7   Autonomy

Autonomy is an important value in liberal societies. It is protected and cherished in both legal and popular culture. Some people argue that a life without it would not be worth living. In his speech to the Second Virginia Convention on March 23, 1775, Patrick Henry is said to have persuaded people to support the American Revolutionary War with a rousing speech that ended with the line, "Give me liberty or give me death." Similar thoughts are echoed in the New Hampshire state motto with its bold imperative to "live free or die." And it's not just in the United States that these sentiments find a welcome audience. The Greek national motto is "Eleftheria i Thanatos," which translates as "liberty or death." Furthermore, commitment to individual liberty and autonomy has often been a source of solace for those who find themselves in difficult circumstances. Nelson Mandela, for example, consoled himself and other inmates during his imprisonment on Robben Island by reciting William Ernest Henley's poem "Invictus." Written in the late 1800s while Henley was recovering from multiple surgeries, the poem is a paean to self-mastery, independence, and resilience, closing with the immortal lines, "I am the master of my soul / I am the captain of my fate."

Given the cherished status of individual autonomy in a liberal society, we must ask how AI and algorithmic decision-making might impact upon it. When we do, we find that there is no shortage of concern about the potentially negative impacts. Social critics of the technology worry that we will soon become imprisoned within the "invisible barbed wire" of predictive algorithms that nudge, manipulate, and coerce our choices.[1] Others such as the Israeli historian Yuval Noah Harari argue that the ultimate endpoint for the widespread deployment of AI is a technological infrastructure that replaces and obviates rather than nudges and manipulates autonomous

human decision makers.[2] But are they right? Is autonomy really imperiled by AI? Or could AI be just the latest in a long line of autonomy-enhancing technologies?

In this chapter, we step back from the hype and fear-mongering and offer a more nuanced guide to thinking about the relationship between AI and individual autonomy. We do this in four stages. First, we clarify the nature and value of autonomy itself, explaining what it is and how it is reflected in our legal systems. Second, we consider the potential impact of AI on autonomy, asking in particular whether the technology poses some novel and unanticipated threat to individual autonomy. Third, we ask whether the negative impacts of AI on autonomy are more likely to emanate from the private sector or the public sector, or some combination of the two. In other words, we ask, "Who should we fear more: big tech or big government?" And fourth, we consider ways in which we can protect individual autonomy in the era of advanced AI.

The overall position put forward in this chapter is that, although we must be vigilant against the threats that AI poses to our autonomy, it is important not to overstate those threats or assume that we, as citizens, are powerless to prevent them from materializing.

A Three-Dimensional Understanding of Autonomy

In order to think properly about the impact of AI on autonomy, we need to be clear about what we understand by the term "autonomy." The opening paragraph mingled together the ideals of autonomy, liberty, independence, and self-mastery. But are these all the same thing or are there important differences between them? Philosophers and political theorists have spent thousands of years debating this question. Some of them have come up with complex genealogies, taxonomies, and multi-dimensional models of what is meant by terms like "liberty" and "autonomy."[3] It would take several books to sort them all out and figure out who provides us with the best understanding of the relevant terms. Rather than do that, what we propose to do in this chapter is offer one specific model of autonomy. This model is inspired by long-standing philosophical debates, and those familiar with those debates will be able to trace the obvious influences, but you won't require any prior knowledge of them to follow the discussion. The goal is to provide readers with an understanding of autonomy that is reasonably

straightforward and self-contained but sufficiently nuanced to enable them to appreciate the multiple different impacts of AI on autonomy.

So what is this model of autonomy? It starts by acknowledging the common, everyday meaning of the term, which is that to be "autonomous" one must be able to choose one's own path in life, engage in self-rule, and be free from the interference and manipulation of others. This common understanding is a good starting point. It captures the idea that to be autonomous requires some basic reasoning skills and abilities—specifically, the ability to pick and choose among possible courses of action—and some relative independence in the exercise of those skills and abilities.

The legal and political philosopher Joseph Raz adds some further flesh to the bones of this model of autonomy in a famous definition of what it takes to be autonomous.

