How to Achieve Both Transparency and Accuracy in Predictive Decision Making: an Introduction to Strategic Prediction

Ben Edelman
bedelman@g.harvard.edu

Chara Podimata
podimata@g.harvard.edu

Yo Shavit
yonadav@g.harvard.edu

Harvard

Tutorial outline

Part I

- What is strategic decision-making and who are the key actors?
- The robustness perspective

Strategic prediction

→ improve GPA
→ retake GRE / pay for classes
→ change schools

Business

Student tracking, secret scores: How college admissions offices rank prospects before they apply

Before many schools even look at an application, they comb through prospective students' personal data, such as web-browsing habits and financial history

ML algorithms making consequential decisions are almost everywhere nowadays.

The New York Times

Is an Algorithm Less Racist Than a Loan Officer?

Digital mortgage platforms have the potential to reduce discrimination. But automated systems provide rich opportunities to perpetuate bias, too.

→ increase # credit cards
→ increase # bank accounts
→ improve credit history

HireVue Platform Why HireVue Hiring Resources

Your end-to-end hiring platform with video interview software, conversational AI, and assessments.

Build a faster, fairer, friendlier hiring process with HireVue's end-to-end hiring platform. Together, we can improve the way you discover, engage, and hire talent.

→ dress a certain way
→ hide piercings / tattoos
→ change way you talk

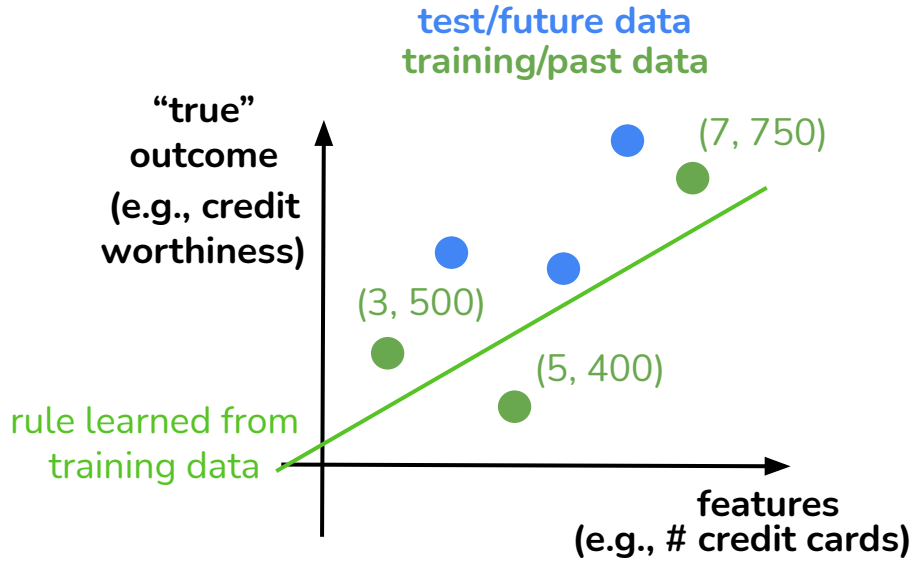
But **why** use automated decision making/ML?

Patterns in **training/past** data = patterns in **test/future** data

→ **abundance of people's data** and the **heart of ML paradigm**

Standard ML

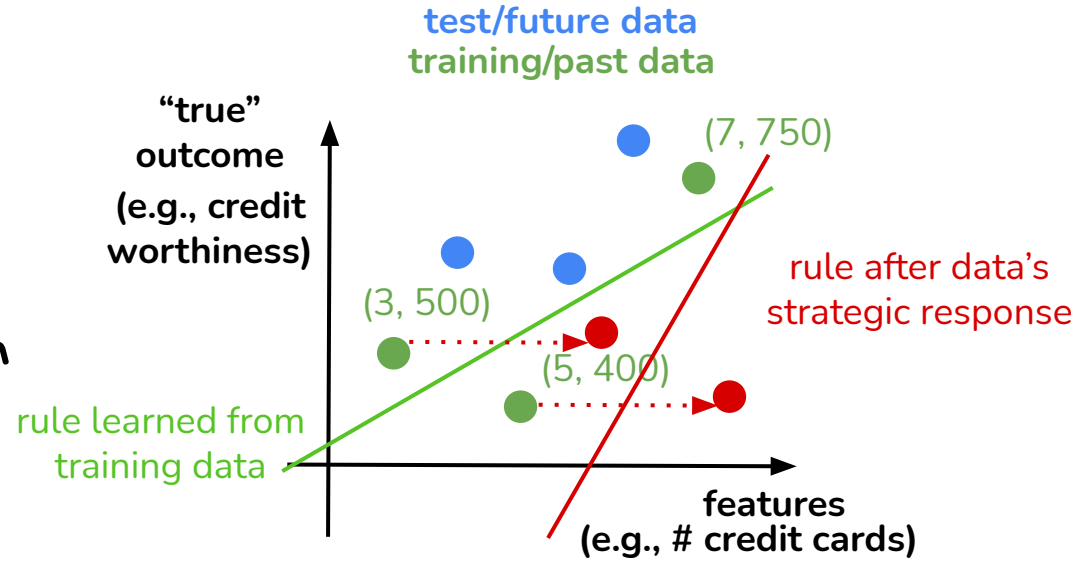
Patterns in **training/past** data = patterns in **test/future** data



- Features: e.g., age, education level, ZIP code, # credit cards, # bank accounts, past credit history, etc.
- "True" outcome: e.g., current credit score, loan/mortgage qualification, etc.

