SPIA 586a: Machine Learning with Social Data Week 13: Machine Learning in Policy and Practice

Brandon Stewart[1]

Princeton

April 21, 2025

[1]Thank you to Rebecca Johnson and Simone Zhang for sharing their slides!

What is the Bureaucratic Counterfactual? Categorical versus Algorithmic Prioritization in U.S. Social Policy

Rebecca A. Johnson*
McCourt Policy School, Georgetown University
Washington, DC, USA
rj545@georgetown.edu

Simone Zhang*
Department of Sociology, University of Notre Dame
Notre Dame, Indiana, USA
szhang27@nd.edu

ABSTRACT

There is growing concern about governments' use of algorithms to make high-stakes decisions. While an early wave of research focused on algorithms that predict risk to allocate punishment and suspicion, a newer wave of research studies algorithms that predict "need" or "benefit" to target beneficial resources, such as ranking those experiencing homelessness by their need for housing. The present paper argues that existing research on the role of algorithms in social policy could benefit from a counterfactual perspective that asks: given that a social service bureaucracy needs to make some decision about whom to help, what status quo prioritization method would algorithms replace? While a large body of research contrasts human versus algorithmic decision-making, social service bureaucracies target help not by giving street-level bureaucrats full discretion. Instead, they primarily target help through pre-algorithmic,

KEYWORDS

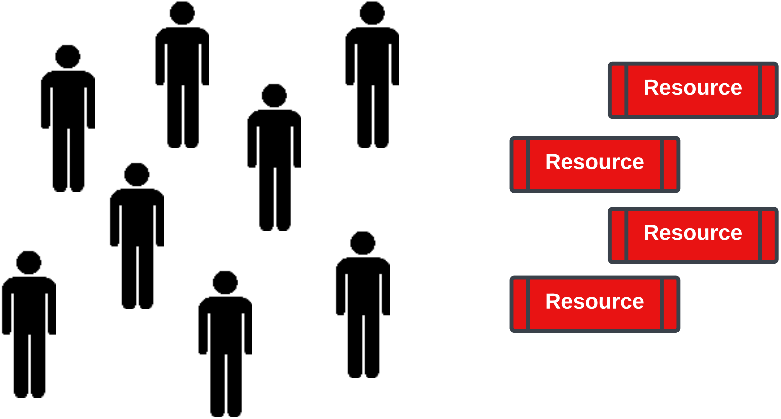
fairness and transparency, social policy, resource allocation

ACM Reference Format:
Rebecca A. Johnson and Simone Zhang. 2022. What is the Bureaucratic Counterfactual? Categorical versus Algorithmic Prioritization in U.S. Social Policy. In *2022 ACM Conference on Fairness, Accountability, and Transparency (FAccT '22), June 21–24, 2022, Seoul, Republic of Korea.* ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3531146.3533223

1 INTRODUCTION

As governments turn to algorithms to help them decide whom to help and whom to punish, recent work has investigated implications for inequality [9, 22, 25, 52, 53]. The majority of research on algorithmic inequality has focused on algorithms of two types, following the framework in Brayne [9]. First is work on inequality resulting

Tradeoffs in Policymaking Due to Scarcity



Main Idea: The Bureaucratic Counterfactual

Counterfactual is usually not:

- Case-by-case review by humans with significant discretion

Main Idea: The Bureaucratic Counterfactual

Counterfactual is usually not:

- Case-by-case review by humans with significant discretion

Counterfactual often is: **categorical prioritization**

- Manually devised rules built on placing people into categories

Main Idea: The Bureaucratic Counterfactual

Counterfactual is usually not:

- Case-by-case review by humans with significant discretion

Counterfactual often is: **categorical prioritization**

- Manually devised rules built on placing people into categories

Challenges:

1) scarcity: more people need help than political will/funding
2) high-dimensionality: people are different in many multi-dimensional ways

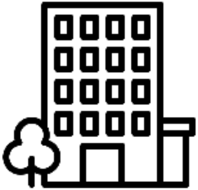
Three Steps in Categorical Prioritization

Step	Examples
1. Manually select attributes	Mother; veteran; low-income; looking for work
2. Coarsen continuous attributes into categories	Income $\leq$ threshold
3. Deliberate on decision rules for mapping category combinations to priority	Adult and looking for work; veteran or mother; mother without felony conviction

