

Why not Rule by Algorithms?*

Abstract

The rise of artificial intelligence (AI) poses new and pressing challenges to society. One such challenge is the increasing prevalence of AI systems in political decision-making, which is often considered as a threat to democracy. But what exactly is lost when certain aspects of political decision-making are handed over to AI systems? To answer this question, I discuss an extreme case in which all political decisions are made by intelligent algorithms that function without human supervision. I will call this case Rule by Algorithms. I consider the epistocratic argument for Rule by Algorithms according to which as long as algorithms can be expected to produce better outcomes than human rulers, we have a good reason to abandon democracy for algorithmic rule. Some authors attempt to resist such conclusions by appealing to the notion of public justification. I argue that these attempts to refute Rule by Algorithms are ultimately unsuccessful. I offer an alternative argument according to which Rule by Algorithms should be rejected because it imposes impermissible constraints on our freedom. The discussion of this extreme case provides valuable insights into the challenges of AI in politics.

Keywords: artificial intelligence, democracy, epistocracy, public reason, domination, freedom

* An earlier version of this paper was given at the 2019 workshop *Artificial Intelligence: Philosophical Issues*, part of the *Action and Context* series organized by the Department of Sociology and Communication, Budapest University of Technology and Economics (BME) and the Budapest Workshop for Language in Action, Department of Logic, Institute of Philosophy, Eötvös University (ELTE). This research was supported by the Higher Education Institutional Excellence Grant of the Ministry of Human Capacities entitled *Autonomous Vehicles, Automation, Normativity: Logical and Ethical Issues* at the Institute of Philosophy, ELTE Faculty of Humanities.

I. INTRODUCTION

Artificial intelligence (AI) systems exert a profound impact on today's society.¹ They fundamentally transform commerce, travel and communication, culture and learning. In recent years AI also started to shape government and public policy. Although one may argue that AI technologies contribute to the efficiency of political decision-making, many view the increasing prevalence of AI and automation in political decision-making as a threat to democracy. In this paper I discuss this problem by considering an extreme case. Imagine that at some point in the future intelligent algorithms that are able to function without human supervision completely take charge of political decision-making. What exactly would be objectionable about such an arrangement? What values would *Rule by Algorithms* undermine? To answer this question, I present in the first section the epistocratic argument for Rule by Algorithms, according to which if decision-making algorithms can be expected to make far better decisions than humans, then we have a good reason to replace democracy with Rule by Algorithms. Some authors might try to resist such a conclusion by appealing to arguments from *public reason liberalism*. In the second section I discuss this type of argument and conclude that it is ultimately unsuccessful. In the third section I present an alternative consideration against Rule by Algorithms, based on the concepts of freedom and domination.

II. THE EPISTOCRATIC CASE FOR RULE BY ALGORITHMS

It is widely accepted that political decisions ought to be made *democratically*, i.e. either directly voted on by citizens or authorized through some such vote.² Yet many argue that democracy should not enjoy this default status, and other forms of government might be preferable to it. One such proposed alternative is epistocracy, i.e. rule by experts or knowers. Its advocates argue that those who have expertise that pertains to political decision-making are in a better position to make *good* decisions than those who lack such expertise. Therefore, if political decision-making were left only to experts, its quality could be expected to increase. And, as Jason Brennan, one of today's leading advocates of epistocracy, argues, citizens have a fundamental right to competent government (Brennan 2011). Therefore, insofar as epistocracy can be expected to be more competent than democracy, we have a pro tanto obligation to replace the latter with the former.³

¹ For a discussion on the definition of, current research on AI as well as its potential future impact on society see Russell and Norvig (2016) and Boden (2016).

² For a discussion on the definition of democracy, see Waldron (2012) and Goldman (2015).

³ For a more detailed discussion on epistocracy see Gunn (2014), Brennan (2016a), and Moraro (2018).

This epistocratic argument, if successful, provides a *prima facie* case for introducing a form of government in which political decisions are made exclusively by extremely intelligent and competent AI systems. If some AI systems could be developed which reliably emulated those cognitive and deliberative faculties by means of which humans make political decisions – except in a much faster, more accurate, and more efficient manner – and if these AI systems thereby attained a level of competence unavailable to humans or groups of humans, whether a democratic public or an expert panel, simply because of their natural limitations, then the epistocratic argument – being focused exclusively on the quality of political outcomes – would favour these AI systems as rulers over any human, expert or not.

One can conceive of Rule by AI in many different ways. For example, one may imagine that AI experts at one point create a superintelligent machine which then establishes itself as the robotic overlord of society (Bostrom 2014, 95). While these scenarios of superintelligent artificial dictators may be interesting, here I would rather focus on a different, no less fanciful, but perhaps somewhat more relevant case. AI systems, particularly machine learning algorithms are already in use in many areas of government and public policy. Such algorithms already support policymaking through data mining, they help optimizing the provision of public services, they provide risk-assessment data for criminal sentencing, they control traffic lights, and carry out many more tasks previously done by humans (Wirtz–Weyerer–Geyer 2019; Oswald 2018; Lepri et al. 2017; Coglianese–Lehr 2017). Suppose that as these algorithms become more sophisticated and efficient, we gradually hand over more and more tasks to them until all aspects of legislation, government, and perhaps even judicial tasks are handled by intelligent algorithms without human supervision. I call this case Rule by Algorithms.

