

Transparency and Fairness in Criminal Justice and Housing

Cris Moore, Santa Fe Institute

Interdisciplinary Working Group on Algorithmic Justice



Melanie Moses
CS, UNM / SFI



Cris Moore
SFI



Kathy Powers
Poli Sci, UNM/ SFI



Alfred Mathewson
Law, UNM



Sonia Rankin
Law, UNM



Mirta Galesic
SFI



Josh Garland
Arizona State



Matthew Fricke
CS, UNM



Gabe Sanchez
Poli Sci, UNM



Tina Eliassi-Rad
CS, NEU/SFI



Mahzarin Banaji
Psych, Harvard/SFI



Trilce Estrada
CS, UNM



Nadiyah Humber
Law, UConn

High-Stakes Decisions

Algorithms are being used in both the public and private sector to make decisions that have long-term effects on people's lives:

Employment (automated hiring)

Health care and social services

Housing: lending, tenant screening, public housing waiting lists

Criminal justice: pretrial, sentencing, parole

How can we tell whether these algorithms work?

How can we have an informed democratic discussion about whether and how they should be used, and whether we should spend taxpayer \$\$ on them?

Transparency vs. Black Boxes

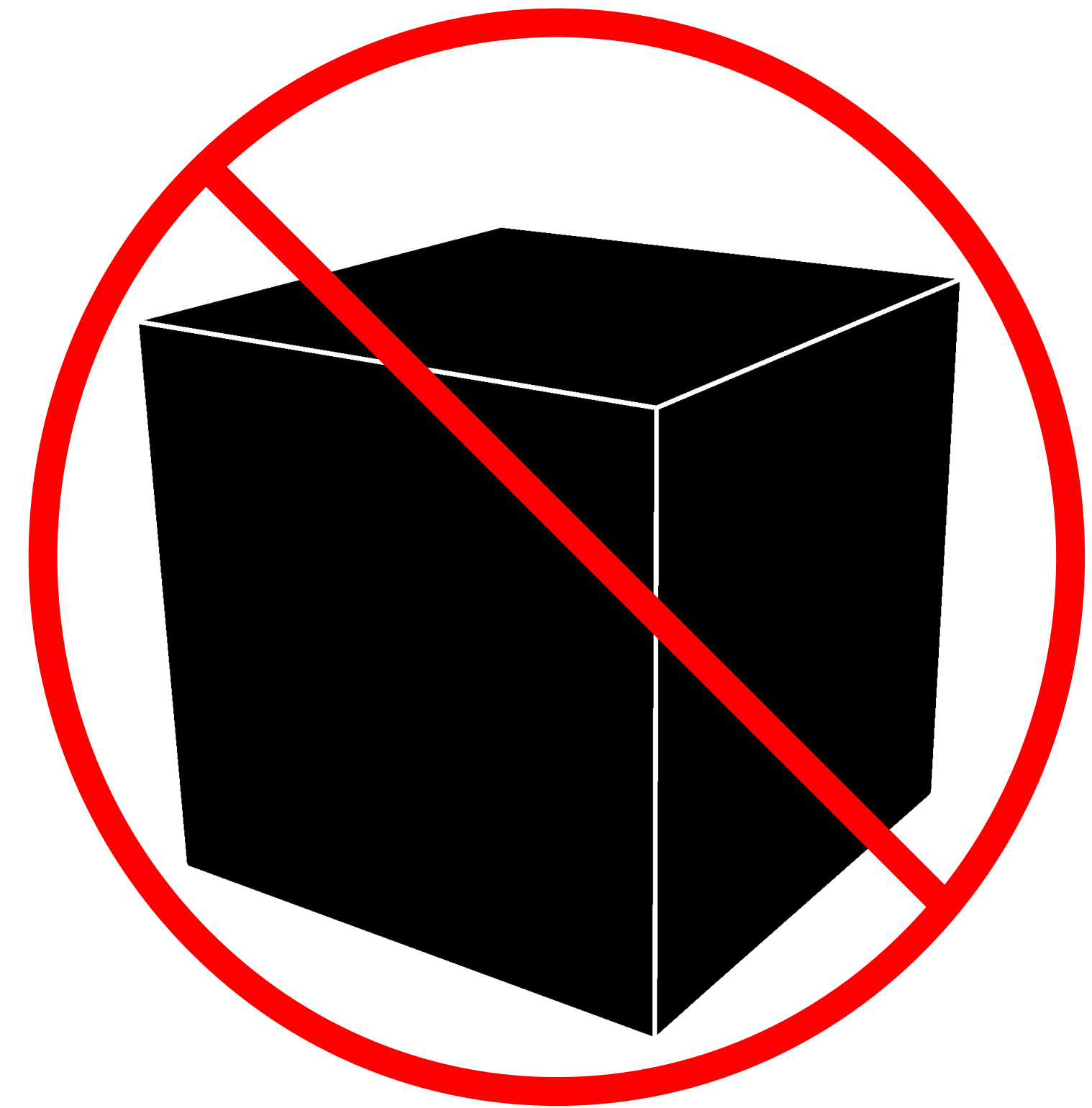
What data does the algorithm use about a defendant or applicant?

How does it weight and combine these factors?

