1 Introduction

Digital systems control many aspects of our lives. These decisions vary enormously in consequence; sometimes they determine whether forum posts we make are made in good faith [16, 14]; sometimes they instruct judges whether it would be safe to allow us to roam free in advance of an upcoming court trial [7]. The applications that we see in society are many and diverse, but one aspect of these algorithmic systems seems to come through across all channels: surprise and — quite often — frustration with these systems over the decisions they make.

Researchers have spent some time documenting this frustration, mostly as isolated incidents. Sometimes it manifests when on–demand workers have to work to show their eligibility to work [3, 15, 9]; other times, it arises as a result of precarious negotiations with bots which operate within and alongside online platforms [13, 6]; further afield, researchers have written extensively about the many dimensions of frustration people feel when algorithmic systems mediate their news feeds in unexpected ways [10, 11, 2, 4]. But this body of work has generally discussed these cases as separate, even independent, phenomena.

In this paper, we'll explore these issues as one problem space, brought into alignment by focusing on and talking about algorithmic systems as occupying the role of "street–level bureaucracies". By unpacking these systems in the moments surrounding consequential (sometimes controversial) decisions, and as substantive agents of policymaking themselves, we hope to show that useful insights, questions, and agendas naturally arise. Or to put it another way: we'll try to give researchers some "foothold" from which they can begin to climb this otherwise sheer cliff of a domain.

This paper will sketch out a number of ways that algorithmic systems fail in their designed roles, and ultimately show that what we'll call "street–level algorithms" will inevitably continue to fail. We'll look at a number of cases from a wide range of settings: (1) in YouTube's content labeling and monetization systems; (2) in the perceived algorithmic bias in bail–setting recommendations; (3) in Facebook's news feed and in the decisions that get made regarding promotion and moderation; and (4) in the algorithmic management of on–demand work and workers themselves.

While this paper will draw a line from Lipsky's work to what we're seeing today, we won't ultimately end with the same predictions and recommendations that he did. Namely, while Lipsky rejected the viability of street–level bureaucrats being subsumed by algorithmic systems, we'll take it as a given that this has already happened, and deal with the outcomes. Specifically, we'll formalize some of the issues that have emerged by replacing this human work with algorithms, and proceed to recommend ways that we can mitigate the negative effects of "street–level algorithms".

2 Bureaucracy

A large body of work on bureaucracies already exists, describing the subject both in theory [17] and in practice cite lots of ethnographies. For our narrowly scoped needs, we'll let it suffice to say that bureaucracies are organizations which have goals, and whose goals are broken down into tasks. To accomplish those tasks, roles are defined, and those roles are occupied by people known as bureaucrats. Bureaucrats occupy all levels of bureaucracies, but Lipsky specifically calls attention to the agents who make decisions "on the street", which he calls the "street–level bureaucracy" [12].

2.1 Street–level bureaucracy

Street–level bureaucracies are the subset of larger bureaucratic organizations that everyday people interact with. The judicial system is a bureaucracy, but the subset of people with whom the public interact — a judge, a prosecutor, and other functionaries of the judicial system — represents the street–level bureaucracy of the court system. Law enforcement organizations may be enormous bureaucratic institutions, but police officers, parking enforcement officers, and others constitute the "street–level bureaucracies" of law enforcement organizations.

Arguably one of the most powerful tools street–level bureaucrats have is the ability to exercise discretion. They can do this in a number of ways: a police officer may choose not to pull over someone who's speeding; a judge may pass a more lenient sentence upon a defendant; an instructor may waive a prerequisite for a course. Strictly speaking, exercising discretion means working against the general instructions which inform their work, but street–level bureaucrats exercise discretion to further the goal of the organization, even if it means not performing their task.

It's critically important that street–level bureaucrats not be rigidly prescribed, and that their latitude to exercise discretion remains intact, for a number of reasons. For one thing, human circumstances vary so widely and unpredictably that anticipating every circumstance, prescribing a course of action in advance, is infeasible. But even in circumstances that appear predictable, it's important for people to have faith that the institution operates in their interest, or they'll lose hope that the bureaucracy will benefit them in general, and they'll find ways to work around the system. "Street–level bureaucrats …at least [have] to be open to the possibility that each client presents special circumstances and opportunities that may require fresh thinking and flexible action." [12].

