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, rule-based methods. In this paper, we outline social policy’s current status quo method—categorical prioritization—where decision-makers manually (1) decide which attributes of help seekers should give those help seekers priority, (2) simplify any continuous measures of need into categories (e.g., household income falls below a threshold), and (3) manually choose the decision rules that map categories to priority levels. We draw on novel data and quantitative and qualitative social science methods to outline categorical prioritization in two case studies of United States social policy: waitlists for scarce housing vouchers and K-12 school finance formulas. We outline three main differences between categorical and algorithmic prioritization: is the basis for prioritization formalized; what role does power play in prioritization; and are decision rules for priority manually chosen or inductively derived from a predictive model. Concluding, we show how the counterfactual perspective underscores both the understudied costs of categorical prioritization in social policy and the understudied potential of predictive algorithms to narrow inequalities.

CCS CONCEPTS
• Applied computing → Sociology; Law; • Social and professional topics → Governmental regulations.

KEYWORDS
fairness and transparency, social policy, resource allocation

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

ACM Reference Format:

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 Braye [9]. First is work on inequality resulting from the use of algorithms of suspicion, tools like risk assessments in criminal justice or child welfare that classify individuals according to the risk they present to law and public safety [12, 20, 26, 27, 35]. Second is work on inequality emerging due to algorithms of selection or valuation, tools like credit scores, college admissions, or hiring assessments that classify individuals according to the value they present to organizations [8, 18, 23, 39, 40, 53, 57, 61]. Here, we focus on algorithms used to steer the allocation of scarce, beneficial resources to those deemed most deserving, or what Braye [9] labels algorithms for care. While other work investigates inequalities embedded within algorithms to classify need in contexts like homelessness [22], civil legal aid [33], and healthcare [49], we ask: what are the counterfactual, pre-algorithmic methods that bureaucracies use to allocate help and how do predictive algorithms compare to these existing bureaucratic counterfactuals?1

The remainder of the paper proceeds in five parts. In Section 2, we outline the challenge that policymakers face as they decide whom to help: faced with a high-dimensional set of markers that could indicate a person deserves help, how do policymakers decide on a yes or no “give help” decision? Social policy has long relied on one approach to dimensionality reduction, what we call categorical prioritization. Categorical prioritization involves three steps: deciding which attributes indicate someone deserves help, grouping people into categories, and deliberating on rules that map

∗Both authors contributed equally to this research.

1In the original framework, algorithms for care are also used to coordinate service delivery. The label of algorithms for “care” in Braye [9] also encompasses the distribution of goods beyond direct care-related services (e.g., medical care or home health aids). Instead, this group of algorithms includes ones that classify individuals “according to their need” and “potential as clients” (p. 1002).

As we discuss later, this perspective is similar to what Kasy and Abebe [35] call the “causal impact of algorithms” on outcomes. However, while those authors formalize that framework in a context-agnostic way where the decision-makers focus on operationalizing some idea of “merit,” we focus on concrete social policy contexts where the operating principle is some notion of “deservingness” or “need.”
different combinations of categories to decisions of whom to help. After outlining categorical prioritization, we discuss its prevalence as a dominant method for United States social policy prioritization. In Section 3 we ask: what changes when social policies shift from categorical prioritization to algorithmic prioritization, or when social policy uses models trained to predict a chosen policy-relevant target outcome to select features and infer decision rules? We identify three core differences: (1) is the basis of prioritization formalized into a label that operationalizes ideas about who deserves help or left unspecified; (2) what role does the relative power and political voice of different groups play in prioritization; and (3) does a deliberative body select the decision rules for combining categories into priority or does a predictive model inductively learn these rules? In Section 4, to illustrate the ethical tradeoffs between categorical prioritization and algorithmic prioritization more concretely, we draw upon our original empirical research to present two case studies of categorical prioritization in U.S. social policy: waitlists for scarce housing vouchers and pupil weights in K-12 school finance formulas. For each case, we outline what algorithmic prioritization might look like and its possible tradeoffs relative to the status quo of categorical prioritization. Concluding, we connect our framework to related work on the fairness of algorithms in high-stakes decisions by governments and point to a need for future research on the impact of categorical versus algorithmic prioritization on various forms of inequality.

2 COUNTERFACTUAL TO ALGORITHMS IN SOCIAL POLICY: CATEGORICAL PRIORITIZATION

A large body of research shows that the U.S. social welfare state, rather than providing people universal entitlements to social assistance, offers help through a patchwork of conditional programs that target specific subgroups [63]. This targeting often involves restricting access to those deemed “deserving,” conceptualized in a variety of ways. Policymakers sometimes conceptualize deservingness through assessments of who is in most dire need or who stands to benefit most from assistance. More often, deservingness in the U.S. is bound up with broadly-circulated ideas about the moral status of different groups [36, 58, 64, 71]. Notably, people popularly thought to face hardships through no fault of their own (e.g., the elderly, children, and people with disabilities) or who follow specific behavioral norms (e.g., marriage, labor force participation) are more commonly regarded as deserving of assistance in the U.S. context [47, 62].

Yet even with a general conception of what makes one “deserving,” policymakers face a dimensionality reduction problem when deciding whom to help. People are multidimensional: they have many attributes that could indicate deservingness. Categorical prioritization thus involves reducing these many dimensions into formally-defined eligibility criteria. This involves three steps that recur across many social policies.