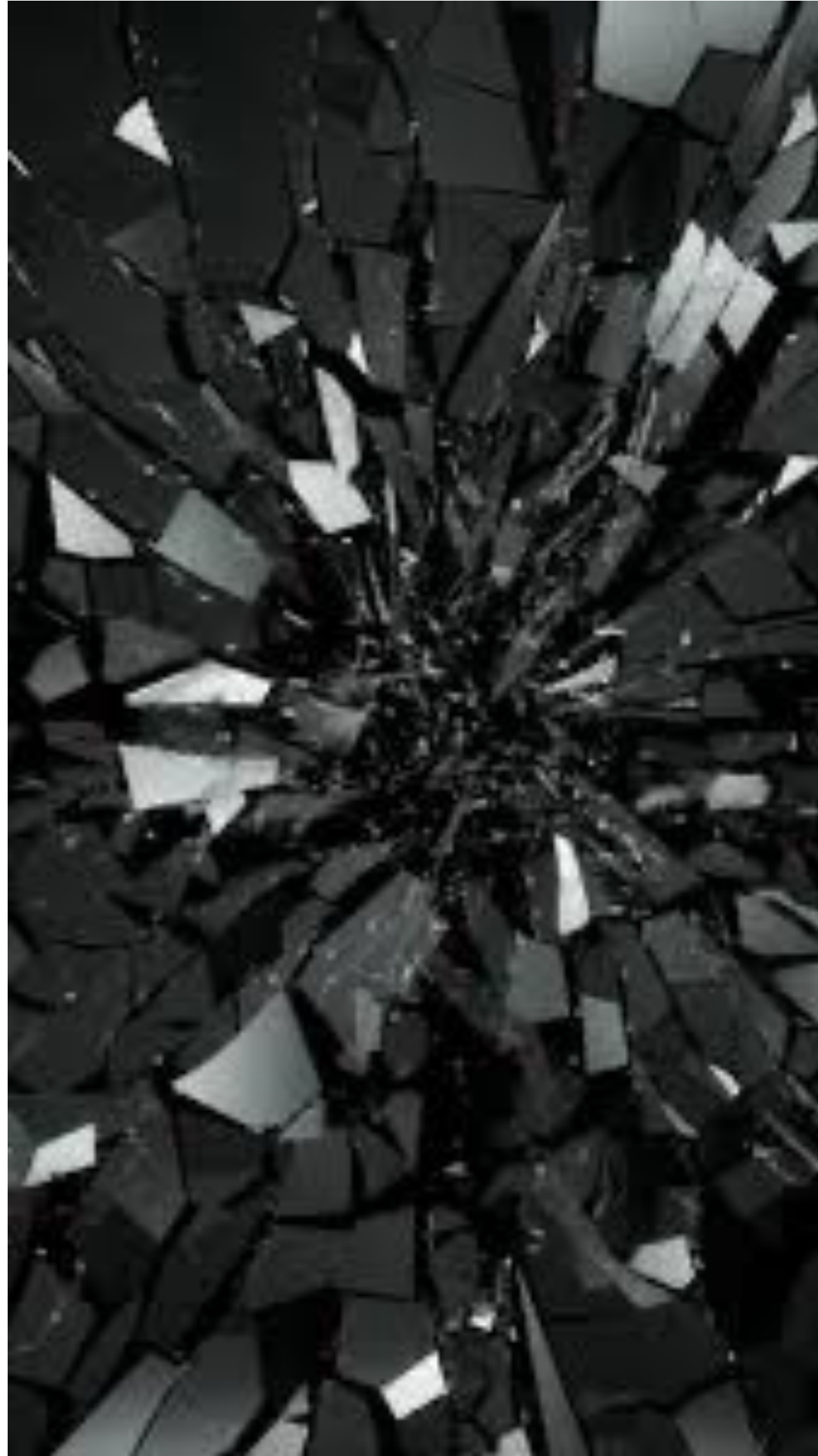
Where does this data come from?

How was it collected and curated?

How was the algorithm designed or trained?



Transparency vs. Black Boxes



How do decisionmakers interpret an algorithm's outputs?

Do they understand how its scores are derived, and what kinds of errors it can make?

What does "high risk" mean:
How much risk, and risk of what?

Algorithms in Criminal Justice

Big Data to the Rescue?



Anne Milgram, former New Jersey Attorney General



Machine Bias

There's software used across the country to predict future criminals. And it's biased against blacks.

by Julia Angwin, Jeff Larson, Surya Mattu and Lauren Kirchner, ProPublica

May 23, 2016

Two competing algorithms or “risk assessment tools”

COMPAS: Northpointe / equivant

137-item questionnaire and interview

Proprietary (secret) formula

Public Safety Assessment (PSA): Arnold Foundation

In 40 jurisdictions (and rising)

Just 9 factors from criminal record

Simple, publicly known formula

Public Safety Assessment (PSA) (Arnold Ventures)

Bernalillo County,
Second Judicial District

Simple point system,
publicly known weights

Past convictions, not arrests

Doesn't use juvenile record

Uses age, but not gender,
employment, education,
or environment

PUBLIC SAFETY ASSESSMENT RISK FACTORS

RISK FACTOR	WEIGHTS
FAILURE TO APPEAR maximum total weight = 7 points	
Pending charge at the time of the offense	No = 0 Yes = 1
Prior conviction	No = 0 Yes = 1
Prior failure to appear pretrial in past 2 years	0 = 0 1 = 2 2 or more = 4
Prior failure to appear pretrial older than 2 years	No = 0 Yes = 1
NEW CRIMINAL ACTIVITY maximum total weight = 13 points	
Age at current arrest	23 or older = 0 22 or younger = 2
Pending charge at the time of the offense	No = 0 Yes = 3
Prior misdemeanor conviction	No = 0 Yes = 1
Prior felony conviction	No = 0 Yes = 1
Prior violent conviction	0 = 0 1 or 2 = 1 3 or more = 2
Prior failure to appear pretrial in past 2 years	0 = 0 1 = 1 2 or more = 2
Prior sentence to incarceration	No = 0 Yes = 2
NEW VIOLENT CRIMINAL ACTIVITY maximum total weight = 7 points	
Current violent offense	No = 0 Yes = 2
Current violent offense & 20 years old or younger	No = 0 Yes = 1
Pending charge at the time of the offense	No = 0 Yes = 1
Prior conviction	No = 0 Yes = 1
Prior violent conviction	0 = 0 1 or 2 = 1 3 or more = 2

Conditions of Release Matrix (Bernalillo County until 2023)