Categorical Prioritization is Common

Examples:
Medicaid; Medicare

Examples:
Categories in finance formulas; lottery plus factors



Examples:
Categories for voucher waitlists

Algorithmic prioritization:

- Model a specific outcome (e.g. eviction, grade repetition)
- Prioritize according to greatest predicted risk of adverse event or expected benefit

NB: This gets called 'predictive optimization' in Wang and Kapoor et al 2023

Contrasting Prioritization Approaches

Categorical	**Algorithmic**
No target outcome	Target outcome

Contrasting Prioritization Approaches

Categorical	**Algorithmic**
No target outcome	Target outcome
Manually pick a few attributes	High-dimensional input
Coarsen continuous attributes	Mix of attribute types

Contrasting Prioritization Approaches

Categorical	Algorithmic
No target outcome	Target outcome
Manually pick a few attributes	High-dimensional input
Coarsen continuous attributes	Mix of attribute types
Manually chosen rules	Learned rules

How might shifting to algorithmic prioritization impact inequality in access to help in an unequal world?

Case Study: Housing Voucher Waitlists

Housing Choice Voucher Program

- Funding only covers $\approx$ 1 in 4 eligible renter households (Schwartz 2014)
- Housing authorities — bureaucracies that distribute the vouchers — can advance some groups to the front of long waitlists

Case Study: Housing Voucher Waitlists

Housing Choice Voucher Program

- Funding only covers ≈ 1 in 4 eligible renter households (Schwartz 2014)
- Housing authorities — bureaucracies that distribute the vouchers — can advance some groups to the front of long waitlists

Mixed methods study: 22 interviews with housing authority officials + review of 1,398 local policies

Sample Local Prioritization Policy

10.1 PREFERENCES

The Jeffersonville Housing Authority will select families based on the following preferences based on our local housing needs and priorities:

A. Substandard Housing as determined by local housing code / Involuntarily Displaced by Government Action, Declared Disaster at the local level or sale/loss of property by landlord, or Victims of Domestic Violence (20 points)

B. Elderly/Handicapped/Disabled OR working a minimum of 20 hours weekly (20 points)

C. Applicants who are working and/or living within the City Limits of Jeffersonville/Clarksville (30 points)

E. Veterans (honorably discharged) (5 points)

A family may qualify for more than one (1) preference. The family with the most cumulative preferences will be offered housing first based upon availability.

The date and time of application will be noted and utilized to determine the sequence within the above-prescribed preferences.

Sample Local Prioritization Policy

Most points: residents

10.1 PREFERENCES

The Jeffersonville Housing Authority will select families based on the following preferences based on our local housing needs and priorities:

A. Substandard Housing as determined by local housing code / Involuntarily Displaced by Government Action, Declared Disaster at the local level or sale/loss of property by landlord, or Victims of Domestic Violence (20 points)

B. Elderly/Handicapped/Disabled OR working a minimum of 20 hours weekly (20 points)

C. Applicants who are working and/or living within the City Limits of Jeffersonville/Clarksville (30 points)

E. Veterans (honorably discharged) (5 points)

A family may qualify for more than one (1) preference. The family with the most cumulative preferences will be offered housing first based upon availability.

The date and time of application will be noted and utilized to determine the sequence within the above-prescribed preferences.

Sample Local Prioritization Policy

Fewer points: living in substandard housing *OR* involuntariy displaced *OR* victim of domestic violence *OR* elderly *OR* person w/ disability *OR* *working*

10.1 PREFERENCES

The Jeffersonville Housing Authority will select families based on the following preferences based on our local housing needs and priorities:

A. Substandard Housing as determined by local housing code / Involuntarily Displaced by Government Action, Declared Disaster at the local level or sale/loss of property by landlord, or Victims of Domestic Violence (20 points)

B. Elderly/Handicapped/Disabled OR working a minimum of 20 hours weekly (20 points)

C. Applicants who are working and/or living within the City Limits of Jeffersonville/Clarksville (30 points)

E. Veterans (honorably discharged) (5 points)

A family may qualify for more than one (1) preference. The family with the most cumulative preferences will be offered housing first based upon availability.

