The Ethics of Algorithmic Outsourcing in Everyday Life

John Danaher, NUI Galway

(pre-publication draft of a paper set to appear in Yeung, K and Lodge, M. (eds) *Algorithmic Regulation* Oxford: OUP 2019)

Abstract: We live in a world in which 'smart' algorithmic tools are regularly used to structure and control our choice environments. They do so by affecting the options with which we are presented and the choices that we are encouraged or able to make. Many of us make use of these tools in our daily lives, using them to solve personal problems and fulfill goals and ambitions. What consequences does this have for individual autonomy and how should our legal and regulatory systems respond? This chapter defends three claims by way of response. First, it argues that autonomy is indeed under threat in some new and interesting ways. Second, it evaluates and disputes the claim that we shouldn't overestimate these new threats because the technology is just an old wolf in a new sheep's clothing. Third, and finally, it looks at responses to these threats at both the individual and societal level and argues that although we shouldn't encourage an attitude of 'helplessness' among the users of algorithmic tools there is an important role for legal and regulatory responses to these threats that go beyond what are currently on offer.

1. Introduction: Algorithmic Tools and Everyday Life

Today, I went for a run. As per usual, I brought my smartphone with me, plugging in my earphones as I went out the door. The rain was pouring down and the wind was swirling in my face, but I was determined to beat my personal best. A little voice in my ear told me I could do it. It told me how far I had run, what my average speed per kilometre was, whether I was beating my previous split times. With these constant, encouraging reminders I knew not to get disheartened by the conditions. I persevered. In the end I did indeed beat my personal best. The app from which the encouraging voice

emanated pinged me with a gold merit badge as a reward. It also immediately informed all my followers on social media of my success. A congratulatory cycle soon followed.

My experience is one of millions of similar experiences that takes place everyday. We live in a world in which a vast ecosystem of algorithmic tools are used to incentivise, control, nudge and manipulate our behaviour (Pasquale 2015; Wu 2017; Alter 2016; Crawford 2015). Whether it is planning a journey using Google Maps, recording workouts and sleep data with a FitBit, or browsing through film and book choices on Amazon or Netflix, it is hard to avoid the influence of these algorithmic tools. Indeed, so pervasive are they that I even use an algorithmic tool to block access to all the other algorithmic tools that might distract me from my daily tasks. I am using it right now, in a desperate attempt to finish writing this book chapter.

The pervasiveness of these algorithmic tools raises several important social, legal and regulatory questions, many of which are addressed in this book. People worry about the impact they have on our privacy and data protection (Polonetsky and Tene 2013; Hartzog 2018), on biased and unequal access to social services and goods (Noble 2018; Eubanks 2018, O'Neil 2016), and on the transparency and legitimacy of public and political decision-making (Danaher 2016; Danaher et al 2017; Burrell 2016; Pasquale 2015). In this chapter, I want to focus on the impact of these algorithmic tools at a different scale: in the day to day activities of our personal lives. What happens when we outsource or share many of our daily tasks and objectives to digital assistants and other algorithmic tools? In particular, what happens to our autonomy and freedom of choice? Algorithmic tools affect our choice environments: they pre-filter and highlight options; they 'ping' us with reminders or incentives; and they sometimes even make choices on our behalf, giving us (if we are lucky) the power to veto those choices.

Does this mean that individual autonomy is under a significant new, technological threat? How should we as individuals and societies respond? Do we need reformed legal and regulatory systems to address this threat? This chapter defends three claims by way of response. First, it argues that autonomy is indeed under threat in some new and

interesting ways. To be precise, it identifies two interesting mechanisms through which algorithmic tools can manipulate and undermine our autonomy. Second, it evaluates the claim that we shouldn't overestimate these new threats because the technology is just an old wolf in a new sheep's clothing. In other words, it is no different from previous historical threats to autonomy. I challenge this by arguing that there are some distinctive qualities to the threats posed by algorithmic outsourcing tools, even though it is also important to bear in mind the various ways in which algorithmic tools can promote individual autonomy. Third, and finally, it looks at responses to these threats at both the individual and societal level. It argues that we shouldn't encourage an attitude of 'helplessness' among the users of algorithmic tools — there are ways in which they can personally adapt their behaviour to address the new threats — but there may also be an important role for legal and regulatory responses to these threats, ones that go beyond what are currently on offer.

The chapter is divided into four parts. First, I provide a framework for thinking about algorithmic tools and the various ways in which they can affect individual autonomy. This will give us a sense of the diverse strategies employed by the makers of algorithmic devices to control our choice environments and so help us to appreciate some of the complexities in how the users relate to algorithmic tools. Second, I introduce the argument for thinking that these algorithmic tools pose a significant threat to autonomy, focusing on two particular mechanisms for this. Third, I look at the skeptical response to that argument — the one that claims we have seen it all before. And fourth, I consider the possible individual and social responses.

2. Understanding Algorithmic Tools

We need to start by clarifying the phenomenon of interest. This chapter is concerned with the threat that algorithmic tools pose to our autonomy. The use of the word 'tools' to describe the phenomenon is somewhat controversial in this context. A tool is typically understood as something that an agent controls; not something that controls and manipulates an agent. Take a simple example. A hammer is a tool used for

driving nails into hard surfaces. It enhances the autonomy of the user, enabling them to do things that were previously impossible. It does not 'shout' at the user, demanding that they hammer more nails, or 'scold' them if they miss a certain target number of hammered nails for the day. There is, consequently, no sense in which it could be thought to undermine an individual's autonomy.