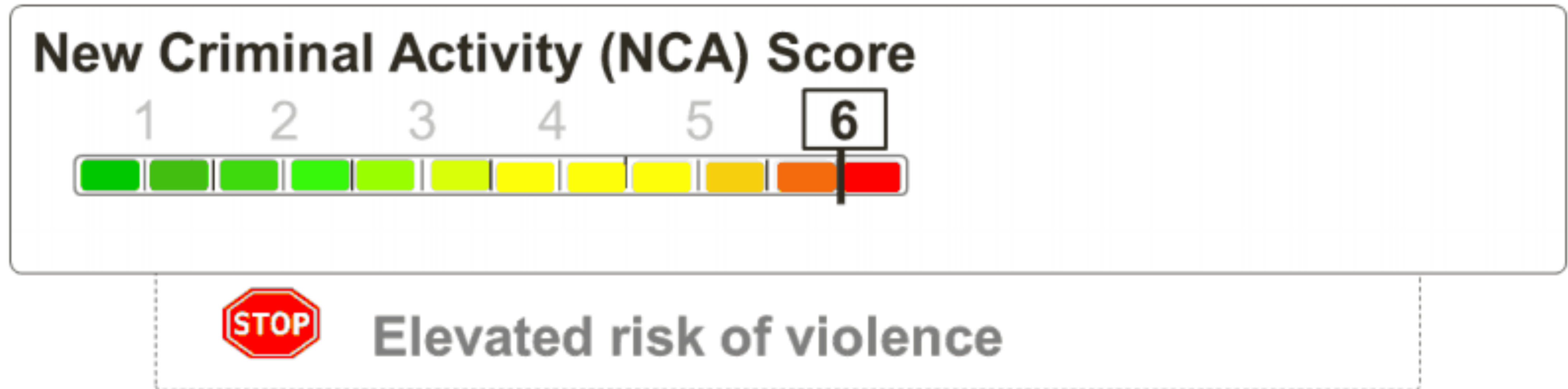
New Criminal Activity (NCA) Score



Elevated risk of violence

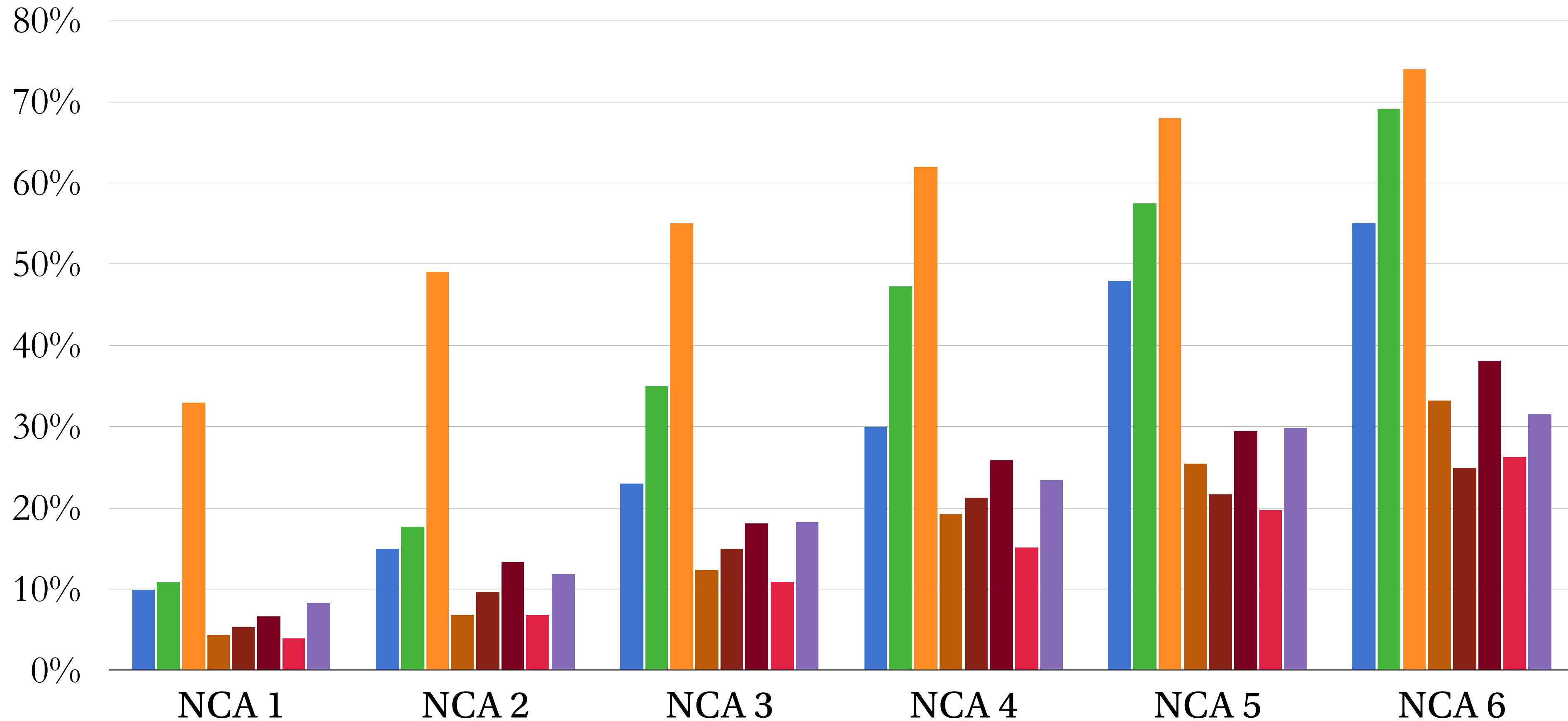
		New Criminal Activity Scale					
		NCA 1	NCA 2	NCA 3	NCA 4	NCA 5	NCA 6
Failure to Appear Scale	FTA 1	(A) ROR	(B) ROR				
	FTA 2	(C) ROR	(D) ROR	(E) ROR-PML 1	(F) ROR-PML 3	(G) ROR-PML 4	
	FTA 3		(H) ROR-PML 1	(I) ROR-PML 2	(J) ROR-PML 3	(K) ROR-PML 4	(L) Detain or Max Conditions
	FTA 4		(M) ROR-PML 1	(N) ROR-PML 2	(O) ROR-PML 3	(P) ROR-PML 4	(Q) Detain or Max Conditions
	FTA 5		(R) ROR-PML 2	(S) ROR-PML 2	(T) ROR-PML 3	(U) Detain or Max Conditions	(V) Detain or Max Conditions
	FTA 6				(W) Detain or Max Conditions	(X) Detain or Max Conditions	(Y) Detain or Max Conditions

What do PSA scores really mean?

