Jacob Sparks* and Athmeya Jayaram

Rule by Automation: How Automated Decision Systems Promote Freedom and Equality

https://doi.org/10.1515/mopp-2020-0066
Published online January 6, 2022

Abstract: Using automated systems to avoid the need for human discretion in government contexts – a scenario we call 'rule by automation' – can help us achieve the ideal of a free and equal society. Drawing on relational theories of freedom and equality, we explain how rule by automation is a more complete realization of the rule of law and why thinkers in these traditions have strong reasons to support it. Relational theories are based on the *absence* of human domination and hierarchy, which automation helps us achieve. Nevertheless, there is another understanding of relational theories where what matters is the *presence* of valuable relationships with those in power. Exploring this further might help us see when and why we should accept human discretion.

Keywords: AI, automation, republicanism, non-domination, social equality

1 Introduction

Enforcing and adjudicating the law involves a great deal of human labor and leaves a great deal of room for human discretion. Because human decisions are subject to many kinds of errors and distortions, one might hope to automate the administration of the law. Governments have already started using various kinds of automated systems to avoid the need for discretion. Red light cameras can issue traffic tickets without the need for traffic cops, machine learning systems can replace government agents in detecting fraud and tax evasion, judicial decisions

Jacob Sparks and Athmeya Jayaram contributed equally to this article.

***Corresponding author: Jacob Sparks**, Philosophy Department, California Polytechnic State University, San Luis Obispo, Bldg. 47, Rm. 37 1 Grand Avenue, San Luis Obispo, CA 93407-0329, USA, E-mail: jasparks@calpoly.edu. https://orcid.org/0000-0002-6571-3314
Athmeya Jayaram, Berman Institute of Bioethics, Johns Hopkins University, Baltimore, MD, USA; and Wellcome Centre for Ethics and Humanities, University of Oxford, Old Road Campus, Oxford OX3 7LF, UK, E-mail: ajayara2@jhu.edu

about bail and parole can be performed by various kinds of algorithms, and all manner of administrative decisions – from immigration to government hiring to regulatory enforcement – can be made by machines instead of people. We'll refer to any system that replaces human discretion in the administration, enforcement or adjudication of the law with an automated process as 'rule by automation.'

The rise of rule by automation has occasioned many significant concerns. Such systems might lack transparency, reinforce biases or obscure responsibility in harmful ways. Nevertheless, it would be a mistake to oppose rule by automation completely. Human discretion in the administration of the law is highly problematic. It leaves us open to domination by public officials and can create objectionable social hierarchies between those who administer the law and those who are subject to their decisions. By replacing human discretion, rule by automation advances important forms of freedom and equality.

The specific forms of freedom and equality we have in mind are associated with neo-republican and social egalitarian political philosophy. Neo-republicanism and social egalitarianism are both relational theories; they analyze their respective values in terms of certain human relations. Republicans see freedom as the absence of dominating human relations. Social egalitarians see equality as the absence of hierarchical human relations. That is why, for both theories, the rule of law is essential. When we are ruled by laws, rather than human discretion, we reduce dominating and hierarchical human relations.

For the same reasons these views support the rule of law, we'll argue, they should also support rule by automation. Automation is a more complete realization of the rule of law ideal (D'Amato 1977). It further reduces the discretionary power of public officials, and it does so without introducing any new social hierarchy.

Despite these advantages, there may still be an intuitive discomfort with having an automated system make decisions that, for instance, determine your immigration status or send you to jail. A society ruled by automation certainly doesn't feel like an unmitigated good from the perspective of freedom and equality. The emphasis that relational theories place on eliminating objectionable human relations, and their relative neglect of the importance of promoting valuable human relations, makes it difficult for them to capture these freedom-and-equality based concerns with rule by automation.

So, we close by exploring a few ways in which it might be important to non-domination or social equality to allow for human discretion in the administration of the law. As advancing technology and a growing awareness of the foibles of human judgement put pressure on governments to automate more processes, appreciating the impact that rule by automation has on our freedom and equality

can help us make better choices about the division of labor between human and machine intelligence.

One point of clarification before we continue: there are many dimensions and gradations in our notion of rule by automation. Automated decision making can involve very simple rules or highly complicated calculations. They can be implemented by single persons, teams of people, or machines. Individuals can have more or less oversight over automated processes and can have more or less of a capacity to reverse their decisions. While advances in AI are the occasion for recent debates about rule by automation, we do not mean to limit the discussion to systems that utilize any particular technology. What matters most is that rule by automation, whatever form it takes, reduces the need for human discretion in the administration of the law.