First, policymakers manually select attributes that they think should qualify or disqualify someone from help. In many targeted programs, these often include attributes like household income, age, the number of dependent children in a household, and a person’s criminal history.

Second, for simplicity, policymakers discretize continuous attributes into categories. For example, many policies draw thresholds on continuous household income to categorize households as “in poverty” or not. Important is that category definitions result from institutionally-negotiated boundaries that can be exclusionary, mismatched with people’s experiences, or developed to serve other applications entirely. For instance, poverty thresholds for social programs initially emerged from (1) calculating the cost of a “standard food plan” that would be needed to feed a family and (2) adjusting these calculations to account for family size and age composition [24]. This example illustrates ethical challenges with categorical prioritization. Historical contingencies in category definitions can weaken the link between membership in a category and need for a resource. A category originally developed for one purpose—calculating poverty with respect to the cost of food—diffused to social policies in other areas like housing and education where other measures might capture more relevant forms of need.

The third step involves deliberating on decision rules to apply to the categories to determine who will ultimately receive assistance. These rules take three forms: (1) and logic, where a person needs to jointly be a member of two categories to receive help (e.g., non-elderly adult and looking for work); (2) or logic, where a person qualifies if they are a member of either category (e.g., a person who is either elderly or an adult who is looking for work); (3) unless logic, where a person qualifies unless they have an excluding characteristic (e.g., someone with a criminal record [16]). Importantly, while these decision rules may resemble the output of a predictive algorithm, deliberative bodies choose the rules rather than a model inductively selecting rules that optimize prediction of a target outcome.

The broad contours of categorical prioritization—many dimensions of deservingness reduced into a sparse set of categories; continuous attributes collapsed into categories; decision rules that specify which pairings of categories qualify or disqualify individuals for help—characterize targeting within most of the major U.S. social programs. Political and legal contestation over the Patient Protection and Affordable Care Act (ACA), for instance, interrogated whether states should be forced to expand Medicaid from a limited set of “morally worthy” categories—the elderly and adults with dependent children—to a new category: low-income adults without dependent children [55]. Social Security takes a continuous attribute—age—and uses a standard threshold to determine who is old enough to benefit, even as demographic groups with lower life expectancy benefit from the cash resources for fewer years as a result. Finally, within the Supplemental Nutrition Assistance Program (SNAP), formerly known as the Food Stamp Program.

---

1In this section, we focus on the way categories define who is eligible for help. When we move to the case studies, we discuss a distinction between two sets of triage decisions: who is eligible for help at all and, among the pool of those eligible, who gets priority to receive help more quickly.

2The age cutoff of 65 was established in the Social Security Act of 1935, the original act authorizing the statute [45]. There have been arguments that due to changing patterns in life expectancy, and growing fiscal pressures on the program, that the threshold needs to be raised. The 1983 amendments to the program changed the structure of benefits administration so that for those born after 1959, full benefits phase in at age 67 rather than age 65 [68]. Yet the need for simplicity means that the eligibility age is still uniform across demographic and occupational groups, despite the fact that some...
U.S. states remain embroiled in debates about and logic, specifically whether able-bodied adults without children should only be allowed access to help if they are also categorized as “working” [72].

While the ACA, Social Security, and SNAP are all highly-visible cases of categorical prioritization, we focus on two less visible examples (Table 1): waitlist priority categories for housing vouchers and weights in K-12 school finance formulas. While housing and K-12 education represent crucial sites of triage, inequalities resulting from their use of categorical prioritization are more subterranean in nature [29], arising from bureaucratic details like whether bureaucracies define homelessness to include doubling up with relatives or only shelter stays. In the next section, we ask: what would change if these policies shifted from their status quo method of prioritization—categorical prioritization—to predictive model-guided prioritization?

3 WHAT CHANGES WITH ALGORITHMIC PRIORITIZATION?

The shift from categorical prioritization of who needs help to algorithmic prioritization involves three changes, each of which reveals tradeoffs of the two methods in terms of inequality. Table 1 highlights the case studies we use to ground the discussion of tradeoffs, with Section 4 expanding upon these cases more fully.

3.1 Difference one: is the basis for prioritization formalized?

First is the formalization of desert: algorithmic prioritization forces decision-makers to operationalize nebulous ideas of who deserves help as a specific measured outcome that can be modeled [1, 2, 23, 35]. Should housing assistance go to those facing the highest risk of homelessness, eviction, costly rent, or another adverse outcome? Should help go to people in the most precarious positions or, as in the mechanism design community have advanced, to those expected to gain the largest marginal benefit from assistance? Categorical prioritization, by contrast, imposes no such constraint on policymakers. It can accommodate different ideas about who policies should help and advance multiple policy objectives.

What does this difference mean for inequality? Formalizing the basis that entitles individuals to help can have both ethical benefits and drawbacks relative to categorical prioritization. One major drawback of algorithmic prioritization that scholars have drawn attention to is that selecting and elevating a particular outcome conflicts with visions of good policy-making as achieving an effective balance between different values that might be important within a policy domain [28, 66]. Returning to the voucher case, given a fixed quantity of vouchers, policymakers may want to give some to existing renters to help them stay housed and others to individuals experiencing chronic homelessness; allocating vouchers by predicting a specific outcome may thus go against allowing a plurality of values. A second concern is that the measured outcomes the algorithms predict may reflect discriminatory, biased, or exclusionary social structures and processes [3, 41]. If the housing-focused model uses predicted risk of homelessness as a factor, but the homelessness measure only reflects those who stay at formal homeless shelters, then the outcome measure may exclude unhoused individuals who feel unsafe in those institutional settings or who have other living arrangements [37, 67]. Alternately, if the housing-focused model targets resources to those who are expected to benefit the most from assistance [2], the model may only flag individuals who resemble those who have accepted assistance in the past, thus codifying existing disparities. Distributing help according to a model trained using poorly-measured outcome or moderator data could thus lock out needy individuals whose life circumstances and identities do not align with the assumptions of state records and data systems [48].