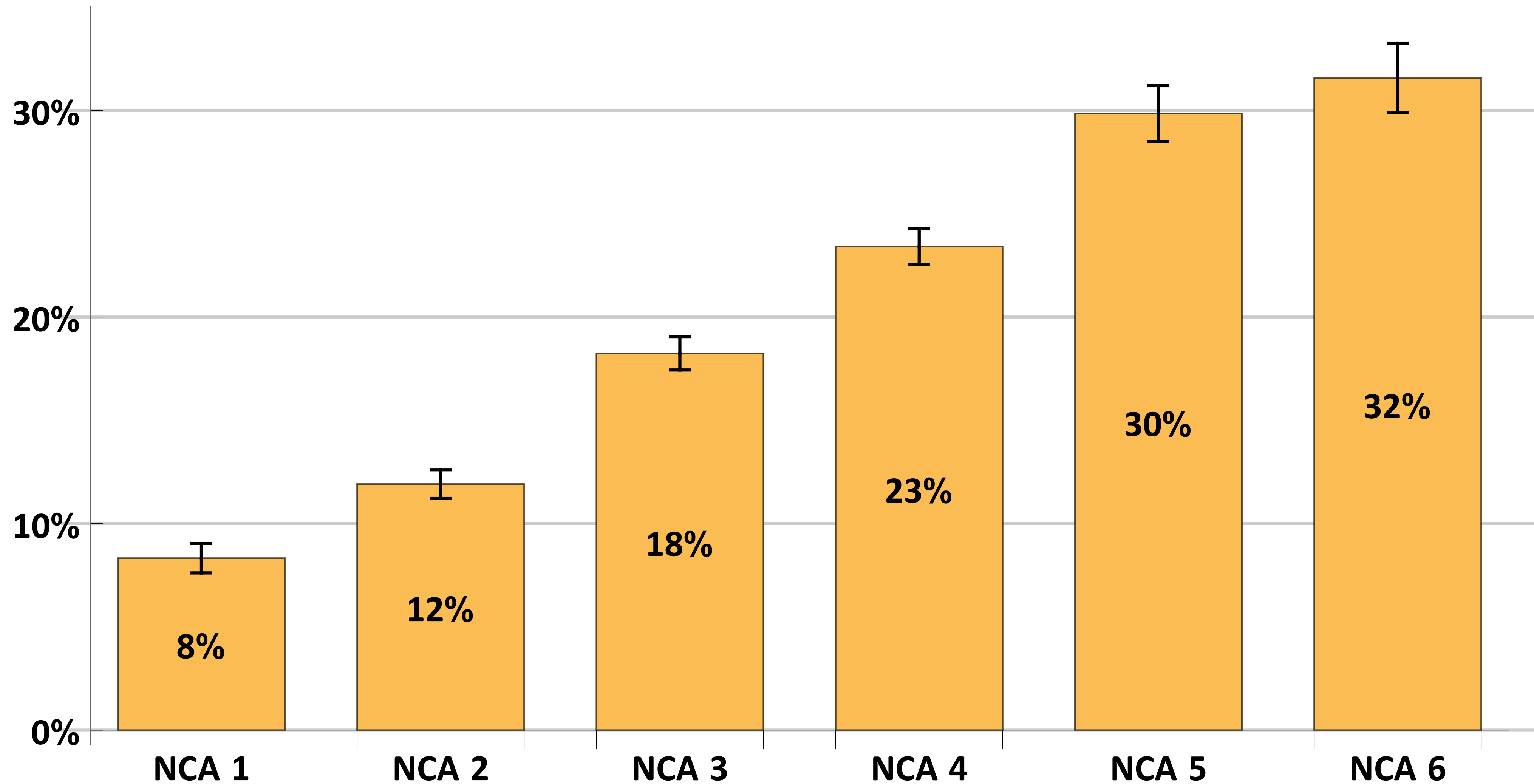


How much risk, and risk of what?

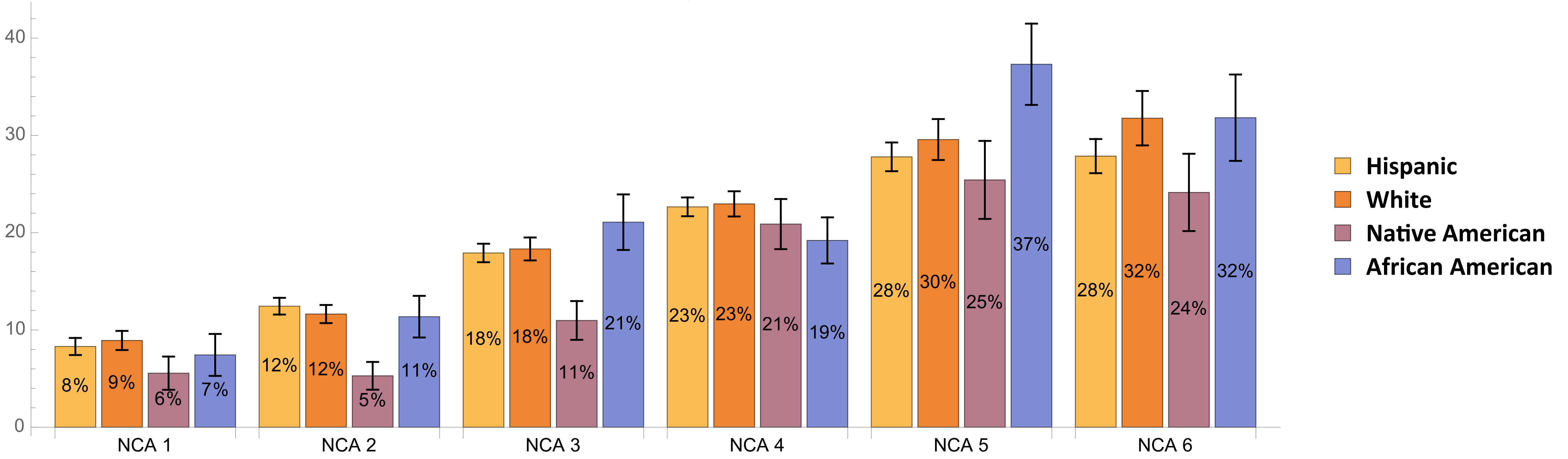
The Need for Local Revalidation Studies: Each Jurisdiction is Different



