM), http://blogs.wsj.com/law/2016/07/13/court-judges-can-consider-Predictive-algorithms-in-sentencing (last accessed Nov. 25, 2018).

2. Cost-Bearing

In this section, we describe the impact of these systems on three parties: the individuals they judge, the organization deploying them, and society as a whole. We attempt to weigh the positive benefits to each party against the negative benefits to each party for each system. Unsurprisingly, in many cases, a positive impact in one area must be traded-off against a negative impact in another area. The tool itself is usually agnostic to these trade-offs; a policy must be created that determines which incentives to favor.

2.1 Unified Passenger (UPAX)

We use Unified Passenger (UPAX), and its predecessor, the Automated Targeting System-Passenger (ATS-P), as an example of a very extreme risk assessment tool—one designed for law enforcement and national security, and therefore exempt from many laws that would normally limit the use of a risk assessment system, such as the Privacy Act of 1974.[5] UPAX, developed and maintained by U.S. Customs and Border Protection (CBP), "aggregates data in a consolidated, automated interface to provide continuous vetting of foreign nationals from application through the duration of their visa."[6] UPAX was originally developed and maintained separately from the ATS, and had a different purpose as well. ATS, originally created in 2007, was supposed to inspect cargo and passengers at the country's border, whereas UPAX was a tool for analyzing passengers, and was not limited to the border. However, in 2013, the older ATS-Passenger (ATS-P) interface was replaced by UPAX.[7] For the purposes of this paper, we use "UPAX" to refer to the system as it is now, even if referring to something that was originally introduced in ATS-P, and we analyze UPAX from a border-control perspective only.

When assessing the risk of an individual, UPAX checks the individual's status on watch lists and warrants, as well as matching "patterns of suspicious activity identified through past investigations and intelligence."[8] This information informs CBP officers where to spend more resources for additional inspection. DHS also claims that decisions are never based *solely* on the

---

[5] *See infra* § 4.1.

[6] DHS Office of Inspector General, *DHS Tracking of Visa Overstays is Hindered by Insufficient Technology* 35 (2017), https://www.oig.dhs.gov/sites/default/files/assets/2017/OIG-17-56-May17_0.pdf

[7] Department of Homeland Security Privacy Office, *2014 DHS Data Mining Report to Congress* 6 (2015) https://www.dhs.gov/sites/default/files/publications/2014%20DHS%20Data%20Mining%20Report%20Signed_1.pdf

[8] Department of Homeland Security Privacy Office, *2016 Data Mining Report to Congress* 23 (2017) https://www.dhs.gov/sites/default/files/publications/2016%20Data%20Mining%20Report%20FINAL.pdf

results of automated analysis. This demonstrates the reluctance of DHS to be seen as relying solely on an automated risk score, probably to ensure the department's legitimacy in the eyes of the public. However, even if DHS's decisions are not based solely on automated analysis, it is clear that the automated risk assessment is a large step toward a final decision. We choose UPAX as our first case study because it is an extreme example of a risk assessment tool with strong opacity needs. UPAX claims exemptions from the parts of the Privacy Act that normally allow individuals to see the records pertaining to them, for fear that this mechanism would be abused by would-be criminals and terrorists.[9]

One of the central questions of our analysis of UPAX is whether or not "big data blacklisting" can be considered a primary harm. *Big data blacklisting* is a term coined by Professor Margaret Hu to refer to the harm incurred when individuals are considered "guilty until proven innocent" by virtue of data analytics and pattern matching.[10] All automated risk assessment systems can be considered big data blacklists under this definition. UPAX is an example of *confidential* big data blacklisting: a big data blacklisting program where either the data inputs, the analysis method, or both, are kept secret. Many of the data sources ingested into UPAX are classified or semi-classified[11] and the methods for analysis are kept secret to prevent criminals and terrorists from gaming the system.

Hu argues that big data blacklisting *by itself* is a constitutional harm to the citizens of a nation if due process rights are reconsidered for a digital age.[12] If big data blacklisting is a primary harm, then the question of harm to an individual becomes "whether or not big data blacklisting occurred," instead of "what rights the individual was denied as the result of the big data blacklisting." She identifies ten suspicious implications of big data blacklisting,[13] one of which is "*ex ante* and preventive or precrime objectives," pointing out that the goal of systems like UPAX is to prevent crime or terrorism before it happens. This can significantly obstruct freedoms and cause stigma, and it also violates our ideals of justice when it punishes an

[9] ATS claims exemptions from 5 U.S.C. § 552a (c)(3), (d)(1), (d)(2), (d)(3), (d)(4), (e)(1), (e)(4)(G), (e)(4)(H), (e)(4)(I), and (f). *See* 71 FR 212 64546 (2006).
[10] Margaret Hu, *Big Data Blacklisting*, 67 Fla. Law Rev. 1735, 1738 (2015).
[11] *Id.* at 1741 ("Semi-classified" is a term borrowed from Hu, *Big Data Blacklisting* meaning that the existence of the program is public, but all further information about it is classified.).
[12] *Id.* at 1752 ("[B]ig data blacklisting itself is the harm that may be considered unconstitutional when grounded in due process through a conceptualization of what substantive due process rights can and should encompass, in an age when daily existence and governing methods have both been utterly transformed by big data technologies.").
[13] *Id.* at 1756-1761.

individual for "the wrongful choice he was predicted to make" and not the one the individual actually made.[14] Big data blacklisting marks individuals as guilty until proven innocent, and since the "guilt" is determined by checking whether the features of an individual are similar to those of criminals, some unfortunate people are tainted by association and have no way to prove their "innocence." Since the system is very opaque, there is very little opportunity to learn or remediate the consequences.[15]

Big data blacklisting also comes with further harms. The cost of an innocent person being misjudged by the UPAX system is very high to that individual. In the best case, misjudged individuals must follow a long, not-guaranteed redress program[16] before traveling to the United States. At worst, they may be detained.[17] Misclassification of innocent individuals also creates a cost to CBP in resources spent in pursuing or further vetting an individual who posed no threat.[18] The cost of a guilty person being misjudged has the same security/freedom tradeoff seen in other counterterrorism cases. The final determination of what cost is "too high" is often based on very low-confidence estimates of the prevalence of terrorism,[19] and could be the topic of an entire other paper. Certainly the ATS-P and UPAX did not come cheap in terms of dollars or bureaucratic overhead.[20]

---

[14] Slobogin, Christopher, *Principles of Risk Assessment: Sentencing and Policing.*,15 *Ohio St. J. Crim. L.* 583, 592 (2017) (quoting Andrew von Hirsch, *Past or Future Crimes: Deservedness and Dangerousness in the Sentencing of Criminals* 11 (1985) "[U]nless the person actually made the wrongful choice he was predicted to make, he ought not be condemned for that choice - and hence should not suffer punishment for it.").

[15] Hu, *supra* note 10, at 1758-1761 (items (5), (6), (7), and (10)).

[16] *See infra* § 4.1.

[17] CBP Publication No. 0000-0119 A Look at the CBP Traveler Inspection Process, CBP FOIA Library (2009) https://foiarr.cbp.gov/docs/Manuals_and_Instructions/2009/283172355_16/0909291657_A_Look_at_the_CBP_Traveler_Inspection_Process.pdf ("Unless probable cause has been developed, you can request that CBP notify someone of your delay if you are detained more than two hours after the personal search has begun.").

[18] Peter Shane, *The Bureaucratic Due Process of Government Watch Lists*, 75 Geo. Wash. L. Rev. 804 at 807 (2006).

[19] *See generally* John Mueller, & Mark G. Stewart, *Terrorism and Bathtubs: Comparing and Assessing the Risks.* Terrorism and Political Violence, 1-26 (2018).