Yet algorithmic prioritization also presents potential benefits. The requirement that policymakers select an outcome to model could also be an advantage: it could lead to clearer deliberations about the goals of providing assistance [1]. To return to the housing assistance case, to decide whether to model eviction or homelessness, policymakers will need to clarify what the policy is trying to achieve: should the policy help existing renters stay in their homes or get unhoused individuals into a home? Public administration scholars have often observed how ill-defined and conflicting goals can hamper the effectiveness of a policy [44]. As a result, debates about what outcome to predict may be beneficial because they can clarify to stakeholders what trade-offs exist and better ensure that the objectives advanced are an actively negotiated choice, rather than a murky default [1]. Similarly, the process of formalization might lead decision-makers to more closely match the prioritization method to the type of assistance provided. So, for instance, the process of deciding what outcome best reflects who deserves housing assistance might lead decision-makers to choose outcomes like homelessness and eviction risk that would align allocation more closely with housing need than do some of the actual categories we document in Section 4.1.

3.2 Difference two: how prioritization approaches intersect with inequalities in power

The use of algorithms to steer decisions in public policy has been critiqued for how it shifts the distribution of power in society [34, 35, 60]. Algorithmic prioritization, as it is often implemented, can concentrate power in the hands of technologists, technocrats, and other already powerful professionals, while further dis-empowering less powerful members of society who may be disproportionately affected by algorithms [11]. Algorithmic prioritization can undermine possibilities for community deliberation, participation, and engagement. It can make it difficult for impacted communities to contest unfair decisions generated by the use of inaccurate or inappropriate input data [13, 70] and to challenge the broader value judgments an algorithm may codify [73]. If those with power choose a basis for prioritization that harms marginalized individuals, there might be limited room for those individuals to contest it.

The common status quo approach—categorical prioritization—however can also intersect within unequal power structures in ways that replicate inequalities. The development of categorical
Algorithmic prioritization could enhance equality if it means policy choices depend less on which social groups are the most politically organized or represented by formal organizations. While categorical prioritization often prioritizes groups that are well-positioned to make effective claims of being deserving of help, algorithmic prioritization would prioritize based on factors determined to be predictive of the selected target outcome. Models could also be trained with greater weight placed on the welfare of those with the least political influence [35]. These changes could make it more possible for less politically visible and stigmatized groups to receive priority. For instance, a predictive model for homelessness might prioritize individuals suffering from substance dependence, a stigmatized group less likely to achieve explicit priority under categorical prioritization [4].

Relatively, algorithmic prioritization could also make resource allocation less of a zero-sum game for groups seeking priority. Under categorical prioritization, triage can pit one group against the other to argue for who deserves priority. For instance, Best [5] shows how disease-specific advocacy groups have competed to frame themselves as the most deserving of biomedical research funding. Since algorithms involve prioritizing based on predicted risk of a target outcome rather than manually selecting discrete categories, algorithmic prioritization could reduce inter-group competition and conflict.

### 3.3 Difference three: are decision rules for prioritization manually chosen or derived from a model?

Finally, with categorical prioritization, decision-makers must deliberate on decision rules for how the categories combine to produce priority. As discussed previously, these decision rules involve using a combination of three logics: (1) and logic (requiring that a help seeker belong to more than one category to qualify for assistance, such as requiring that “able-bodied” adults are also looking for work); (2) or logic (a help seeker can qualify for assistance through membership in either category); and (3) unless logic (a help seeker who is a member of prioritized categories can get assistance unless another factor disqualifies them for priority, such as a criminal history or a particular immigration status). Important in categorical prioritization is that these decision rules are chosen through a deliberative process.

The key shift in algorithmic prioritization is that, rather than a deliberative body choosing these decision rules, the training and validation process for the model inductively selects these rules. To consider a simple example, suppose a government agency uses a decision tree classifier to predict an individual’s risk of homelessness [69]. The process of selecting the most highly predictive model (train-test split, defining a loss function, selecting the model and hyperparameters that minimize prediction error under that loss function) would inductively learn at what splits of age the risk of homelessness increases most dramatically and how attributes like gender and parent status predict risk.

How might this shift from manually to inductively-selected rules impact inequality? In addition to black-boxing the prioritization of stigmatized categories in a potentially ethically beneficial way, an advantage we discuss in the previous section, inductively-chosen