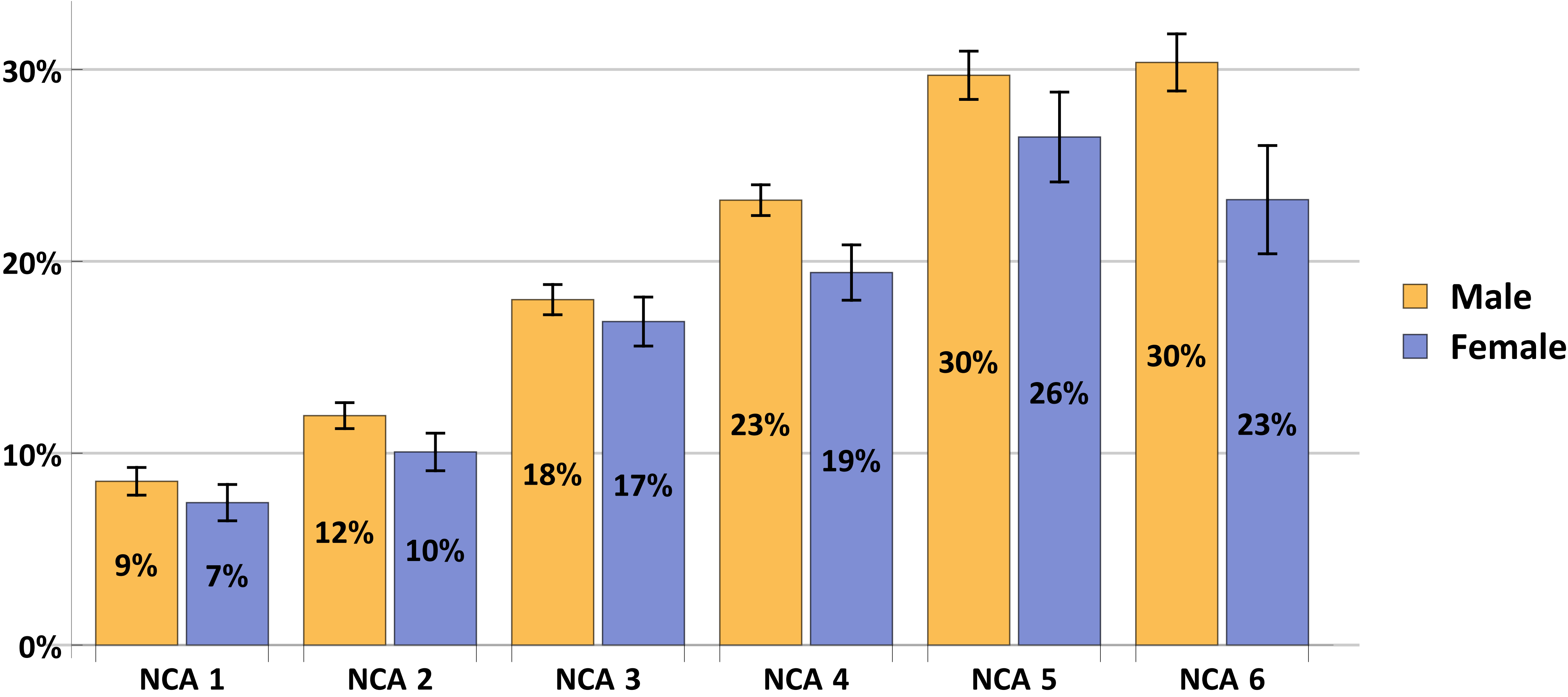
Audit for Accuracy: New Criminal Activity (rearrest) by NCA score



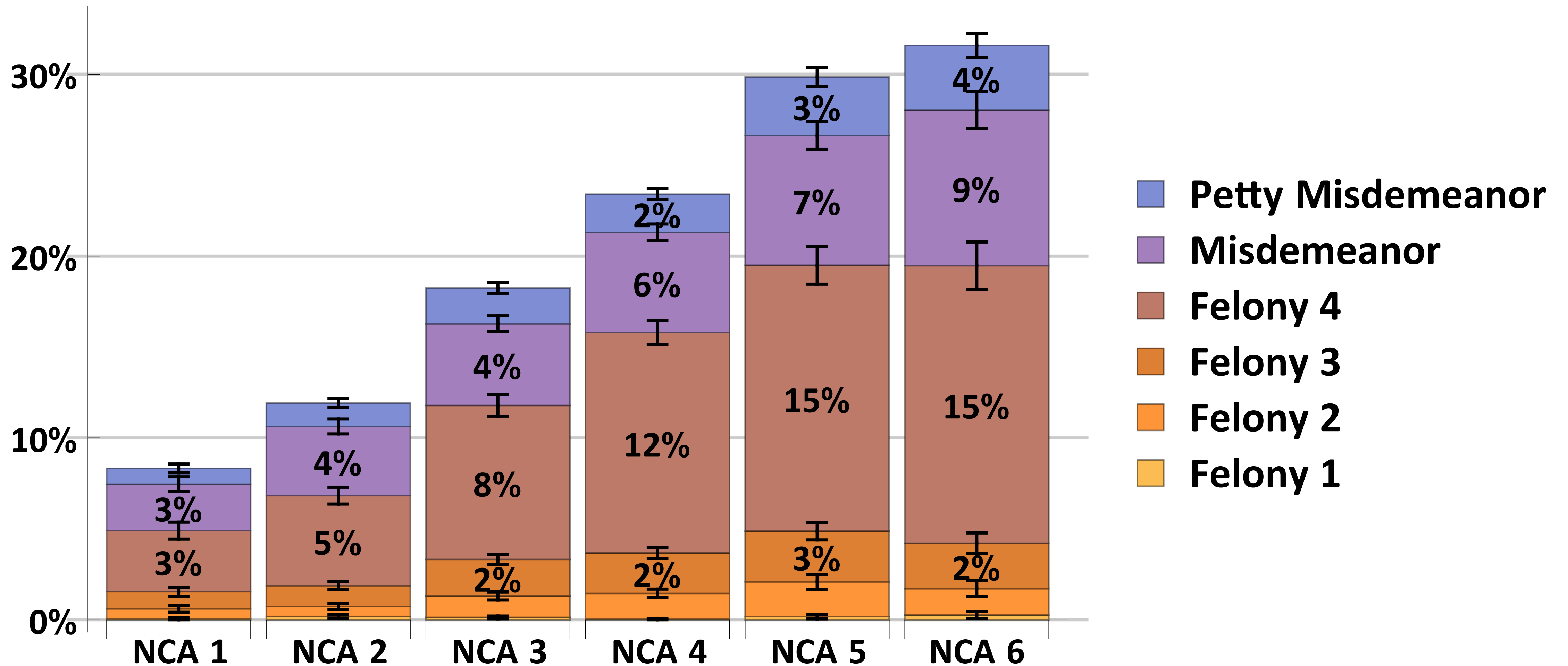
Audit for Fairness: New Criminal Activity by Race