2 Automation and Non-domination

Neo-republicans like Philip Pettit argue that being under the power of others is objectionable when your choices depend on their arbitrary will. The classic example is that of the benevolent slave master. Under liberal theory, the problem with slavery is simply that the slave master interferes with the choices of the slave. The neo-republican response is that this is not the only, or even the main, problem with slavery. The main problem is that, even if the master decides not to interfere with the slave, the master still gets to make that decision. The slave's freedom depends on the master's pleasure – on his arbitrary will. So, even if the slave is not interfered with, she is still 'dominated.'

This example illustrates the three key features of domination (Pettit 1999). Domination occurs when an agent has the capacity to intentionally and arbitrarily interfere with one's choices. The first feature is that the dominator must be an *agent* (Pettit 1999, p. 52). The dominator must be an agent, rather than a system, because domination is fundamentally a relational problem. A capitalist system might create pressures that interfere with one's choices, but one is not therefore dominated by it.

The second feature is that the dominator must have the capacity to *intentionally* interfere with one's choices (Pettit 1999, pp. 52f.). This condition is meant to distinguish domination from the effects of nature and chance, which may interfere with our choices but not by anyone's hand. The third feature is that domination involves being subject to another's *arbitrary* power. Neo-republicans employ different conceptions of arbitrariness, but all agree that an essential aspect of arbitrary power is that it is uncontrolled (Lovett 2012). The slave master is uncontrolled in their power, so they are free to impose their arbitrary will.

The solution to domination is, therefore, to control the exercise of power (Pettit 1999, pp. 57f.). Controls on power address all three features of domination. When our rulers are controlled by laws, for instance, their decisions no longer reflect their intentional and uncontrolled wills. They may still have the capacity to interfere with our choices but only in their role as public servants, not as independent agents. One test of whether power is adequately controlled is what Pettit calls the 'eyeball test' (2012, p. 84). When our fate does not depend on others' arbitrary will, we can look them in the eye without bowing or scraping, without fear or deference. We are not dominated.

The rule of law is therefore an essential part of the solution to domination because it controls the wills of public officials, giving us a 'government of laws and not of men.' When we are all governed equally by laws, we can look others in the eye as equals, not as ruler and ruled. A familiar critique of the rule of law, however, is that laws cannot apply themselves (Hart 2012, ch. VII). Human discretion is needed to decide which laws apply to a situation, to interpret those laws, and then to determine which actions they require. This discretion reintroduces the possibility of domination. If a public official has the discretion to decide whether or how much to punish a citizen, perhaps because the law is vague or does not specify a penalty, then the citizen's choices still depend on the uncontrolled will of the official. In that space, public officials still dominate citizens.

Of course, Pettit (2012) denies that such discretionary power reintroduces domination. He argues that discretionary power is unproblematic if: (1) The decision is contestable and any abuse is punishable (pp. 213ff.); (2) The decision-maker is trusted based on 'local standards of when trust is well-placed' (p. 176); (3) The laws that constrain decision-makers are under the equal influence of citizens (p. 210); and (4) The laws are based on reasons all can accept (or see as relevant) (pp. 253ff.). Under these conditions, Pettit argues, discretionary power is not objectionable. If you don't like the decisions made under such conditions, that's just 'tough luck' (p. 176). It is not a 'sign of a malign will at work against you or your kind' (p. 229). Nor is it 'due to the special influence of those who are richer or electorally better placed or closer to the corridors of power' (p. 177).

Despite Pettit's arguments, we think that discretion – even under these conditions – can leave people at the mercy of the arbitrary will of others. Consider the 'reasonable suspicion' standard for police searches. Suppose the standard is adopted under the equal influence of all citizens, based on reasons all can accept and includes a mechanism for people to contest any decisions made under it. But because the standard is vague, authorities will often defer to the police's decision in applying it. Even if police officers are trusted by local standards, there is a sense in which normal citizens are still dominated. We suspect that anyone who has been searched without their consent will agree. Officers have broad discretion to search

and so their decisions represent intentional choices made by people who are 'closer to the corridors of power.' One might have no reason to suspect their 'malign will,' but the worry about domination is about being subject to the will of others, whether good or ill. Even under the democratic rule of law, discretionary power remains problematic.[1]

Furthermore, even if we suppose that human discretion is not problematic when Pettit's conditions are met, there is still the fact that his conditions are hard to satisfy. In particular, securing trust in agents of the government is an especially thorny problem, as the example above illustrates. For this reason, we might have serious practical worries about human discretion in the administration of the law, even if we recognize that it is theoretically possible to render such discretion compatible with non-domination.

Pettit may respond that the very inevitability of official discretion renders it legitimate. If discretionary power is necessary for effective administration, then it must be consistent with non-domination. If the laws were not administered effectively, after all, then citizens could freely dominate each other, which would be worse than these circumscribed instances of state domination.