I discuss Rule by Algorithms not because I believe that it can become reality anytime soon. My goal, rather, is to gain insight into the way in which the fundamental values of democracy can come into conflict with the increasing prevalence of AI systems in society and politics, and today the relevant type of AI system is closer to a machine learning algorithm than to a superintelligent digital dictator. Furthermore, certain core features of Rule by Algorithms are particularly interesting in comparison with the digital dictator scenario, as it will become clear in later sections.

There are a few assumptions I make about Rule by Algorithms here for the sake of the argument. First, I assume that Rule by Algorithms can be expected to produce significantly better outcomes than human decision-makers; otherwise the question of its preferability to democracy would not even arise. Second, I assume that ruling algorithms do not form a coherent mind or an *artificial person*⁴

⁴ Here the term “artificial person” does not refer to the legal concept under which corporations and the like also count as artificial persons, but rather to a human-made AI system with

with its own interests, desires, volitions, beliefs, and so on. Rule by Algorithms, therefore, does not mean handing over power to a robotic overlord, but rather to a cluster of intelligent algorithms each carrying out various tasks pertaining to decision-making. The cooperation of these various algorithms emulates the way in which ordinary decision-makers produce outcomes, without constituting a coherent mind; in roughly the same way as various algorithms today (e.g., those used by social media sites or other online platforms) govern much of our lives without necessarily congealing into a single artificial patriarch overseeing our activities.⁵

Third, I assume that the algorithms would be sufficiently independent of their makers not to be thought of as mere tools in the hands of those who create them. Clearly, some human involvement is necessary for setting up and running Rule by Algorithms; someone has to make them, maintain them, etc. But for the scenario to be even worthy of discussion, the algorithms must be conceived of as being able to function on their own to a great extent, without human supervision. Their makers and users cannot have control over or ability to predict the outcome of the functioning of the algorithms in a precise manner.⁶ This assumption is crucially important to distinguish Rule by Algorithms from AI-enhanced epistocracy, or from Rule by Software Engineers.

Assuming, then, that such algorithms could take over political decision-making, should we let them? One may argue that Rule by Algorithms is impossible. It requires human-level AI or Artificial General Intelligence (AGI), which according to many authors cannot be constructed (Boden 2016. 153–155). However, even if AGI is impossible – which it may not be (Turner 2019. 6n19) – it is not immediately clear that Rule by Algorithms requires AGI. A further argument is needed to show that to emulate all the deliberative faculties we use in political decision-making requires the artificial reproduction of the human mind in its entirety. Such a claim cannot simply be presupposed. And, in fact, it seems that in many areas of political decision-making which call for solving coordination problems and allocating resources efficiently – non-AGI-type – algorithms could be expected to do as good if not a better job than humans.

It is true, however, that there is more to political decision-making than solving coordination problems. Government also involves setting long-term goals and settling hard questions of value. But algorithms, the objection goes, could not do this on their own; such goals and core values would have to be ultimate-

all the features that constitute personhood in ordinary humans, e.g., a mind, the capacity for rational deliberation etc.

⁵ One may object that any such cluster of algorithms would be bound to constitute a coherent mind and ultimately an artificial person. I will assume without argument that this is not the case, acknowledging that if it were, my account would have to be adjusted accordingly.

⁶ For more on such algorithms and the ethical issues concerning them see Mittelstadt et al. (2016).

ly supplied by humans. Note, however, that the same is true of human decision-makers. Humans do not conjure long-term goals and values out of thin air; we are socialized by other humans, our reflections on values and goals start with material we receive from parents, teachers, and society in general. Still, as long as we operate on this material in a sufficiently independent manner, we can be thought of as making our own decisions. Similarly, perhaps humans would supply initial material on which algorithms operate for setting goals and making value-judgements; but as long as they function sufficiently independently, emulating those deliberative faculties humans use for setting goals and settling questions of value – which, again, cannot be simply stipulated to be impossible – they may be thought of as ruling on their own.⁷ Thus, while human involvement would not be absent from Rule by Algorithms, ruling algorithms would not be mere pawns of any human being any more than human decision-makers are mere pawns of their parents or teachers.

This short discussion shows that there are no obvious reasons to discard Rule by Algorithms as in principle impossible. There may very well be nonobvious reasons, supported by further arguments, as well as reasons to think of it as practically unfeasible. Indeed, if, due to contingent circumstances, we never arrive at a level of technological advancement where Rule by Algorithms would be possible, reasonably inexpensive, and safe to implement, then introducing it in real life will never be an issue. But this does not affect the main argument of this paper, which is not about future scenarios for the use of AI in government, but rather the philosophical question of what kind of challenge, if any, is posed by AI to democracy. Rule by Algorithms is simply a hypothetical scenario which I use to draw out conclusions about this question; it can fulfil this role without ever being feasible in real life.