Audit for Fairness: New Criminal Activity by Gender

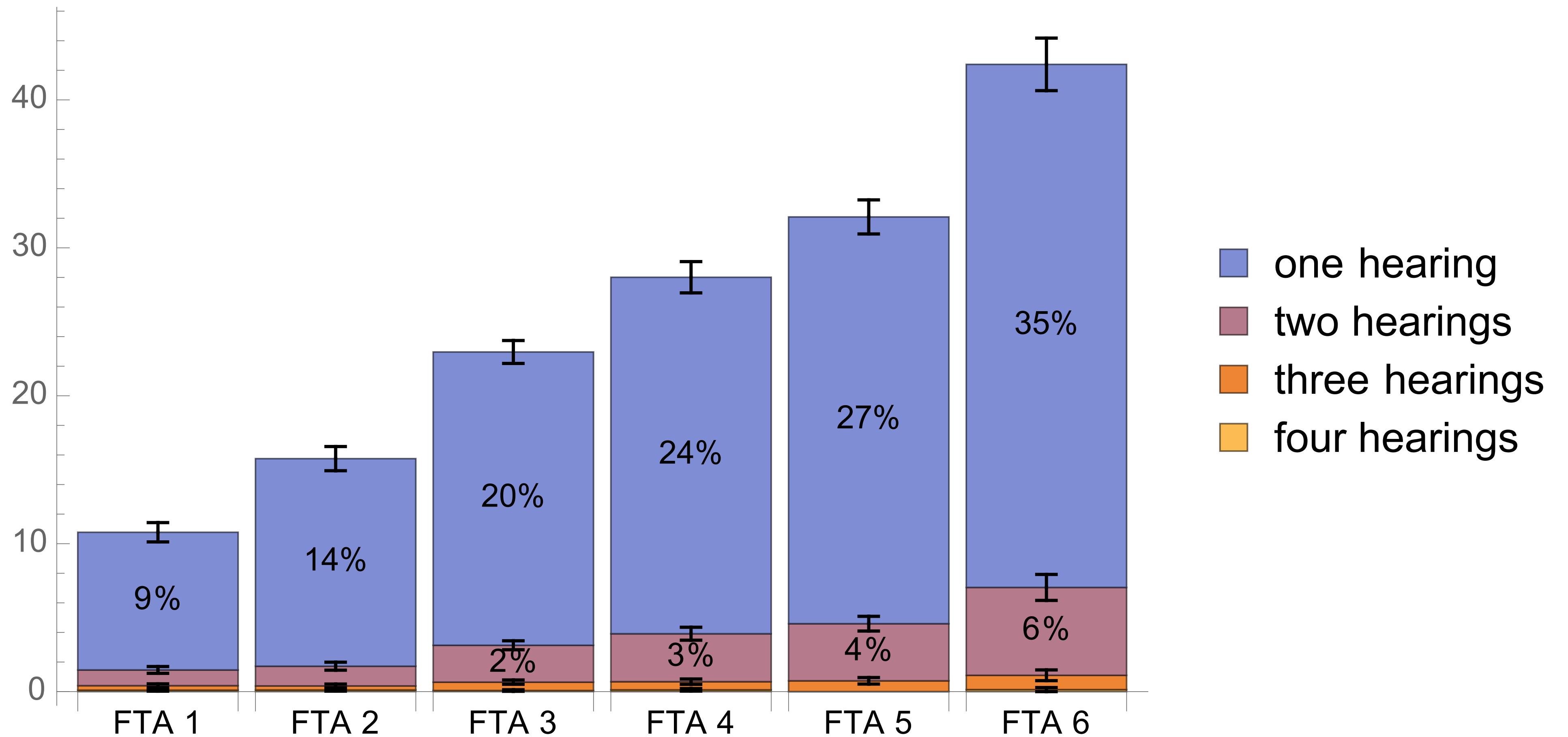


What does “New Criminal Activity” really mean?



Most new charges less severe than source charge
Most are misdemeanors or low-level felonies

What does “Failure to Appear” really mean?



86% of those with FTAs only missed one hearing

56% only missed preliminary hearing

Reminders, transportation, jobs, child care: not flight risks

What does “Accuracy” mean anyway?

Table 2. Successfully predicted crimes under deployed conditions

	ETAS				Analyst				Boost	<i>P</i> -value
	Success	Total	Rate	PAI	Success	Total	Rate	PAI		
Foothill	22	346	6.4%	16.9	11	347	3.2%	8.4	2.0	0.0244
N. Hollywood	21	611	3.4%	4.9	12	732	1.6%	2.4	2.1	0.0170
Southwest	38	981	3.9%	2.9	21	936	2.2%	1.7	1.7	0.0194
Total	81	1938	4.2%	6.8	44	2015	2.2%	3.5	1.9	0.0002

Mohler et al., Randomized Controlled Field Trials of Predictive Policing
Journal of the American Statistical Association (2015)

a 6 month randomized controlled trial found that crime analysts using PredPol technology in addition to their existing tools are **twice as effective** as experienced crime analysts using hotspot mapping alone.



Algorithms can help inform high-stakes decisions if...

People affected by them (e.g. applicants, defendants) understand what data about them is used and how their scores are derived

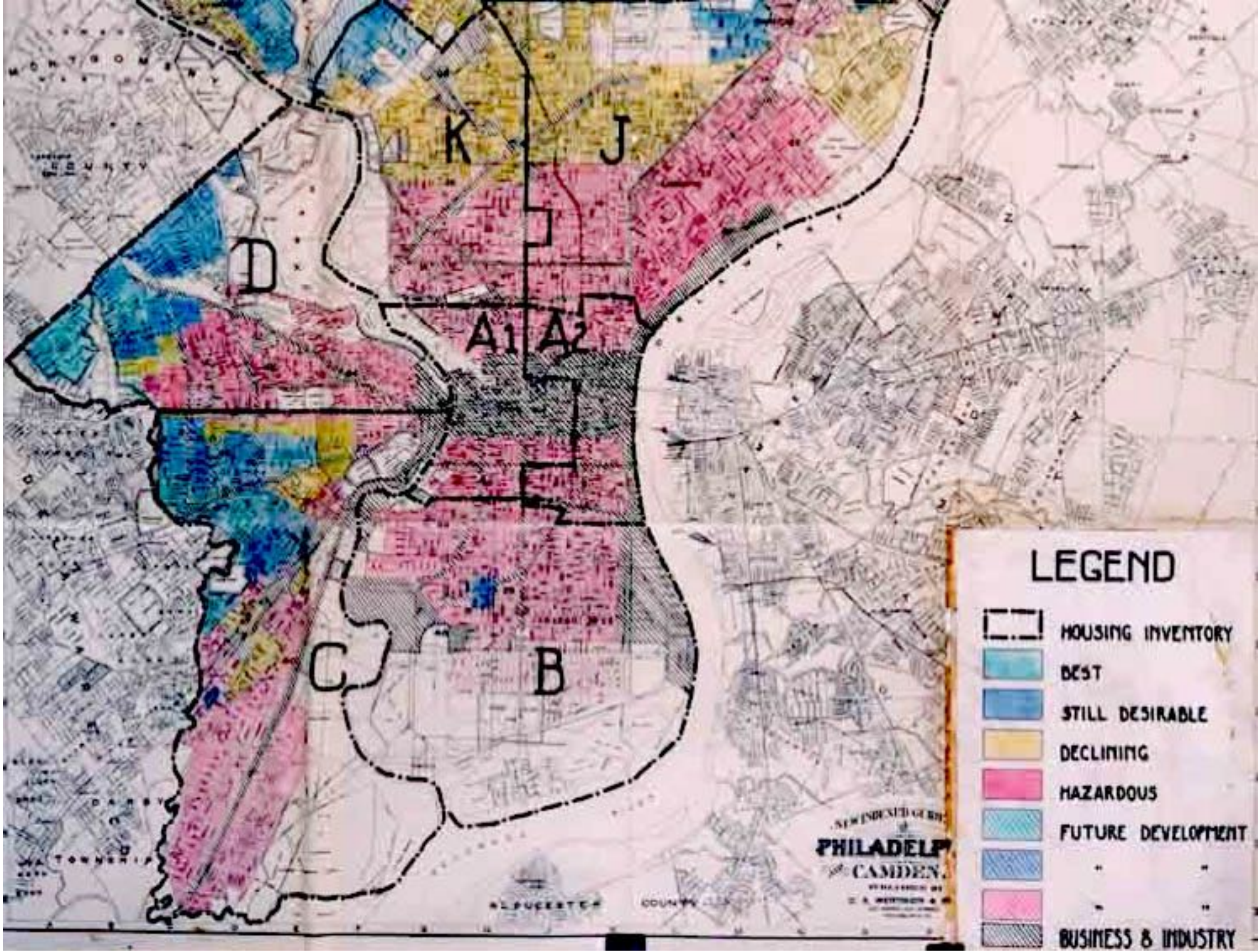
Decision makers advised by them (e.g. judges) understand what they mean and what mistakes they can make, and also take individual information into account