Human discretion is not inevitable, however. There is an alternative: rule by automation. Instead of entrusting a human to apply the rules, rule by automation allows the rules to 'apply themselves.' For instance, a loan application system can consult a standard of credit-worthiness, compare it to your inputs, and decide whether you should get a loan – all without the need for human decision making. We could rely on a similar system for adjudicating laws. The system would identify the appropriate law, see whether the rule applies to this situation, and then render a decision.

Automating decisions reduces the concern with domination in all three aspects we have described. First, automated systems are not agents, and have neither wills nor intentions, so they cannot dominate by Pettit's conception. We have therefore replaced a human agent's power with a system that cannot dominate us, which should be an improvement in terms of non-domination. Second, an

1 Some argue that the discretionary power of public officials is desirable, even for neo-republican theorists. Sharon (2016), for instance, argues that discretion is desirable because it allows for the influence of experts and the application of abstract laws to concrete cases. However, these are arguments that discretion will lead to better decisions, not that it promotes non-domination. We would argue that discretion still places others under one's arbitrary will, even if it leads to better decisions in certain cases. Compare with Douglas's dissent in Terry v. Ohio ('To give the police greater power … is to take a long step down the totalitarian path. Perhaps such a step is desirable to cope with modern forms of lawlessness … Yet if the individual is no longer to be sovereign, if the police can pick him up whenever they do not like the cut of his jib, if they can "seize" and "search" him in their discretion, we enter a new regime.')

automated decision system would adhere to rules without leaving room for discretion. This would not only reduce dependence on others' wills, but also improve consistency and predictability – values that neo-republicans prize (Lovett 2012). Decisions would no longer vary from one official to another and, if the system were sufficiently transparent, we could better anticipate the results. This would allow citizens to know, in advance, the rules that they must comply with and to build a life around them. Third, automated decisions would better serve the goals of the eyeball test. There is no possibility that your fate depends on a system's mood or whether you showed it adequate deference. Under this interpretation of non-domination, then, Pettit and other neo-republicans should support replacing the discretionary powers of public officials with automated systems.

2.1 A Tool for Human Domination?

One might object to the claim that neo-republicans should support rule by automation by arguing that automated systems are just tools of human power, which merely transmit the wills and intentions of those who design or implement them. If officials have a certain discretion in making their decisions, then they can use that discretion to implement an automated system that shares their priorities. Whatever decision that system makes, one might argue, is just an extension of the discretionary and arbitrary will of those officials.

This objection parallels a similar objection one might make to the rule of law. Laws might serve as mere tools of power which simply transmit the will of legislators. If law makers have sufficient discretion in the writing of the law, then those laws do not prevent domination but enable it. Sometimes this possibility – where laws serve as tools of domination – is called rule *by* law (Tamanaha 2004).

There is, however, one sense in which rule by law is better than the direct rule of officials with discretionary power: criteria are laid out in advance and (presumably) do not change from one application to the next. Even though the laws might reflect the will of the legislator, they do not reflect the legislator's will in any particular case. From the perspective of non-domination, reducing case-by-case discretion in this way is already a significant improvement. Consider, for instance, this passage by neo-republican Frank Lovett:

> After the introduction of public rules, the situation has indeed changed, and in an important way. Members of the subordinate groups can now at least know exactly where they stand: they can develop plans of life based on reliable expectations; provided they follow the rules, they need not go out of their way to curry favor with members of the powerful group; and so on…. These are important experiential differences, best captured by saying that the

> introduction of externally effective constraints on the holders of power itself constitutes a reduction of domination, and something to be desired. (2012, p. 147)

So, one response to this 'tools of human power' objection is to say that, even if automated systems merely transmit the will of the officials who put the system in place, it's better to be ruled by their generalized and abstract decisions about how to automate the application of the law than to be ruled by the decisions that officials render in particular cases.

But many neo-republicans would not look favorably on rule by law, nor would they accept rule by automation that merely transmits the will of the administrators. To create laws that truly promote the value of non-domination, Pettit requires that citizens must have an equal and individualized control over the law-making process (2012, p. 187). Control has two elements: influence and direction. Equal influence means, roughly, giving citizens equal votes in competitive elections (2012, p. 210). Individualized influence, however, adds the additional requirement that individuals or groups must be able to contest public decisions to avoid being dominated by the majority will (2012, pp. 213ff.).

The directionality requirement adds another demand on public decision-making. Pettit argues that an equal opportunity to influence representatives does not give citizens control over the direction of state power because the representatives all have a different mandate. Each representative will reflect the priorities of their constituency, and then negotiate those priorities with other representatives. This process gives representatives too much discretion in the final decision, which may not reflect the priorities of the electorate as a whole (2012, p. 246).