Why standard ML is not enough

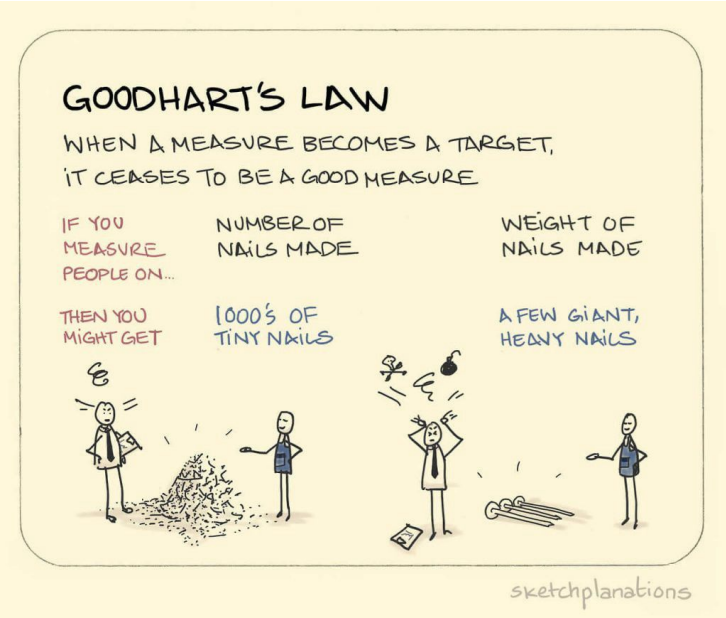
Patterns in **training/past** data = patterns in **test/future** data

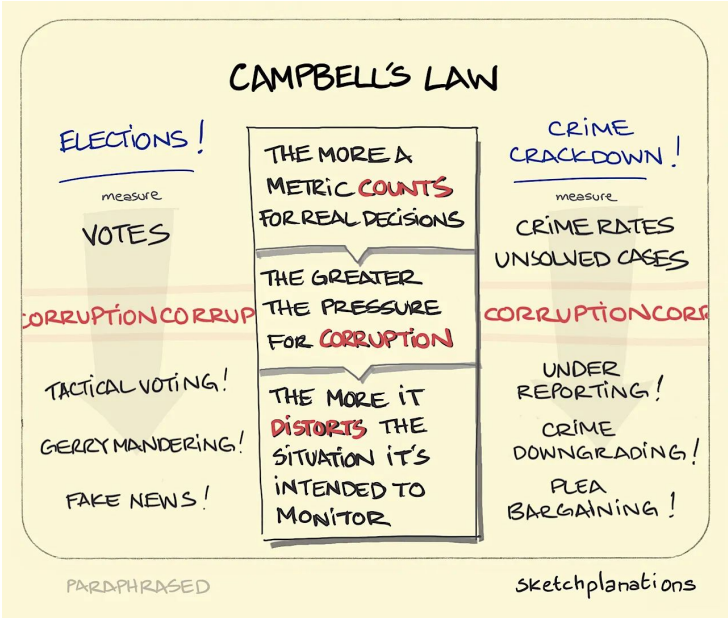


"Strategic" Prediction

Data **corresponds to individuals who have agency** and want to affect the decisions made on them by the ML algorithms.