A final objection to consider is that the epistocratic argument presented above misunderstands the nature of political decision-making. It is false, one might claim, to say that there are better and worse decisions in politics, for political decisions are about values rather than facts, about clashes and compromises between antagonistic interests, rather than puzzles in social engineering where a solution can always be singled out as unambiguously optimal. Even if algorithms could emulate reasoning about goals and values, there is no sense in which their decisions could be better than those of humans, and therefore the epistocratic case for Rule by Algorithms evaporates.

This objection, again, relies on certain non-obvious premises which need to be argued for before the strength of the objection can be assessed. For example, even if in the case of certain value-judgements there is no way to tell which is

⁷ Recall, again, that no single algorithm needs to have the capability to do all this on its own. It is sufficient if the collective functioning of all the ruling algorithms emulates these deliberative faculties without congealing into a single artificial mind.

better, it seems rather implausible to say that *no* distinction between better and worse decisions can be made when it comes to society's final goals and basic values. There is a clear sense in which Nazi Germany's choice of basic values and final political goals were much worse than many alternatives. As authors such as David Estlund (1993), Susan Hurley (2000) and H el ene Landmore (2012) argued, any plausible conception of politics must accept at least some degree of *political cognitivism*, i.e., the view that some political decisions, e.g., ones that promote liberty and prosperity, are better than others, e.g., those that promote destitution and tyranny, as an epistemically accessible objective matter of fact. With these considerations in mind, what should we think of the epistocratic case for Rule by Algorithms?

III. PUBLIC REASON AGAINST RULE BY ALGORITHMS

Some defences of democracy against epistocracy are epistemic in nature. Proponents of *epistemic democracy* argue, for example, that democracy possesses epistemic merits that epistocracy would lack, and is therefore in a better epistemic position to identify good political outcomes.⁸ But since I assumed that Rule by Algorithms would outperform human decision-makers, the democratic answer to this challenge needs to be non-epistemic.⁹ A prominent line of non-epistemic arguments against epistocracy comes from the tradition of *public reason liberalism*. Public reason liberals, such as John Rawls (1993) and Gerald Gaus (1996; 2010), argue that the exercise of political power is legitimate only if it is justifiable to all reasonable points of view, i.e., justified on the basis of reasons that are accessible to all reasonable members of society.¹⁰

David Estlund puts forward one of the most well-known arguments for the claim that epistocracy cannot be justified to all reasonable, or, as he calls them, qualified points of view (Estlund 2008. 48). Epistocracy justifies the exercise of coercive political power by appealing to the better outcomes that experts are able to produce due to their epistemic superiority. But as Estlund's *demographic objection* holds, "it is not unreasonable or disqualified to suspect that there will be other biasing features of the educated group, features that we have not yet identified and may not be able to test empirically, but which do more epistemic harm than education does good" (Estlund 2008. 222). For example, the experts may all come from wealthy families or be members of an otherwise dominant so-

⁸ For more on epistemic democracy see Estlund (2003), Landmore (2012), and Peter (2016).

⁹ There are practical objections to standard epistocracy as well, which I cannot discuss in detail here (Viehoff 2016; Arneson 2009).

¹⁰ For more discussion on public reason liberalism, the criteria of reasonableness, public justification, and other related concepts see Chambers (2010) and Gaus (2015).

cial group which can make them biased in favour of their own group and against others. These distorting factors may detract from their ability to create good outcomes for everyone regardless of their epistemic superiority.

Estlund's argument is not that experts would surely produce biased outcomes. His argument is that it is not unreasonable to suspect that they would. It is also not unreasonable to reject this suspicion. Reasonable people can disagree about whether or not experts would produce the best outcomes. But precisely because this kind of reasonable disagreement is possible, the rule of experts cannot be justified to all qualified points of view potentially subjected to this rule by appealing to the consideration that experts would produce the best outcomes. Some could reject this consideration on reasonable grounds, and thus subjecting the population to the authority of experts would not be publicly justified.

One may argue against Rule by Algorithms in a similar way. Algorithms, however well they compute, can exhibit bias (Barocas–Selbst 2016; Howard–Borenstein 2018); therefore, one may argue that it is not unreasonable to suspect that although algorithms could produce good outcomes, the features that enable them to do so may travel with epistemically countervailing features that hinder this capacity. John Danaher (2016) formulates a related worry. He points out that algorithmic decision-making is often *opaque*, i.e. algorithms' decision-making mechanisms are not always transparent even to their makers or other experts. However, legitimate authority has a *non-opacity requirement*: decisions must be made based on reasons and principles that all reasonable or qualified citizens can endorse; if these reasons and principles cannot be accessed by citizens, not even in principle, then the decisions have no authoritative force and are illegitimate. For it is then never unreasonable for citizens to suspect that opaque decision-making mechanisms appeal to principles and reasons which they could reasonably reject (Danaher 2016. 251–252).