But think about it again. The mere presence of the hammer in your workshop might make the option of driving nails into surfaces more attractive (if only because it becomes less difficult). In other words, the presence of the tool could change the cost/benefit ratio associated with certain options and, consequently, 'nudge' you to make choices that you otherwise would not have made (Morrow 2014). So although talk of algorithmic tools might seem misleading at first glance — since tools are autonomy-enhancing not autonomy-threatening — there is in fact a dual character to even the simplest of tools: they enable, incentivize, and therefore influence, our choices. Furthermore, algorithmic tools are clearly a bit different from traditional tools like hammers. Digital assistants such as Alexa or Siri, or ranking algorithms such as those used by Google and Facebook, are 'tool-like' in the sense that we can use them to accomplish goals, e.g. finding useful information, connecting with our friends, scheduling calls and meetings. But they go beyond this. They don't just enhance our agency; they also do things on our behalf. They constrain the range of options with which we are confronted, make some information more manageable and more salient, and ease some of our cognitive burden (Brundage and Danaher 2017). Sometimes they go even further. Fitness and productivity apps often come equipped with behaviour change algorithms that try to explicitly prompt and encourage us to do certain things. If instead of the traditional, inert hammer, we had a modern 'smart' hammer in our workshops, coded with some behaviour-change algorithm, then we might very well expect it to start shouting at us and demanding that we hit a certain target for the day. It would still be a tool, but it would also much more than that: it would be a device for regulating and governing our behaviour. It would also not just be a self-imposed tool for regulating and governing our behaviour. It is would be one whose software and programming is controlled by third parties. It would not be entirely *ours*; it would be *theirs* as well.

So why stick with the term 'algorithmic tool'? I do so because using the term highlights the fact that algorithms still function as tools — as things that enable us to accomplish our goals — and it is important that we don't lose sight of that fact. Indeed, it is a fact to which I will return below. Nevertheless, the term must be used with some caution, and with the proviso that it is understood that algorithmic tools often have a regulatory or governance dimension. Furthermore, I want to be clear that I am not interested in all possible algorithmic tools. The term 'algorithm' is vague and could be used to describe any set of rules that transforms a defined set of inputs into a set of outputs. A recipe in a cookbook could, consequently, count as an algorithmic tool. I am not concerned with such cases in this chapter. I am concerned with algorithmic tools that are incorporated into information technology devices (personal computers, smartphones, smartwatches and so on). The algorithmic tools embedded in such devices typically rely on data-capture and data-mining and are often networked and cloud-based. To avoid doubt or confusion I will provide a precise definition of the concept of an algorithmic tool that will be used for the remainder of this chapter:

> **Algorithmic tool**: Any computer-coded algorithm embedded in an information technology device that is used to generate and act upon knowledge that can be used to aid decision-making and goal achievement, and that functions in part by regulating and governing behaviour.

This definition blends together a distinction that Yeung (2017b, 3) makes between *algorithmic decision-making* (decision-making that is aided by algorithmically generated knowledge) and *algorithmic regulation* (regulatory systems that utilise algorithmic decision-making). This blending together is deliberate. It is the very fact that these systems combine tool-likeness and with regulation and governance that makes their impact on the day to day activities of our lives, and on individual autonomy so difficult to determine. I discuss this in more detail below.

Now that we have a clearer sense of what an algorithmic tool is (in the abstract) we

can introduce some additional distinctions. For not all algorithmic tools are created equal. They often function in different ways and these differences can be relevant to any discussion of their autonomy-promoting and autonomy-undermining effect. There has been some interesting recent work on this very issue. Gal (2017) argues that there are at least four different kinds of algorithmic tool, each of which has a distinctive effect on individual choice. The four kinds are: (i) '*stated preference' algorithmic tools*, for which the individual specifies exactly what they want the algorithm to do and the algorithm assists in achieving this outcome; (ii) '*menu of preferences' algorithmic tools*, for which the individual doesn't specify their preferred outcome but chooses from a menu of options provided to them by the algorithm; (iii) '*predicted preference' algorithmic tools*, for which the algorithm, based on data-mining (often from other individuals), tries to predict what the user will want and target options at them accordingly; and (iv) '*self-restraint preference' algorithmic tools*, for which the algorithm functions as a pre-commitment device, favouring the user's long-term interests (perhaps stated; perhaps predicted) over their immediate interests. As you might imagine, these different kinds of algorithmic tool have different consequences for individual autonomy. A stated preference algorithm, for example, would seem to be obviously autonomy-enhancing; a predicted preference algorithm much less so.

In a similar, but more complex, vein, Yeung (2017b) describes a 'logical space' of possible algorithmic tools, focusing specifically on their use in regulatory systems that try to enforce certain behavioural standards. This logical space has three dimensions to it, each of which has two 'settings', making for eight possible forms of 'regulation by algorithmic tool'. The first dimension concerns the nature of the algorithmic tool itself. Is it fixed or adaptive? The second dimension concerns the way in which the algorithmic system monitors individual behaviour. Does it 'react' to the user's violation of its behavioural standards or does it try to predict and pre-empt the user? The third dimension concerns the role that human regulators play in the system. Does the system automatically enforce its standards (perhaps giving humans a veto power) or does it simply recommend (perhaps strongly) enforcement options to them? Again, the different settings on each of these dimensions would appear to be relevant when it comes to

assessing the impact of these systems on individual choice and autonomy, even in cases where we impose regulatory tools on ourselves. Intuitively, it seems like a system that anticipates and pre-empts violations of prescribed standards, and that automatically enforces sanctions on those violations, poses more of a threat to autonomy than a system that simply reacts and recommends. Yeung's discussion seems to suppose that these tools are imposed on us by outside regulators, but it seems obvious that we can self-impose them too by 'voluntarily' adopting more tools into our lives.

The taxonomies proposed by Gal and Yeung are useful because they encourage nuance and complexity in how we consider the impact of algorithmic tools on our behaviour. They inspire and guide the analysis that follows.

3. Do Algorithmic Tools Undermine Autonomy?

So what about the impact of algorithmic tools on our day-to-day behaviour? If we make excessive use of them, will they significantly undermine autonomy or could they actually enhance it? To answer that question we need to bear in mind both the complex nature of algorithmic tools – as suggested by both Gal and Yeung – and, perhaps more importantly, the complex nature of autonomy itself as a value.

There is no shortage of material written about the nature of autonomy. Everyone agrees on the basic idea — to be autonomous means that you are, in some sense, the 'author' of your own life — but there are differences when it comes to the critical sub-conditions of autonomy. This is where the complexity comes in. Many sub-conditions have been identified over the years, leading some people to propose quite complicated theories of autonomy as a result (e.g. Killmister (2017) proposes a four-dimensional theory of autonomy). Acknowledging this complexity, in this chapter I will adopt and adapt a theory of autonomy that was first proposed by Joseph Raz back in the 1980s. This theory focuses on three conditions that need to be satisfied if a particular choice/life is to count as autonomous. The following quote from Raz sets them out:

If a person is to be maker or author of his own life then he must have the mental abilities to form intentions of a sufficiently complex kind, and plan their execution. These include minimum rationality, the ability to comprehend the means required to realize his goals, the mental faculties necessary to plan actions, etc. For a person to enjoy an autonomous life he must actually use these faculties to choose what life to have. There must in other words be adequate options available for him to choose from. Finally, his choice must be free from coercion and manipulation by others, he must be independent.