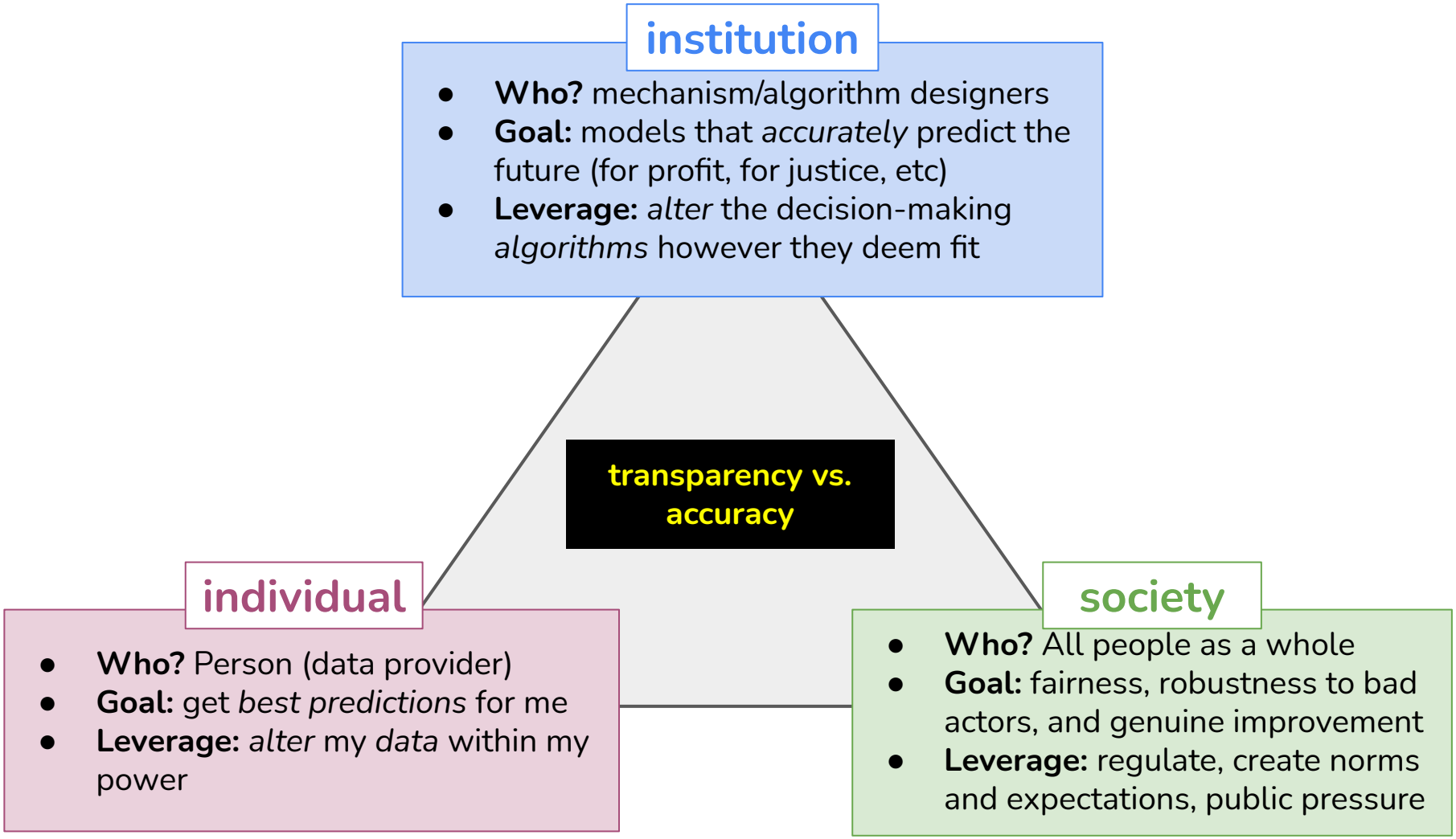
Similar problem, different fields





- School's admission rule: admit anyone who has more than 100 books in their house.
- Students with (say) 90 and more books can "easily" buy (**but need not read!**) 10 more and get admitted.

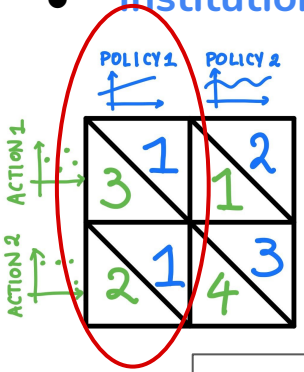
→ defeats the purpose of having the # books as a measure of qualifications


institution
Who? mechanism/algorithm designers
Goal: models that accurately predict the future (for profit, for justice, etc)
Leverage: alter the decision-making algorithms however they deem fit
transparency vs. accuracy
individual
Who? Person (data provider)
Goal: get best predictions for me
Leverage: alter my data within my power
society
Who? All people as a whole
Goal: fairness, robustness to bad actors, and genuine improvement
Leverage: regulate, create norms and expectations, public pressure

Modeling robustness in CS

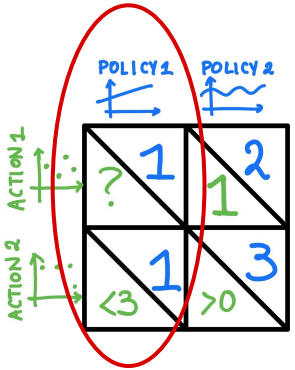
UTILITY = VALUE − COST
(value getting for obtaining outcome A) (effort / time / money spent to achieve it)

- **Individual** measures net gains in terms of a utility function.
- **Institution** has prior information about individual's utility function.



Utility function **fully** known

Utility function **only partially** known



Equilibrium Rule

This decision rule is the most accurate one, if individuals respond strategically according to a specified utility.

[Hardt, Megiddo, Papadimitriou, Wooters, ITCS16]

No-Regret Rule

In hindsight, this decision rule achieves similar results to the best possible policy.

[Dong, Roth, Schutzman, Waggoner, Wu, EC18], [Chen, Liu, Podimata, NeurIPS20], [Ahmadi, Beyhaghi, Blum, Naggita, arXiv20]

Q & A for Part I

Strategic Prediction from the **Individual's** Perspective

Yo Shavit

Tutorial outline

Part I

- What is strategic decision-making and who are the key actors?
- The robustness perspective

Part II

- How do strategy-aware predictors affect individuals?

Strategic prediction from the individual perspective

What is the individual's objective?

To get the best possible decision ("**recourse**")

Not to be forced to do bad things/avoid good things

To minimize the resources I spend achieving recourse

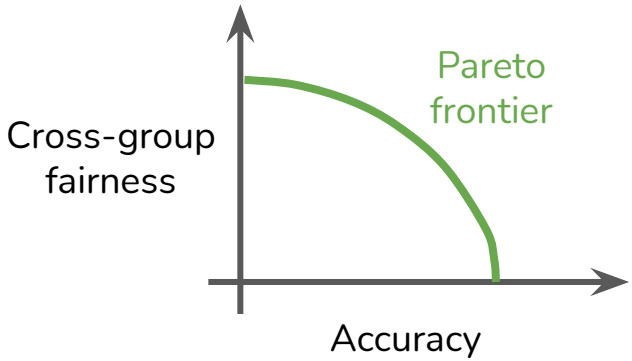
To have access to recourse regardless of my demographic group

Barbara Underwood, “Law and the Crystal Ball”, 1979:

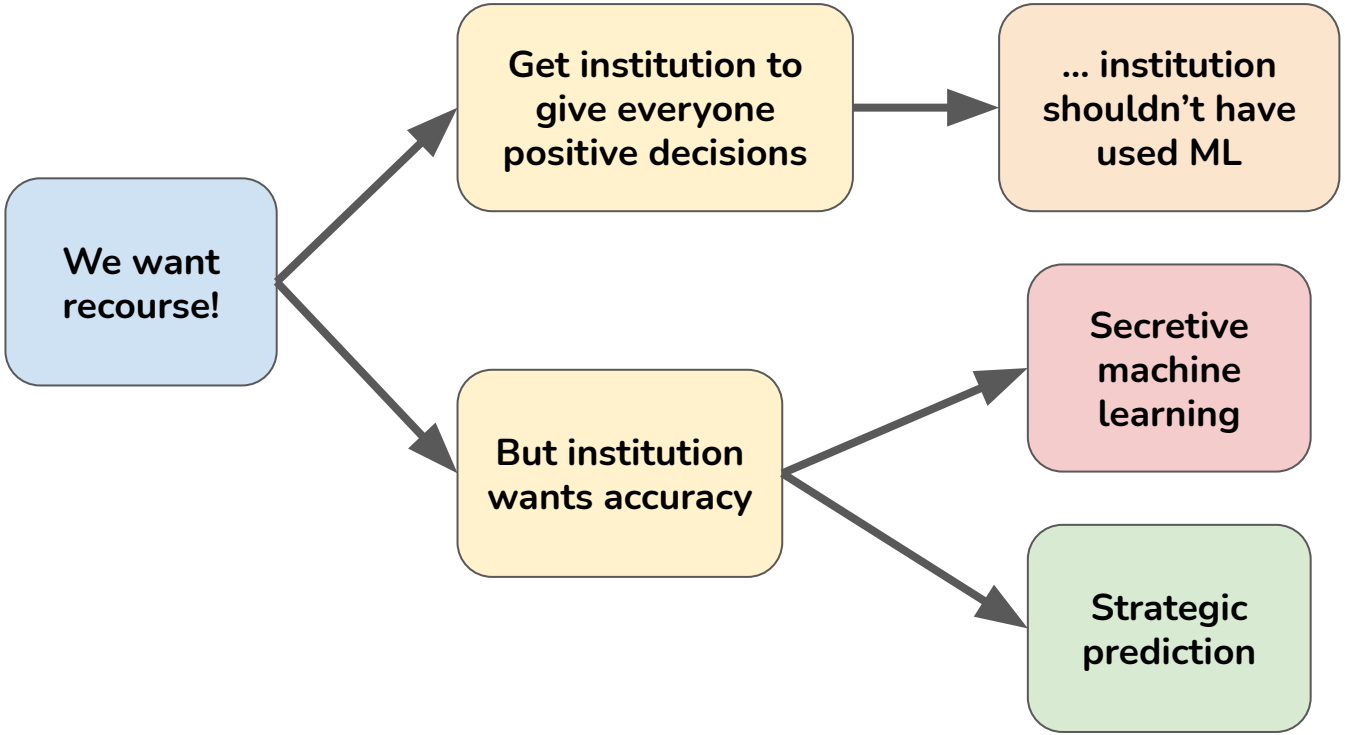
*“The use of **controllable factors** provides an **opportunity** for the individual who really wants to be selected to **choose conduct** that **improves** his predictive score and hence his **prospects for selection**.”*

What's new in today's research?

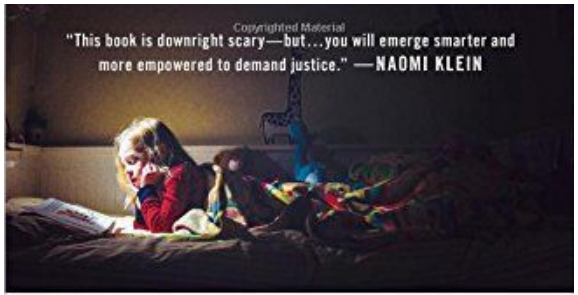
- Too many predictors to scrutinize individually
- Black-box models
- Formalizing objectives in high-dimensions
- Revealing Pareto frontier

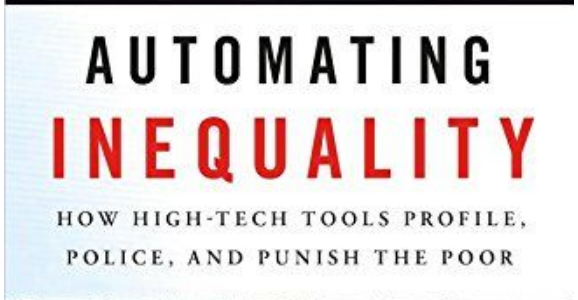


Why recourse may need strategic prediction



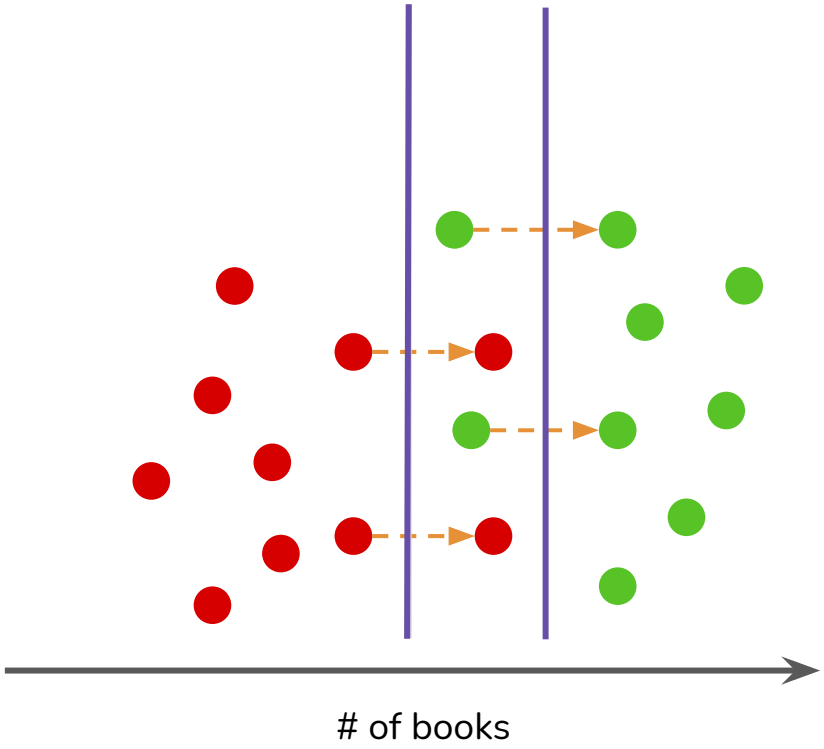
Incentivization harms





SM SNAP Supplemental Nutrition Assistance Program

Strategic predictors can reduce individual utility

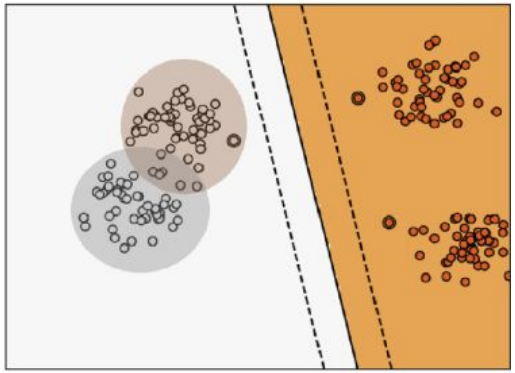


Minimize
the sum of **costs** paid by
truly positive individuals
to get the **right decision**