They are regularly and independently audited for accuracy and fairness, rather than relying on vendor's claims

All this requires transparency

Algorithms in Housing

Proxies and Redlining



“disparate treatment” vs. “disparate impact”

The Markup

Big Tech Is Watching You. We're Watching Big Tech.

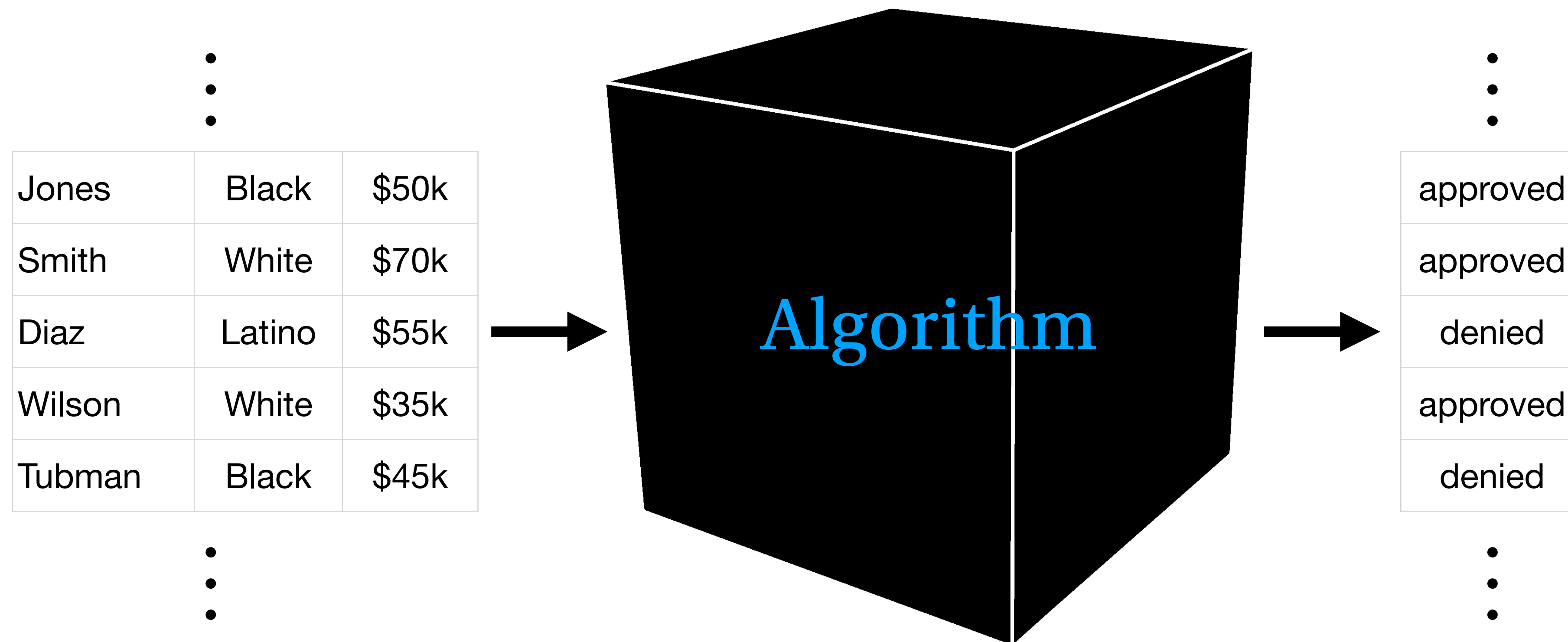
Locked Out

The Obscure Yet Powerful Tenant-Screening Industry Is Finally Getting Some Scrutiny

Reforms have been in the works for years, but a looming eviction crisis has made them urgent