> 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.[4]

Raz's definition breaks autonomy down into three component parts. It says that you are autonomous if (1) you have the basic rationality required to act in a goal-directed way; (2) you have an adequate range of options to choose from; and (3) your choice among those options is independent, that is, free from coercion and manipulation by others. This definition can be applied to both individual decisions and lives as a whole. In other words, following Raz, we can consider whether a person's whole life is autonomous or whether a particular decision or set of decisions is autonomous. Some people like to use different terms to differentiate between the different scopes of analysis. For instance, the philosopher Gerald Dworkin once suggested that we use the term "autonomy" when talking about life as a whole (or some extended portion of life) and "freedom" when talking about individual decisions. This, however, is confusing because, as we shall see below, the term "freedom" has also been applied just to Raz's third condition of autonomy. So, in what follows, we'll just use the one word—"autonomy"—to refer to the phenomenon in which we are interested. Furthermore, we won't really be

discussing the impact of AI on life as a whole but rather its impact on specific decisions or specific decision-making contexts.

It's tempting to think about Raz's three components of autonomy as conditions that need to be satisfied in order for a decision to count as autonomous. If you satisfy all three, then you are autonomous; if you do not, then you are not. But this is too binary. It's actually more helpful if we think about the three components as *dimensions* along which the autonomy of a decision can vary. We can call these the rationality, optionality, and independence dimensions, respectively. There may be a minimum threshold that needs to be crossed along each dimension before a decision will qualify for autonomous status, but beyond that threshold decisions can be more or less autonomous.

To make this more concrete, compare two different decisions. The first decision involves you choosing among movie options that are recommended to you by your Netflix app (or other video streaming app). The app gives you ten recommended movies. You read the descriptions of them and choose the one in which you are most interested. The second decision involves you choosing a walking route that has been recommended to you by Google Maps (or some other mapping service). The app gives one recommended route, highlighted in blue, and one other similar route in a less noticeable grey. You follow the highlighted route. Are both of these choices autonomous? Probably. We can assume that you have the basic rationality needed to act in a goal-directed way; we can see that you have been given a range of options by the apps; and the apps do not obviously coerce and manipulate your choices (though we'll reassess this claim below). But is one decision more autonomous than the other? Probably. On the face of it, it seems like the Netflix app, by giving more options and information and not highlighting or recommending one in particular scores higher along the second and third dimensions. It gives you more optionality and more independence.

Thinking about autonomy in this three-dimensional way is useful. It allows us to appreciate that autonomy is a complex phenomenon. It encourages us to avoid simplistic and binary thinking. We can now appreciate that autonomy is not some simple either/or phenomenon. Decisions can be more or less autonomous, and their autonomy can be undermined or promoted in different ways. In this respect, it is important to realize that there may be practical tradeoffs between the different dimensions and that we need to balance them in order to promote autonomy. For example, adding more options to a decision may only promote autonomy *up to a point.*

Beyond that point, adding more options might be overwhelming, make us more confused, and compromise our basic rationality. This is something that psychologists have identified as "the paradox of choice" (more on this later).[5] Similarly, some constraints on options and some minimal coercion or interference might be necessary to make the most of autonomy. This has long been recognized in political theory. Thomas Hobbes, for example, in his famous defense of the sovereign state, argued that we need some minimal background coercion by the state to prevent us from sinking into bitter strife and conflict with one another.