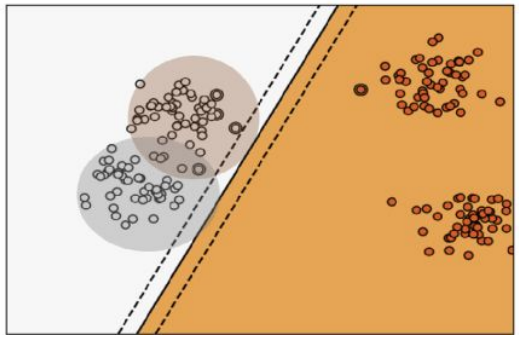
"The Social Cost of Strategic Classification",
Milli, Miller, Dragan, & Hardt, FAccT 2018

Recourse unfairness: unequal access to recourse

Unequal recourse



Equal recourse



Minimize

|| mean recourse on group A

-

mean recourse on group B ||

From “Equalizing Recourse Across Groups”,
Gupta, Nokhiz, Roy, & Venkatasubramanian, arXiv 2019

Many more challenges

- Can subsidies help with cross-group recourse?
  - Hu, Immorlica, & Wortman-Vaughan, FAccT 2018
- Choosing counterfactual explanations under strategic predictors
  - Tsirtsis & Gomez-Rodriguez, NeurIPS 2020
- What if decisions are randomized?
  - Braverman & Garg, FORC 2020
- What if decision rule isn't fully transparent?
  - Bechavod, **Podimata**, Wu, & Ziani, arXiv 2021
- Much more!

General trend?

Transparency, **Accuracy**, **Individual Utility**

Pick two… ?

Q & A for Part II

The Causal Perspective
Beyond Robustness

Ben Edelman

Tutorial outline

Part I

- What is strategic decision-making and who are the key actors?
- The robustness perspective

Part II

- How do strategy-aware predictors affect individuals?

Part III

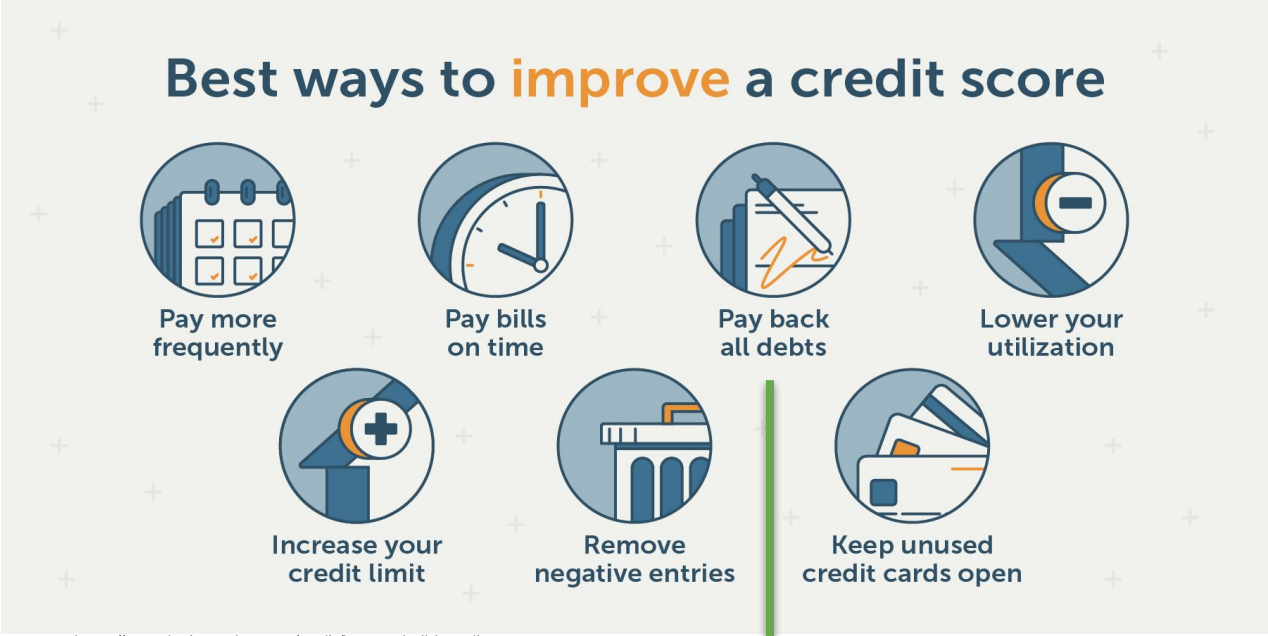
- Beyond robustness: the causal perspective
- A zoo of goals
- performative prediction

A caricature of an institution

If my model isn't completely **secret** and **opaque**, then individuals will **game** it by manipulating their features. This makes my job harder.

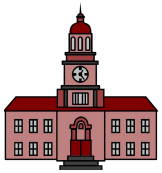
institution

Is it all just “gaming”?



source: https://www.lexingtonlaw.com/credit/how-to-build-credit

ability to pay back future loans



institution

Model

> Prediction

Features

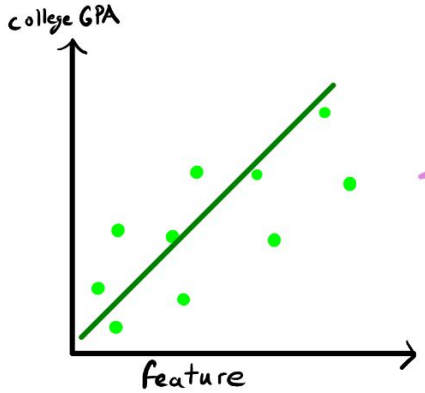
True label

GPA

CollegeBoard Advanced Placement Program

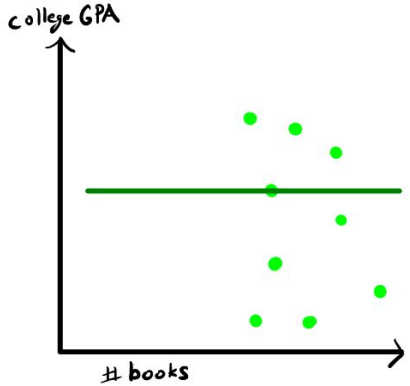
individual

Causation or just correlation?