(Raz 1986, 373)

The three conditions of autonomy embedded in this quoted passage are: (a) the autonomous person must have the minimum rationality to plan actions that will allow them to achieve their goals; (b) they must have an adequate range of options available to choose from; and c) they must be independent, which Raz takes to mean free from *coercion* and *manipulation* when making and implementing their choices, but which we can take to mean freedom from interference or domination in a broader sense (Pettit 2014).

What I like about Raz's theory is that it captures both the essence of autonomy, while at the same time allowing for considerable differences in how the conditions are interpreted or understood. For example, you could understand them as minimal or threshold conditions. In other words, you could say that once a minimal degree of rationality is achieved, once there is an adequate range of options available, and once there is independence from others (in some appropriately defined sense) then a particular choice (or set of choices) is, indeed, autonomous. That's all there is to it. But it is also possible, I believe, to understand these three conditions as defining a dimensional space of possible degrees of autonomy. In other words, to hold that the higher the degree of rationality and independence, and the more options available to you, the more autonomous you become.

That said, if one embraces this dimensional view of autonomy, one must be sensitive

to the possibility that once you go beyond a certain level of independence, rationality, and optionality, you may start to experience *diminishing marginal returns*. Indeed, at extreme levels it is possible that there is an inverse relationship between an increase along those dimensions and autonomy. An example of this inverse relationship might be the so-called 'paradox of choice' (Schwartz 2004). This is an effect that has been discussed by psychologists which suggests that if you have too much choice (too wide a range of adequate options) you become unable to make a decision. Inability to decide is clearly not compatible with autonomy. I return to this point below.

Relatedly, it could be that some infringement of one of the conditions is necessary if we are to achieve the right overall balance of autonomy. We can see this most clearly with the independence condition. This states that in order to live an autonomous life you must be free from interference and domination by others. While this certainly seems like a plausible requirement, it should be noted that some degree of external interference may be necessary in order to achieve one's goals. This is, in fact, a central paradox underlying liberal modes of government. Many prominent liberal authors argue that some coercion by the state is necessary if you are to be free from coercion and manipulation by your fellow citizens. This is, for example, the classic Hobbesian argument about the need for a coercive state. There is undoubtedly something faintly paradoxical about this but its truth seems to be borne out by our collective history: a lingering threat of coercive interference by the state (or other institutional actor) seems to be integral to maintaining peace and stability. Furthermore, not all autonomy-undermining interferences and dominations are going to come from other people. Sometimes we are our own worst enemies. We have cognitive and motivational biases that draw us away from doing what we most want to do. Weakness of the will and procrastination are obvious examples of this. It's possible that by consenting to some degree of external constraint we may be able to overcome these internal interferences and allow ourselves to do what we most want to do. My use of an algorithm to block access to other algorithmic tools is an example of this. I know that my distractibility and lack of focus are my own worst enemies and so I deliberately restrict myself in order to accomplish my goal of finishing this chapter. It is a classic pre-commitment device that promotes rather than undermines autonomy.

We can reinterpret the Razian account to incorporate these considerations and use it to formulate a test for autonomy:

> **Autonomy**: If you have (a) the rationality to select the appropriate means to your desired ends; (b) an adequate range of options; and (c) are free from most external and internal sources of interference with, and coercion and manipulation of, your will, then you are autonomous. Indeed, the more you have of each, the more autonomous you are (up to some point of diminishing returns).

We can then use this test to assess the threat posed by algorithmic tools to our autonomy. Whenever we are confronted by a particular tool we can ask ourselves the three questions prompted by the Razian definition: does the tool undermine minimum rationality? Does it undermine optionality? And does it undermine independence? If the answer is yes to all three, then the tool clearly undermines autonomy. If it is yes to some but no to others, then we have some tricky and nuanced judgments to make. How significantly does the tool affect the relevant conditions? How important are those conditions, overall, to our understanding of autonomy? The answers will determine our assessment of the impact of the tool on autonomy.

Given the way in which I am talking about it, it's pretty obvious that I don't think a global assessment of the threat that algorithmic tools pose to autonomy is possible. They have to be considered on their own merits. They could pose a threat, sometimes, depending on the context in which they are used and the way in which they function. It all depends. Nevertheless, it is possible to get a sense of the possibilities and trends by comparing the impact of algorithmic tools on each of the three conditions identified above. In doing so, I think it is possible to identify some of the particular, and distinctive mechanisms through which these tools might pose a threat to our autonomy.

First, let's consider the impact that they might have on the rationality condition. This consideration can be brief: I am generally inclined to discount any claimed impact that

algorithmic tools might have on this condition. Why so? Well, because it seems unlikely to me that algorithmic tools of the sort under consideration could completely undermine our capacity to plan and execute complex intentions. Personal digital assistants, and the like, are all premised on the fact that we have complex intentions/goals that we need machine assistance to realise. They work with a presumption of minimal rationality, not against. They provide us with both the information and the means we need to make use of that rationality. That said, if we follow the arguments of Nicholas Carr (2015) — who claims that the persistent use of tools might degenerate our cognitive faculties — or the arguments of Frischmann and Selinger (2018) — who claim that such tools may actually program us to act as simple stimulus-response machines — there may well be some negative impact on our capacity for rationality. Nevertheless, I think it is much more plausible to suppose that automating technologies will impact on the optionality and independence conditions, either by limiting and constraining our options or interfering (directly or indirectly) with our exercising of those options. It's there that the real effects on autonomy are likely to be felt.