Although all aspects of the three-dimensional model of autonomy deserve scrutiny, it is worth focusing a little more attention on the independence dimension. This dimension effectively coincides with the standard political concepts of "freedom" and "liberty." The key idea underlying this dimension is that, in order to be autonomous, you must be free and independent. But what does that freedom and independence actually entail? Raz talks about it requiring the absence of coercion and manipulation, but these concepts are themselves highly contested and can manifest in different ways. Coercion, for instance, usually requires some threatened interference if a particular option is not chosen ("do this or else!"), whereas manipulation can be more subtle, often involving some attempt to brainwash or condition someone to favor a particular option. In modern political theories of freedom, there is also an important distinction drawn between freedom understood as the absence of *actual* interference with decision-making and freedom understood as the absence of *actual* and *potential* interference. The former is associated with classical liberal theories, such as those put forward by John Locke and Thomas Hobbes; the latter is associated with republican theories of freedom, such as those favored by Machiavelli and, more recently, the Irish philosopher Philip Pettit (not to be confused with the Republican party in the US).[6]

We can explain the distinction with a simple example taken from Pettit. Imagine a slave, that is, someone who is legally owned and controlled by another human being. Now imagine that the slave owner just happens to be particularly benevolent and enlightened. They have the legally recognized power to force the slave to do as they will, but they don't exercise that power. As a result, the slave lives a relatively happy life, free from any actual interference. Is the slave free? Proponents of the republican theory of freedom would argue that the slave is not free. Even though they are

not actively being interfered with, they are still living in a state of *domination*. The benevolent slave master might only allow the slave to act within certain arbitrarily defined parameters, or they might change their mind at any moment and step in and interfere with the slave's choices. That's the antithesis of freedom. The problem, according to the republicans, is that the classical liberal view cannot account for that. We need to consider both actual and potential interferences with decision-making if we are to properly protect individual liberty. A person cannot be free if they live under the potential domination of another.

For what it's worth, we think that this is a sensible position to adopt. The bottom line, then, is that we need to ensure that our three-dimensional model of autonomy includes both actual and potential interferences in its approach to the independence dimension.

That's enough about the nature of autonomy. What about its value? Why would some people rather die than do without it? We can't exactly answer that question here, but we can at least map out some of the ways it *could* be answered. We'll leave it up to the individual reader to determine how important autonomy is to them.

Broadly speaking, there are two ways to think about the value of autonomy. The first is to think of autonomy as something that is *intrinsically* valuable, that is, valuable in itself, irrespective of its consequences or effects. If you take that view, then you'll think of autonomy as one of the basic conditions of human flourishing. You'll probably believe that humans cannot really live good lives unless they have autonomy. You may even go so far as to think that a life without autonomy is not worth living. The second way to think about it is as something that is *instrumentally* valuable, that is, valuable because of its typical consequences or effects. If you take this view, you might highly prize autonomy but only because you think autonomy is conducive to good outcomes. This is a common view and features widely in political and legal justifications for autonomy. The idea is that people are better off if they are left to do their own thing. They can pick and choose the things that are most conducive to their well-being rather than having the state or some third party do this for them. The counterpoint to this, of course, is the *paternalistic* view, which holds that people don't always act in their best interests and sometimes need a helping hand. The battle between paternalism and autonomy is rife in modern political debates. At one extreme, there are those who decry the "nanny state" for

always stepping in and thinking it knows best; at the other extreme, there are those that lament the irrationality of their fellow citizens and think we would all become obese addicts if left to decide for ourselves. Most people probably take up residence somewhere between these two extremes.

It is possible, of course, to value autonomy for both intrinsic and instrumental reasons, to think that it is valuable in and of itself and more likely to lead to better outcomes. Even more exotic views are available too. For example, one of the authors of the present book has defended the claim that autonomy is neither instrumentally nor intrinsically valuable but rather something that makes good things better and bad things worse.[7] Imagine, for example, a serial killer who autonomously chooses to kill lots of people versus a serial killer who has been brainwashed into killing lots of people. Whose actions are worse? The obvious answer is the former. The latter's lack of autonomy seems mitigating. It is still bad, of course, for many people to have died as a result of his actions, but it seems worse, all things considered, if they died as a result of an autonomous choice. It suggests a greater evil is at work. The same is true for good things that happen accidentally versus those that happen as the result of autonomous action.