However, if each participant relies only on reasons that are acceptable to all citizens, then the final decision will at least be acceptable to all citizens, even if it is not optimal for many. This acceptability requirement gives each citizen as much control as is consistent with equal control for all. Every citizen has equal access to deliberate on how to use state power and the deliberation reveals reasons that all citizens can accept. Those reasons are then the basis for the making and administration of the laws, giving citizens equal and individualized influence over the direction of public decisions.

Importantly, however, non-domination does not require that citizens actually deliberate on each decision (2012, pp. 267f.). All that matters is that decisions are made based on acceptable reasons, known through deliberation at sites of 'opposition and contestation,' such as electoral campaigns and judicial hearings (2012, p. 261). This allows for the making and administration of the law without the deliberation or influence of citizens on each decision.

Taken together, then, laws do not transmit the discretionary will of their creators when they are made by representatives chosen by the equal votes of citizens,

when individuals and groups can contest them, and when they are based on reasons that all citizens can accept.

Some automated decision-making systems would not meet these standards. Consider, for instance, a bank that implements an algorithmic system to decide who should get a loan. If we imagine that the bank officials are not chosen under the equal influence of citizens, customers cannot contest its decisions, and the decisions are not based on reasons that all could accept, then the algorithm would simply transmit the will of its creators. The system merely encodes the bankers' interests, which, for Pettit, means this automated system would fail to promote non-domination.

However, rule by automation could reduce domination under the right conditions. Like the laws, the automated system could be implemented by the people's representatives (or those they delegate authority to). Its decisions could be contestable by citizens, whether by appeal to another automated system or to a human authority. And it could be designed to only give weight to factors all can accept. Under these conditions, the automated system would not transmit the discretionary will of its creators or implementors.

Of course, one could argue that the ability to contest the decision by appeal to a human authority merely reintroduces human discretion into the decision. That is true, and that may be a reason to extend automation into the contestation of decisions as well. But even if automation is only used at one level of the decision-making process, it still serves the purpose of reducing human discretion and therefore domination. An initial decision still has consequences, even if those consequences are later reversed, so it matters that the decision is made without dependence on an arbitrary human will.

Automation is therefore not only a more complete realization of the rule of law; it is also a more complete realization of the requirement for equal influence and control over state power. Equal influence requires administrators to execute the will of the people, without giving any weight to their own will. Equal control requires administrators to decide based on acceptable reasons, without using those reasons as cover for their own agenda. Lacking a will or an agenda, automated systems *are* merely a tool; but they can be a tool of public purposes. Designed properly, they can be the ideal public servants.

2.2 Domination Without a Will?

A second objection is that domination does not necessarily involve subjection to another's will, so removing that will is not necessarily an improvement. Recent versions of neo-republican theories have argued that there can be domination even

without a dominator – without the presence of an agent who can intentionally interfere with one's choices. Dorothea Gadeke (2020) offers the example of a mugger in a park in a sexist society. A mugger has the capacity to interfere with both men and women in the park. So, if domination is just a capacity to interfere, then both are equally dominated. However, she argues that, 'male visitors to the park are not dominated, even if the gunman could also overpower them. But if he did, he would be effectively sanctioned; so his power does not express an asymmetry of standing vis-à-vis male victims' (p. 208).

What actually makes a difference to domination, then, is the laws that create an asymmetry of standing, whether or not a mugger walks into the park. Gadeke calls this 'systemic domination,' which is the 'systematic disempowerment the dominated suffer over and beyond their relation to a particular dominator' (p. 200). If Gadeke's interpretation of non-domination is right, then domination does not require agents or intentions and neo-republicans would have no special reason to favor rule by automation.

Other neo-republicans have similarly argued that non-domination is about equal status and not about controlling the power of other agents. For instance, Vincent Chiao reinterprets the paradigm example of domination between slave master and slave as an issue of status rather than of controlling power.

> What we should object to in [the master/slave relationship] is not what the law does (fails to protect the slave from the possibility of unilateral interference), but rather what the law says (that the slave is not the social equal of the master). The law describes one person as the subordinate of another and is objectionable for that very reason, not because it leaves one person subject to the mere possibility of interference by another. (Chiao 2016, p. 103)

At times, Pettit also acknowledges the importance of equal status or standing. He writes, 'to enjoy the relevant freedom of nondomination is to be someone who commands a certain standing amongst your fellows' (2012, p. 91).

If domination is about unequal status rather than about being under the power of an arbitrary will then, as theorists like Gadeke argue, systems can dominate us. Systems dominate when they express or assign someone a subordinate status. One can see the difference between the two senses of domination in the contrast between the slave master example and another paradigm of domination: a caste system. In a caste system, domination occurs as much through inequalities of status as through the capacity to interfere with other citizens. The problem is not that upper castes make the rules and rule based on their whims. It is rather that the rules in place assign the lower castes a subordinate role.