The date and time of application will be noted and utilized to determine the sequence within the above-prescribed preferences.

Sample Local Prioritization Policy

10.1 PREFERENCES

The Jeffersonville Housing Authority will select families based on the following preferences based on our local housing needs and priorities:

A. Substandard Housing as determined by local housing code / Involuntarily Displaced by Government Action, Declared Disaster at the local level or sale/loss of property by landlord, or Victims of Domestic Violence (20 points)

B. Elderly/Handicapped/Disabled OR working a minimum of 20 hours weekly (20 points)

C. Applicants who are working and/or living within the City Limits of Jeffersonville/Clarksville (30 points)

Fewest points: veterans

E. Veterans (honorably discharged) (5 points)

A family may qualify for more than one (1) preference. The family with the most cumulative preferences will be offered housing first based upon availability.

The date and time of application will be noted and utilized to determine the sequence within the above-prescribed preferences.

Sample Local Prioritization Policy

10.1 PREFERENCES

The Jeffersonville Housing Authority will select families based on the following preferences based on our local housing needs and priorities:

A. Substandard Housing as determined by local housing code / Involuntarily Displaced by Government Action, Declared Disaster at the local level or sale/loss of property by landlord, or Victims of Domestic Violence (20 points)

B. Elderly/Handicapped/Disabled OR working a minimum of 20 hours weekly (20 points)

C. Applicants who are working and/or living within the City Limits of Jeffersonville/Clarksville (30 points)

E. Veterans (honorably discharged) (5 points)

A family may qualify for more than one (1) preference. The family with the most cumulative preferences will be offered housing first based upon availability.

The date and time of application will be noted and utilized to determine the sequence within the above-prescribed preferences.

Points are summed, more points -> earlier access

Sample Local Prioritization Policy

10.1 PREFERENCES

The Jeffersonville Housing Authority will select families based on the following preferences based on our local housing needs and priorities:

A. Substandard Housing as determined by local housing code / Involuntarily Displaced by Government Action, Declared Disaster at the local level or sale/loss of property by landlord, or Victims of Domestic Violence (20 points)

B. Elderly/Handicapped/Disabled OR working a minimum of 20 hours weekly (20 points)

C. Applicants who are working and/or living within the City Limits of Jeffersonville/Clarksville (30 points)

E. Veterans (honorably discharged) (5 points)

A family may qualify for more than one (1) preference. The family with the most cumulative preferences will be offered housing first based upon availability.

No points: application order

The date and time of application will be noted and utilized to determine the sequence within the above-prescribed preferences.

Three Morally-Relevant Differences

Q: How might shifting to algorithmic prioritization impact inequality in access to help in an unequal world?

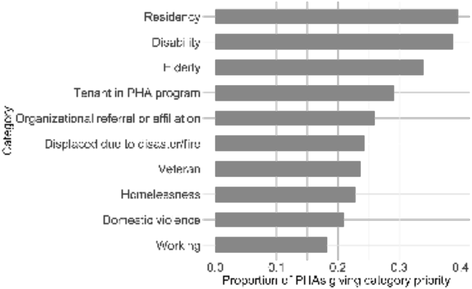
1. Is the basis for prioritization formalized?
2. What role does power play in prioritization?
3. Are decision rules manually chosen or derived from a model?

Difference One: Formalization

Categorical prioritization benefit

Supports value pluralism

- Allows expression and balancing of multiple values



- Affirm community ties, earlier commitments
- Address dire housing need, hardship
- Promote self-sufficiency, help those deemed struggling through "no fault of their own"

Difference One: Formalization

Categorical prioritization drawback

In practice, policy goals often remain murky defaults

- Choices lack a traceable basis

Officials interviewed:

- Some struggled to offer rationale: "just what we have", "before my time"
- Some offered minimal account: "obvious", "straightforward"
- Some described policy as aligned with objective or "common sense" assessments of applicants' deservingness

Difference one: formalization

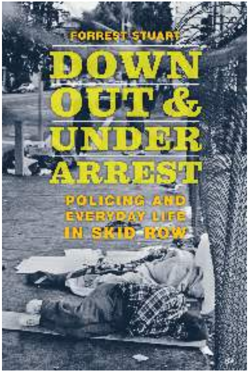
What might change if shifted to algorithmic prioritization?