Whichever view you take, there are additional complexities to consider. For starters, you need to consider where in the hierarchy of values autonomy lies. We value lots of things in life, including health, sociality, friendship, knowledge, well-being, and so on. Is autonomy more or less important than these other things? Or is it co-equal? Should we be willing to sacrifice a little autonomy to live longer and happier lives (as the paternalists might argue)? Or do we agree with Patrick Henry that the loss of liberty is a fate worse than death? How we answer those questions will play a big role in how seriously we take any threat that AI may pose to our liberty.

An analogy might help. We discussed the importance of privacy in chapter 6. Lots of philosophers, lawyers, and political theorists think that privacy is important, but some people question our commitment to privacy. After all, the lesson of the internet age seems to be that people are quite willing to waive their right to privacy in order to access fast and efficient digital services. This leads some to argue that we might be transitioning to a post-privacy society. Could something similar be true for autonomy? Might we be willing to waive our autonomy in order to gain the benefits of AI? Could we be transitioning to a post-autonomy society? This is something to take seriously as we look at the potential threats to autonomy.

Finally, you should also consider the relationship between autonomy (however it is valued) and other basic legal rights and freedoms. Many classic negative legal rights, for example, have a firm foundation in the value of autonomy. Examples would include freedom of speech, freedom of movement, freedom of contract, freedom of association, and, of course, privacy and the right to be left alone. Although there are economic and political justifications for each these freedoms, they can also be justified by the value of autonomy, and their protection can help to promote autonomy. In short, it seems like a good case can be made for the view that autonomy, however it is valued, is foundational to our legal and political framework. We all have some interest in it.

Do AI and Algorithmic Decision Making Undermine Autonomy?

Now that we have a deeper understanding of both the nature and value of autonomy we turn to the question at hand: does the widespread deployment of AI and algorithmic decision-making undermine autonomy? As mentioned, critics and social commentators have already suggested as much, but armed with the three-dimensional model, we can undertake a more nuanced analysis. We can consider the impact of these new technologies along all three dimensions of autonomy.

Doing so, we must admit upfront that there is some cause for concern. Consider the first of the three dimensions outlined above: basic rationality. Earlier chapters in this book have highlighted various ways in which this might be affected by AI and algorithmic decision-making tools. Basic rationality depends on our ability to understand how our actions relate to our goals. This requires some capacity to work out the causal structure of the world and to pick actions that are most likely to realize our goals. One of the problems with AI is that it could prevent us from working these causal relationships out. Consider the problem of opacity and the lack of explainability we discussed in chapter 2. It's quite possible that AI systems will just issue us with recommended actions that we ought to follow without explaining why or how those options relate to our goals. We'll just have to take it on faith. In a recent book entitled *Re-Engineering Humanity,* Brett Frischmann and Evan Selinger articulated this fear in stark terms. They argue that one of the consequences of excessive reliance on AI would be the re-programming of humans into a simple "stimulus-response" machines.[8]

Humans will see the recommendations given to them by AI and implement the response without any critical reflection or thought. Basic rationality is compromised.

Similar problems arise when we look at the optionality dimension. One of the most pervasive uses of AI in consumer-facing services is to "filter" and constrain options. Google, for example, works by filtering and ordering possible links in response to our search queries. It thereby reduces the number of options that we have to process when finding the information we want. Netflix and Amazon product recommendations work in a similar way. They learn from your past behavior (and the behavior of other customers) and give you a limited set of options from a vast field of potential products. Often this option-filtering is welcome. It makes decisions more manageable. But if the AI system only gives us one or two recommendations, and gives one of them with a "95 percent" confidence rating, then we have to query whether it is autonomy preserving. If we have too few options, and if we are dissuaded from thinking critically or reflectively about them, then we arguably lose something necessary for being the authors of our own lives. Chapter 5 on control is really a particular illustration of this problem.