Let's consider the limitation of options. David Krakauer has expressed some fears about this, specifically in relation to the widespread deployment of AI decision-support tools. He uses the Homeric story of the Lotos Eaters to make his point:

> *In Homer's The Odyssey, Odysseus's ship finds shelter from a storm on the land of the lotus eaters. Some crew members go ashore and eat the honey-sweet lotus, "which was so delicious that those [who ate it] left off caring about home, and did not even want to go back and say what happened to them." Although the crewmen wept bitterly, Odysseus reports, "I forced them back to the ships…Then I told the rest to go on board at once, lest any of them should taste of the lotus and leave off wanting to get home." In our own times, it is the seductive taste of the algorithmic recommender system that saps our ability to explore options and exercise judgment. If we don't exercise the wise counsel of Odysseus, our future won't be the dystopia of Terminator but the pathetic death of the Lotus Eaters.*

(Krakauer 2016)

As evidenced in this quote, Krakauer is particularly concerned about AI 'recommender' systems that block us from exploring an adequate range of options. He thinks we are fed recommendations by such systems on a daily basis and that this 'saps us' of our will. This certainly chimes with my everyday experience. When I go shopping on Amazon, I am immediately presented with a list of 'recommended' products, allegedly based on my past purchases (and the purchases of other, like-minded, people); when I go to watch a movie on Netflix, I am fed different menus of options with titles like: 'currently trending', 'because you watched Fargo…', 'films with a strong female lead', and so on; and when I am trying to find my way around a strange city, Google maps tells me which route to take, and which routes have a similar ETA. This is just a small sample of recommender systems. Similar systems are also widely deployed in business and government. Each of these systems could have a negative impact on the optionality condition. They all work by designing the 'choice architecture' in which we select our preferred options (Thaler and Sunstein 2009). The major problem is that the range of options made available to us through such recommender systems might be quite limited, not 'adequate' to our preferred self-conception. It might be more reflective of the needs and interests of the corporations who control the recommender systems than ourselves. In extreme cases, the choice architecture might be so limited as to give us only one 'live' option to choose from thereby effectively 'forcing' a single choice out of us. Google Maps sometimes has this effect due to the way in which it highlights a preferred route. In even more extreme cases, the system might block us completely from making a choice among available options and simply make the choice for us. This can happen if the default setting of the algorithmic tool is to *act* rather than to *omit*.

It is important, however, not to overstate the impact of algorithmic tools on the optionality condition. As hinted at above, there are ways in which AI recommender systems could help to promote autonomy by filtering and pre-selecting options. It's long been noted that having a range of valuable options is no good by itself: you have to be able to search through those options and exercise judgment in selecting the ones you

prefer. Too many options could actually hinder us from doing this. I could be presented with ten million randomly arranged movies on Netflix and quickly get overwhelmed. I might get 'stuck' in the choice architecture, hovering between all the possibilities, like a modern-day version of Buridan's Ass. The psychologist Barry Schwartz calls this the 'paradox of choice' and he and others have documented it in a series of experimental studies (Schwartz 2004; Scheibehenne et al 2011). Some of these studies suggest that it has a significant impact on our capacity to choose; others suggest it has a more modest effect (Scheibehenne et al 2011). Either way, AI recommender systems could help to overcome the paradox of choice by limiting our options and making it easier for us to navigate through the choice architecture. They could also help us to sustain and develop our capacity for choosing among valuable options by reducing our cognitive burden in certain tasks (Brundage and Danaher 2017).

That leaves us with the independence condition. At first glance it might seem like the personal use of algorithmic tools is entirely consistent with our independence — after all most of the time *we* voluntarily choose to use them to accomplish *our* goals. But things are not so straightforward. As noted in the previous section, the tools are not just ours; they are *theirs* too. We don't retain full ownership and control once we start using them. The companies and programmers who designed them continue to monitor and update their functionality, and they often transfer (with our explicit consent or otherwise) the data they collect to other interested parties. As we get comfortable with outsourcing some of the cognitive burden of choice to the AI assistant, we become more trustworthy. This trust can be abused. This can lead to some important negative effects on autonomy.

Some of these are pretty obvious. Some algorithmic tools can be designed to coerce you into acting in a certain way, i.e. threaten you with some adverse outcome if you fail to act in a prescribed way. This is already beginning to happen to some extent. Health and car insurance companies already 'offer' better deals to people if they agree to use algorithmic monitoring tools that track their driving and health related behaviours. We might initially view these 'offers' as autonomy-enhancing (and hence not a form of coercion) but given the way in which insurance markets function (i.e. given the

widespread adverse selection and information problems), a failure to consent to algorithmic monitoring could very quickly take on significance in the calculation of premiums. This could turn what seemed like an offer into a coercive threat. There is also even the possibility of turning this from purely financial coercion into actual physical coercion, if the algorithmic tool is joined up with some robotic technology that can physically interfere with the human user. This is not completely far-fetched. The Pavlok behaviour change bracelet, for example, shocks its user if they don't follow through on certain commitments. For the time being, this system is something that an individual chooses to impose on themselves; not something that is imposed on them by some outside force. Consequently it may not undermine freedom. Nevertheless, it is not that hard to imagine an insurance company (or a more nefarious third party) coopting such a system to physically coerce its users in an autonomy-undermining way.

More important than explicit coercion, however, are the subtle forms of manipulation and decisional interference that are made possible through the use of algorithmic tools. Yeung (2016) argues that modern algorithmic tools enable 'hypernudging', which is a kind of behaviour change technique that operates beneath the radar of conscious awareness and happens in a dynamic and highly personalised fashion. Nudging is a concept that was made popular by Cass Sunstein and Richard Thaler (2009). It involves using insights from behavioural science to construct choice architectures that 'nudge' people towards actions that are welfare-maximising or that serve the common good. For example, setting the default on retirement savings to 'opt-out' rather than 'opt-in', or placing healthy foods at eye level and unhealthy ones below or above, makes it more likely that people will choose options that are in their long-term interests. Nudges usually operate on subconscious biases in human reasoning. Sunstein and Thaler maintain that nudging is not freedom-undermining because it is still possible for people to identify and reject the 'nudges'. Others are more doubtful and argue that nudges are highly manipulative (Sunstein 2016). Whatever the merits of nudging, Yeung's point is that algorithmic outsourcing technologies bring nudging to an extreme level of precision. Instead of creating a one-size-fits-all choice architecture that is updated slowly, if ever, you can create a highly personalised choice architecture that learns and adapts to an

individual user. Hypernudging consequently looks like a distinctively algorithmic mechanism for undermining autonomy. It is definitely not something that is possible with older forms of technology.