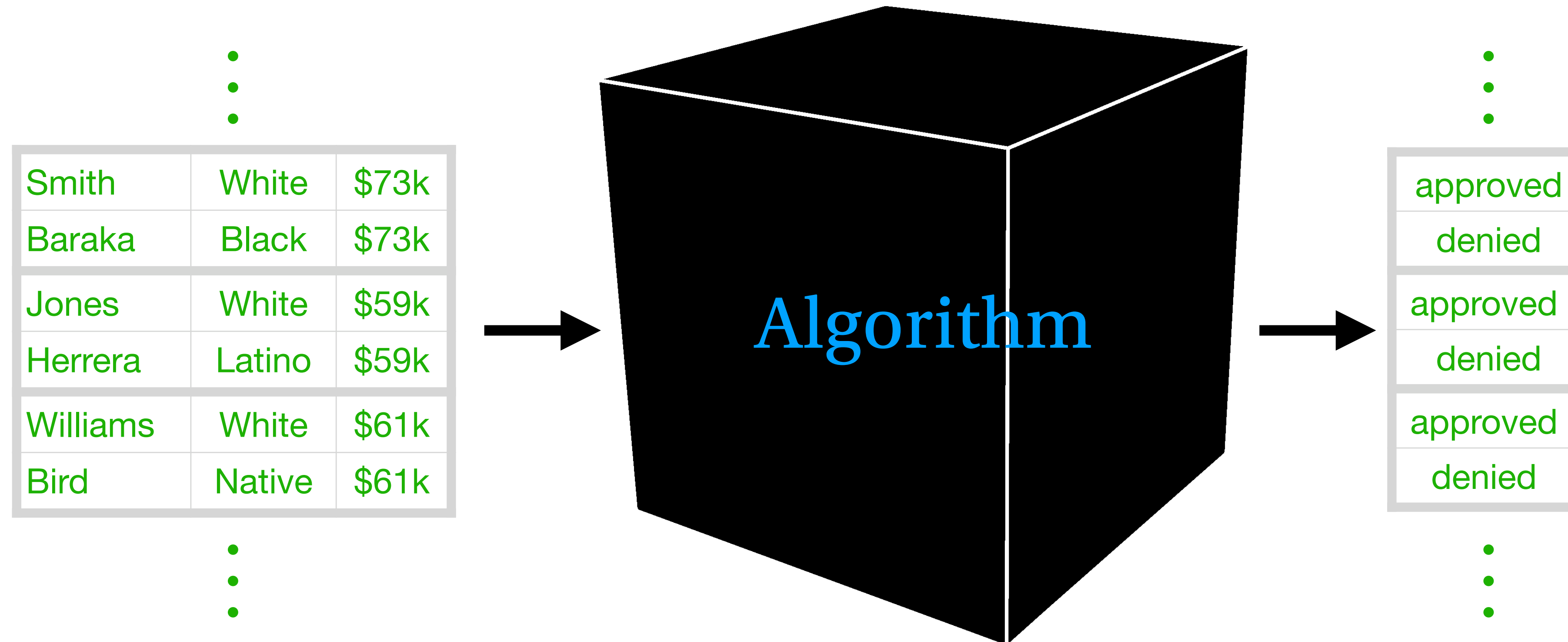
By [Lauren Kirchner](#)

Levels of Transparency: Historical Data



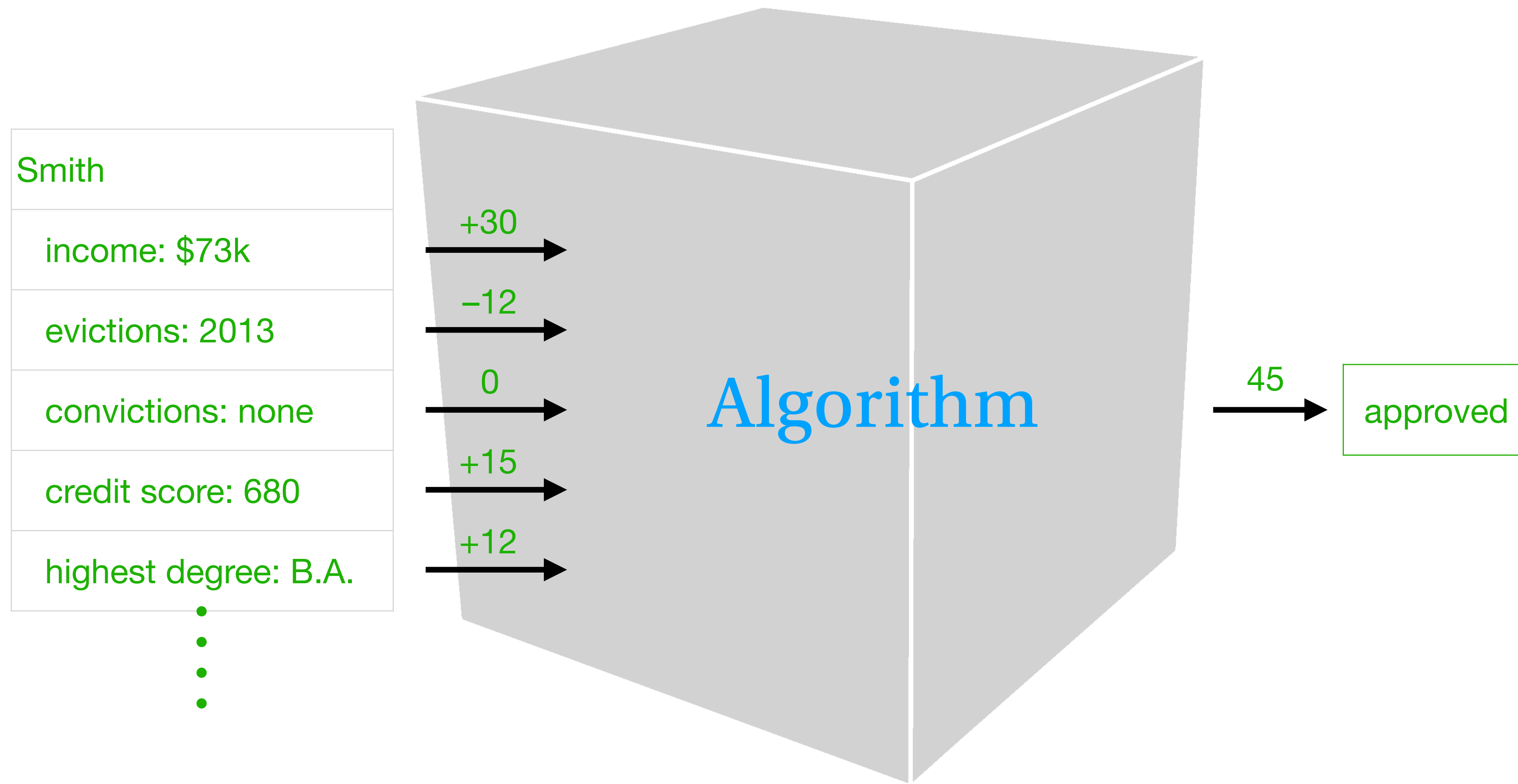
doesn't establish causality
no disclosure law (HMDA) for tenant screening

Levels of Transparency: Black-Box Testing



violates terms of service: Facebook ads
hard to create fictional profiles

Levels of Transparency: Internal Math/Logic



trade secrets, intellectual property

Legislative Ask:

Transparency in Public Sector Algorithms

California SB36: pretrial risk assessment tools must be transparent and revalidated every three years

California AB331 (proposed): impact statements for “automated decision tools” used in “consequential decisions” (including in the private sector)

Vermont H410, Connecticut SB1103: transparency, impact assessments, and testing for bias in algorithms used or procured by the state

Transparent Automated Governance (TAG) Act (Peters, Braun, Lankford) would require government agencies to notify people when AI systems are used, and provide an appeals process with human oversight

Proposal: require transparency and auditability for any algorithm that state or local governments use to make or inform life-altering decisions

Questions?