Difference one: formalization

What might change if shifted to algorithmic prioritization?

Algorithmic prioritization drawbacks

- Could disrupt balancing of values (Green and Chen 2021)
- Reliance on measured outcomes produced by exclusionary, biased, or harmful social structures and processes (Barocas and Selbst 2016, Jacobs and Wallach 2021)



Difference one: formalization

What might change if shifted to algorithmic prioritization?

Algorithmic prioritization benefits

- Enforce deliberation about the goals of assistance
- Clarify basis for choices

(Abebe et al. 2020)

Should assistance help people *get into* housing? Or *prevent displacement* from housing?

Should it target those at greatest risk of an adverse outcome? Or those expected to benefit most?

Difference two: interactions with power inequalities

Categorical prioritization benefit

More accessible site for community engagement

Difference two: interactions with power inequalities

Categorical prioritization benefit

More accessible site for community engagement

Categorical prioritization drawback

In practice, can reflect relative elites in the room

- Unequal civic participation, advocacy

Difference two: interactions with power inequalities

Categorical prioritization benefit

More accessible site for community engagement

Categorical prioritization drawback

In practice, can reflect relative elites in the room

- Unequal civic participation, advocacy

Little to no community member input

Choices more often shaped by

- Leaders of community organizations
- Idiosyncratic commitments and beliefs of PHA staff and board members in the room

"...the executive director actually happens to be a veteran himself. My father, [also] a veteran. So creating a special preference for veterans, that took no thought at all."

Difference two: interactions with power inequalities

What might change if shifted to algorithmic prioritization?

Algorithmic prioritization benefits

- Stigmatized and hidden groups could be prioritized even if they lack political representation

Algorithmic prioritization drawbacks

- Shift power toward the already powerful (Kalluri 2020; Kasy and Abebe 2021; Burrell and Fourcade 2021)

Difference three: manual vs inductive decision rules

Categorical prioritization benefit

Rules that "make sense"

- Expressive function
- Realistic practically

Difference three: manual vs inductive decision rules

Categorical prioritization benefit

Rules that "make sense"

- Expressive function
- Realistic practically

Categorical prioritization drawback

Arbitrary or biased choices

- Gut instincts, folk theories of poverty

"To me, those working preferences are like incentives... If you explain to them the reason why their number is fluctuating because of the preference... they sometimes go and get jobs, just so they can move them up. So, it's motivation.

Difference three: manual vs inductive decision rules

What might change if shifted to algorithmic prioritization?

Algorithmic prioritization benefits

- Identify interactions, hidden attributes that associated in complex, nonlinear ways with need/marginal benefit

Algorithmic prioritization drawbacks

- May produce procedures that are more complex, difficult to interpret or implement

- Social service bureaucracies frequently face question: how do we decide who needs or deserves help most?[2]
- In US social policies, counterfactual is often not human with full discretion; instead, categorical prioritization
- Social policies have long history of simplifying need and being informed by constructions of “deservingness” (Steensland 2006; Katz 2013; Watkins-Hayes and Kovalsky 2017)

In the paper, two case studies of categorical prioritization:

1. voucher waitlists
2. pupil weights for school finance formulas

[2] More research on: *risk:* Chouldechova et al. 2018; Eckhouse et al. 2019; Goel et al. 2021; Green and Chen 2019; *merit:* Binns et al. 2018; Dodge et al. 2019; Langer et al. 2019; O'Brien and Kiviat 2018; Saxena et al. 2019

Johnson and Zhang (2022) highlights a variety of benefits and drawbacks of categorical vs algorithmic prioritization.

Further research needed to evaluate distributional and procedural consequences grounded in concrete social policy contexts.