[20] Jena B McNeill, *Air Cargo Security: How to Keep Americans Secure without Harming the Economy*, 2422 Heritage Foundation Backgrounder (2010) (discusses the economic costs of cargo screening, not passenger screening, but a better estimate for the cost of passenger screening was not found).

2.2 COMPAS Risk & Need Assessment System

The COMPAS Risk & Needs Assessments tool,[21] or COMPAS, for short, is a tool developed by Equivant[22] that predicts an individual's likelihood of recidivism. It is one of many tools[23] that incorporate data about an individual from many sources and use statistical methods to attempt to predict an outcome, in this case, whether or not the individual is likely to commit a crime if released. Originally used only for parole decisions, COMPAS and other such tools are now used in sentencing in many states.[24] As with many other risk assessment tools, it accomplishes this by weighing features[25] about the person, or about the person's background. We pick recidivism tools as our second case study because they are still a matter of criminal law, but are usually not subject to concerns of national security. We choose COMPAS as a well-scrutinized recidivism tool that has been subject to litigation.[26]

The practice of performing risk assessments in criminal justice is not as new as the recent advancement of algorithmic tools would make it appear. California had a method for determining sentence length based on the likelihood of committing a new crime in 1917.[27] These practices were largely abandoned in the 1970s "in favor of 'truth' in sentencing: fixed periods of confinement based on backward-looking appraisals of an offender's culpability for crime already committed."[28] Over the last few decades, America's prison population has exploded[29] and the

---

[21] Case Management for Supervision, *supra* note 2 (Originally known as "Correctional Offender Management Profiling for Alternative Sanctions", now called "COMPAS Risk & Needs Assessments" as part of the "Northpointe Suite")

[22] *Id.*, Formerly Northpointe, Inc.

[23] Algorithms in the Criminal Justice System, Electronic Privacy Information Center https://epic.org/algorithmic-transparency/crim-justice/ (last accessed Nov. 20, 2018) (The "Public Safety Assessment Tool" (Justice System Partners), the "Level of Service Inventory-Revised" (Multi-Health Systems), and COMPAS (Equivant) are the most widely-used parole and sentencing tools.).

[24] *Supra* note 4.

[25] *See infra* § 3.2.

[26] *See infra* § 3.2.

[27] John Monahan, *Risk Assessments in Sentencing*, Reforming Criminal Justice Volume 4: Punishment, Incarceration, and Release 79 (2017) http://academyforjustice.org/wp-content/uploads/2017/10/Reforming-Criminal-Justice_Vol_4.pdf. *See also* John Monahan and Jennifer Skeem, *Risk Redux: The Resurgence of Risk Assessment in Criminal Sanctioning* in 26 Federal Sentencing Reporter at 1 (2013) (on the introduction of parole statues, a precursor to risk assessment thinking, in 1876).

[28] Monahan and Skeem, *supra* note 27, at 1.

[29] The Sentencing Project, *Trends in U.S. Corrections*, http://sentencingproject.org/wp-content/uploads/2016/01/Trends-in-US-Corrections.pdf (last accessed Nov. 21, 2018) (In 1972 approximately 200,000 people were in the U.S. state and federal prison population; by 2010 that number was more than 1,500,000.). *See also* Jenny Roberts, *Why Misdemeanors Matter: Defining Effective Advocacy in the Lower*

court system cannot keep up with the pace.[30] Risk assessment tools in modern criminal justice cannot be evaluated without the context of a long-bloated prison system.

The cost to individuals of over-estimating risk of recidivism fits into a much broader literature about the costs of over-policing and over-punishing minor offenses:[31] Jailing an individual makes it much more difficult to get a job or additional education, and impacts family life.[32] These issues also disproportionately affect the poor[33] and racial minorities.[34]

The cost of recidivism risk assessments to local courts is much simpler; the bulk of it can be assigned a dollar value. The use of risk assessments is primarily intended to alleviate the burden on an already overworked court staff, and is thus primarily a benefit to these deployers. However, what these tools make up to the staff in time, they cost in dollars: a local Wisconsin court paid $1,765,334 for a two-year contract with Northpointe for the COMPAS software.[35]

The courts also typically face asymmetrical costs for different types of errors: if they wrongly jail an individual who would not have recidivated, no one will know, and so the only cost of this type to the courts is the rare appeal on the basis of a risk assessment such as *State v. Loomis*.[36] However, if a court wrongly releases an individual who does recidivate later, then it has effectively "spun the wheels," going through an entire trial process for nothing as it now must judge the individual *again*, and in the meantime its credibility was harmed by releasing the individual. Since wrongly releasing an individual is more costly to a court than wrongly holding one, the court is incentivized to err on the side of jail.

Criminal Courts, 45 UC Davis Law Rev 277 (The volume of misdemeanor cases nationwide rose from five to more than ten million between 1972 and 2006.).

[30] Roberts, *supra* note 29, at 277 ("Many individuals charged with low-level crimes receive representation from defense attorneys with overwhelming caseloads, in a criminal justice system singularly focused on rapid finality in the large numbers of docketed cases.").

[31] Babe K. Howell, *Broken Lives from Broken Windows: The Hidden Costs of Aggressive Order-Maintenance Policing*, 33 N.Y.U. Rev. L. & Soc. Change 271 at 274 (2009) ("policing minor offenses so aggressively creates significant hidden costs that undermine the legitimacy of the criminal justice system, create substantial burdens for poor people (the majority of those arrested for order offenses), and erect barriers to education and employment.").

[32] *Id.*

[33] *Id. See also* Jason Sunshine, and Tom R. Tyler, *The Role of Procedural Justice and Legitimacy in Shaping Public Support for Policing*, 37 Law & Soc. Rev. at 513 (2003).

[34] Stephen J. Schulhofer, *et al.*, *American Policing at a Crossroads: Unsustainable Policies and the Procedural Justice Alternative*, 101.2 The Journal of Criminal Law & Criminology 336 (2011) ("Public order successes have been achieved at great cost to politically powerless communities… our laws and the way they are enforced have resulted in public attitudes sharply polarized along racial lines, a division that is scarcely surprising in a nation marked by conspicuous racial disparities in its prison populations.").

[35] Electronic Privacy Information Center, *Routing Slip - Internal*, https://epic.org/algorithmic-transparency/crim-justice/EPIC-16-06-23-WI-FOIA-201600805-FY16Contract.pdf (last accessed Nov 21, 2018).

[36] *See infra* § 3.2.

This feedback loop, which led to the necessity of efficient risk assessment tools in the first place, also poses costs to the community and society. Over-policing minor offenses de-legitimizes the criminal justice system in the eyes of society[37] and can ultimately increase crime and disorder.[38] Furthermore, unnecessarily keeping individuals in the system places a greater strain on the resources of the courts, which in turn lead to even more incorrect convictions.[39]

Under-sentencing also poses costs to society. Four of the main purposes of sentencing as described in the National Center for State Courts guide[40] are "(1) Punishment proportional to the seriousness of the offense and the degree of offender culpability (i.e., 'just deserts'); (2) Enhancing public safety through offender risk reduction and management involving considerations of specific deterrence, rehabilitation, incapacitation, and control; (3) Restitution to the victim and/or restitution to the community; and (4) Enhancing public safety through general deterrence."[41] Note that goal (1) is retrospective and the other three are prospective. When a recidivism algorithm under-estimates risk for an individual, we take the resulting unfulfillment of the prospective goals to be the cost to society.