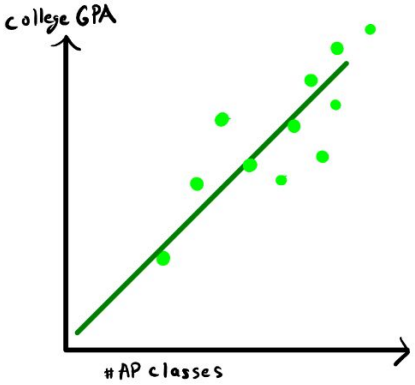


before strategic response

?



gaming



improvement: no cost to transparency

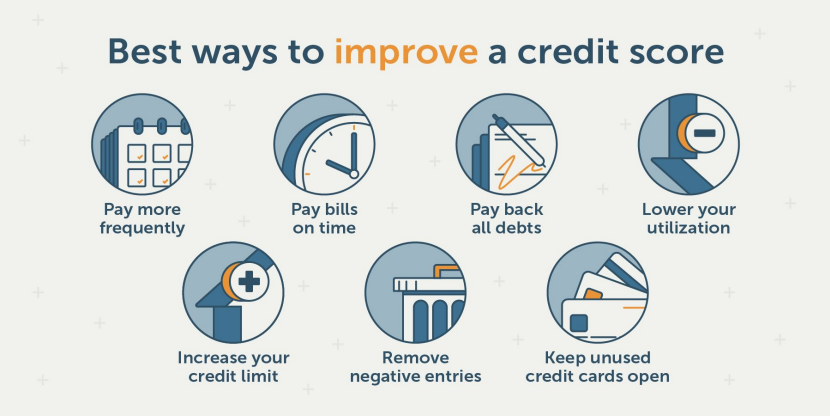
after strategic response

Gaming vs. improvement

Strategic Classification is Causal Modeling in Disguise

John Miller Smitha Milli Moritz Hardt

February 19, 2020



Improvement

A *benefit* of strategic prediction, from the **institution**'s perspective

The institution's mechanism design challenge:

How to encourage improvement without encouraging gaming?

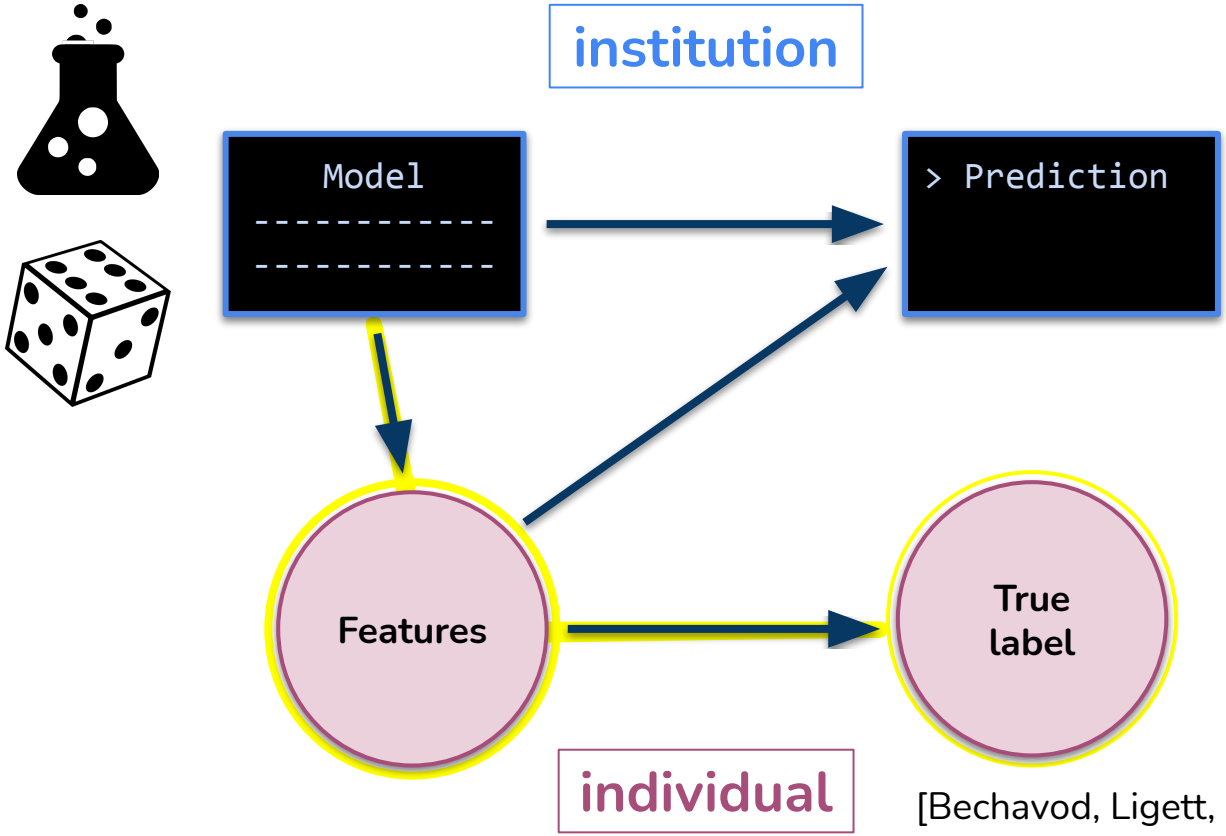
[Kleinberg & Raghavan, EC 2019]