1. Bureaucratic Counterfactual and Categorical Prioritization

2. Against Predictive Optimization

1. Bureaucratic Counterfactual and Categorical Prioritization

2. Against Predictive Optimization

Against Predictive Optimization: On the Legitimacy of Decision-making Algorithms That Optimize Predictive Accuracy

ANGELINA WANG and SAYASH KAPOOR, Princeton University
SOLON BAROCAS, Microsoft Research; Cornell University
ARVIND NARAYANAN, Princeton University

We formalize predictive optimization, a category of **decision-making algorithms** that **use machine learning (ML)** to **predict future outcomes** of interest about **individuals**. For example, pre-trial risk prediction algorithms such as COMPAS use ML to predict whether an individual will re-offend in the future. Our thesis is that predictive optimization raises a distinctive and serious set of normative concerns that cause it to fail on its own terms. To test this, we review 387 reports, articles, and web pages from academia, industry, non-profits, governments, and data science contests, and we find many real-world examples of predictive optimization. We select eight particularly consequential examples as case studies. Simultaneously, we develop a set of normative and technical critiques that challenge the claims made by the developers of these applications—in particular, claims of increased accuracy, efficiency, and fairness. Our key finding is that these critiques apply to each of the applications, are not easily evaded by redesigning the systems, and thus challenge whether these applications should be deployed. We argue that the burden of evidence for justifying why the deployment of predictive optimization is not harmful should rest with the developers of the tools. Based on our analysis, we provide a rubric of critical questions that can be used to deliberate or contest specific predictive optimization applications.[1]

CCS Concepts: • **Social and professional topics → Computing / technology policy**; **Government technology policy**; • **Computing methodologies → Machine learning;**

Additional Key Words and Phrases: Automated decision-making, validity, machine learning, optimization

ACM Reference format:
Angelina Wang, Sayash Kapoor, Solon Barocas, and Arvind Narayanan. 2024. Against Predictive Optimization: On the Legitimacy of Decision-making Algorithms That Optimize Predictive Accuracy. *ACM J. Responsib. Comput.* 1, 1, Article 9 (March 2024), 45 pages.
https://doi.org/10.1145/3636509

Velogica claims to use "human underwriting on less than 5% of applications"

Upstart claims that over 75% of loan decisions are fully automated

Hirevue claims it is "Fast. Fair. Flexible."

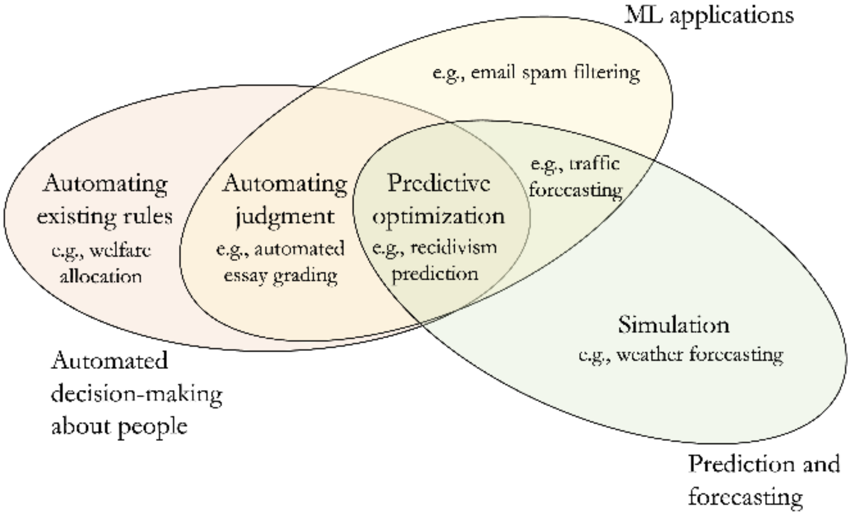

ML applications
e.g., email spam filtering
Automating existing rules
e.g., welfare allocation
Automating judgment
e.g., automated essay grading
Predictive optimization
e.g., recidivism prediction
e.g., traffic forecasting
Simulation
e.g., weather forecasting
Automated decision-making about people
Prediction and forecasting

Table 1. Positive and Negative Examples to Illustrate the Definition of Predictive Optimization

Application	Example	Uses ML?	Decisions about individuals?	Predicts future outcome?	Predictive optimization?	Reason
Predictive policing. Decides geographical areas where police should be deployed.	PredPol	✓	✗	✓	✗	Decisions are not made about individuals.
Welfare allocation. Automates hand-coded rules for deciding whether an applicant is eligible for a public service, such as Medicaid.	Indiana welfare eligibility	✗	✓	✗	✗	Decisions are not made using ML. Decisions do not predict the future.
Automated essay grading. Uses past decisions by human graders to learn decision rules for grading.	TOEFL	✓	✓	✗	✗	Decisions use past judgments instead of predicting future outcomes.
Traffic prediction. Uses information about current traffic to predict estimated time of arrival.	Google Maps	✓	✗	✓	✗	Predictions are about route timings rather than an individual's outcomes.
Pre-trial risk prediction. Uses past data about individuals to predict future arrests or failure to appear in court.	COMPAS	✓	✓	✓	✓	Satisfies all three criteria for predictive optimization.