2.3 FICO Score

We pick credit scores as our third case study because, unlike UPAX or COMPAS, they are governmentally-regulated risk assessment mechanisms which touch almost every member of American society. Many credit scoring variations are used all around the world in order to determine risk posed to lenders by loan-seekers but in the United States, the FICO score (named for its parent corporation, Fair, Isaac and Company) is the premier metric. The FICO score is used for more than 90% of lending decisions,[42] affecting dozens of millions of people. Although

---

[37] Schulhofer, *et al.*, *supra* note 33, at 338 ("Local policing practices currently favored in much of America not only have hidden costs for effective crime prevention but also can directly undermine sound responses to the threat of terrorism… [we propose] a shift to [an approach in which] compliance with the law and willingness to cooperate with enforcement efforts are primarily shaped not by the threat of force or the fear of consequences, but rather by the strength of citizens' beliefs that law enforcement agencies are legitimate.").

[38] Howell, *supra* note 31, at 274.

[39] Roberts *supra* note 29, at 333.

[40] Pamela M. Casey *et al.*, *Using Offender Risk and Needs Assessment Information at Sentencing*, National Center for State Courts 11 (2011) https://www.ncsc.org/~/media/Microsites/Files/CSI/RNA%20Guide%20Final.ashx

[41] *Id. See also* Monahan, *supra* note 27, at 80 ("Many legal scholars have argued that any workable theory of sentencing must address both retributive and utilitarian concerns, rather than just one of them.").

[42] FICO® Scores Used in Over 90% of Lending Decisions According to New Study, https://www.mercatoradvisorygroup.com/Press_Releases/FICO%C2%AE_Scores_Used_in_Over_90__of_Lending_Decisions_According_to_New_Study/ (last accessed Nov 26, 2018).

FICO scores are not calculated by a government entity itself—instead coming from the company, FICO—the scores still still draw on information provided from three national (but publicly-traded) credit bureaus: Equifax, Experian and TransUnion. Furthermore, FICO scores are audited and regulated by the federal government, specifically the Consumer Financial Protection Bureau (CFPB).

The FICO score was initially developed in 1956 by a mathematician and an engineer to leverage the use of data to "improve business decisions."[43] The history is important because from the very start, its purpose was to suit the institutions making the lending decisions. Before the credit score, the decision-making process consisted of a loan officer sitting down and giving you an interview. Understandably, loan officers are limited by insufficient information and will make decisions that maximize profit for their institutions.

Financial institutions would like to predict perfectly which of the available loan-seekers is the best "bet." For this end, the ability to make data-driven decisions has been wildly beneficial to financial institutions. According to a RAND study, institutions have seen an average increase in profit of "roughly 1,000 dollars per loan,"[44] totalling a 42% increase. FICO markets its score as a mechanism to increase profits not only through identifying the best lending opportunities, but also through active crime and fraud prevention.

The FICO score system uses historical data in order to determine how likely and quickly a loan-seeker will repay a loan. FICO gives a non-binary score that is scaled relative to population as a whole. An institution must then employ heuristic qualification thresholds depending on the people who come through seeking a loan within a time period. Most financial institutions will make decisions based on what is best at the time, even if those decisions would be riskier given another time and different context.

The impact to loan-seekers is massive because your potential to obtain new credit is determined by your past history, incentivizing educated seekers to try to improve their scores through a variety of means, the simplest of which is paying back debt quickly. But FICO scores can affect individuals in ways other than risk assessment for loans. Around 47% of employers

[43] FICO History, https://www.fico.com/en/about-us#history (last accessed Nov 26, 2018).
[44] Einav, Liran, Mark Jenkins, and Jonathan Levin. "The impact of credit scoring on consumer lending." *The RAND Journal of Economics* 44.2 (2013): 249-274.

use credit checks for screening job applicants.[45] This practice has been becoming more and more commonplace even though Eric Rosenberg—director of state government relations for TransUnion—told legislators that there is no "research to show any statistical correlation between what's in somebody's credit report and their job performance or their likelihood to commit fraud."[46] Going beyond the workplace, Experian offers landlords the ability to view tenant credit reports at no cost as part of their review process[47]. Simple mistakes in the past for individuals could mean not being able to secure proper housing.

Under a societal lens, the FICO score also facilitates some financial institutions to take advantage of loan seekers through predatory lending, or when "lenders impose excessive or unnecessary fees or steer borrowers into expensive loans." Predatory Lending disproportionately targets women, black people, and low-income communities.[48] The American Bar Association has found that this targeting is exacerbated by institutions using credit scores.[49] And, as a result, wealth gaps between white people and people of color have grown.

For many consumers, it's a cycle of cumulative disadvantage.[50] Not having access to loans due to a deficient credit report might make it harder to access the resources necessary to secure stable employment, feeding back into the cycle. And for individuals, there are many ways to end up with a deficient credit report without having done anything "wrong." The CFPB found that "43 million Americans have overdue medical debt on their credit reports."[51] To contextualize further, 52% of all debt that exists on credit reports is due to medical debt and 7% of consumers with medical debt have no other type of debt.

---

[45] The Use of Credit Background Checks in Hiring Decisions, https://www.shrm.org/hr-today/trends-and-forecasting/research-and-surveys/Pages/creditbackgroundchecks.aspx (last accessed Nov 26, 2018).
[46] Millions Need Not Apply, https://www.nytimes.com/2011/05/30/opinion/30mon3.html?_r=0 (last accessed Nov 26, 2018).
[47] Landlord Credit Check, https://www.experian.com/connect/landlord.html (last accessed Nov 26, 2018).
[48] Predatory Lending: The New Face of Economic Injustice, https://www.americanbar.org/publications/human_rights_magazine_home/human_rights_vol32_2005/summer2005/hr_summer05_predator/ (last accessed Nov 26, 2018).
[49] *Id.*
[50] How Your Credit Score Impacts Your Financial Future, http://www.finra.org/investors/how-your-credit-score-impacts-your-financial-future (last accessed Nov 26, 2018).
[51] CFPB Spotlights Concerns with Medical Debt Collection and Reporting, https://www.consumerfinance.gov/about-us/newsroom/cfpb-spotlights-concerns-with-medical-debt-collection-and-reporting/ (last accessed Nov 26, 2018).

3. Data Inputs

In this section we seek to answer questions regarding the inputs to risk assessment systems. A system with many data inputs is useless if all of those inputs are low-quality, as represented by the saying "garbage in, garbage out." As risk assessments continue into the future, they may also start incorporating other risk assessments as inputs, thus perpetuating whatever features are already present. The information ingested also affects judged individuals' ability to change their scores: a system that ingests only information about activities the individual has done in the last year is very easy to influence, whereas a system that uses features out of a person's control (e.g. family information) will be more difficult to change, often to the individuals' detriment.

3.1 Unified Passenger (UPAX)

Over the last several years, UPAX has gained many new sources to use as inputs to its analysis. Thirty-four disparate data sources were approved for unification into ATS in 2012[52], at least 16 of which are large databases of individuals maintained by other agencies or sub-agencies that are ingested directly into ATS. An additional 8 large government databases are accessed but not ingested, and ATS also imports private databases maintained by commercial data aggregators. Some of these databases contain Passenger Name Record (PNR) data, merely showing that a certain person was registered to fly on a certain plane. However, some of them also provide data analytics. CBP is tight-lipped about the specific companies, but we know that at least indirectly, Palantir provides inputs to UPAX.[53] UPAX also mines "open source" information[54] including social media information, and has entered into a contract with an unknown "private database vendor" to analyze this information and evaluate its use.