Algorithmic tools can also enable distinctive forms of domination. Domination is a kind of power. It arises from asymmetrical relationships between two or more individuals or groups of individuals. The classic example of such an asymmetrical relationship is that between a slave and his/her master, but there are many others. If you live your life constantly in the shadow of such a master, then you cannot be free, even if the master never explicitly interferes with or manipulates your choices. This is the central insight of modern neo-republican political theory, which argues that 'non-domination' is an essential condition for individual autonomy (Pettit 2001, 2011 & 2014). Recently, some scholars have argued that in making use of algorithmic tools we may be letting algorithmic masters into our lives. Hoye and Monaghan (2015) and Graf (2017), for instance, both argue that the systems of mass surveillance and big data analytics that undergird modern AI tools are domination-facilitating. The idea they share is that by welcoming such tools into our lives we facilitate the construction of a digital panopticon around our activities. We are then all watched over by algorithmic tools of loving grace, each of which is standing in wait to nudge us back on track if we ever try to escape. Evgeny Morozov (2013) has captured this problem in an evocative metaphor, suggesting that the Big Data systems that surveil and govern our lives, function like 'invisible barbed wire'. They enclose and constrain our space of autonomy, without us even realising.

I believe it is important that we understand the precise mechanisms of this algorithmic domination. It is not a crude and explicit form of domination — like that of the master over the slave. It is more subtle than that. lt is what I would call a form of 'algocrithmic micro-domination'. 'Micro-domination' is a concept that I take from the work of Tom O'Shea (2018), who uses it to understand the forms of domination experienced by people with disabilities. He argues that people with disabilities often suffer from many small-scale instances of domination. If they live in an institutional setting, or are heavily reliant on care and assistance from others, then large swathes of

their daily lives may be dependent on the good will of others. They may need these others to help them when they wake up, when they go to the bathroom, when they eat, when they go outside, and so on. Taken individually, these cases may not seem all that serious, but aggregated together they take on a different guise:

The result is often a phenomenon I shall call 'micro-domination': the capacity for decisions to be arbitrarily imposed on someone, which, individually, are too minor to be contested in a court or a tribunal, but which cumulatively have a major impact on their life.

(O' Shea 2018, 136)

The pervasive use of algorithmic tools across all domains of our personal lives can, I believe, give rise to a similar phenomenon. Many small-scale, arguably trivial, choices in our everyday lives might be executed with the help of an algorithmic tool: what route to drive, what news stories to read, who to talk to on social media, what film to watch next and so on. The resulting network of algorithmic tools would monitor and track our behaviour and send us prompts and reminders. This means that we would then be the 'subjects' of many algorithmic masters. They would surveil our lives and create a space of permissible/acceptable behaviour. Everything is fine if we stay within this space. We can live happy and productive lives (perhaps happier and more productive than our predecessors), and to all intents and purposes, these lives may appear to be autonomous. But if we step out of line we may be quick to realise the presence of the algorithmic masters. There are already some cautionary tales to this effect. Consider, Janet Vertesi's experiences in trying to 'hide' her pregnancy from the algorithmic tools that enable online shopping and that undergird social media (Vertesi 2014). Vertesi, an expert in Big Data, knew that online marketers and advertisers like to know if women are pregnant. She decided to conduct an experiment in which she would hide her own pregnancy from them. This turned out to be exceptionally difficult. She had to avoid all credit card transactions for pregnancy-related shopping and all social media mentions of her pregnancy. In the end, her attempt to avoid algorithmic domination led to her behaviour

being flagged as potentially criminal. Suddenly, the domination was made visible.

Taken together this assessment of the impact of algorithmic tools on autonomy seems to be negative. Although there is hope when it comes to the rationality and optionality conditions, there is a reasonable case to be made for thinking that algorithmic tools can negatively impact on the independence condition for autonomy. Furthermore, there might some distinctive and novel mechanisms — hypernudging and algorithmic micro-domination — through which this negative impact is realised. That makes it sound like we should avoid algorithmic tools if we care about our autonomy.

4. But is it any different this time round?

There is, however, a skeptical view that is worth considering. This skeptical view maintains that the analysis given above is overhyped and overstated. Algorithmic tools are just an old wolf in new clothing. As Ord and Bostrom have argued (2006), it is essential in any debate about a new technology (or indeed any new policy), to avoid unwarranted status quo bias. This arises whenever you fixate on identifying the risks and benefits of the new technology but ignore the risks and benefits of the status quo (and the best available alternative). The status quo is never value neutral – this is one of the central insights of Thaler and Sunstein (2009) in their defence of 'nudging'. It could be that the status quo is much worse than the new reality that is being offered to us, even if that new reality has its flaws.

Applying this insight to the argument about autonomy, we have to accept that the world we currently inhabit is filled with threats to autonomy. Our decisions are constantly coerced, manipulated and dominated by others. We are being bought and sold competing ideologies all the time. We are pushed into alliances and factions that are not of our own choosing at our births. We persuade and overwhelm one another's intellectual resistance. We have been at this for centuries. Nothing about the new technology changes that. Algorithmic tools are just that: tools that humans use to do what they have always done. If we fixate on the 'threats' posed by these tools – or, indeed, the fancy new mechanisms

through which they act – we will lose sight of this larger reality and dangerously overstate and overhype the risk to autonomy. This is could be made even worse if we fixate on the threats while ignoring the opportunities for enhanced rationality and/or optionality made possible by these tools, coupled with the fact that they might eliminate forms of manipulation or decisional interference that were once in the ascendancy. This is not purely hypothetical. The information made accessible through algorithmic behaviour change tools, for example, can wean us off of old, irrational habits, or cause us to question routines and beliefs that have been handed down to us by others. In other words, algorithmic tools, even if they do somehow negatively affect autonomy might be simply replacing or offsetting other autonomy-undermining practices. This doesn't mean that we should welcome them with open arms, of course, but it does suggest that we should be wary of overstating the problem.

In response to this skeptical view, we need to consider whether there is something genuinely different about the ways in which algorithmic tools undermine autonomy. These 'genuine differences' must be differences in effect, not just differences in form. You can use two different tools to achieve the same result; it's only if the results differ that we should really be concerned. Are there any such 'genuine differences'? There are three that might be worth considering:

Differences of scope, scale or speed: algorithmic tools undermine autonomy at a scale, scope and/or speed that is truly unprecedented. The tools are everywhere; they are used across many domains of activity; and they operate (and update) at a speed that is completely incomprehensible to human beings.