An example is considered predictive optimization only if it satisfies all three criteria: It uses ML, it makes decisions about individuals, and the target variable for the ML system is a future outcome of interest.

Application	Developer	What is being predicted (construct)	Proxy for the prediction (target)	Decision made based on prediction
COMPAS	Northpointe/ Equivant	Pretrial risk	Re-arrest in two years or failure to appear in court	Whether to release a defendant pre-trial or what bail amount to set
AFST	Allegheny County	Child maltreatment	Placement into foster care or multiple referrals within two years	Whether to investigate a family for child maltreatment
Hirevue	Hirevue	Job performance	Performance based on an industry-dependent performance indicator (varies by firm)	Whether to hire someone or invite them to the next round of interviews
Navigate	EAB	School dropout	Varies by school. E.g., "enrollment until next fall", "graduation within four years", or "graduation at any point of time"	Whether to offer targeted interventions to aid students
Upstart	Upstart	Creditworthiness	Repayment or future salary	Whether to offer a loan to someone and at what rates
Facebook suicide prediction	Facebook (Meta)	Suicide risk	Whether someone was assessed to be at high risk of suicide	Whether to refer someone for a welfare check
ImpactPro	Optum	Medical risk	Healthcare costs	Whether to put a patient in the high-risk health program
Velogica	SCOR	Life insurance risk	Mortality or policy lapse	Whether to offer a life insurance policy and at what

Modeling step	Activity	Limitation	Description	Difference with automating judgment
Algorithm design	Recast decision problem as prediction problem	**Prediction vs. intervention**	Optimal predictions may not result in optimal interventions	Not formulated as prediction problem
	Operationalize construct of interest by selecting an observable proxy as the target (e.g., GPA as proxy for scholastic success)	**Target-construct mismatch**	No proxy can perfectly encapsulate construct	No target variable needed
Data collection	Select training samples collected under previous policy (e.g., students admitted in previous years; no rejected students)	**Selection bias**	Training sample doesn't match target population	Training sample includes both accepted and rejected instances
Training	Build a model to predict target variable	**Limits to prediction**	The future isn't determined yet; achievable predictive accuracy is inherently limited	Does not rely on prediction
		Disparate performance	Model may perform worse for one group or have lower rate of positive classification	Bias is an issue, but the sources and interventions tend to be different
Deployment	Make decisions using the model	**Contestability**	May be difficult due to lack of explanation of decision	Fallback to human judgment
		Goodhart's law	Decision subjects may adapt in a way that defeats goals of system	Human decision makers have some ability to notice and respond to adversarial adaptation

Prediction	Case study	Intervention vs. prediction	Target-construct mismatch	Distribution shifts	Limits to prediction	Disparate performance	Lack of contestability	Goodhart's Law
Pre-trial risk	COMPAS [131]	●	●	●	●	●	●	◑
Child maltreatment	AFST [50]	●	●	●	◑	●	●	●
Job performance	HireVue [87]	◑	◑	●	◑	◑	●	●
School dropout	EAB Navigate [56, 63]	◑	◑	◑	◑	◑	●	●
Creditworthiness	Upstart [182]	◑	◑	◑	●	◑	◑	◑
Suicide	Facebook [44]	●	◑	◑	◑	◑	●	◑
Medical risk	Optum ImpactPro [136]	◑	●	●	◑	◑	◑	◑
Life insurance risk	Velogica [73]	◑	◑	◑	◑	◑	◑	◑

Table 2. A matrix with rows representing the eight consequential applications and columns representing the seven shortcomings of predictive optimization. ● represents concrete evidence that an application suffers from a limitation; ◑ represents partial or circumstantial evidence. Detailed explanations of each cell are included in Appendix B. Our key finding is that this matrix is dense.