[Miller, Milli, Hardt, ICML 2020]

[Haghtalab, Immorlica, Lucier, Wang, IJCAI 2020]

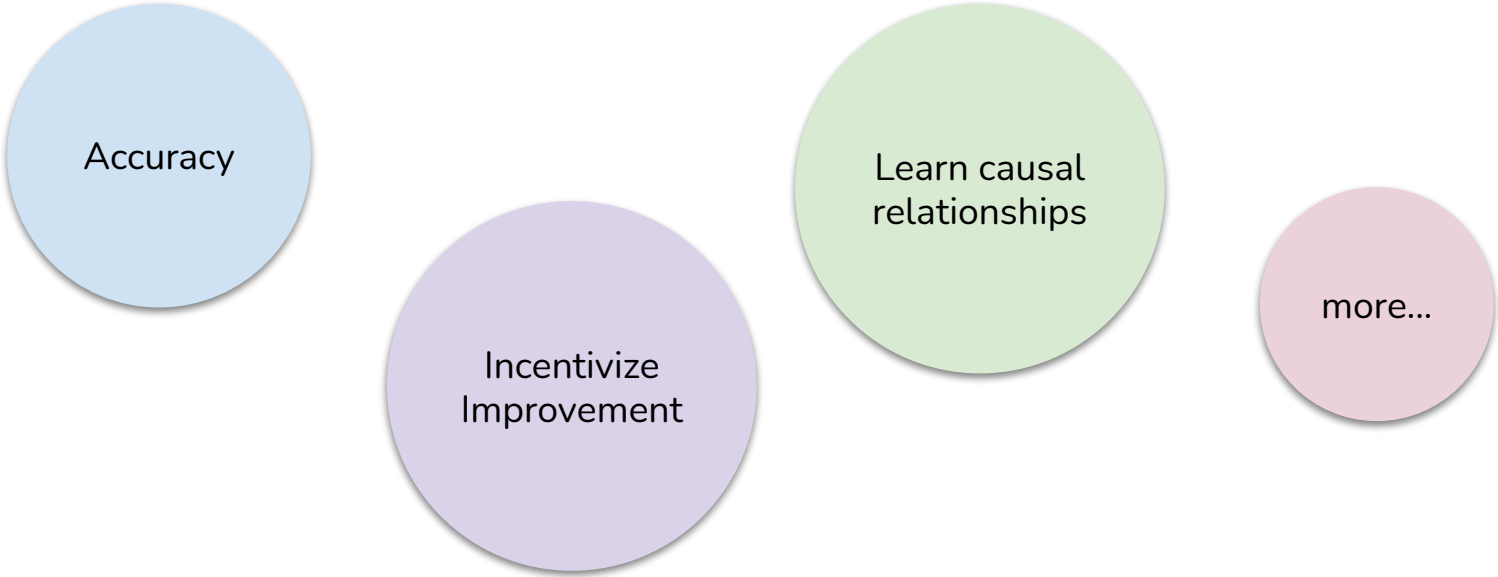
[Shavit, Edelman, Axelrod ICML 2020]

Connections to information economics:
principal-agent models



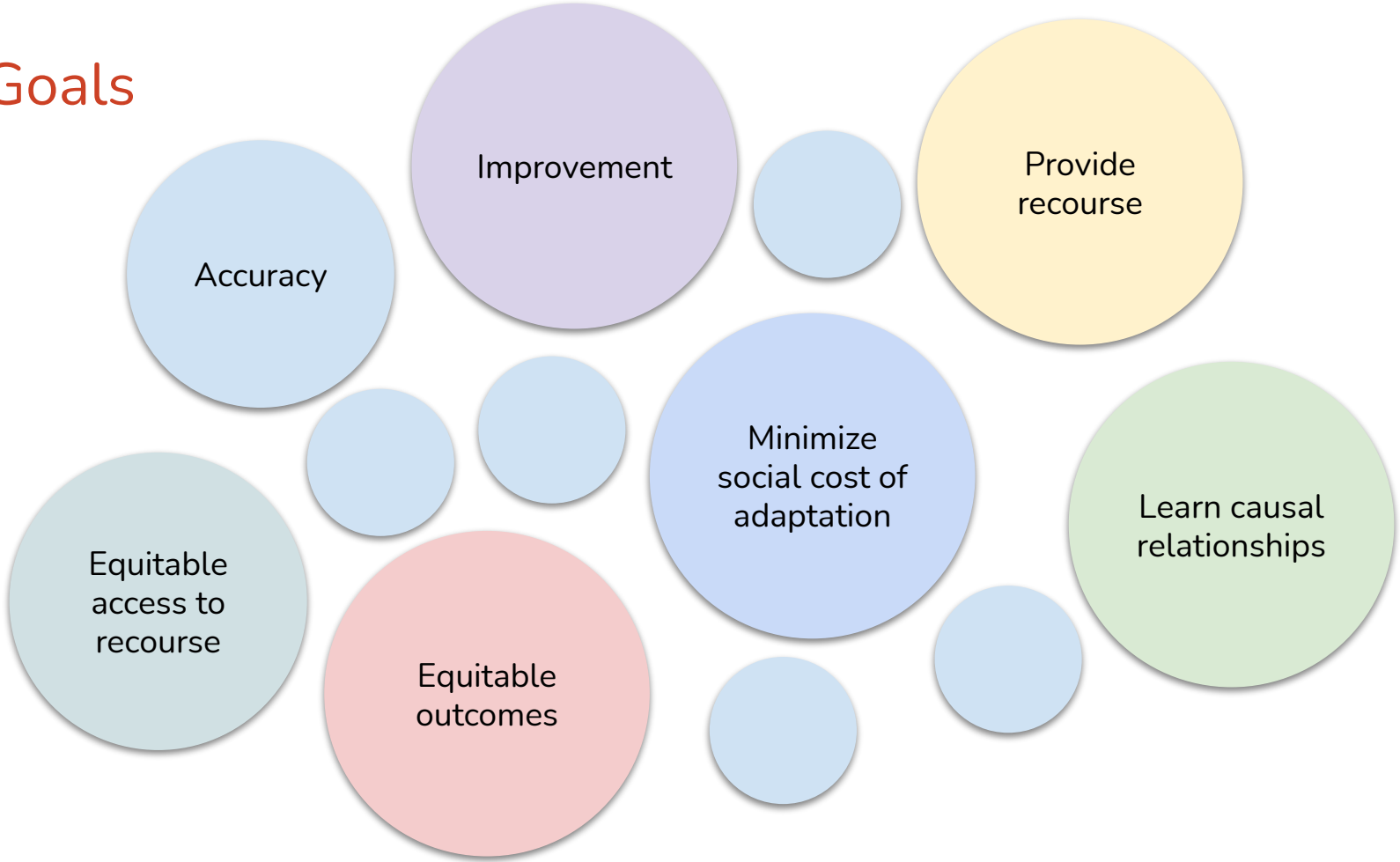
[Bechavod, Ligett, Wu, Ziani, AISTATS 2020]
[Shavit, Edelman, Axelrod ICML 2020]