Differences having to do with the centralisation/monopolisation of power: algorithmic tools are controlled by relatively few actors (i.e. large tech enterprises like Amazon, Google, and Facebook) and so when they undermine autonomy they do so by enabling and growing the power of these few institutional actors.

Differences having to do with the distinction between personalisation and

manipulation: algorithmic tools tend toward personalisation (i.e. they try to adapt to and predict our preferences) rather than outright manipulation (i.e. the imposition of external preferences over our own).

Each of these has a whiff of plausibility so it's worth considering them in some more detail.

With regards to differences of scope, scape and speed, there is definitely something distinctive here. This is, essentially, what Yeung (2017) is getting at in her description of hypernudging. We are used to dealing with interferences in our decision-making that emanate from other human beings or institutions, not from a superfast technological infrastructure. We are also used to interferences that manifest and act at human speeds. When this happens we are able to spot the interferences, and potentially respond to them before it is too late (or before the interference shifts and takes on a new form). With algorithmic interference, this is no longer the case. Our brains just can't compete and we can't easily escape the oversight of the algorithmic masters. Our natural, evolved forms of resistance are overwhelmed. This means the net effect of algorithmic interference might be far more pernicious than the traditional and more familiar kinds.

With regards to the centralisation of power, there is also definitely something to this. Companies like Google, Facebook and Amazon (and other tech giants) really do control large swathes of our digital lives. They are empowered and enriched when we make use of their algorithmic tools. Our individual lives can seem very small by comparison. You could argue that when human beings are manipulating and interfering with one another there is at least greater competition for and fragmentation of power. And so there are differences in effect with important moral consequences. The differences here are not so much to do with how the algorithmic tools affect us on an individual scale, but how the empowerment of the tool-makers could have other spillover effects, particularly relating to the inequality of wealth and political power. That said, I tend to think the concern about the centralisation of power is not that compelling. There have been some very centralized, and very powerful underminers of autonomy in human history. Churches and

states have, historically, wielded outsized influence over human actors. Sometimes that influence has been extremely pervasive and invasive. At one point in time, the Catholic Church played a significant role in constructing the choice architectures of many Europeans. It did this from dawn to dusk and birth to death. The same was true for other religious organisations and doctrines elsewhere on the globe. Although it is right to be concerned about the outsized influence of tech giants, it's not clear that they threaten something genuinely new and different. On the contrary it seems like the choice architectures of the typical, modern European (to stick with that example), even if they are the result of algorithmic tools, are much more diverse than used to be the case, and that the power landscape much more fragmented than it once was.

Finally, with regards to the personalisation of algorithmic tools, it is hard to know whether that is a net positive or net negative difference. On the one hand, you could argue that the personalisation that is made possible by algorithmic tools is exactly what we would want if these tools are to respect and promote individual autonomy. Being authentic to oneself and acting in ways that are consistent with one's higher order preferences is generally thought to be conducive to autonomy (Frankfurt 1971**)** (and, probably, well-being if we go with the general presumption that we are the best judges of what is in our interest). If algorithmic tools are just enabling this, then what's the problem? On other hand, you could argue that algorithmic tools typically don't work by enabling us to act in a manner that is consistent with our higher order preferences. On the contrary, they are customised to our lower order instincts and practices. They perpetuate us at our worst, not at our best. In addition to this, through personalisation, these apps undermine solidarity and the sense of collective identity (Harari 2018, 67). Instead of being the common victims of shared masters we each become the unique victims of a personalised algorithmic master. This can be disempowering and thus disabling of another natural form of resistance to the corrosion of autonomy. (It is also, however, worth noting that the actual degree of personalisation made possible by algorithmic tools can be overstated since most of these tools are trained by making statistical generalisations from patterns spotted in very large, aggregated and de-personalised datasets; this can actually lead to systematic oppression – Noble 2018).

In summary, despite the skeptical view, it may genuinely be different this time. The actual effects of algorithmic tools (and not just their mechanisms of action) may be different. Still, even if we take this insight onboard, it is difficult to make a global assessment about the impact of these tools on our autonomy since the net effects could be both positive and negative.

5. How should we respond?

How should we respond to this? If the widespread use of algorithmic tools poses new and distinctive threats to autonomy (at the same time as it introduces some opportunities for autonomy enhancement) what should we do to make the best of it? It's worth considering the potential responses at a both a personal and social level.

Let's start at the personal level. It's important not to overlook this. There is a tendency in discussions of algorithmic governance and algorithmic regulation to presume that we are helpless victims of technological overlords.[1] This may be true in some contexts, but when it comes to the algorithmic tools under discussion in this chapter – the digital assistants, and productivity, health and fitness apps that we use to accomplish our daily goals – this is not true. We still exercise considerable discretion when it comes to the acceptance of these tools into our lives. Some of this discretionary power is boosted by existing legal regimes, such as the revised General Data Protection Regulation (GDPR), but even when those regimes are in place it is important that we don't forget that it is we as individuals that have the ultimate power to accept or reject the influence of these algorithmic tools. To presume otherwise is to foster an attitude of helplessness, which will never be conducive to autonomy. So we need to empower individuals to make autonomy-enhancing decisions about when to accept algorithmic tools into their everyday lives. To do this, we need to encourage them to think about the overall importance of the goals they want to achieve with the help of the tools (of course) and whether the tools do

[1] Some of my previous work has tended toward this view. See Danaher (2016) for more.

this in an autonomy-undermining or promoting way. Sensitivity to the different dimensions and conditions of responsibility will be critical in this regard. If an individual has a goal of significant importance to them, and if to achieve that goal they must navigate a complex choice architecture with many thousands of options, some algorithmic assistance could be quite beneficial. It might make them less confused and more rational. This may come at the expense of some independence, but this could be worth it when we think about the other conditions of autonomy. On the other hand, if the individual uses an algorithmic tool to accomplish a goal of little overall importance, and that wouldn't be that difficult to achieve in the absence of the tool, they may be sacrificing their independence for minimal gain. They should think twice about running the risk. In between these two extreme cases, there will be plenty of room for individual judgment to be exercised.