Note, again, that the argument does not presuppose that ruling algorithms are bound to be biased or to appeal to unacceptable reasons in their opaque decision-making. The argument only claims that it is not unreasonable to suspect that they would. Again, there may be reasonable disagreement on these worries. For example, reasonable people may argue that there are satisfactory safeguards against algorithmic bias which ultimately may even prove to be more successful in eliminating unfairness in political decision-making than any kind of human intervention (Zarsky 2011. 312; Zarsky 2016. 126). The point is that reasonable disagreement is possible on this matter, which undermines the legitimacy of Rule by Algorithms, as it undermines the legitimacy of standard epistocracy.

Are these arguments successful? I have my doubts. It certainly seems plausible that “as democratic citizens have the right to scrutinise and hold [to] account the exercise of political power, so algorithmic constituents have the right to scrutinise and hold account the exercise of algorithmic power” (Binns 2018. 553). But political power wielded by humans and algorithmic power wielded

by nonpersons without human supervision are fundamentally different. First, it is unclear if ruling algorithms would make and enforce rules in the same sense human holders of political power do, i.e., by creating authoritative directives which we have to obey. Algorithms on social media do not issue directives as to which advertisements we must watch or which news we must read. Rather, they shape our digital environment in such a way that we cannot help but act in certain kinds of ways (Yeung 2017). How similar this form of governance is to the traditional exercise of coercive political power is far from clear. But let us set aside this issue for the sake of the argument.

A more important difference between ordinary political rule and Rule by Algorithms is that the former requires that someone or some group be placed in a position of authority. In that position they are granted rights to treat others in ways which are *prima facie impermissible*. Normally we are not allowed to issue commands and coerce our fellow human beings to obey them. This would involve treating them not as moral equals, but as inferior beings subject to our private will. Only when the exercise of coercive power *by some persons over others* is publicly justified, is this threat averted. In other words, the demand of public justification is based on the fact that one person wielding coercive power over another carries an extremely high risk of moral injury, i.e., that of treating others as non-equals.

The reason why many authors endorse democracy is precisely because insofar as it distributes political power equally it does not threaten but rather affirms individuals' standing as equals (Christiano 2008). In a well-functioning democracy no one has the power to subject the polity to their private will, for citizens have an equal say in shaping political outcomes; in effect, no one rules over anyone (Kolodny 2014. 227). In contrast, non-democratic arrangements, e.g., epistocracy, as Estlund notes, "introduce an extra element of rule of some by others, and that element is subject to the qualified acceptability requirement, whereas its absence is not" (Estlund 2008. 219). Note, however, that the rule of some by others is also absent under Rule by Algorithms. In Rule by Algorithms the ruling is done by nonpersons, and therefore no one is threatened with being subjected to anyone else's private will.¹¹ Algorithms have no private will. For this reason, Rule by Algorithms seems to be on a par with democracy at least insofar as it also does not introduce the "extra element of rule of some by others".

This clarifies why I focus on the special case of Rule by Algorithms rather than Rule by AI more generally. If Rule by AI meant simply creating an artificial person with a single mind and extraordinary decision-making capabilities to rule over others, similarly to Bostrom's dystopia, then the case would not be significantly different from ordinary epistocracy. If artificial dictators are subject

¹¹ Recall the distinction between Rule by Algorithms and Rule by Software Engineers from the previous section.

to the same moral requirements as human ones, then it seems that Estlund's anti-epistocratic argument, if successful, would reject this form of Rule by AI as well.¹² The same is not true of Rule by Algorithms, however. Under Rule by Algorithms citizens are not ruled by anyone, for the ruling algorithms – not persons themselves – are sufficiently independent and intelligent not to be thought of as mere extensions of any person who have been involved in their making. Citizens are instead ruled by impersonal mechanisms which can be expected to reliably produce good political outcomes. Why should we reject such a proposition?

Of course, the claim that neither democracy nor Rule by Algorithms involves unjustified power hierarchies does not imply that one can never object to democratic or algorithmic decisions. If either democracy or Rule by Algorithms produced overt injustices, their rule should be condemned and resisted. Political decision-making, whether automated or not, should always take place in an institutional environment where strong, e.g., constitutional, guarantees guard against the worst injustices, protect human rights, individual liberty, and so on. Democracy, epistocracy and Rule by Algorithms should always be subjected to such restraints. Here we do not discuss the legitimacy of the unrestrained absolute dictatorship of algorithms, experts or majorities, as these are non-starters for any theory of legitimate authority.