<table>
<thead>
<tr>
<th>Case study</th>
<th>Resource allocated</th>
<th>Need for triage</th>
<th>Categories used</th>
<th>Potential prioritization bases</th>
</tr>
</thead>
<tbody>
<tr>
<td>Waiting lists for housing vouchers</td>
<td>Housing Choice Vouchers which cap rent at 30% of income</td>
<td>Funding only covers vouchers for 1 in 4 eligible renters</td>
<td>Housing-related categories (displacement; rent burden; homelessness), traditional welfare state categories (e.g., elderly; disability), and categories reflecting those with ties to the community and local bureaucracies (local residents; previous tenants terminated from program due to funding cuts; people receiving social services)</td>
<td>Risk of homelessness; risk of eviction; risk of high rent burden</td>
</tr>
<tr>
<td>Pupil weights in school finance formulas</td>
<td>Money from state budgets gets to local districts; money from local districts to individual schools</td>
<td>Students benefit from individualized attention but funding constraints involve sharing resources</td>
<td>Common categories reflect poverty, limited English proficiency, disability, and homelessness</td>
<td>Risk of scoring low on standardized tests; risk of repeating a grade; risk of non-independent living</td>
</tr>
</tbody>
</table>
rules may better reflect the multidimensionality of need. Manual rules simplify the multidimensional nature of need into relatively simple combinations. For instance, as we further detail in Section 4.2, many school finance formulas use manually-chosen or logicals: a district gets extra money for each student that is growing up in poverty or experiencing homelessness or who has limited English proficiency. Yet these student characteristics may combine in complex, nonlinear ways to produce higher need. Algorithmic prioritization would ideally detect this multidimensionality; deliberative bodies may miss it, especially amidst unequal landscapes of advocacy.

4 TWO CASE STUDIES

The previous section highlighted three differences between categorical and algorithmic prioritization, each of which might affect inequality in access to help: (1) is the basis for prioritization formalized; (2) how the approach intersects with inequalities in power; and (3) are decision rules for prioritization manually chosen or found inductively through the model?

While each distinction is relevant for inequality in access to help, precisely how inequality plays out depends on the concrete social and policy contexts in which algorithmic prioritization might supplement or replace categorical prioritization. These contexts shape the ethical tradeoffs of algorithms [35, 60], as, for instance, the distribution of power with respect to algorithms depends on "contingent historical developments and political choices" [35, p.581].

We turn to two case studies previewed in Table 1 to illustrate how these differences play out: waitlists for housing vouchers and pupil weights in school finance formulas. Each discussion draws on broader original research we have conducted of each categorical prioritization system. Here, we summarize key findings and discuss contrasts with how algorithmic prioritization could be used in each context.6

4.1 Waitlists for housing vouchers

4.1.1 The resource being allocated and the need to triage. The first case study of categorical prioritization is the Housing Choice Voucher (HCV) program, a federally-funded, locally-administered housing assistance program that is only funded enough to support 1 in 4 eligible families [59]. Housing vouchers are allocated by over 2,000 local public housing authorities (PHAs), organizations that are given a fixed number of vouchers by the federal U.S. Department of Housing and Urban Development (HUD). These organizations are comprised of street-level bureaucrats [44] who interact with people seeking vouchers and administer assistance, senior management and Executive Directors who oversee program administration and often propose policy changes, and boards of commissioners who provide oversight and steer high-level strategic decisions.

PHAs face a daunting prioritization task. Unlike programs like Temporary Assistance for Needy Families (TANF) or Medicaid, where assistance is given to all individuals who meet eligibility criteria, people seeking vouchers go through two screens for assistance. First, households must be found eligible for help, a process with categorical criteria that involve comparing the household’s income to the area median income and unless criteria that exclude certain categories of non-citizens and people with certain felony convictions. Then, among eligible households, housing vouchers are not an entitlement. Instead, at many PHAs, eligible families are placed on waitlists. These waitlists not only govern how long families must wait for a voucher, but also can govern whether they get help at all, with many waitlists closed and wait times in large metropolitan areas often exceeding a decade [54].

4.1.2 Categorical prioritization on housing voucher waitlists. PHAs rely on various methods to triage help: lotteries, first-come first-served, and categorical priority ranking systems. Figure 1 shows an example of a priority ranking system from Jeffersonville Housing Authority, a PHA in a small city in Indiana, that assigns priority categories different numbers of points.

Figure 1: Example of categorical prioritization on a voucher waitlist: policy for Jeffersonville Housing Authority

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)
D. 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 from based upon availability.

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

Priority systems like that of Jeffersonville Housing Authority arise from the three-step process of categorical prioritization discussed in Section 2. First, decision-makers within local PHAs decide what factors they think make an applicant more deserving of assistance. Does it matter how much of the applicant’s monthly income is spent on rent? What about whether households have elderly members or children? Then, PHAs simplify continuous attributes into categories—for instance, if a PHA decides to prioritize applicants who are rent burdened, it typically relies on spending more than 30% of income on rent as a cutoff. Finally, PHAs devise decision rules for aggregating these categories into priority. Some PHAs use "flat" systems that give equal priority to members of any category. Others like Jeffersonville Housing Authority depicted in Figure 1 use tiered systems that give more points and a higher waitlist position to certain categories or pairings of categories. In that example, being a resident of the local community is the category prioritized most highly, assigned 30 points; the next tier prioritizes categories like victims of domestic violence or applicants who are “involuntarily displaced” due to issues like natural disasters, as well as individuals who are either working or elderly or have a disability. Finally, honorably discharged veterans receive some but lower priority. The example illustrates the three-step process: deliberating about dimensions of deservingness; simplifying those dimensions...
into categories; and deciding on decision rules (in this case, allowing applicants who fall into multiple categories to sum the point values attached to each category and then ranking applicants based on cumulative points).

In a separate paper with additional details on the data and methods [74], we investigate two sets of questions. First, which categories do PHAs use to grant priority? Second, how are prioritization policies developed and why do PHAs adopt the prioritization policies that they do? We investigate the former by collecting and coding the prioritization policies of 1,398 PHAs and the latter through interviews with 22 housing authority officials.