Because UPAX unites many existing data sources but does not directly collect information from individuals themselves, it has few requirements for providing notice to individuals.[55] This is in keeping with several other U.S. government methods that focus on the

---

[52] Automated Targeting System, System of Records. Docket No. DHS-2012-0019 May 22, 2012. 77 FR 99. https://www.gpo.gov/fdsys/pkg/FR-2012-05-22/html/2012-12396.htm

[53] Palantir contributes data to other CBP databases ICM and FALCON, both of which are ingested into ATS. *See* Civ. Action No. 17-2684 ¶16, 20, 22-27, https://epic.org/foia/ice/palantir/1-Complaint.pdf

[54] In the U.S. governmental sphere, "open source" refers to data about an individual that is accessible on the public Internet, such as social media information. *See* 2017 PIA, at 71.

[55] *See* 2012 PIA, at 22-23

legality of data collection as opposed to data analysis. Contrast this to the European Union's GDPR,[56] which maintains strict restrictions on new uses and analyses of data even if it was legitimately collected by another party.[57] This change over the last decade to bring more and more data sources spanning a longer and longer amount of time into the past fits well into major changes in American systems of detecting international terrorists over the last decade: moving away from collecting physical evidence and toward pattern-based and rule-based discovery, especially from analysis of electronic records.[58] ATS-P was primarily designed to identify individuals and compare them against several watch lists. Starting with the changes described in the 2012 Privacy Impact Assessment update, the system did more analysis. Rather than merely identifying an individual's presence on a watch list, UPAX now performs data analysis to identify people who have never had contact with law enforcement before,[59] and it indirectly alters its input watch lists via reuse by other DHS departments.[60]

Because UPAX ingests data from so many different sources, and maintains records for 15 years, DHS acknowledges a risk to data integrity "because the aggregated information may be stale or inaccurate due to an unupdated source system."[61] Because the UPAX database draws from other databases, errors from those databases will propagate to UPAX, which may in fact propagate it to others as well. CBP acknowledges the data integrity problem, but does not provide a satisfactory way to address it. CBP officers are required by policy[62] to correct the data if they become aware of an inaccuracy, but simply ingesting the data from so many disparate sources puts the onus of correcting information onto the source databases. But since the source databases are often less visible to the individuals than UPAX, it will be even more difficult to

---

[56] General Data Protection Regulation (EU) 2016/67 Art. 22(1) and 22(2)(b) ("The data subject shall have the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her.... This does not apply if the decision is made by a Member State law, but even then, there must be "suitable measures to safeguard the data subject's rights and freedoms and legitimate interests").

[57] Even before GDPR was enacted in 2018, many European nations had constitutional rights to determine what could be done with individuals' data, e.g. the German "right to informational self-determination" (Bignami, *European Versus American Liberty: A Comparative Privacy Analysis of Antiterrorism Data Mining*, BCL Rev. 641 (2007), *quoting* Kommers, *The Constitutional Jurisprudence of the Federal Republic of Germany*, 323.).

[58] Sergio Koc-Menard, *Trends in Terrorist Detection Systems*, Journal of Homeland Security and Emergency Management 1 (2009). *See also* Michelle Mittelstadt *et al*, *Through the Prism of National Security: Major Immigration Policy and Program Changes in the Decade since 9/11*, Migration Policy Institute (2011).

[59] *See* 2017 PIA, at 4.

[60] *See* 2012 PIA, at 26 (specifically allowing reuse within DHS).

[61] *See* 2017 PIA, at 13.

[62] *Id.*

correct misinformation in the source databases than it would be to change it in UPAX. At all parts of the system, fixing incorrect inputs is "someone else's problem," which means that incorrect information will likely never be corrected. When a system is required to remain opaque, the need for ensuring data quality in-house rises dramatically.

3.2 COMPAS Risk & Need Assessment System

The COMPAS risk assessment tool uses a number of demographic factors and official records in its predictions of recidivism likelihood.[63] The full list includes demographic information, including gender and marital status, current and past charges, family history (e.g. whether parents had drug or alcohol problems, or were divorced), peers, drug use, residential history, education, employment, leisure activities, social isolation, personality, anger, and attitudes.[64]

Two major features stand out about COMPAS's risk assessment methods: First, COMPAS uses many variables outside the judged individual's control, e.g. family history. Second, COMPAS, like all correlation-detecting tools, conflates similarity with equality, meaning that unfortunate innocent individuals who share attributes with criminals are likely to be misclassified as criminals themselves. These features do not make COMPAS and other algorithmic tools unsuitable for use in a courtroom; similar actuarial analyses that have been incorporated in the sentencing process for many years.[65] Interestingly, though actuarial analysis is often considered expert testimony,[66] the output of a risk assessment is not.[67] This is surprising,

---

[63] *Official Response to Science Advances*, Equivant (2018) *(*A different part of the COMPAS tool, which determines the most likely "needs" of an individual not to recidivate, from a criminological perspective, uses 137 features, but the risk assessment tool uses only six.) http://www.equivant.com/blog/official-response-to-science-advances . *Contra* Bloomberg *et al., Validation of the COMPAS Risk Assessment Classification Instrument* (However, this validity study of COMPAS says it uses 22 "scales" that were created via interviews with the individuals.) http://criminology.fsu.edu/wp-content/uploads/Validation-of-the-COMPAS-Risk-Assessment-Classification-Instrument.pdf

[64] ProPublica, *Sample COMPAS Risk Assessment COMPAS CORE* https://www.documentcloud.org/documents/2702103-Sample-Risk-Assessment-COMPAS-CORE (last accessed Nov. 21, 2018) [*hereinafter* "COMPAS CORE"].

[65] Tracy Fass *et al., The LSI-R and the COMPAS: Validation Data on Two Risk-Needs Tools*, 35 Criminal Justice and Behavior 1095 (2008).

[66] Frederick W. Kilbourne, George W. McCauslan, and Murray A. Segal, *Actuarial Expert Testimony* (1999) https://www.soa.org/library/proceedings/record-of-the-society-of-actuaries/1990-99/1999/january/rsa99v25n363pd.pdf (last accessed Dec. 7, 2018).

[67] *See In re Commitment of Simmons*, 821 NE 2d 1184 - iii: Supreme Court (2004) ("actuarial principles are not the least bit novel and therefore are not subject to Frye". Frye refers to *Frye v. United States*, 293 F. 1013 (D.C.Cir. 1923) which determined that "a court must determine whether the scientific technique used as evidence is generally

since performing a recidivism risk assessment is certainly not a "layman" task.[68] Were COMPAS outputs considered expert testimony, its results would need to be the "product of reliable principles and methods," and it would need to be "reliably applied" to the facts of the specific case.[69] However, as we will discuss further in section 4.2, the methods of COMPAS are unknown because they are proprietary. This means that any assessment of its reliability must be done in a black-box way, which is much more difficult.

3.3 FICO Score

The FICO score takes into consideration only the data which can be found within one of your credit reports. Credit reports are reported by three separate national credit bureaus. The information at each credit bureau might differ if one or more of the bureaus are not updated with the latest information. As such, depending on which credit report is used, your FICO score may differ.

For the general population, the input data is weighted with the following breakdown: "payment history (35%), amounts owed (30%), length of credit history (15%), new credit (10%) and credit mix (10%)."[70] However, these weights are not federally mandated nor enforced. In fact, even though these are the "default," the algorithm shifts these weights depending on the data, sometimes drastically. "Payment history" refers to factors such as if the individual has paid off past debt, the rate at which the debt was paid off, and public records like lawsuits. Some of the things that "amounts owed" refers to are the total debt that the individual has, on what types of accounts that debt lies, and an individual's credit utilization ratio. "Length of credit history" considers things like the average age of your credit accounts, the age of the oldest account, and how often you use your various accounts. The "credit mix" category includes how many types of accounts you have, the distribution of those accounts, and even the reasons you closed your past accounts. Finally, the "new credit" category considers the frequency at which accounts are

---

accepted as reliable within the relevant scientific community", People v Megnath, 27 Misc. 3d 405 - NY: Supreme Court (2010).)