Note that even if the dangers of absolute algorithmic dictatorship are averted, reasonable worries remain that Rule by Algorithms would not produce the best outcomes. But these alone, absent potentially problematic asymmetric power relations, are insufficient to disqualify Rule by Algorithms the same way they disqualify epistocracy. Democracy can be reasonably suspected not to produce the best outcomes as well. Even if one is convinced of the wisdom of the crowds, one is not necessarily compelled to think that crowds are always the wisest. Thinking that non-democratic forms of decision-making would be epistemically superior to democracy is not an unreasonable view. But the reasonable suspicion of epistemic suboptimality only has illegitimizing force when it is coupled with the introduction of asymmetric power relations. Estlund's argument, in the end, is that reasonable people can suspect that under epistocracy they would sacrifice their political equality for nothing; for epistocracy may fail to deliver the great outcomes it promises. But what exactly would be sacrificed under Rule by Algorithms? Certainly not equality, which seems to trigger the demand of public justification in Estlund's argument. Then what? In the next section I will provide an answer to this question.

¹² Many authors challenge Estlund's argument (e.g., Lippert-Rasmussen 2012). Even if it fails, however, my point holds for the general approach of using public reason liberalism against Rule by Algorithms.

IV. DOMINATION BY ALGORITHMS

In my view, the test Rule by Algorithms fails is not that of equality but that of freedom. Political philosophers generally agree that political institutions should cater to citizens' freedom to some extent. For many authors, this means primarily the protection of certain basic liberties which carve out, for each citizen, spheres of non-interference within which said citizen's freedom can be exercised.¹³ Ruling algorithms should, in principle, have no problem with securing these basic liberties for citizens and thus one might believe that there is no reason to think of them as particularly grave threats to freedom.

However, as philosophers – especially within the so-called *republican* tradition (Lovett–Pettit 2009) – pointed out, freedom is threatened not only by *interference* in what is usually thought of as individuals' *private affairs*, guarded by their liberty rights, but also by relations of *domination*. Domination is a complex idea, but it may be initially defined – although this definition will be revised later – as subordination to an alien will (Pettit 2012a. 79) exemplified most clearly by the relationship of the master and the slave. A slave remains subordinated to the will of the master even when said master decides out of benevolence never to interfere with the life of the slave.

Similarly, citizens may remain unfree in significant ways if certain subordination relations obtain, which may be the case even if their liberty rights are never breached. For example, even if a benevolent dictator – or a panel of experts under epistocracy – were to define liberty rights exactly in the desirable ways, e.g., granting free speech, free association, occupational freedom and all other important liberties, citizens may still be thought of as not having their freedom sufficiently protected, insofar as their liberties depend entirely on the benevolence of the dictator or the goodwill of the experts. Non-domination, as republicans often argue, requires not only that laws protect citizens' liberty rights, but also that citizens exercise *control* over legislation and political decision-making in general through democratic institutions (Pettit 2012b).¹⁴

Rule by Algorithms may seem immune to this challenge. Domination is most often thought of as a matter involving two persons, one dominated and one dominating. But as I noted so emphatically in the previous sections, algorithms are not persons, and as such, one may argue, they cannot dominate. I would dispute this point. In my view, domination or something very similar to it is possible without dominating agents. Some authors forcefully deny this claim (Lovett 2010. 47–49), while others point to oppressive social structures, such as

¹³ A view roughly along these lines is presented, for example, in Rawls's *A Theory of Justice* (Rawls 1999. 177–178).

¹⁴ Note that in this paper I do not endorse the republican doctrine that domination is the *only* threat to freedom, and that there is nothing more to freedom than non-domination. I only claim that non-domination is an important aspect of freedom.

patriarchy or perhaps capitalism as potential non-agential sources of domination (Gourevitch 2013; Einspahr 2010). I would like to focus on a different case of non-agential domination proposed by Gwilym Blunt (2015), who invites us to imagine a scenario in which an unjust apartheid system is set up

by a legislator who then promptly dies. The laws are impartially enforced, not by the privileged group, but by a series of automatons; they enforce the law impartially and cannot be reprogrammed. In this case, all groups have no influence over their status, even though one group is privileged, they cannot be said to dominate the others since they do not have systemic or interactional arbitrary power. They do not even act as agents of domination. The automatons cannot be said to dominate since they are not agents, but only machines with no will of their own. The legislator cannot be said to dominate after laying down the law, since he is dead and has no agency. It seems at least possible that this would be an instance of ‘pure’ systemic domination. (Blunt 2015. 18)

One may dispute that this is in fact domination. Still, it is clearly a worrying case of restricting individuals’ freedom. Even if it were shown that some terms other than domination would be more appropriate for the analysis of this case, this would not affect my argument greatly, as it is not premised on a domination-only view of freedom. For this reason, I will continue using the term “domination”, acknowledging that if other terms are proven to be more suitable for this discussion, they should be adopted instead.