Figure 2 presents the prevalence of different categories across the PHAs in our sample. It indicates that housing authorities often prioritize a number of groups popularly regarded as the “deserving poor” [36], such as people with disabilities, the elderly, veterans, and working individuals, as well as people with ties to the community, other local organizations, or to the housing authority itself. Notably only some of the commonly used categories directly relate to housing need.

In what follows, we draw on this research to ask: what forms of inequality emerge under categorical prioritization and how might algorithmic prioritization alter those?

4.1.3 Formalized basis for prioritization. Focusing on categorical prioritization’s lack of a formalized basis of prioritization, we see the prioritization of categories that reflect a high degree of value pluralism about who deserves housing assistance. Our interviews suggested that categories like “is a local resident” are driven by a mix of ideas that members of the local area deserve help before “outsiders” and ideas that deservingness should also be tempered by practicality. As an example of practical concerns shaping prioritization, some of our interviewees noted concerns about wasting resources on families who may not be willing to live in the area long-term. Categories like disability and elderly reflect more traditional deservingness considerations, with some interviewees discussing how these groups are morally worthy of assistance because they struggle through no fault of their own. Meanwhile, prioritizing existing tenants in other PHA programs reflects ideas that housing authorities ought to make good on their earlier commitments and better position the families they already help for upward mobility. Notably, amidst this value pluralism, categories that directly reflect housing need are not the most prevalent across PHAs. The category that is arguably most reflective of severe housing need—homelessness—is ranked below categories like disability.

In our interviews with housing authority officials, some interviewees also struggled to offer a rationale for the categories they chose, with several indicating simply that the ones they chose were “obvious” or “straightforward.” Just as some scholars have critiqued algorithms for making rankings seem objective and a matter of “natural necessity” [35, p. 581], some interviewees viewed their PHA’s system of categorical prioritization as a natural outgrowth of objective assessments of deservingness rather than as a negotiated construction.

In contrast to categorical prioritization, algorithmic prioritization would replace this vagueness about the basis for prioritization with a single target outcome to model, such as risk of eviction, risk of homelessness, or another outcome. This could collapse the multitude of values and goals that currently inform voucher prioritization policies. It could be a drawback if the status quo—categories on waitlists reflecting a mix of normative ideas—are valuable since they allow the policy to express heterogeneous values. Conversely, the formalization required by algorithmic prioritization could also be seen as a benefit, as it would help spur localities to clarify the often murky goals of local voucher programs and provide a consistent, traceable basis for prioritizing some over others.

4.1.4 How power inequalities shape prioritization. Categorical prioritization in the HCV program revealed several ways that inequalities in power and representation shaped prioritization choices. Nearly universally across our interviews, housing authority officials, shared that, despite outreach efforts by some, no current or prospective voucher holders attended meetings to provide input on prioritization policies. More commonly, prioritization choices were shaped by what the leaders of local community organizations advocated for. Officials described how organizations serving particular sub-populations often argued that those sub-populations had unique needs and ought to be prioritized. This is reflected in the large proportion of housing authorities with preferences related to an applicant’s relationship with other organizations.

Prioritization choices also often reflected the particular commitments and beliefs of the people who tended to be in the room when decisions were made—housing authority staff and board members. The personal connections that these individuals happen to have to certain social groups, as well the negative stereotypes that they hold of other social groups, have an outsized influence over what categories are selected for priority. Thus, policies tend to reflect the views of comparatively powerful actors.

Algorithmic prioritization, by contrast, would have less clearly-defined constituencies. While categorical prioritization could lead to zero-sum games in triage—advocates for homeless households making claims; advocates for victims of domestic violence making counterclaims—the core decision in algorithmic prioritization would be choosing a target outcome to model. Lack of clear constituencies around a target outcome could be both a benefit and a drawback. On the benefit side, it may reduce lobbying over priority, since the constituency around a target outcome of “risk of eviction” might overlap a great deal with the constituency around an alternate outcome of “risk of high rental burden” in a way that reduces the risk of deep divisions over the basis for priority. Yet on the drawbacks side, it could risk further ceding the prioritization process to technocrats distant from impacted communities.

4.1.5 Manually versus inductively-chosen decision rules. Our interviews with housing authority officials suggested some benefits and drawbacks of manual decision rules. For some, the ability to manually rank different groups served an expressive function in itself. It offered a chance for housing authorities to show off their understanding the local community and its unique needs and preferences. The decision rules that officials devised moreover tended to be tailored to organizational realities—what schemes would be feasible for overworked housing authority staff to implement—and were often kept simple so as to be easier to explain to applicants and the public.

At the same time, housing authorities settled on decision rules for priority in often arbitrary ways. Interviewees suggested that
deliberations led to a general sense that some groups should be given more points for priority than others, but specifically how many was often based on guesswork or intuition. Indeed, a number of interviewees said they were not sure how the points attached to different categories were determined. Some interviews further indicated that category rankings were shaped by key stakeholders’ folk theories of poverty. One interviewee, for instance, drawing on ideas that poverty is due to a lack of effort, described giving “working” a certain point value because she thought it might be sufficient incentive to motivate an unemployed applicant to get a job. Inductive decision rules in this setting thus are often informed by gut instinct rather than systematic efforts to link policy choices to policy goals.

In contrast, algorithmic prioritization would learn weights on applicant attributes most predictive of the target outcome. An ethical benefit is that this would make these weights less arbitrary. A drawback, however, could be that algorithmic prioritization could result in decision rules that are more complex to implement and challenging for the public to understand.