These alternative understandings of the value of non-domination all point to the importance of equal status among citizens, not just controlling the power of an arbitrary will. To understand the case for rule by automation on the assumption

that these alternative views are correct, we should turn to social egalitarianism: the tradition that emphasizes the importance of social equality.

3 Automation and Social Equality

Social egalitarians oppose social hierarchies, where some people are afforded greater standing, respect, consideration, power or access in virtue of their class, caste or social rank. An obvious source of social hierarchy is the state, where public officials are afforded greater power than citizens. What makes the greater power of public officials consistent with social equality? An important part of the answer is the rule of law. By looking at the reasons social egalitarians support the rule of law, we can see why they would also support rule by automation.

According to Gowder (2013), the rule of law has three features that promote equality. First, the rule of law requires regularity: the use of coercive force by agents of the state must be reliably constrained by good faith and reasonable interpretations of pre-existing, reasonably specific, legal rules. Second, the rule of law requires publicity: the rules that authorize coercion must be promulgated; citizens must be offered reasonable explanations for the application of the rules in particular cases;[2] and those subject to state coercion must be able to offer arguments about the applicability of those rules to their particular cases. Third, the rule of law requires generality: neither the rules authorizing coercion, nor the interpretations of those rules by agents of the state can make irrelevant distinctions between those subject to the law. (A distinction is irrelevant when it is not justified by public reasons, in the Rawlsian sense.) To the extent that laws realize the regularity and publicity requirements, Gowder argues, they prevent social hierarchies between citizens and agents of the state. And to the extent that laws realize the generality requirement, they prevent social hierarchies among citizens.

Looking closely at Gowder's requirements, we can see how the ability of the rule of law to promote social equality depends, not just on the content of the law, but also on the way the laws are interpreted and enforced. The regularity of the law depends on human officials making good faith and reasonable interpretations of the rules. Those that make such interpretations will uphold the rule of law. But those that make bad faith interpretations to serve their own ends claim greater power over us, creating a social hierarchy between themselves and citizens. Similarly, generality depends on humans relying only on public reasons in

2 The explanatory requirement is real, but perhaps not as extensive as some have thought. See Cohen (2015).

drawing distinctions. Those that fail to do so, or do so inconsistently, will contribute to social hierarchy among citizens.

Rule by automation can alleviate these dangers. An automated system will make reasonable or unreasonable interpretations depending on its design, but it will never act in bad faith. It has no other ends to serve. Similarly, if generality depends on adhering to public reasons, an automated system can be reliably constrained to a set of acceptable factors in a way that a human cannot. By reducing the role of humans in interpreting and enforcing the law, then, automated systems promote social equality.

We can see another important way in which automation promotes social equality by looking at the creation of the laws. Social equality requires that everyone have an equal opportunity to influence the making of the laws. So, if one group has more influence than others, there is an objectionable social hierarchy.

For Kolodny (2014), this concern should lead us to favor 'merely equal' over 'positive' democratic procedures. Merely equal procedures would be compatible, for instance, with a lottery. We all have equal influence over the outcome by having no influence at all. Positive procedures require, instead, that people have some input into the decision-making process. Different contexts may call for one or the other kind of democratic procedure; lotteries might be the right procedure for a military draft, while a vote might be the best way to choose an executive. But in general, we have a reason to prefer merely equal procedures over positive procedures: positive procedures introduce the possibility of some interest capturing the process and thereby creating an objectionable social hierarchy.

For this reason, Kolodny writes favorably of the fact that many of our laws are inherited, an aspect of the rule of law that he calls rule by the 'dead hand of the past' (2014, p. 312). Because we have no ongoing social relations with the dead, there is no question of them being our social superiors. So, there's a sense in which inherited laws are better than laws created by our contemporaries; the *merely equal* rule of the dead avoids the risk of social hierarchy that positive procedures allow.

There is, of course, no possibility of the dead enforcing or adjudicating the law. But we could achieve the same benefit by automating that process. Just as there is no question of the dead being our social superiors, there is no question of automated systems being our social superiors (Kolodny 2019, p. 107, 112). So, for the same reasons that social egalitarians support the rule of law and especially the rule of inherited laws, they should also support rule by automation.

4 Unease with Automation

We've argued that relational theories should favor rule by automation for the very same reasons they favor the rule of law. When we reduce human discretion in the administration of the law, we cut off a source of dominating and hierarchical social relations. We are inclined to agree with these theories; we think that rule by automation has an important role to play in creating a free and equal society. But we also suspect that anyone who imagines, for instance, being convicted of a crime and sentenced to a prison term by some automated system will feel a profound sense of unease. Even if we imagine the sentence to be perfectly just, there is something uncomfortable about this impersonal process.