[68] Experts may be used only if the "untrained layman" would be unqualified to make the decision. *See* Mason Ladd, *Expert Testimony*, 5 Vanderbilt L. Rev. 414 418 (1951).

[69] Fed. R. Evid. 702(c) and (d). *See also* President's Council of Advisors on Science and Technology, *Report to the president: Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods* 54 (2016) (regarding requiring "foundational validity" for expert testimony).

[70] What's in my FICO® Scores, https://www.myfico.com/credit-education/whats-in-your-credit-score (last accessed Nov 26, 2018).

opened and how often lenders inquire about your credit. For each of these categories, the examples were meant to be illustrative, not exhaustive.

On the other hand, FICO maintains that it does not consider any data that's not in an individual's credit report.[71] That includes factors such as age, race, color, religion, national origin, sex, marital status, salary, occupation, title, employer, or employment history. It goes as far as to say that it does not consider "any information that is not proven to be predictive of future credit performance."[72] However, it's important to note that even through some factors are not looked at explicitly, those factors may affect an individual's credit score because of historic and systemic reasons. For example, FICO "does not [directly] consider race" when computing the credit score. Nevertheless, race indubitably affects credit scores of non-white communities because of decades-old impacts of redlining and implicit segregation.[73] While the Fair Credit Reporting Act (FCRA) requires that credit reports remove any delinquencies after seven years, that policy alone is not enough to erase the systemic impacts. The mechanism through which race affects credit feeds into the cycle of cumulative disadvantage that was discussed in Section 2.3.

4. Transparency

The term "transparency" has many definitions, especially when it comes to the transparency of algorithms.[74] We use "transparency" to capture a small number of distinct, well-defined meanings, all of which stem from the basic democratic idea that rulers should be held accountable for their actions by citizens.[75] Analogously, in keeping with basic democratic

---

[71] What's not in my FICO® Scores, https://www.myfico.com/credit-education/credit-scores/whats-not-in-your-credit-score/ (last accessed Nov 26, 2018).

[72] *Id.*

[73] How Algorithms Can Bring Down Minorities' Credit Scores, https://www.theatlantic.com/technology/archive/2016/12/how-algorithms-can-bring-down-minorities-credit-scores/509333/ (last accessed Nov 26, 2018).

[74] Zachary Lipton, *The Mythos of Model Interpretability*, ICML Workshop on Human Interpretability in Machine Learning *(2016).* (describing different definitions of "interpretability" for machine learning, motivated by an attempt to identify "trust" in the outcomes of the system, or to understand the "cause" of a specific outcome. One of these is "transparency", which Lipton defines as the answer to the question "what is the model doing?")

[75] Philippe C. Schmitter and Terry Lynn Karl, *What Democracy Is… and Is Not*, 2 Journal of Democracy no. 3 75-88 (1991) (This idea is of course expressed in many prior and future studies of democracy as well, but the Schmitter and Karl wording is meaningful and concise, and and defined a new, extremely well-cited standard of the meaning

principles, it should be possible to learn policy-relevant information about risk assessment tools that are used on the public[76] and to hold the deployers of those tools "accountable" for those decisions in a meaningful way. In this section we first investigate how the public can learn information about the system or its quality. Note that question remains even in the presence of the "best possible" risk assessment tools - the transparency properties of the system would still be necessary even if the system were perfectly accurate, to ensure that the system was not lying about the mechanism by which it achieves such a result. Second, we ask how an individual who was incorrectly judged by the system can request redress or correction.

4.1 Unified Passenger (UPAX)

UPAX is one of many American institutions that is deliberately secretive and opaque, since its purpose is to fight crime and terrorism. However, even these institutions are not completely exempted from all scrutiny. Public scrutiny is limited, but UPAX is audited periodically by the Government Accountability Office and from a technical perspective by the DHS Inspector General. For the purposes of this section, we will take it as a given that these audits are sufficient to ensure that the methods used by UPAX are consistent with CBP's public claims, though we will re-investigate the auditing problem in section 5. But even if we take the validity of the UPAX methods as a given without additional scrutiny, there are four other transparency properties that we can ask of UPAX without challenging its opacity for national security reasons.

The first transparency property is a better redress mechanism. Because UPAX is a government-run system that collects, maintains records on, and matches information about U.S. citizens and permanent residents, it is governed by the Privacy Act of 1974.[77] Section (g) of the Privacy Act contains "civil remedies," which grant U.S. citizens, permanent residents, and those covered by the Judicial Redress Act the right to file suit to force the agency to allow them to

---

of modern democracy: "Modern political democracy is a system of governance in which rulers are held accountable for their actions in the public realm by citizens, acting indirectly through the competition and cooperation of their elected representatives.")

[76] Hollyer *et al., Democracy and Transparency* 73 The Journal of Politics 41 at 1193 (2011) ("The willingness of a government to release policy-relevant information… Without adequate provision of policy-relevant information, the public is unlikely to be able to hold the government to full account for its actions.")

[77] 5 U.S.C. § 552a(g).

access or amend their records.[78] However, UPAX claims a law enforcement exemption[79] from this requirement. It further claims exemptions from the rights of individuals to access the records, because it would interfere with law enforcement and counterterrorism investigations.[80] Non-citizens are limited to requesting redress via the Freedom of Information Act.[81] Neither of these remaining mechanisms require the information to be corrected, meaning that the only requirement for a CBP officer to change incorrect information is policy, not law.[82]

This distinction is important. Until courts began allowing individuals standing to sue to have their names removed from watch lists (many of which are imported into UPAX), DHS didn't take much action to figure out what to do about people who had been wrongly included.[83] The same principle applies to UPAX itself. At present, there is no legal recourse for individuals who believe their information in UPAX is wrong, especially if they are not U.S. citizens. UPAX also claims exemptions from its requirement to inform persons or agencies about corrections or disputes made to the info,[84] so even if the records concerning an individual are corrected, that individual may never find out that the information was changed.

A second reasonable transparency goal would be to share information not about the analysis methods, but about the data integrity methods. As mentioned in section 2.1, UPAX has a very difficult problem of data correctness, since it ingests so many disparate data sources. DHS has "internal quality assurance procedures"[85] to attempt to mitigate the chance of using false

---

[78] *See* 2017 PIA, at 42. *See also* Exec. Order 13,768, 82 FR 8799 (Jan. 25, 2017) (explicitly forbidding non-resident non-citizens from requesting redress via the Privacy Act).

[79] Appendix C to 6 CFR 5 (claims exemptions at 5 U.S.C. § 552a(j)(2) and (k)(2) "exempt records are the risk assessment analyses...and a pointer to the data from the source system of records"; The exemption at (j)(2) states exemptions on information for identifying or investigating individuals for agencies that have enforcement of criminal laws as their principal function; The exemption at (k)(2) is a blanket exemption for law enforcement purposes, but "if any individual is denied any right, privilege, or benefit ... for which he would otherwise be eligible, as a result of the maintenance of such material, such material shall be provided to the individual").

[80] Appendix C to 6 CFR 5(5)(c) (claims exceptions from 5 U.S.C § 552a(d)(1) through (4)).

[81] *See* 2012 PIA, at 42.

[82] *Id.* at 17.

[83] Anya Bernstein, *The Hidden Costs of Terrorist Watch Lists*, 61.3 *Buffalo Law Review* 461, 461 (2013) (quoting *Latif v. Holder*, 28 F. Supp. 3d 1134 8489 (2014) "The government still appears 'stymied' by the 'relatively straightforward question' of what people who 'believe they have been wrongly included on' [the No Fly List] should do. In recent months, courts have haltingly started to provide their own answers, giving some individuals standing to sue to remove their names or receive additional process.").