So while it's important that street–level bureaucrats have latitude, it's equally important to reflect on the effective power this consequently gives them. By exercising their discretion and not issuing tickets for cars speeding only marginally over the speed limit, police officers effectively create a policy that raises the speed limit. If an instructor waives a prerequisite for their course for all inquiring students, they are effecting a policy that negates the prerequisite. If a judge hands down more

lenient sentences for certain categories of crimes, or certain kinds of defendants, their actions instantiate a policy that didn't exist before. Most importantly, the nature of the work street–level bureaucrats do makes scrutiny of their decisions either prohibitively difficult or actually impossible.

3 Case studies

3.1 YouTube

YouTube employs a number of algorithms which determine myriad characteristics about videos uploaded to the site, such as whether the content is protected by an existing copyright [8], or violates YouTube's "community guidelines" [18]. While it generally makes the right calls, in certain cases YouTube's monetization algorithms have made systematically wrong decisions: YouTube has been discovered labeling videos uploaded from members of the LGBTQ community as "sexually explicit", which deprived them of income they would otherwise have earned [5]; at other times, it's failed to identify deeply inappropriate algorithmically–generated content [1].

Thinking of YouTube as a sort of policing force — one which must work with and ultimately regulate public performers — might help us bring some of the dynamics into focus. Many cities have laws which restrict busking (street performance) in certain contexts; the details of enforcement of those laws is naturally left to police officers, affording them substantial discretion. In negotiating a fair balance with the busking community, police may elect not to enforce restrictions on some public performances while enforcing those restrictions on others. The decisions these street–level bureaucrats make about when, where, and why they enforce public performance and solicitation laws effectively becomes a policy that hasn't been legislated, but instead came into being by way of their action (or in this case, inaction).

In much the same way that police make decisions about what kinds of street performance to allow and flourish, YouTube's monetization algorithms make decisions about what sort of content ought to thrive. By demonetizing videos about LGBTQ issues, YouTube effects a policy which tangibly stymies performers who talk about gender and sexuality. Without instruction derived from any particular policy that YouTube's engineers or leadership have articulated, the algorithms that operate within YouTube itself become policymakers in their own right.

But the street–level bureaucracy that is YouTube's monetization system has substantially less capacity for discretion for a handful of reasons. On the street, police and performers can negotiate and come to mutual understandings about boundaries. Police can interpret the circumstances and the content of the performance and make reasonable decisions even without prior exposure to a performance of that nature. Algorithmic systems can be trained on existing data, but do poorly when trying to make sense of new dimensions to which they haven't been made aware. In other words, they can't be adequately prepared for novel situations, and as a result can never exercise discretion in precisely the context that it's needed — in marginal, ambiguous, or unprecedented circumstances.

More importantly, however, street–level bureaucrats know what they're doing and why, which is currently beyond the limits of state–of–the–art artificial intelligences. A police officer may arrest a street performer engaging in a "three–card Monte" trick — a widely known confidence scam — but not arrest a musician. To the police officer, it's clear that arresting the con artist serves the police department's goal, while arresting the musician does not. An algorithmic system attempting to replace a police officer in this scenario wouldn't be able to differentiate between benign street performers and malicious ones until it had been trained to understand characteristics that differentiate the two.

YouTube's monetization algorithms fail in the same way in both of the aforementioned cases. Without ever having seen such a thing before, YouTube's algorithms could never have anticipated that sexually explicit language might have meaning that's about sex, but not sexual. Similarly, YouTube's monetization algorithms had never been exposed to children's television material reformulated by adversarial systems in bizarre, sometimes deeply troubling ways. YouTube's algorithms lack the capacity even to understand what they're doing, to say little of when and where to fail the task in favor of the goal, and to say even less of whether it understands the substance of the content it's evaluating, which it can only passably do when it has seen similar content before.

With every novel idea online performers imagine, YouTube gets better at understanding that particular novel idea — but the nature of transient performance art (and perhaps the nature of people in general) is to continue to push limits in new, radical ways. With each new edge case, the frontier of that edge pushes further outward, making the prospect that algorithms will ever be enough in this context unlikely. So, being unable to deeply understand the task, let alone the organization's goal, and not being meaningfully capable of reflecting on the content or the circumstances surrounding the content itself, and being largely inadequate to negotiate boundaries with performers, for whom the only consistent characteristic may be a tendency to seek the unconventional — it should come as little surprise that YouTube's monetization algorithm continues to represent a frustrating adversarial agent to many performers.