With some notable exceptions, the relational theories we've discussed lack the resources for vindicating this unease. This is because these theories tend to focus on the elimination of objectionable social relationships. But a valuable society consists of more than just the absence of domination and social hierarchy – it requires the presence of certain kinds of politically valuable social relations as well. Pettit comes close to expressing this idea when he writes,

> To enjoy this freedom presupposes relationships with others and consists in relating to them on a pattern that rules out domination. It requires the absence of domination, not as such, but in the presence of relationships that make domination saliently possible and non-domination correspondingly desirable. (Pettit 2012, p. 91)

Elizabeth Anderson also recognizes the importance of not just preventing negative social relationships, but promoting positive ones as well. She says that

> Negatively, egalitarians seek to abolish oppression – that is, forms of social relationship by which some people dominate, exploit, marginalize, demean and inflict violence upon others … Positively, egalitarians seek a social order in which persons … live together in a democratic community. (Anderson 1999, p. 313)

Building off this idea about the importance of promoting politically valuable human relations, we might ask: is there any aspect of the relationship between a sentencing judge and a person convicted of a crime, or between a police officer and the citizen they may or may not choose to search, or between any human administrator of the law and those subject to their discretionary power, that constitutes a 'democratic community'? Do these relationships in any way contribute to non-domination and social equality? What is it that makes some of us uncomfortable being sentenced by a machine? In the final section, we explore possible sources of this unease by looking at multiple dimensions of 'role reversibility.'

5 Role-Reversibility

'Role-reversibility' is both a classical republican and a neo-republican value, drawn from Aristotle's conception of democracy as ruling and being ruled in turn. In its classical form, role-reversibility refers to the requirement that those who make the law should also be subject to it. This is supposed to serve as a check on power because when decision-makers are subject to their own decisions, they are more likely to make choices that are either beneficial or not especially harsh to those that are subject to them. After all, next time, they could be on the other end of that decision (Pettit 2012, p. 204). This kind of role-reversibility might explain some of our unease with rule by automation, since automated systems are not, in any meaningful sense, subject to their own decisions.

However, in many societies, cases of actual role reversal are relatively rare. Judges and administrators often occupy their positions for long periods and make decisions for people who are much less fortunate or well-connected than themselves, or people in circumstances that they are unlikely to be in. So, while this kind of role-reversibility might explain part of our unease with rule by automation, it won't be the whole story, unless we're willing to accept the implication that a great deal of current human-administered law is similarly problematic.

Perhaps another kind of role-reversibility can help. Even if decision makers will never actually find themselves in the position of people subject to their decisions, there is still something valuable about the fact that they *might have* had their roles reversed. Brennan-Marquez and Henderson argue that the mere *possibility* of role reversal helps to constitute a democratic community. When both parties know that their roles could have been reversed, they can see the decision, not as the product of anyone's will, but as issuing from a democratic community. They write,

> Role-reversibility enables decision-makers to respect the gravity of decision-making from the perspective of affected parties. This, in turn, allows the act of judgment to be understood as the vindication of values shared by a broader moral community—a community of equals that includes both the decision-maker and the affected party, as well as many other people who were not involved in the decision but equally might have been, and who, in any case, share responsibility for the decision's consequences … It is not merely that a decision-maker should be able to imagine accepting the same judgment in reverse. It is that she should be able to say: this decision is an outcome of democracy; it reflects a constraint we have decided to impose on ourselves, and in this case, it just so happens that another person, rather than I, must answer to it. (Brennan-Marquez and Henderson 2019, 140ff.)

We agree it's important that decisions about the application of the law can by understood as 'the vindication of values shared by a broader moral community.'

And we also agree that if Brennan-Marques and Henderson are right, then rule by automation is problematic, since we can't imagine reversing roles with an automated system. So, this is one way to understand the value of human administration of the law: only where those who apply the law are sufficiently like us can we imagine swapping roles with them and thereby seeing their decisions as issuing from a democratic community.

However, we're doubtful that the mere possibility of switching roles is sufficient. We could, for instance, imagine a society where people are selected at a young age, via some random process, to be lifetime rulers with wide discretion in interpreting the law. In such a society, rulers and ruled *could have* had their positions reversed. Nevertheless, we would have a hard time understanding their decisions as an outcome of democracy.

Brennan-Marquez and Henderson might admit as much. On their view, role-reversibility comes in degrees. In our imagined world, there is less democratic community than in a world where people move in and out of powerful decision-making roles. But with the exception of Brennan-Marquez and Henderson's favored example – trial by jury – there are few places in government that enjoy the full measure of role-reversibility. So again, if we explain our unease with rule by automation with this kind of role-reversibility, we'll have to admit that there is something similarly problematic about much human-administered law.