[84] Appendix C(5)(b) to 6 CFR 5 (claims an exemption to 5 U.S.C. 552a(c)(4), the exemption states "because certain records in this system are exempt from the access and amendment provisions of subsection (d), this requirement to inform any person or other agency about any correction or notation of dispute that the agency made with regard to those records, should not apply")

[85] Appendix C(5)(h) to 6 CFR 5.

information, but of course, without knowing what those are, it is infeasible to learn whether they are sufficient.

Third, as with COMPAS, the issue of proprietary methods arises. Because Palantir and other companies are involved with these analysis tools,[86] hiding the analysis methods becomes a commercial interest. This will be discussed further in section 4.2, but it should be unacceptable to protect the nature of a private asset used in the public sector by claiming a trade secret; it should be held to the same standards of transparency as if it had been developed by the government entity that uses it.[87] In the case of UPAX, much information would likely remain secret because of law enforcement and national security concerns. But some information about UPAX operation is likely not covered by these concerns, and is merely proprietary. If this status were removed, there would be an immediate practical impact: it could be requested via the Freedom of Information Act.[88] This information includes the value and nature of the contracts between DHS and the companies, and it may also include the companies' analysis methods.

Finally, we note that new technology has made it possible to improve the transparency of the system while still releasing no information about ongoing investigations. A recent paper by Frankle et al. describes the use of a cryptographic tool called a *zero-knowledge proof* that would enable proving properties of data use and exchanges that would reveal no information about ongoing investigations, but would force an agency to "commit" to the data used, and only later reveal the data itself to the public.[89] This could happen after the retention period of the data has passed anyway (15 years) so it cannot interfere with the data's use in ongoing law enforcement investigation, removing some of the justification for exemptions to the Privacy Act.

4.2 COMPAS Risk & Need Assessment System

The transparency properties of the COMPAS recidivism prediction tool are extremely policy-relevant because they cause dissonance between the U.S.'s democratic principles of due

---

[86] Either directly or by contributing to ingested databases. *See* Mijente, *Who's Behind ICE?*, https://mijente.net/wp-content/uploads/2018/10/WHO%E2%80%99S-BEHIND-ICE_-The-Tech-and-Data-Companies-Fueling-Deportations-_v1.pdf (last accessed Nov. 23, 2018).

[87] *See infra* note 101. *See also* Whittaker, et al., *AI Now Report 2018*, 4 (2018) https://ainowinstitute.org/AI_Now_2018_Report.pdf (recommendation 4) (last accessed Dec. 7, 2018).

[88] *Department of Justice Guide to the Freedom of Information Act* 263 (2009), https://www.justice.gov/oip/foia_guide09/exemption4.pdf (Exemption 4).

[89] Jonathan Frankle *et al.*, *Practical Accountability of Secret Processes*, Proceedings of the 27th USENIX Security Symposium, (2018).

process and its extremely protective trade secrets laws. Requests for viewing proprietary the COMPAS algorithm have been repeatedly denied,[90] leaving people to guess at the algorithm's functionality with the use of an outdated field manual.[91] In the case *Loomis v. Wisconsin*,[92] the defendant argued that it is a violation of due process to be sentenced using a risk score that was generated using a secret, proprietary algorithm, specifically because the validity of an unknown algorithm cannot be challenged.[93] Loomis's argument had been based on two prior due process cases, *Gardner v. Florida*[94] and *State v. Skaff*.[95] But the Wisconsin court ruled that Loomis's analogy was not correct; Loomis could review and challenge the risk score itself, even without knowing the algorithm.[96]

The case was appealed to the U.S. Supreme Court, but ultimately Loomis's petition was denied after the federal government filed a brief of amicus curiae[97] describing its views on the matter. However, this brief did not adequately address the proprietary nature of the algorithm.[98] It cites a publicly available COMPAS practitioner's guide[99] which lists some factors that influence scores, e.g. criminal history, history of substance abuse, and criminal associates. The brief concluded that Loomis "was not sentenced based 'on information [that he] did not have any

---

[90] *See e.g.*, *infra* note 93, *Loomis*, 881 N.W.2d at 761

[91] Northpointe, *Practitioners Guide to COMPAS*, (2012) http://www.northpointeinc.com/files/technical_documents/FieldGuide2_081412.pdf

[92] The original case *State v. Loomis* in the Wisconsin court questioned only the use of a risk assessment tool in general, not the fact that its mechanisms were secret. The ruling by the Wisconsin court to allow risk assessment tools (though with a warning statement to the judge that one had been used) drew criticism from some legal scholars. See *e.g.* Katherine Freeman, *Algorithmic Injustice: How the Wisconsin Supreme Court Failed to Protect Due Process Rights in State v. Loomis*, 18 N.C. J.L. & Te 75 (2016). *See also* 130 Harv. L. Rev. 1530 (2017).

[93] *State v. Loomis*, 881 N.W.2d 749 (Wis. 2016) (also separately challenged because the tool incorporated race and gender as inputs).

[94] *Gardner v. Fla.*, 430 U.S. 358 (1977) ("[T]he defendant ha[d] a legitimate interest in the character of the procedure which leads to the imposition of a sentence even if he may have no right to object to a particular result of the sentencing process").

[95] *State v. Skaff*, 447 N.W.2d 84 (Wis. Ct. App. 1989). (Access to a Presentence Investigation Report was "an essential factor of due process, i.e., a procedure conducive to centensing based on correct information.").

[96] *State v. Loomis supra* note 93 ("[A]lthough Loomis cannot review and challenge how the COMPAS algorithm calculates risk, he can at least review and challenge the resulting risk scores set forth in the report attached to the [presentence investigation report]. At the heart of Gardner and Skaff is the fact that the court relied on information the defendant did not have any opportunity to refute, supplement or explain. That is not the case here.").

[97] Brief for the United States as amicus curiae, *Loomis v. Wisconsin*, no. 16-6387 http://www.scotusblog.com/wp-content/uploads/2017/05/16-6387-CVSG-Loomis-AC-Pet.pdf (last accessed Dec. 7, 2018).

[98] The brief only focused on the use of risk assessment tools in general, comparing them to previously-used actuarial tools, and emphasized the fact that the COMPAS risk assessment agreed with the court's independent assessment.

[99] *See* Northpointe, *supra* note 91.

opportunity to refute, supplement, or explain.'"[100] which makes it different from the cases with which Loomis was trying to make an analogy.[101]

The brief misses a subtlety, namely that validating the inputs to the risk assessment tool is not the same as validating the decisions made on the basis of those inputs. Legal scholars have criticized the brief's reasoning, and argue that Loomis was trying to argue for a "right to explanation,"[102] not a "right to information." as the court's ruling implies, and that being denied the right to an explanation is an infraction upon due process. The odd situation arises because if the court had developed the risk analysis tool itself, it would be held to public scrutiny. But instead it is developed by a company that claims it as a trade secret. A compelling argument states that private companies that are paid to indirectly provide a public service should be held to the same transparency standards as that public service.[103] The reasoning is twofold: First, trade secrets may "leave us by default with policies and practices that would not stand up to public scrutiny if the policies were made by legislatures in an open deliberative fashion."[104] Second, even if the defendant is aware of the inputs and outputs of the system, no party but Equivant can know to what extent the output is based on the factors.[105]

---

[100] *See Loomis v. Wisconsin* amicus brief, *supra* note 97, at 8, 9.

[101] *See Gardner v. Fla. supra* note 94, at 362.