None of this is to deny that individuals may need help. If, as was argued above, algorithmic tools act at a speed, scope and scale that is truly unprecedented, and thus wears down our natural forms of resistance, some external assistance may be required to promote individual autonomy. This assistance could take a number of different forms. It could come direct from the marketplace for algorithmic tools. In other words, if there are fears about the erosion of autonomy, and these are taken seriously, we might expect the marketplace to self-correct (to some extent) and for companies that offer tools that are autonomy-preserving to gain a competitive advantage. This has already happened with respect to privacy. Apple, for example, now markets itself as a privacy-protecting company, relative to its peers in Google, Facebook and Amazon, which rely more heavily in data-mining and surveillance to make their money. Could we see something similar happening with respect to autonomy? Perhaps, though I wouldn't be too confident of this. Many of the tech giants work off advertising and marketing-based business models. These business models have always depended on the capacity to manipulate and persuade potential consumers. Consequently these companies are deeply committed to practices that are, shall we say, not always respectful of autonomy: they want to capture our attention and influence our behaviour (Wu 2017, Bartholomew 2017, Alter 2017; Crawford 2015). If they don't do this, they won't make money. Unless the successfulness

of this business model is completely undermined, we can't expect the market to easily correct itself.

This means we may have to look to other legal and regulatory responses. Recent changes to data protection law in Europe are certainly helpful in this regard insofar as data is the fuel that feeds the fire of algorithmic manipulation, but other legal responses may need to be contemplated. Changes to laws on monopoly power and anti-competitiveness might be needed to break-up the powerful tech giants and address concerns about the centralization of power. More precise regulations on the speed, scope and scale of algorithmic interference might also be required. Perhaps algorithmic tools that update and adapt at speeds that erode natural forms of resistance will need to be banned (in much the same way as subliminal advertising was banned in the past)?[2] Perhaps stated preference algorithms (which we might expect to map onto our higher order preferences) should be encouraged and promoted, while predictive preference algorithms (which we might expect to map onto our lower order preferences) should be discouraged? This is to say nothing of bans or limitations on the more extreme and coercive forms of algorithmic interference, e.g. those made available through insurance markets.

Perhaps we should also consider recognizing new fundamental rights? Frischmann and Selinger (2018) have argued as much, suggesting that we need to recognise the 'right to be off' if we are to retain our sense of agency and autonomy in the 21st century. The right to be off is, as the name suggests, a right to live a life without any algorithmic tools being foisted upon you. In a similar vein, I have argued that we may need to recognise a right to attentional protection to address the impact of algorithmic tools on autonomy.[3] The idea behind this is that attention is the thing that tech companies try to capture and control through their algorithmic tools (Wu 2017, Bartholomew 2017, Alter 2017; Crawford 2015**).** Without capturing our attention, and slowly eroding our

[2] See Bartholomew 2017, p 11 and pp 112-117 on this point.
[3] Danaher 2017, working off of Tran 2016. My argument is partly influenced by the right to data protection which has only been identified and recognized in the recent past.

capacity to ignore digital distractions, these companies would be much less successful and much less able to interfere with our autonomy. By targeting our legal interventions at the capacity to pay attention, we can address one of the root causes of the potential problem. This is not a completely unprecedented idea. As Tran (2016) notes, there are many legal doctrines that already target attention. For example, informed consent laws are, at least in part, about requiring us to pay attention to certain things, and giving us the right to ignore others. By creating a new right to attentional protection you can unify and bring order to these disparate areas of law, and address the threats to autonomy posed by algorithmic tools.

We may also need to consider more radical social responses. Throughout this article I've been assuming that individual autonomy is something that we want to protect and preserve, but maybe that is not the case. Maybe autonomy competes with other important values (e.g. individual well-being and flourishing) or maybe it is only valuable to the extent that it helps bring about other values. Some prominent legal philosophers have argued as much, holding that freedom is only instrumentally valuable to the extent that it improves well-being (Leiter 2016). If that's the case, maybe the algorithmic threat is exactly the excuse we need to push us toward a post-autonomy society (Harari 2016 and 2018), just as digital convenience and mass surveillance have encouraged some people to argue for a shift to a post-privacy society (e.g. Brin 1998).

6. Conclusion

It is time to draw together the strands of argument from the preceding discussion. In brief, I have argued that algorithmic tools play an increasingly important role in how we live our daily lives. I have suggested that they are tools – insofar as they enable us to fulfill our goals – but that they also have a regulatory and governance dimension – insofar as they can be used by ourselves and third parties to manipulate, nudge, incentivize and control our behaviour. Consequently, I believe they can have an impact on our autonomy. I have tried to clarify the nature of that impact, without overstating it. I have used Joseph Raz's three conditions of autonomy (rationality, optionality and

independence) to assess the likely impact of algorithmic tools on autonomy. In doing so, I identified two novel mechanisms through which algorithmic tools can undermine our independence -- Yeung's 'hypernudging' and O'Shea's 'micro-domination' – whilst also emphasising the positive impact these tools could have on our optionality and rationality. I have also considered the extent to which algorithmic tools threaten something genuinely new and concluded that they may do so due to their speed, scope and scale and the way in which this overwhelms our natural forms of resistance to outside interference. Finally, I have argued that we need to consider responses to these threats at both an individual and societal level. Individuals should be empowered to make informed choices about when using an algorithmic tool might undermine their autonomy. And societies should consider new regulatory responses and fundamental rights (such as the right 'to be off' or the right to 'attentional protection), as well as the possible need for a radical shift in values to a post-autonomy society.

Acknowledgements: I would like to thank Karen Yeung for her helpful comments on a previous draft of this paper. I would also like to thank the attendees at a Whitaker Institute seminar at NUI Galway for their criticisms and questions about the main arguments given above.

Bibliography

Alter, A. (2017). *Irresistible.* London: The Bodley Head.

Bartholomew, M. (2017). *Adcreep: The Case Against Modern Marketing*. Stanford, CA: Stanford University Press.

Brin, D. (1998). *The Transparent Society*. New York: Basic Books.