Finally, the independence dimension can also be compromised by the use of AI. Indeed, if anything, the independence dimension is likely to be the most obvious casualty of pervasive AI. Why so? Because AI offers many new opportunities for actual and potential interferences with our decision-making. Outright coercion by or through AI is one possibility. An AI assistant might threaten to switch off a service if we don't follow a recommendation, for example. Paternalistic governments and companies (e.g., insurance companies) might be tempted to use the technology in this way. But less overt interferences with decision-making are also possible. The use of AI-powered advertising and information management can create filter bubbles and echo chambers.[9] As a result, we might end up trapped inside technological "Skinner boxes," in which we are rewarded for toeing an ideological party-line and thereby manipulated into a set of preferences that is not properly our own. There is some suggestion that this is already happening, with many people expressing concern about the impact of AI-mediated fake news and polarizing political ads in political debates.[10] Some of these ads potentially exemplify the most subtle and insidious form of domination, what Jamie Susskind refers to as *perception-control*.[11] Perception-control is, literally, the attempt to influence the way we perceive the world. Filtering is the lynchpin

of perception-control. The world is messy and bewildering, so our transactions with it have to be mediated to avoid our being swamped in detail. Either we do this sifting and sorting ourselves, or, more likely, we rely on others to do it for us, such as commercial news outlets, social media, search engines, and other ranking systems. In each case, someone (or *something*) is making choices about how much is relevant for us to know, how much context is necessary to make it intelligible, and how much we need to be told. Although in the past filtering was done by humans, increasingly it is being facilitated by sophisticated algorithms that appeal to us based on an in-depth understanding of our preferences. These preferences are reliably surmised through our retail history, Facebook likes and shares, Twitter posts, YouTube views, and the like. Facebook, for example, apparently filters the news it feeds you based on about 100,000 factors "including clicks, likes, shares [and] comments."[12] In the democratic sphere, this technology paves the way to active manipulation through targeted political advertising. "Dark" ads can be sent to the very people most likely to be susceptible to them without the benefit—or even the possibility—of open refutation and contest that the marketplace of ideas depends on for its functioning. The extent of perception-control that new digital platforms make possible is really a first in history, and likely to concentrate unprecedented power in the hands of a few big tech giants and law-and-order-obsessed state authorities (see below).[13]

On top of all this, the widespread use of mass surveillance by both government and big tech companies creates a new form of domination in our lives. We are observed and monitored at all times inside a digital panopticon, and although we might not be actively interfered with on an ongoing basis, there is always the possibility that we might be if we step out of line. Our position then becomes somewhat akin to that of a digital slave living under the arbitrary whim of our digital masters. This is the very antithesis of the republican conception of what it means to be autonomous.[14]

But there is another story to tell. It is easy to overstate the scare-mongering about AI and the loss of autonomy. We have to look at it from all sides. There are ways in which AI can promote and enhance autonomy. By managing and organizing information into useful packages, AI can help us sort through the complex mess of reality. This could promote rationality, not hinder it. Similarly, and as noted, the filtering and constraining of options, provided it is not taken too far, can actually promote autonomy by making decision-making problems more cognitively manageable. With too many options we

get stuck and become unsure of what to do. Narrowing down the field of options frees us up again.[15] AI assistants could also protect us from outside forms of interference and manipulation, acting as sophisticated ideological spam filters, for example. Finally, and more generally, it is important to recognize that technologies such as AI have a tendency to increase our power over the world and enable us to do things that were not previously possible. Google gives us access to more useful (and useless!) information than any of our ancestors; Netflix gives us access to more entertainment; AI assistants like Alexa and Siri allow us to efficiently schedule and manage our time. The skillful use of AI could be a great boon to our autonomy.

There is also the danger of status quo bias or loss aversion to consider.[16] Often when a new technology comes onboard we are quick to spot its flaws and identify the threats it poses to cherished values such as autonomy. We are less quick to spot how those threats are already an inherent part of the status quo. We have become desensitized to those. This seems to be true of the threats that AI poses to our autonomy. It is impossible to eliminate all possible threats to autonomy. Humans are not perfectly self-creating masters of their own fates. We depend on our natural environment and on one another. What's more, we have a long and storied history of undermining one another's autonomy. We have been undermining rationality, constraining options, and ideologically manipulating one another for centuries. We have done this via religious texts and government diktats. But we have also created reasonably robust constitutional and legislative frameworks that protect against the worst of these threats. Is there any reason to think that there is something special or different about the threats that AI poses to autonomy?