There is another kind of role-reversibility that we think is important. It's not that the administrators of the law will actually find themselves in the place of those who are subject to their decisions, and it's not the mere possibility that their role and ours *could have been* swapped. Instead, a more abstract kind of role – as evaluator and evaluated – is actually reversed.

In applying the law to our case, human administrators evaluate us and our actions. But we are able to evaluate the decision-maker in turn. We are able to direct various kinds of evaluative attitudes towards them: blame, censure, praise, critique, approval, reproach, etc. This aspect of our relationship with public officials furthers equality in two ways. First, the appropriateness of directing evaluative attitudes toward decision-makers is, we suggest, part of relational equality. Consider an extreme alternative: a caste system. In a caste system, the inequality is not only manifested in the concentration of power with the upper castes. It is also reflected in the fact that it is inappropriate to blame or resent them or their decisions. The upper castes are not responsible for their superior power, which is given to them by nature. And because of their superiority, their decisions are beyond criticism or reproach. They can evaluate the lower castes, but the lower castes cannot felicitously evaluate them. This absence of *evaluative* role-reversibility is yet another way in which the norms in a caste system express the inequality of the lower castes.

Evaluative role-reversibility also promotes equality by fostering democratic community between rulers and ruled. In the lawmaking arena, this occurs when representatives make decisions for the public and the public, in turn, evaluates their representatives both informally in public discussion and formally during elections. In the administration of the law, however, the ruled have different forms of democratic influence over their rulers. One form of influence, as Pettit argues, is the ability to contest decisions made about us, by appealing or simply criticizing those decisions.

But there is another, more subtle form of influence that we have over administrative decisions: the ability to direct reactive attitudes toward the decision-makers and the expectation that they will be received. Reactive attitudes are moralized forms of evaluative attitudes that target the quality of will of the other person (Strawson 1962). Our ability to deploy such attitudes allows for a form of influence that issues from our emotional entanglement with other humans. Human officials are susceptible, not just to our criticism, but to our blame and resentment as well. This represents a valuable kind of relationship where we, in part, help to negotiate the details of how the law will apply in our case and where we and those who administer the law share a mutual – even if not perfectly equivalent – kind of vulnerability. In this sense, we achieve non-domination, not by avoiding vulnerability to another's will, but by making that vulnerability reciprocal. While public officials might be our political superiors, owing to their outsized influence on the application of the law, they are our moral-evaluative inferiors: their actions are our business, and they are subject to our blame or censure. We are, in a sense, their subjects; but they are our public servants. This kind of morally loaded evaluative role-reversibility might, we suggest, be an important part of democratic community and therefore of relational equality.

We don't claim that *moral-evaluative role-reversibility* is the whole of freedom or equality, only that it may be a necessary part. Consider, for example, the way most marriages involve moral-evaluative role-reversibility. Each spouse generally cares deeply about the quality of will of their partner; they are ready to deploy and are susceptible to each other's reactive attitudes. But, as the history of marriage shows, this situation is compatible with extreme inequalities and forms of domination. Nevertheless, moral-evaluative role-reversibility might still be essential. Imagine a marriage where partners have equal influence over important decisions, are able to contest those decisions, and where decisions are based on reasons each partner can accept, but where the partners are not ready to deploy and submit to one another's reactive attitudes. Even if they have some forms of influence over each other's actions and can criticize or complain, they are lacking an important kind of emotional entanglement. There seems to be something essential missing from this 'domestic community.'

Rule by automation can achieve evaluative role-reversibility but not moral-evaluative role-reversibility. Moralized evaluations that target the quality of will of other people (e.g. blame, resentment, reproach) are not the sorts of attitudes we can felicitously direct at an automated process.[3] But less moralized forms of evaluation (e.g. criticism, appraisal, assessment) are. Which kinds of evaluations are important for achieving a democratic community? We're not entirely sure. Maybe it is enough if we can criticize or complain about an automated system, even if we can't blame or resent it. Society, after all, isn't like a marriage.

On the other hand, moralized forms of mutual evaluation – the deployment of and susceptibility to reactive attitudes – lend a richness and texture to social life that might be required for democratic community. Consider the difference between being able to complain about the decisions of low-level bureaucrats and being able to blame them. Once you recognize that the bureaucrat is not responsible for the decision – that they were just following rules that left them no room for discretion – you may still complain or criticize, but you don't have anyone to blame. Anyone who has experienced the frustration of a bureaucracy can understand the sense in which not having someone to blame reduces our sense of control and community.