Brundage, M. and Danaher, J. (2017). Cognitive Scarcity and Artificial Intelligence: How Assistive AI Could Alleviate Inequality. *Philosophical Disquisitions* 15 May 2017,

available at http://philosophicaldisquisitions.blogspot.com/2017/05/cognitive-scarcity-and-artificial.html

Burrell, J. (2016). How the machine thinks: Understanding opacity in machine learning systems. *Big Data and Society*. DOI: 10.1177/ 2053951715622512. Available at: http://journals.sagepub. com/doi/abs/10.1177/2053951715622512

Carr, N. (2014) *The Glass Cage: Where Automation is Taking Us* (London: The Bodley Head)

Crawford, M. (2015). *The World Beyond Your Head*. New York: Farrar, Strauss and Giroux.

Danaher, J., Hogan, M. and Noone, C. et al (2017). Algorithmic Governance: Developing a research agenda through the power of collective intelligence. *Big Data and Society* 4(3) DOI: 10.1177/2053951717726554

Danaher, J. (2017). The Right to Attention in an Age of Distraction. *Philosophical Disquisitions* 19 May 2017, available at http://philosophicaldisquisitions.blogspot.com/2017/05/the-right-to-attention-in-age-of.html

Danaher, J. (2016). The Threat of Algocracy: Reality, Resistance and Accommodation. *Philosophy and Technology* 29(3): 245-268.

Eubanks, V. (2018). *Automating Inequality: How High-Tech Tools Profile, Police and Punish the Poor*. New York: St. Martin's Press.

Frankfurt, H. (1971). Freedom of the will and the concept of a person. *Journal of Philosophy*, 68, 5–20

Frischmann, B. and Selinger, E. (2018). *Re-Engineering Humanity*. Cambridge: Cambridge University Press.

Gal, M. (2017). Algorithmic Challenges to Autonomous Choice. *Michigan Journal of Law and Technology* – available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2971456##

Gräf, E. (2017). When Automated Profiling Threatens Freedom: a Neo-Republican Account. *European Data Protection Law Journal* 4: 1-11.

Harari, Y.N. (2016). *Homo Deus*. London: Harvill Secker.

Harari, Y.N. (2018). *21 Lessons for the 21st Century*. London: Jonathan Cape.

Hartzog, W. (2018) *Privacy's Blueprint*. Cambridge, MA: Harvard University Press.

Hoye, JM, and Monaghan, J. (2015). Surveillance, Freedom and the Republic. *European Journal of Political Theory*. DOI: 10.1177%2F1474885115608783

Killmister, J. (2017). Taking the Measure of Autonomy: A Four-Dimensional Theory of Self-Governance. London: Routledge.

Kitchin R (2017) Thinking critically about and researching algorithms. *Information, Communication and Society* 20(1): 14–29

Krakauer, D. (2016). Will AI Harm Us? Better to Ask How We'll Reckon with Our Hybrid Nature. *Nautilus* 6 September 2016 - available at **http://nautil.us/blog/will-ai-harm-us-better-to-ask-how-well-reckon-with-our-hybrid-nature** (accessed 29/11/2016)

Leiter, B. (2016). The Case Against Free Speech. *Sydney Law Review* 38: 407-439.

Morrow, D.R. (2014). When technologies makes good people do bad things: another argument against the value-neutrality of technologies. *Science and Engineering Ethics* 20(2):329-43

Morozov, E. (2013). The Real Privacy Problem. *MIT Technology Review*. Available at http://www.technologyreview.com/featuredstory/520426/the-real-privacy-problem/ (accessed 29/11/16)

Noble, S. (2018) *Algorithms of Oppression*. New York: NYU Press.

Ord, T. and Bostrom, N. (2006). The Reversal Test: Eliminating Status Quo Bias in Applied Ethics. *Ethics* 116: 656-679.

O'Neil, C. (2016). *Weapons of Math Destruction*. London: Penguin.

O'Shea, T. (2018). Disability and Domination. *Journal of Applied Philosophy* 35(1): 133-148

Pasquale F (2015) *The Black Box Society*. Cambridge, MA: Harvard University Press.

Pettit, P. (2001). Republicanism: A Theory of Freedom and Government. Oxford: OUP.

Pettit, P. (2011). The Instability of Freedom as Non-Interference: The Case of Isaiah Berlin. Ethics 121(4): 693-716

Pettit, P. (2014). Just Freedom: A Moral Compass for a Complex World. New York: WW Norton and Co.

Polonetsky J and Tene O (2013) Privacy and Big Data: Making ends meet. *Stanford Law Review* 66: 25–33.

Raz, J. (1986). *The morality of freedom*. Oxford: OUP

Rouvroy, A., (2013). Algorithmic Governmentality and the End (s) of Critique. *Society of the Query*, 2.

Rouvroy, A., (2015). Algorithmic governmentality: a passion for the real and the exhaustion of the virtual. All watched over by algorithms. Berlin. January 29th.

Scheibehenne, B., Greifeneder, R., & Todd, P. M. (2010). Can there ever be too many options? A meta-analytic review of choice overload. *Journal of Consumer Research*, 37, 409–425

Schwartz, B. (2004). *The paradox of choice: Why less is more.* New York, NY: Harper Collins.

Sunstein, C. (2016). *The ethics of influence*. Cambridge, UK: Cambridge University Press.

Thaler, R., & Sunstein, C. (2009). *Nudge: Improving decisions about health, wealth and happiness*. London: Penguin.

Tran, J. (2016). The Right to Attention. *Indiana Law Journal* 91(3): 1023-1062

Vertesi, J. (2014). Internet Privacy and What Happens When You Try to Opt Out. *Time* 1 May 2014.

Wu, T. (2017). *The Attention Merchants.* London: Atlantic Books.

Yeung, K. (2017a). ‘Hypernudge’: Big data as a mode of regulation by design. *Information, Communication and Society*, 20(1), 118–136

Yeung, K. (2017b). Algorithmic Regulation: A Critical Interrogation. *Regulation and Governance* DOI: 10.1111/rego.12158

Zarsky, T. (2012). Automated predictions: perception, law and policy. Communications of the ACM, 15(9), 33–35.

Zarsky, T. (2013). Transparent prediction. *University of Illinois Law Review*, 4, 1504

Danaher, J. (2016b). The threat of algocracy: Reality, resistance and accommodation. *Philosophy and Technology*, 29(3), 245–268