Perhaps. Although threats to autonomy are nothing new, AI does create new modalities for realizing those threats. For example, one important concern about AI is how it might concentrate autonomy-undermining power within a few key actors (e.g., governments and big tech firms). Historically, autonomy-undermining power was more widely dispersed. Our neighbors, work colleagues, friends, families, states, and churches could all try to interfere with our decision-making and ideologically condition our behavior. Some of these actors were relatively powerless, and so the threat they posed could be ignored on a practical level. There was also always the hope that the different forces might cancel each other out or be relatively easy to ignore. That might no longer be true. The internet connects us all together and creates an environment in which a few key firms (e.g., Facebook,

Google, Amazon) and well-funded government agencies can play an out-sized role in constraining our autonomy. What's more, on the corporate side, it is quite possible that all the key players have a strong incentive to ideologically condition us in the same direction. The social psychologist and social theorist Shoshana Zuboff, in her book, *The Age of Surveillance Capitalism*, argues that all dominant tech platforms have an incentive to promote the ideology of surveillance capitalism.[17] This ideology encourages the extraction of value from individual data and works by getting people to normalize and accept mass digital surveillance and the widespread use of predictive analytics. She argues that this can be done in subtle and hidden ways, as we embrace the conveniences of digital services and normalize the costs this imposes in terms of privacy and autonomy. The same is true when governments leverage AI. The Chinese government, for example, through its social credit system (which is facilitated through partnerships with private enterprises) uses digital surveillance and algorithmic scoring systems to enforce a particular model of what it means to be a good citizen.[18] The net result of the widespread diffusion of AI via the internet is that it gives a handful of actors huge power to undermine autonomy. This might be a genuinely unprecedented threat to autonomy.

AI could also erode autonomy in ways that are much more difficult for individuals to resist and counteract than the traditional threats to autonomy. The idea of the "nudge" has become widespread in policy circles in the past decades. It was first coined by Cass Sunstein and Richard Thaler in their book, *Nudge*.[19] Sunstein and Thaler were interested in using insights from behavioral and cognitive psychology to improve public policy. They knew that decades of research revealed that humans are systematically biased in their decision-making. They don't reason correctly about probability and risk; they are loss averse and short-termist in their thinking.[20] As a result, they often act contrary to their long-term well-being. Knowing all this seems like a justification for paternalistic intervention. Individuals cannot be trusted to do the right thing so someone else must do it for them. Governments, for example, need to step in and put people on the right track. The question posed by Sunstein and Thaler was "how can we do this without completely undermining autonomy?" Their answer was to nudge, rather than coerce, people into doing the right thing. In other words, to use the insights from behavioral psychology to gently push or incentivize people to do the right thing, but always leave them with the option of rejecting

the nudge and exercising their own autonomy. This requires careful engineering of the "choice architectures" that people confront on a daily basis so that certain choices are more likely. Classic examples of "nudges" recommended by Sunstein and Thaler include changing opt-in policies to opt-out policies (thereby leveraging our natural laziness and loyalty to the status quo), changing how risks are presented to people to accentuate the losses rather than the gains (thereby leveraging our natural tendency toward loss aversion), and changing the way in which information is presented to people in order to make certain options more salient or attractive (thereby leveraging natural quirks and biases in how we see the world).

There has been much debate over the years about whether nudges are genuinely autonomy-preserving. Critics worry that nudges can be used manipulatively and nontransparently by some actors and that their ultimate effect is to erode our capacity to make choices for ourselves (since nudges work best when people are unaware of them). Sunstein maintains that nudges can preserve autonomy if certain guidelines are met.[21] Whatever the merits of these arguments, the regulatory theorist Karen Yeung has argued that AI tools facilitate a new and more extreme form of nudging, something she calls "hypernudging."[22] Her argument is that persistent digital surveillance combined with real time predictive analytics allows software engineers to create digital choice architectures that constantly adapt and respond to the user's preferences in order to nudge them in the right direction. The result is that as soon as the individual learns to reject the nudge, the AI system can update the choice architecture with a new nudge. There is thus much reduced capacity to resist nudges and exercise autonomy.