4.2 Pupil weights in K-12 school financing formulas

In this second case study, we shift to another example of categorical prioritization: school financing formulas that allocate funding to districts and schools based on categorical measures of student need.

4.2.1 The resource being allocated and the need to triage. The resource being triaged in school finance formulas is either money that states allocate to Local Education Agencies (LEAs; we use the term school district for short) or money that school districts allocate to individual schools. Each of these pools of money is finite, and decision-makers face difficult tradeoffs between many dimensions of need that could entitle some students to extra support. Resource shortages in schools, combined with students with many dimensions of need, require triage and prioritization.

4.2.2 Categorical prioritization in school finance formulas. This prioritization often takes the form of “pupil weights” or “weighted school financing” schemes [19, 42, 56]. To illustrate pupil weights, imagine two districts with the same actual count of enrolled students: District A and District B each have 1,000 students. The idea behind pupil weights is that depending on who those 1,000 students are, districts might deserve more resources if they have higher concentrations of students who need extra support. So if

---

This support can take the form of smaller class sizes, more paraprofessionals like school counselors or social workers, tutoring, and other educational inputs that require resources.

These are both used by states to shape allocations to each school district in the state and by especially large metropolitan school districts to allocate money to individual schools.

In practice, counting students is its own challenge with formulas varying in whether they use measures of average enrollment, measures of average attendance, or attendance on a particular day.
District A’s 1,000 students have a higher prevalence of learning-related disabilities, a higher prevalence of students growing up in poverty, and a higher prevalence of students still learning English, then the district’s funding should be equivalent to 1,500 students (as an example). In contrast, if District B has an affluent student population with little disability and high levels of English proficiency, the district’s funding remains calibrated to 1,000 students. Put differently, students are upweighted and count as needing funds for more than one student if they fall in certain categories of need.

States and districts select these categories and their weights using the three-step process of categorical prioritization discussed in Section 2. First, decision-makers decide which dimensions of student need are salient. As Duncombe and Yinger [19] describe, formulas vary nationwide but often prioritize students with disabilities, students who are growing up in poverty, and students who are English language learners (ELL). These represent different dimensions of why students might deserve extra help, ranging from medically-caused learning-related challenges (disability) to the stresses of growing up in poverty to challenges with English instruction as someone who is not a native speaker (ELL). While the remedies for each type of challenge might be different, formulas upweight students who belong in these categories so that districts and schools can fund these remedies.

School finance formulas also reflect the second and third steps of categorical prioritization: simplifying continuous forms of need into categories and deciding on decision rules for how those categories combine. To make this more concrete, we turn to one specific pupil weighting scheme that the District of Columbia (DC) uses to allocate local funding to over 200 public and public charter schools: the Uniform Per Student Funding Formula (UPSFF) [6].

UPSFF starts with a “foundation level,” or weight, which is the amount of money thought to be needed to educate a student with few extra needs. UPSFF then has pupil weights intended to cover the extra costs of serving students in certain categories. The three main categories consist of i) students with disabilities, ii) students who are English language learners, and iii) students in a DC-specific “at risk of academic failure” category, which includes students whose families are receiving needs-based governmental assistance through TANF or SNAP, students experiencing homelessness, who are involved in the child welfare system, or who are in high school and over-age for their grade (likely indicating earlier grade repetition). Table 2 shows the amounts for each category in the 2022 fiscal year.

Returning to step two of categorical prioritization, the formula simplifies multidimensional need into categories. For instance, students in the category of English language learners can have very different continuous levels of proficiency. But for simplicity, that varying need is collapsed into a single category. In terms of step three, the funding formula also reflects manually-chosen decision rules about how categories combine. Focusing on the “at-risk” category, the formula uses an or logic instead of an and logic. A student who has been involved in the child welfare system, whose family qualifies for SNAP, and who is over-age for their grade receives the same funding weight and priority as a student with only one of these attributes. Indeed, a recent report critiqued the equity of this manually-chosen or logic, arguing that the category should be updated to give higher priority to students in rarer groups within a category (e.g., those involved in the child welfare system), give more funding for students who display two or more characteristics, or include other changes that would reduce the within-category homogenization of need [51]. Yet doing so would undo some of the “at-risk” category’s simplicity.

Finally, returning to step three of categorical prioritization—manually deciding on decision rules—the funding amounts in Table 2 reflect political deliberations within the DC budget process. While the supplemental weights are meant to be calibrated to the added costs of educating students in that category, there are large between-category differences in estimates of these costs; schools receive the most for students with disabilities, roughly half that amount for English language learners, and between one quarter and one-fourteenth of that amount for at-risk students.

How might algorithmic prioritization alter school finance formulas and what are the potential implications for inequality?

4.2.3 Formalized basis for prioritization. First, the categories in formulas like UPSFF reflect different ideas about which students deserve extra help, and by extension, which schools and districts need extra funding to help them. The disability category reflects ideas that students who face learning challenges that are traceable to a defined disability are entitled to extra support, reflecting ideas of compensatory justice at the student level [32]. Students who are still learning English are also cast as deserving of extra help, but in UPSFF and many other formulas, often generate less of a funding boost for their schools. Meanwhile, the at-risk category shifts the unit of analysis for deservingness from the individual student to his or her family unit, reflecting how family poverty and stress related to experiences like homelessness can spill over onto the student’s academic outcomes. Similar to the housing case, categorical prioritization allows for a high degree of value pluralism about who deserves state help.