[102] Iñigo De Miguel Beriain, *Does the use of risk assessments in sentences respect the right to due process? A critical analysis of the* Wisconsin v. Loomis *ruling*, 17.1 Law, Probability, and Risk (2018) 45-53. ("[Loomis] argued that this right [to due process] was violated because the proprietary nature of COMPAS prevents a defendant from challenging the scientific validity of the risk assessment. … In doing so, Loomis' defence appealed to what Wachter, Mittelstadt and Floridi have called a 'right to explanation'").

[103] Alyssa Carlson, *The Need for Transparency in the Age of Predictive Sentencing Algorithms.* 103 Iowa Law Rev. 303 (2017) ("Because private companies are benefiting financially by providing a public service, they should be required to conform to the same transparency requirements as public agencies."). *See also* De Miguel Beriain, *supra* note 102 ("The developers' interests do not always coincide with the social interest, and business logics do not always coincide with the need for scientific accuracy. … If the mechanism makes a mistake by recommending custody or imprisonment for what a defendant who does not need such measures, this mistake would hardly be detected, and, even if it were, it is unlikely that any huge social scandal would be created. Therefore, the developers' commercial interests would remain safe.") and Freeman *supra* note 92 at 92 ("The court misapplied the precedent and failed to account for the realities of for-profit businesses … Northpointe is a for-profit company with a $1,765,334 contract at stake in Wisconsin's use of their services."). *See also* Whittaker et al. *supra* note 87.

[104] *Id.* at 322, quoting David S. Levine, *The People's Trade Secrets?*, 18 Mich. Telecomm & Tech L. Rev. 61, 100 (2011). *See also* Natalie Ram, *Innovating Criminal Justice*, *112 Nw. U. L. Rev.* 686-687(2018) ("Chief among these [practical harms on the criminal justice system] is that such secrecy may give rise to worse algorithms by impeding effective oversight of the validity and reliability of these tools. … Outside experts repeatedly have identified algorithmic weaknesses and outright errors in proprietary source code revealed in litigation.").

[105] *See* Freeman, *supra* note 92, at 93-94 ("Neither Loomis nor the courts know to what 'extent' Northpointe based the risk assessment off of those [known] factors…. [The practitioner's guide] does not explain the value given to each factor, nor does it include a specific breakdown of every factor used in the algorithm. Rather, the company

This second reason is especially compelling because COMPAS is developed and deployed by two different entities: Equivant and local court systems respectively. Even the deployer of the system—the court system—is not privy to the inner workings of the tool.[106] And since the code is hidden even to court-appointed experts, the court has no way to validate or even measure the system involved.[107] Natalie Ram notes that, "even the technologists who [normally] eschew calls for source code transparency recommend … requiring transparency—at least to courts—of code and inputs."[108] These authors agree with several legal scholars[109] who have written opinions that trade secrecy should not be used to hide creations used in and paid for by the public sector, when the public has a vested interest in confirming the validity, accuracy, and fairness of the results, such as when a significant loss of liberty occurs when the results are of poor quality. This lack of transparency severely limits options for redress—defendants may challenge the correctness of the system, but not the correctness of the mechanism itself.

4.3 FICO Score

Regarding transparency, the FICO score is interesting because unlike UPAX, it does not necessarily deal with impacts on the level of national security and terrorism. Therefore, it's expected that people should be able to have more access to the data inputs that go into the algorithm (which they do) and the algorithm itself (which they don't).

The FICO score's transparency is also interesting by virtue of the fact that most Americans have a credit report and such, a FICO score. Unlike UPAX or COMPAS, whose most serious consequences primarily affect populations who are likely to be disadvantaged and lacking privilege, credit scores affect everyone across a wide spectrum of privilege. The pervasiveness of the FICO score means that people actively and frequently seek out its algorithm to find ways they can improve their scores.

---

hides details of its algorithm by declaring that it is a 'core piece of [its] business' and, as such, the company maintains it must shield the code from examination because of its proprietary nature.").

[106] *See State v. Loomis*, *supra* note 93, at 27.

[107] Ram, *supra* note 104, at 688 ("Other mechanisms intended to ensure the accuracy, validity, and reliability of criminal justice algorithms fall short in the absence of outside code review. For instance, one group of authors has counseled reliance on validations studies to ensure accuracy and fairness… but reliance on validation studies in place of source code access, rather than alongside it, is likely insufficient to verify that software has performed as its designer claims.").

[108] *Id.* at 690.

[109] *See supra* notes 102-105.

Individuals can access their credit report at no cost once a year from each of the three national credit bureaus. These reports do not, however, directly include FICO scores. Instead, individuals will typically have access to their FICO score through banks like American Express or Citibank which will give them a calculated score based off of one of their credit reports. The data used for the FICO score—in an individual's credit report—is extremely transparent.

On the other hand, the formula used to compute the FICO score is kept secret. The score is calculated by the FICO company who profits off of it, which means that the algorithm is a trade secret and not disclosed. Rather, there is the expectation from the financial institution that FICO's algorithm will be the most profitable for the institution, and loan-seekers expect that FICO's algorithm is "fair" in terms of their credit histories. In reality, fairness for the loan-seekers may be compromised due to factors out of their control—such as the effects of outstanding medical debt and systemic oppressions as discussed in sections 2.3 and 3.3.

FICO is generally clear about the categories of items that are considered to calculate the score (see Section 3.3), but the weights of the data points are unavailable to the public for individuals. The weights are released for the general population, but they are shifted based off of the specific report of the individual using undisclosed mechanisms. Therefore, there's little information at all readily available about the FICO score algorithm and is not transparent.

There are two levels of transparency surrounding auditing of a risk assessment mechanism such as this one. The first level is how easily can the auditing bodies get information about the system's algorithm and the inputs. In the case of the FICO score, federal agencies such as the Bureau of Consumer Protection can request and receive that information quite easily. The second level of transparency is how much people know about the agency that audits the system. It's easily found that the CFPB is primarily responsible for holding FICO responsible, but it's not necessarily clear to the public how these audits take place or the procedure to seek redress. This procedure is especially hard to educate about in cases like the FICO score where the content is numerical, complicated, and requires specialized education. A Financial Capability Study found that almost two-thirds of Americans cannot pass a financial literacy test.[110]

Under the FCRA, in the case of erroneous reporting by a credit reporting agency, an individual can seek redress as well as punitive damages. This is contrasting to UPAX and

[110] Lin, Judy T., et al. "Financial capability in the United States 2016." *Washington, DC: FINRA Investor Education Foundation* (2016).

COMPAS, which have no redress mechanisms regarding incorrect data inputs. Nevertheless, these redress policies are not very clear when it comes to how to find out if your data is erroneous. This is a problem because a Federal Trade Commission study looking into the US credit reporting industry found that 25% of consumers "identified errors on their credit reports that might affect their credit scores."[111] Of those, one-fifth of consumers had serious errors that could result in them paying more for auto loans and insurance. People should be checking their credit reports manually and often, but they often don't know they have to do so or don't have the knowledge of *how* to do so. In 2017, the Consumer Financial Protection Bureau ordered Equifax and Transunion "to pay more than $23 million in fines and restitution" for taking advantage of Americans.[112]

5. Conclusions and Proposals

This section contains conclusions on the properties of each system we analyzed. In section 5.4 we describe our generalized proposals for all governmentally-regulated risk assessment systems.

5.1 Unified Passenger (UPAX)

As we described earlier, a major concern about UPAX and other confidential big data blacklisting systems is that the individuals being judged by the system do not have the ability to force redress. This is an issue for all those within the UPAX system, but it hits especially hard to those who come to the U.S. fleeing persecution in their previous homes. This exacerbates a problem where those the most in need of correcting missing documentation are likely to have the fewest resources to correct their inputs to the system.