Now imagine that in Blunt’s example the automatons are in fact algorithms that do not uphold an unjust apartheid system, but rather produce good outcomes; furthermore, they do not enforce the will of a deceased legislator, but rather act upon their own determination as autonomous AI systems. Would domination still obtain? In other words, do those characteristics of Blunt’s example that engender domination carry over to Rule by Algorithms as I describe it? In my view, they do. The automatons in Blunt’s example do not dominate because they uphold an unjust apartheid system. As the example of the benevolent dictator shows, domination may obtain under a relatively just system as well; the dictator may defend basic liberties, introduce fair distributive policies, and so on – still, citizens remain dominated under the dictator’s rule.

The automatons also do not dominate because they carry out the will of a deceased person. Although previously I defined domination as subordination to an alien will, it is important to see exactly what it is about such subordination that threatens freedom. Michael Blake notes in a seminal paper that domination violates “the autonomy of the individual by replacing that individual’s chosen plans and pursuits with those of another” (Blake 2001. 272). This replacement or substitution of wills seems to be an essential element of domination. There are two aspects to domination so understood, however. One is the *muting* of the individual’s will, the rendering ineffective of the dominated person’s choice to

pursue certain plans. The other is the replacement of that will with the will of someone else. I would argue that the first is sufficient for domination, or at least for the type of loss of freedom that is relevant both in Blunt's example and in the case of Rule by Algorithms.

Imagine an evil scientist who implants a chip in my brain. Every time I would make a decision about my career, for example, the chip turns off my deliberative faculties and selects a choice randomly. The scientist does not choose my career for me; it is possible that the scientist and I never cross paths again, she has no idea how I live my life and has no way of interfering with my choices anymore. She does not control me in any sense of the word, nor does she replace my will with hers; rather she merely mutes my will handing over this aspect of my life to pure chance. Still, this clearly subtracts from my freedom. I cannot pursue occupations that I find rewarding, establish work-life balance on my own terms, and so on, for my will regarding these matters is rendered weightless.¹⁵

In some cases, therefore, we can be unfree simply because a situation is so arranged that our will does not matter. This doesn't require that anyone else's matter *more*. There may be no one whose will does, and yet we remain unfree. Could intelligent algorithms render us dominated or at least unfree in this sense? They certainly could. Under Rule by Algorithms the rules that govern the shared life of the polity are made without the contribution of citizens; indeed, they are made without the contribution of anyone. Citizens' will regarding the terms of social cooperation is entirely irrelevant, impersonal mechanisms take care of settling all political matters.

Naturally, one is not always unfree when one's will does not matter. It does not matter that I wish to be able to breathe in outer space without aid; I simply cannot. This fact does not detract from my freedom one bit. As Isaiah Berlin famously remarks, "mere incapacity to attain a goal is not lack of political freedom" (Berlin 2002. 169). But citizens' inability to influence political decisions under Rule by Algorithms is not mere incapacity. They are subject to an artificial, human-made arrangement that they have brought into being and are able to change, even if that takes a major effort. Under this arrangement, although there is no one ruling over them, citizens also have no opportunity to weigh in on certain decisions which, many argue, would be crucially important for establishing their standing not only as equal, but also as free citizens of society. For as Ronald Dworkin notes,

¹⁵ None of this is to say that the scientist should not be held *responsible* for what has happened to me. It seems very plausible to me that the scientist is indeed responsible and should be blamed for all the misfortunes that ensued from her operation. It is the scientist's *fault* that I am unfree, but I am not unfree because she substitutes her will for mine.

We cannot make our political life a satisfactory extension of our moral life unless we are guaranteed freedom to express our opinions in a manner that, for us, satisfies moral integrity. [...] But the demands of agency go beyond expression and commitment. We do not engage in politics as moral agents unless we sense that what we do can make a difference, and an adequate political process must strive, against formidable obstacles, to preserve that potential power for everyone. (Dworkin 2002. 201–202)

Delegating all political decisions to impersonal mechanisms would threaten our standing as free moral agents capable of and entitled to reason and make choices about the most fundamental aspects of our shared life in society. This, I believe, is sufficient reason to resist transition to Rule by Algorithms.¹⁶

If this is right, then the reason why Rule by Algorithms should be rejected is not that we can raise reasonable doubts about its effectiveness but that it imposes impermissible constraints on our freedom. For freedom is not something we can trade for greater economic efficiency or growth, more innovation, or whatever else is promised by the superior political decision-making of Rule by Algorithms. Rule by Algorithms threatens to take away control over certain aspects of our shared social and political life without which we cannot view ourselves as free and equal citizens living in a just society.

V. CONCLUSION

I do not believe that algorithms will rule us anytime soon. Still, there are important conclusions to be drawn from the discussion above. The rise of algorithmic decision-making, not only in government and politics, but also in other areas of life, raises innumerable questions and problems. Some of these, such as opacity or bias, serious as they are, can be expected to be mitigated via technical and institutional solutions (Lepri et al. 2018; Danaher 2016. 258–265). Others, however, force us to carefully reflect upon the basic principles and values according to which we organize society. The automation of decision-making through AI systems, including intelligent algorithms, promises to increase efficiency and the quality of outcomes if we are willing to give up control. Control over certain aspects of our lives, however, is constitutive of our freedom both as private individuals and as democratic citizens.