Just as algorithmic prioritization would flatten this value pluralism in the voucher case, algorithmic prioritization for student need would need to settle on a single target outcome. This outcome would make transparent the basis for prioritization. Should students be prioritized if they are at highest risk of grade repetition or failure on required state exams? Or should students most likely to experience the greatest academic gains from extra help be prioritized? Formalizing this basis would likely lead to a different ranking of student priority than the rankings in many formulas.

In contrast, categorical prioritization allows the prioritization basis to remain nebulous. In a study considering revisions to the financing formula, the DC government stated that their formula should achieve four goals: “simplicity: the option considered is easily explained to impacted stakeholder groups; impact: the change

10 The prevalence of these categories likely relates to the fact that they are also defined within federal statutes used for other parts of school financing. Disability is defined with reference to an Individualized Education Plan (IEP) under the Individuals with Disabilities Education Act (IDEA). Poverty is often defined as students who qualify for Free or Reduced Price Lunch, either through family income below 185% of the federal poverty line (roughly $48,000 in School Year 2020-2021 for a family of four) or through receiving public benefits like SNAP or TANF.
11 This analysis draws upon publicly-available documents summarizing the origins of UPSFF and changes over time.
12 These funding amounts are discussed in DC’s legislative code, here: https://code.dccouncil.us/us/dc/council/code/titles/38/chapters/29/.
results in funds going to students that need it the most; accountability: outcomes of funding changes can be measured over time; aligned incentives” [51, see slide 44]. Yet goals like prioritizing students “that need it [the funds] the most” leave the precise basis for prioritization unspecified given many ideas about student need. Choosing a target outcome would result in increased transparency about the rationale for prioritizing some students over others.

4.2.4 How power inequalities shape prioritization. Would algorithm-derived rankings of students widen or reduce inequality relative to the category-based rankings? The answer likely depends on how the power of parents and advocacy groups shaped the rankings under categorical prioritization. Notably, the category receiving the least supplemental funding—the “at-risk” category—is comprised of students whose families are likely low in political power, with research beyond DC documenting low political mobilization of public benefits recipients [10, 46]. In contrast, the highest-ranked category—disability—is comprised of students of families with a wide range of income levels and political power. Therefore, the weights may reflect a mix of true need for funds and the political power to make claims on key decision-makers to recognize those needs.

Similar to the housing case, algorithmic prioritization could make prioritization less of a power struggle between parents of and advocates for students in different categories—for instance, disability advocacy groups pitted against groups advocating for immigrant parents. Instead, the key decision—what measure of student risk or need do we predict—would be the target of contestation and deliberation. Since it may not be clear who is prioritized under different target outcomes, there may be less opportunity for inequality-amplifying advocacy.

4.2.5 Manually versus inductively-chosen decision rules. The power of different groups to shape rankings leads to the final distinction between categorical and algorithmic prioritization. Are the rankings chosen by a deliberative process or through modeling an outcome? Focusing on categorical prioritization in UPSFF, as Bhat [6] documents, the dollar amounts attached to each category, and thus the ranks between them, were set by a two-stage process. An independent commission first developed proposals for how much funding each category should receive for the district to achieve educational adequacy. Then, the final amounts were set in

Table 2: Example of categorical prioritization in school finance formulas: DC’s UPSFF categories and weights. The first two columns show the supplemental weights given to three categories in FY 2022: students with disabilities; English language learners; and an “at risk” category reflecting different forms of poverty and hardship or potential challenges with high school completion. The last column, based on Bhat [6], shows whether categories were funded at a level higher or lower than what experts studying the formula between FY 2014 and FY 2015 recommended.

<table>
<thead>
<tr>
<th>Category</th>
<th>Supplement to foundation funding for uncategorized general education student</th>
<th>Actual funding relative to expert-recommended funding</th>
</tr>
</thead>
<tbody>
<tr>
<td>Special education (disability)</td>
<td>$11,378 - $40,938 (four levels of service intensity)</td>
<td>Higher</td>
</tr>
<tr>
<td>English language learner (ELL)</td>
<td>$5,865 (PK3-5th grade); $8,798 (grades 6-12)</td>
<td>Lower</td>
</tr>
<tr>
<td>“At-risk” (benefits receipt; homelessness; child welfare involvement; high school over-age for grade)</td>
<td>$2,815 (with $704 supplement for at-risk students who are behind grade level in high school)</td>
<td>N/A (new category in FY 2015)</td>
</tr>
</tbody>
</table>

the DC budgetary process. Bhat [6] shows how the commission-recommended funding amounts for the categories compare to the final amounts decided in the FY 2015 budget that followed the study. These comparisons highlight two patterns. First, the English language learner category was funded at a rate lower than the independent study’s recommendation: the independent study recommended funding at ~ 61% of the base amount (~ $6,400), and the budgetary process resulted in an amount 12 percentage points lower (~ $5,172). Second, the special education category was funded at a rate slightly higher than the study’s recommendations. While a detailed account of the budget negotiations that produced those deviations is beyond the scope of the present paper, the funding amounts for categorical prioritization resulted from the combination of an expert-driven “technocratic” process of projecting expenditures needed to achieve education adequacy for each group and a budgeting process where the relative power of different constituencies—advocates for students with disabilities; non-English speaking parents; non-parent, taxpayer residents—may have played a role.