Institution's goals in *causal* strategic prediction

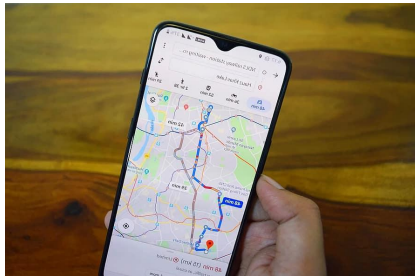


[Shavit, Edelman, Axelrod ICML 2020]

Goals


Improvement
Provide recourse
Accuracy
Minimize social cost of adaptation
Learn causal relationships
Equitable access to recourse
Equitable outcomes

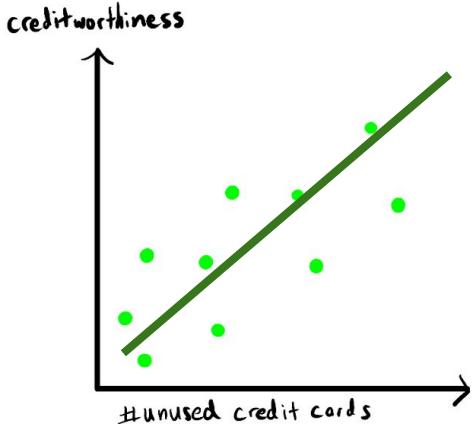
Performativity beyond strategic feature manipulation

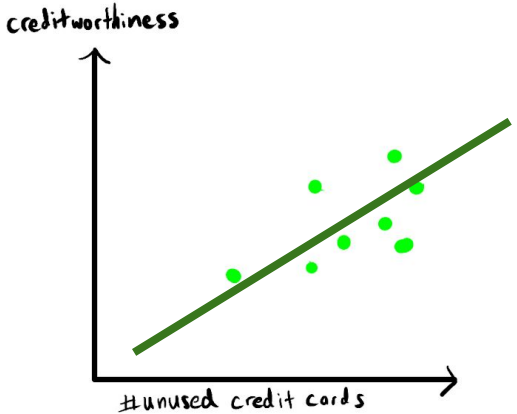


Performative Prediction
When predictions influence the data
(not just strategic prediction)

[Perdomo, Zrnic, Mendler-Dünner, Hardt, ICML 2020]

The simplest accuracy algorithm?





...

Repeated Retraining

What algorithms are better than retraining?

Takeaways

The interplay between **transparency** and **accuracy** is nuanced.

Accuracy and transparency can often be achieved simultaneously

Strategic prediction is a rich subject of study: many settings, many objectives

....Let's broaden this beyond computer science!

Lots of recent, exciting works

Thank you!

- **Robustness**: [Hardt, Megiddo, Papadimitriou, Wooters, ITCS16], [Dong, Roth, Schutzman, Waggoner, Wu, EC18], [Chen, Liu, Podimata, NeurIPS20], [Ahmadi, Beyhaghi, Blum, Naggita, arXiv20], [Sundaraman, Vullikanti, Xu, Yao, arXiv21], [Ghalme , Nair, Eilat, Talgam-Cohen, Rosenfeld, arXiv21]
- **Fairness:** [Milli, Miller, Dragan, Hardt, FAT*18], [Hu, Immorlica, Vaughan, FAT*18], [Liu, Wilson, Haghtalab, Kalai, Borgs, Chayes, FAT*19], [Braverman, Garg, FORC20]
- **Recourse/Incentivizing Effort:** [Ustun, Spangher, Liu, FAT*19], [Kleinberg and Raghavan, EC19], [Khajehnejad, Tabibian, Scholkopf, Singla, Gomez-Rodriguez, arXiv19], [Gupta, Nokhiz, Roy, Venkatasubramanian, arXiv19], [Chen, Wang, Liu, arXiv20], [Tsirtsis, Gomez-Rodriguez, NeurIPS20], [Haghtalab, Immorlica, Lucier, Wang, IJCAI20], [Bechavod, Podimata, Wu, Ziani, arXiv21]
- **Causality:** [Miller, Milli, Hardt, FAT*19], [Shavit, Edelman, Axelrod, ICML20], [Bechavod, Ligett, Wu, Ziani, AISTATS21]
- **Performative Prediction:** [Perdomo, Zrnic, Mendler-Dunner, Hardt ICML20], [Mendler-Dunner, Perdomo, Zrnic, Hardt NeurIPS20], [Miller, Perdomo, Zrnic arXiv21]

Thank you!

benjaminedelman.com

charapodimata.com

yonadavshavit.com