¹⁶ Some authors doubt that democratic participation rights are in fact constitutive of citizens' freedom in society (Brennan 2016b). For a defence of the position that they are, see Gould (1990), Hanisch (2013), and Rostbøll (2015).

We need to understand what we risk when we hand over some or all of these aspects to impersonal mechanisms which promise to take better care of us than we ourselves could. Note that this promise does not have to be false. Decision-making algorithms may prove to be good stewards to our interests, even outstanding ones. And yet we may incur serious losses for which the gifts of hypercompetent algorithmic government, full or partial, may not be able to compensate. Here I discussed one such potential loss, i.e., the loss of freedom, and suggested that the introduction of forms of algorithmic decision-making that threaten with this kind of loss should be resisted. This is an important insight even if full Rule by Algorithms is but a distant possibility, for even the partial automation of political decision-making can diminish democratic control in ways which threaten citizens' freedom. Further research is needed for spelling out how these considerations translate into more tangible regulatory measures. Still, I hope that I made some steps toward the right direction and drew attention to some of the important aspects of these problems.

REFERENCES

- Arneson, Richard J. 2009. The Supposed Right to a Democratic Say. In Thomas Christiano – John Christman (eds.) *Contemporary Debates in Political Philosophy*. London, Blackwell. 197–212.
- Barocas, Solon – Andrew Selbst 2016. Big Data's Disparate Impact. *California Law Review*. 104. 671–732.
- Berlin, Isaiah 2002. Two Concepts of Liberty. In Henry Hardy (ed.) *Liberty*. Oxford, Oxford University Press. 166–217.
- Binns, Reuben 2018. Algorithmic Accountability and Public Reason. *Philosophy and Technology* 31. 543–56.
- Blake, Michael 2001. Distributive Justice, State Coercion, and Autonomy. *Philosophy and Public Affairs*. 30. 257–296.
- Blunt, Gwilym David 2015. On the Source, Site and Modes of Domination. *Journal of Political Power*. 8. 5–20.
- Boden, Margaret A. 2016. *AI: Its Nature and Future*. Oxford, Oxford University Press.
- Bostrom, Nick 2014. *Superintelligence: Paths, Dangers, Strategies*. Oxford, Oxford University Press.
- Brennan, Jason 2011. The Right to a Competent Electorate. *Philosophical Quarterly*. 61. 700–724.
- Brennan, Jason 2016a. *Against Democracy*. Princeton/NJ, Princeton University Press.
- Brennan, Jason 2016b. Democracy and Freedom. In David Schmidtz – Carmen E. Pavel (ed.) *The Oxford Handbook of Freedom*. Oxford, Oxford University Press. 335–349.
- Chambers, Simone 2010. Theories of Political Justification. *Philosophy Compass*. 5. 893–903.
- Christiano, Thomas 2008. *The Constitution of Equality: Democratic Authority and Its Limits*. Oxford, Oxford University Press.
- Coglianesi, Cary – David Lehr 2017. Regulating by Robot: Administrative Decision Making in the Machine-Learning Era. *Georgetown Law Journal*. 105. 1147–1223.
- Danaher, John 2016. The Threat of Algocracy: Reality, Resistance and Accommodation. *Philosophy and Technology*. 29. 245–268.