How might model-selected decision rules impact inequality? Investigating this question is complicated by the endogeneity of student outcomes to school expenditures under categorical prioritization. More specifically, suppose the English language learner is “underfunded” relative to those students’ need for help. This underfunding could depress the educational outcomes of students in the category [31]. A predictive model trained to predict outcomes like student outcomes to school expenditures under categorical prioritization.

5 CONCLUSION
The present paper argues that questions of algorithmic fairness should ask: what is the bureaucratic counterfactual? While a large body of research compares algorithms to one counterfactual method—a human with full discretion [e.g., 15]—in U.S. social policy, the counterfactual is often categorical prioritization. We showed how categorical prioritization involves three steps common across many policies. First, policymakers must reduce the dimensionality of need

13These differ from the FY 2022 values in the middle column of the table because of changes in funding amounts between FY 2015 and FY 2022.
and decide which dimensions are relevant for aid. Then, continuous attributes are often simplified into categories that homogenize different levels of need. Finally, policymakers deliberate on decision rules—or logic; and logic; unless logic—that determine how categories shape eligibility and priority for help.

Shifts in social policy to algorithmic prioritization involve three contrasts with categorical prioritization. First, algorithms, through the choice of a target outcome to predict, require formalizing the basis for prioritization. This basis for prioritization may be absolute risk of some outcome a beneficial resource is attempting to avert, such as homelessness or high school dropout, or the chance of some good outcome in the absence of the resource, such as high test scores. Or it may be the beneficial resource’s marginal impact on an outcome, as in, for instance, Athey and Wager [2]’s work discovering attributes predictive of larger treatment effects from a beneficial resource. On the one hand, each of these formalizations may flatten important value pluralism that guides the choice of different categories. On the other hand, many categorical prioritization schemes may be guided less by value pluralism and more by nebulous ideas of need that strike observers as prima facie plausible but that are not interrogated. In the housing voucher case, many officials saw the categories their organization prioritized as self-evident; in the school finance case, vague ideas of student need and educational equity can conceal tradeoffs between different conceptions of need.

Second is the role of power in prioritization. While algorithmic prioritization is critiqued for placing power in the hands of technocrats rather than impacted communities [34, 35, 60], categorical prioritization can create zero-sum dynamics as categories with defined constituencies jostle for relative priority. In the voucher case, few impacted residents attend the meetings where important prioritization decisions were made; in the school finance case, categories and their ranks are more contested but may also reflect power differentials between parents. Third is whether the decision rules for prioritization are manually selected by those deliberating or inductively selected through the predictive modeling process. On the one hand, manually-selected rules can make it clear who achieves priority and make it easier for those impacted by a prioritization method to understand or challenge it. On the other hand, they often simplify multidimensional forms of need and lack a traceable rationale. In the voucher case, decision-makers often had little recollection of why certain categories received more weight and higher priority; in the school finance case, expert-recommended weights were altered during budgeting processes.

These cases illustrate two important directions for future research in algorithmic fairness. First, as algorithms migrate from ranking “risk people pose to society” to ranking “help people need from society,” the counterfactuals to algorithms shift. Social policy has a long tradition of simplifying the complexity of human need into categories and deliberating about which categories are more deserving, an institutional context that requires its own branch of ethical analysis. Second, our analysis shows the value of empirically-guided studies of the decision-making methods that predate algorithms. Whether algorithms’ formalization of desert, for instance, widens or narrows inequality depends on the specifics of governments’ previous prioritization method. Existing empirical research on the causal impact of algorithms has focused on contrasts between an algorithm-guided decision and a human with full discretion, using either randomized controlled trials of algorithms [e.g., 14, 30] or quasi-experimental-designs [e.g., 38, 65]. While these studies examine the impact of algorithms on the choices of individual decision-makers, categorical prioritization is a bureaucracy-level policy. Therefore, studies of the impact of shifts between categorical and algorithmic prioritization will need to examine distributional outcomes either exploiting pre-post shifts within the same organization, relying on between-organization variation, or using simulation [e.g. 50]. Finally, and as the growing literature on “humans in the loop” of algorithmic decision-making illustrates [e.g. 7, 17], prioritization in practice may involve combining elements of categorical and algorithmic prioritization. For instance, coarsening continuous forms of need into categories—e.g., disability above a certain severity level—may already occur in the input data used to train algorithms, meaning that algorithms lose the advantage of learning nonlinearities in the relationship between continuous features and an outcome. Studies of decisions about algorithms in the pre-deployment stage—for instance, ethnographies that document how time pressures shape biases in input data—can reveal the ways in which issues with older forms of prioritization can seep into prioritization guided by predictive models.

In sum, our analysis joins other calls to avoid “abstraction traps” in the study of algorithmic fairness [60], grounding the rise of algorithms in social policy in a long historical trajectory of ranking need for the purpose of triage.

ACKNOWLEDGMENTS

The authors would like to thank audience members at the 2019 Information Ethics Roundtable at Northeastern University and the Stewart Lab for helpful feedback. This research was partially supported by the Center for Health and Wellbeing at Princeton University. RJ acknowledges funding support from the following sources: the American Bar Foundation/JPB Foundation Access to Justice Scholars Program in the 2020-2021 cohort and a 2021 CompX Grant from the Dartmouth College Neukom Institute for Computational Science.

REFERENCES


[74] Simone Zhang and Rebecca Johnson. 2022. Hierarchies in the Decentralized Welfare State: Prioritization in the Housing Choice Voucher Program. osf.io/xxs8m