Of course, you could blame the system designer, just as you could blame the bureaucrat's boss or boss's boss. Some human is ultimately responsible for laying down the rules that automated systems and bureaucrats apply to your case. However, the more the designer is removed from the actual decision, the more they can claim that they couldn't foresee all the consequences of their design decisions. In some contexts, and with some systems, this claim of innocence will be justified. In other cases, it will be a poor excuse. Whatever is ultimately true about how responsibility resolves in these complex scenarios, rule by automation makes it more difficult to find targets of our reactive attitudes who will feel their full force, who will be directly vulnerable to our blame, resentment and reproach, and who will take on the complex emotional negotiation that occurs when agents of the law have discretionary powers. Rule by automation stretches and strains our capacity to hold each other responsible and thereby reduces both the expressive value of the reactive attitudes and their ability to influence the decisions made for us.

6 Conclusions

We have argued that relational theorists – including neo-republicans who worry about agent-domination and social egalitarians who worry about social

3 This claim would be false if we imagine rule by automation as implemented by highly advanced AI that has something equivalent to a will. But this scenario is remote from our present concerns.

hierarchies – have strong reasons to support rule by automation. Automated systems serve the same goals as laws, but avoid the discretionary power that remains under the rule of law. By eliminating human discretionary power, rule by automation reduces our dependence on the intentional and arbitrary will of other agents and reduces social hierarchy. Rule by automation thus represents a powerful tool in our goal of achieving a free and equal society.

Nevertheless, we recognize that there is a deep unease with rule by automation. Republican and social egalitarian thinkers often focus on the elimination of objectionable human relations and therefore have trouble vindicating this unease. We suggest that the unease is rooted in a lack of moral-evaluative role-reversibility, which is an important feature of democratic community. So, if there is any kind of freedom-and-equality based objection to rule by automation that relational theorists can endorse, it will depend on embracing the importance of democratic community and accepting, for some moralized notion of 'judgement,' the Nat King Cole snowclone: 'the greatest thing society will ever learn, is just to judge and be judged in return.'

Acknowledgements: The authors would like to thank the Institute for Practical Ethics (IPE) at UC San Diego for its support. We would also like to thank John Evans, Craig Callender, Ava Wright, Robert Wallace, participants at the IPE Workshop and two anonymous reviewers for helpful comments.

References

Anderson, E. S. 1999. "What is the Point of Equality?" *Ethics* 109 (2): 287–337.

Brennan-Marquez, K., and S. Henderson. 2019. "Artificial Intelligence and Role-Reversible Judgment." *Journal of Criminal Law & Criminology* 109 (2): 137–64.

Chiao, V. 2016. "Discretion and Domination in Criminal Procedure: Reflections on Pettit." *Politics, Philosophy & Economics* 15 (1): 92–110.

Cohen, M. 2015. "When Judges Have Reasons Not to Give Reasons: A Comparative Law Approach." *Washington & Lee Legal Review* 72 (2): 483–571.

D'Amato, A. 1977. "Can/Should Computers Replace Judges?" *Georgia Law Review* 11 (5): 1277–302.

Gädeke, D. 2020. "Does a Mugger Dominate? Episodic Power and the Structural Dimension of Domination." *Journal of Political Philosophy* 28 (2): 199–221.

Gowder, P. 2013. "The Rule of Law and Equality." *Law and Philosophy* 32: 565–618.

Hart, H. L. A. 2012. *The Concept of Law*. Oxford: OUP.

Kolodny, N. 2014. "Rule Over None II: Social Equality and the Justification of Democracy." *Philosophy & Public Affairs* 42 (4): 287–336.

Kolodny, N. 2019. "Being Under the Power of Others." In *Republicanism and the Future of Democracy*. 1st ed., edited by Y. Elazar, and G. Rousselière, 94–114. Cambridge: Cambridge University Press.

Lovett, F. 2012. "What Counts as Arbitrary Power?" *Journal of Political Power* 5 (1): 137–52.

Pettit, P. 1999. *Republicanism*. Oxford: Oxford University Press.
Pettit, P. 2012. *On the People's Terms: A Republican Theory and Model of Democracy*. Cambridge: Cambridge University Press.
Sharon, A. 2016. "Domination and the Rule of Law." In *Oxford Studies in Political Philosophy*, Vol. 2, edited by D. Sobel, P. Vallentyne, and S. Wall, 128–55. Oxford: Oxford University Press.
Strawson, P. 1962. "Freedom and Resentment." *Proceedings of the British Academy* 48: 1–25.
Tamanaha, B. 2004. *On the Rule of Law: History, Politics, Theory*. Cambridge: Cambridge University Press.
Terry v. Ohio. 1968. 392 U.S. 1, 88 S. Ct. 1868, 20 L. Ed. 2d 889 (Douglas, J. dissenting).