- Dworkin, Ronald 2002. *Sovereign Virtue*. Cambridge/MA, Harvard University Press.
- Einspahr, Jennifer 2010. Structural Domination and Structural Freedom: A Feminist Perspective. *Feminist Review*. 94. 1–19.
- Estlund, David 1993. Making Truth Safe for Democracy. In David Copp – Jean Hampton – John Roemer (eds.) *The Idea of Democracy*. Cambridge, Cambridge University Press. 71–100.
- Estlund, David 2003. Beyond Fairness and Deliberation: The Epistemic Dimension of Democratic Authority. In Thomas Christiano (ed.) *Philosophy and Democracy: An Anthology*. Oxford, Oxford University Press. 69–94.
- Estlund, David 2008. *Democratic Authority*. Princeton, Princeton University Press.
- Gaus, Gerald 1996. *Justificatory Liberalism*. Oxford, Oxford University Press.
- Gaus, Gerald 2010. *The Order of Public Reason*. Cambridge, Cambridge University Press.
- Gaus, Gerald 2015. Public Reason Liberalism. In Steven Wall (ed.) *The Cambridge Companion to Liberalism*. Cambridge, Cambridge University Press. 112–140.
- Goldman, Alvin 2015. What Is Democracy (and What Is Its Raison D’Etre)? *Journal of the American Philosophical Association*. 1. 233–256.
- Gould, Carol 1990. *Rethinking Democracy: Freedom and Social Cooperation in Politics, Economy, and Society*. Cambridge, Cambridge University Press.
- Gourevitch, Alex 2013. Labor Republicanism and the Transformation of Work. *Political Theory*. 41. 591–617.
- Gunn, Paul 2014. Democracy and Epistocracy. *Critical Review*. 26. 59–79.
- Hanisch, Christoph 2013. An Autonomy-Centered Defense of Democracy. *International Philosophical Quarterly*. 53. 371–384.
- Howard, Ayanna – Jason Borenstein 2018. The Ugly Truth About Ourselves and Our Robot Creations: The Problem of Bias and Social Inequity. *Science and Engineering Ethics*. 24. 1521–1536.
- Hurley, Susan 2000. Cognitivism in Political Philosophy. In Roger Crisp – Brad Hooker (eds.) *Well-Being and Morality: Essays in Honour of James Griffin*. Oxford, Oxford University Press. 177–208.
- Kolodny, Niko 2014. Rule Over None I: What Justifies Democracy? *Philosophy and Public Affairs*. 42. 195–229.
- Landemore, H el ene. 2012. *Democratic Reason*. Princeton, Princeton University Press.
- Lepri, Bruno – Nuria Oliver – Emmanuel Letouz e – Alex Pentland – Patrick Vinck 2018. Fair, Transparent, and Accountable Algorithmic Decision-Making Processes: The Premise, the Proposed Solutions, and the Open Challenges. *Philosophy and Technology*. 31. 611–627.
- Lepri, Bruno – David Sangokoya – Emmanuel Letouz e – Nuria Oliver 2017. The Tyranny of Data? The Bright and Dark Sides of Data-Driven Decision-Making for Social Good. In Tania Cerquitelli – Daniele Quercia – Frank Pasquale (eds.) *Transparent Data Mining for Big and Small Data*. Berlin, Springer. 3–24.
- Lippert-Rasmussen, Kasper 2012. Estlund on Epistocracy: A Critique. *Res Publica*. 18. 241–258.
- Lovett, Frank 2010. *A General Theory of Domination and Justice*. Oxford, Oxford University Press.
- Lovett, Frank – Philip Pettit 2009. Neorepublicanism: A Normative and Institutional Research Program. *Annual Review of Political Science*. 12. 11–29.
- Mittelstadt, Brent Daniel – Patrick Allo – Mariarosaria Taddeo – Sandra Wachter – Luciano Floridi 2016. The Ethics of Algorithms: Mapping the Debate. *Big Data & Society*. 3. 1–21.
- Moraro, Piero 2018. Against Epistocracy. *Social Theory and Practice*. 44. 199–216.

- Oswald, Marion 2018. Algorithm-Assisted Decision-Making in the Public Sector: Framing the Issues Using Administrative Law Rules Governing Discretionary Power. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*. 376. 1–20.
- Peter, Fabienne 2016. The Epistemic Circumstances of Democracy. In Michael S. Brady – Miranda Fricker (eds.) *The Epistemic Life of Groups: Essays in the Epistemology of Collectives*. Oxford, Oxford University Press. 133–49.
- Pettit, Philip 2012a. Freedom. In David Estlund (ed.) *The Oxford Handbook of Political Philosophy*. Oxford: Oxford University Press. 77–97.
- Pettit, Philip 2012b. *On the People's Terms: A Republican Theory and Model of Democracy*. Cambridge, Cambridge University Press.
- Rawls, John 1993. *Political Liberalism*. New York, Columbia University Press.
- Rawls, John 1999. *A Theory of Justice*. Cambridge/MA, Harvard University Press.
- Rostbøll, Christian 2015. The Non-instrumental Value of Democracy: The Freedom Argument. *Constellations*. 22. 267–278.
- Russell, Stuart – Peter Norvig 2016. *Artificial Intelligence: A Modern Approach*. 3rd ed. New York, Pearson Education.
- Turner, Jacob 2019. *Robot Rules: Regulating Artificial Intelligence*. London, Palgrave Macmillan.
- Viehoff, Daniel 2016. Authority and Expertise. *Journal of Political Philosophy*. 24. 406–426.
- Waldron, Jeremy 2012. Democracy. In David Estlund (ed.) *The Oxford Handbook of Political Philosophy*. Oxford, Oxford University Press. 189–203.
- Wirtz, Bernd – Jan C. Weyerer – Carolin Geyer 2019. Artificial Intelligence and the Public Sector – Applications and Challenges. *International Journal of Public Administration*. 42. 596–615.
- Yeung, Karen 2017. ‘Hypernudge’: Big Data as a Mode of Regulation by Design. *Information Communication and Society*. 20. 118–136.
- Zarsky, Tal 2011. Governmental Data-Mining and Its Alternatives. *Penn State Law Review*. 116. 286–330.
- Zarsky, Tal 2016. The Trouble with Algorithmic Decisions: An Analytic Road Map to Examine Efficiency and Fairness in Automated and Opaque Decision Making. *Science Technology and Human Values*. 41. 118–132.