UPAX also suffers from a lack of accountability. Complete transparency to the public is a tall order for a risk assessment designed for counterterrorism and law enforcement purposes, but there is a middle ground that alleviates significant costs without sacrificing security. While many

[111] In FTC Study, Five Percent of Consumers Had Errors on Their Credit Reports That Could Result in Less Favorable Terms for Loans, https://www.ftc.gov/news-events/press-releases/2013/02/ftc-study-five-percent-consumers-had-errors-their-credit-reports (last accessed Nov 26, 2018).
[112] Two Major Credit Reporting Agencies Have Been Lying to Consumers, https://www.theatlantic.com/business/archive/2017/01/credit-scores-cfpb/512162/ (last accessed Nov 26, 2018).

of the Privacy Act exemptions claimed by UPAX serve reasonable security needs, some do not. Specifically, the UPAX exemption from 5 U.S.C. § 552a(d)(3) and (4),[113] which would normally require CBP to justify refusals to redress requests, do not seem to serve a legitimate law enforcement or national security purpose.

Furthermore, the auditing done by the Government Accountability Office is not frequent enough to judge major changes in the system; UPAX replaced ATS-P in early 2017, but still no audit of it has been done to the knowledge of the public. ATS-P has been called for additional auditing by congress in the past, but relying on congress or the administration to proactively take the actions required to hold all big data blacklists accountable is not viable in a rapidly shifting political climate.

5.2 COMPAS Risk & Need Assessment System

COMPAS, like all the systems we examined, can sometimes cause a bad self-fulfilling prophecy in which information about a defendant's families and peers is used to heighten the defendant's risk score, thus raising the risk score of innocent individuals and their descendants.

Unlike UPAX, however, the methods of risk assessment tools in sentencing need not be kept hidden for national security reasons. The courts seem to have accepted the opacity of these tools without knowing how they work or the technical analyses they have undergone. Using a proprietary algorithm causes harm to all parties involved except the algorithm vendor - the individuals judged cannot adequately challenge the tool, the courts are spending large sums of money on the tools, and society is harmed by the increased polarization of national public trust of policing.[114]

Not all recidivism risk assessment tools are created by for-profit companies. The Public Safety Assessment (PSA)[115] is developed openly by a group of academics and non-profits.[116]Using an openly-developed algorithm like the PSA over a proprietary system like

---

[113] *See supra* note 10.

[114] Jim Norman, *Confidence in Police Back at Historical Average*, Gallup (2017) https://news.gallup.com/poll/213869/confidence-police-back-historical-average.aspx ("Overall Rise in Confidence Masks Drop Among Hispanics, Liberals, Younger Adults").

[115] *Supporting the Implementation of the PSA Tool*, Justice System Partners, https://justicesystempartners.org/projects/supporting-the-implementation-of-the-psa-tool/ (last accessed Nov. 25, 2018).

[116] *Public Safety Assessment: Risk Factors and Formulas*, Laura and John Arnold Foundation

COMPAS removes many of the hurdles to the due process issues, and the system itself has many more incentives to perform better since it is under more scrutiny. Using a proprietary algorithm over a public alternative is costly to the courts in terms of both dollars and public trust. Secret algorithms also limit the options of individuals who believe they were incorrectly judged. Mandating public access—or at least court access—to risk assessment algorithms used in sentencing would help restore public confidence in the criminal justice system and would alleviate some public distrust of these tools.

5.3 FICO Score

The ubiquity of the FICO score means that almost every American is affected by it, across social and economic classes. While the FCRA attempts to correct for systemic injustices (via seven year delinquency "expirations"), the FICO score is still affected by historical racial oppression and perpetuates cumulative disadvantage. Nevertheless, the credit score's predecessor system—an interview with a loan officer—is limited by insufficient information about the loan-seeker and will inevitably input personal biases and prejudices. The FICO score is fairer, as long as it is applied non-frivolously. This is an area in which legislation can intervene, and with existing precedent. For example, New York City has outlawed checking credit scores for employment checks.[117]

Despite nominally affecting those in higher economic classes as well, the FICO score is more easily manipulated by people with access to knowledge and resources, contributing further to cumulative disadvantage. While the FCRA offers redress functionality, education programs for seeking redress are sparse, despite their importance. Regardless of redress functionality, national credit reporting agencies need to be audited more often as there are large numbers of unchecked inaccuracies within their systems with sometimes disastrous consequences.

---

https://www.arnoldfoundation.org/wp-content/uploads/PSA-Risk-Factors-and-Formula.pdf (last accessed Nov. 25, 2018). *See also* Danielle Kehl *et al.*, *Algorithms in the Criminal Justice System: Assessing the Use of Risk Assessments in Sentencing.* Responsive Communities Initiative, Berkman Klein Center for Internet & Society, Harvard Law School at 28 https://dash.harvard.edu/bitstream/handle/1/33746041/2017-07_responsivecommunities_2.pdf (last acccessed Nov. 25, 2018) ("Academic researchers and governments, by contrast [to for-profit companies] tend to have more incentives to make the details of their algorithms publicly available and ensure that they are subject to appropriate scrutiny and oversight.").

[117] New York City Just Outlawed Running Credit Checks on Job Applicants, https://www.thenation.com/article/new-york-city-just-outlawed-running-credit-checks-job-applicants/ (last accessed Nov 26, 2018).

5.4 Generalized Proposals

Using the analysis of the three case studies in the paper, we propose three policy-based safeguards to place checks against the general class of governmentally-regulated risk assessment mechanisms. These three checks consider the three axes discussed in this paper as well as a general notion of fairness to the affected parties. Generally, we will call these proposals requests that mechanisms be rectifying, reparative, and responsible.

First: rectifying. Regarding the data inputs, there need to be easy channels for individuals to fix incorrect information. Easy refers to the process with an individual finds out information about the algorithm's inputs, the input data, and also the redress mechanism. For systems with inputs known to the individual (FICO and COMPAS), individuals should be able to request redress via a mechanism resembling the Fair Credit Reporting Act. For systems with secret inputs (UPAX), as discussed in section 4.1, individuals should have standing to sue CBP to request corrections to their information as part of a DHS Traveler Redress Inquiry Program, much like they have standing to sue the Terrorist Screening Center for failing to respond to the same redress program.[118] As a matter of data integrity, these opaque risk assessment systems should also propagate corrections to the databases they uses as inputs. And, as much as possible, they should attempt to detect discrepancies between their data sources, and investigate potential mismatches.

Second: reparative. Regarding cost-bearing, risk assessment mechanisms must be mandated to consider past institutional oppression/harm. Many mechanisms perpetuate cumulative disadvantage through their use. In fact, all three of the case studies analyzed disproportionately impact marginalized communities. The communities most affected are usually the ones without resources to seek redress. This problem is made worse by the fact that risk assessments rely on similarity, creating a self-perpetuating cycle that artificially heightens the risk scores of these communities.

Finally: responsible. Even with the implementation of the previous two safeguards, there needs to be more accountability down the line. This could be executed either via transparency of data inputs or methodologies, or via frequent comprehensive audits by one or more trusted entities which specifically protect the assessed cohorts. The less transparent a system is, the more the system-controller has a responsibility to address the first two issues itself, since others do not

[118] *Latif v. Holder*, 28 F. Supp. 3d 1134 8489 (2014).

have sufficient information to do so. Recognizing that some mechanisms cannot afford complete transparency due to national security or law enforcement concerns, removing trade secret exemptions from these public services would go a long way toward improving the accountability of these systems. The U.S. government is using risk assessments for more purposes as time goes on, and thus we must ensure that these assessments, like other government functions, are accountable to the people.