In a similar vein, AI enables a far more pervasive and subtle form of domination over our lives. Writing about this in a different context, the philosopher Tom O'Shea argues that there is such a thing as "micro-domination."[23] This arises when lots of small-scale, everyday decisions can only be made under the guidance and with the permission of another actor (a "dominus" or master). O'Shea gives the example of a disabled person living in an institutional setting to illustrate the idea. Every decision they make—including when to get up, when to go to the bathroom, when to eat, when to go outside, and so on—is subject to the approval of the caretakers employed by the institution. If the disabled person goes along with the caretakers' wishes (and the institutional schedule), then they are fine, but if they wish to deviate from that, they quickly find themselves unable to do what they

want. Taken individually, these decisions are not particularly significant—they do not implicate important rights or life choices—but taken together all these "micro" instances of domination constitute a significant potential interference with individual autonomy.

The pervasive use of AI could result in something similar. Consider a hypothetical example: a day in the life of Jermaine. Just this morning Jermaine was awoken by his sleep monitoring system. He uses it every night to record his sleep patterns. Based on its observations, it sets an alarm that wakes him at the optimal time. When Jermaine reaches his work desk, he quickly checks his social media feeds where he is fed a stream of information that has been tailored to his preferences and interests. He is encouraged to post an update to the people who follow his work ("the one thousand people who follow you on Facebook haven't heard from you in a while"). As he settles into his work, his phone buzzes with a reminder from one of his health and fitness apps that it's time to go for a run. Later in the day, as he drives to a meeting across town, he uses Google Maps to plot his route. Sometimes, when he goes off track, it recalculates and sends him in a new direction. He dutifully follows its recommendations. Whenever possible, he uses the autopilot software on his car to save time and effort, but every now and then it prompts him to take control of the car because some obstacle appears that it's not programmed to deal with. We could multiply the examples, but you get the idea. Many small-scale, arguably trivial, choices in Jermaine's everyday life are now subject to an algorithmic master: what route to drive, whom to talk to, when to exercise, and so on. As long as he works within the preferences and options given to him by the AI, he's fine. But if he steps outside those preferences, he will quickly realize the extent of his dependence and find himself unable to do quite as he pleases (at least until he has had time to adjust to the new normal).

Indeed, this isn't a purely hypothetical concern; we already see it happening. The sociologist Janet Vertesi has documented how both she and her husband were flagged as potential criminals when they tried to conceal the fact that she was pregnant from online marketers who track purchasing data, keyword searches, and social media conversations.

> For months I had joked to my family that I was probably on a watch list for my excessive use of Tor and cash withdrawals. But then my husband headed to our local corner store to buy enough gift cards to afford a stroller listed on Amazon. There, a warning sign behind the cashier informed him that the store "reserves the right to limit the daily amount of prepaid card purchases and has an obligation to report excessive transactions to the authorities.[24]

She took one step outside the digital panopticon and was soon made aware of its power.

The net effect of both AI-mediated hypernudging and micro-domination is likely to be a form of learned helplessness. We might want to be free from the power of AI services and tools, but it is too difficult and too overwhelming to rid ourselves of their influence. Our traditional forms of resistance no longer work. It is easier just to comply.

All of this paints a dystopian picture and suggests that there might be something genuinely different about the threat that AI poses to autonomy. It may not pose a new category of threat, but it does increase the scope and scale of traditional threats.

But, again, some degree of skepticism is needed before we embrace this dystopian view. The threats just outlined are all inferred from reflections on the nature of AI and algorithmic power. They are not based on the careful empirical study of their actual effects. Unfortunately, there are not that many empirical studies of these effects just yet, but the handful that do exist suggest that some of the threats outlined above do not materialize in practice. For example, the ethnographer Angèle Christin has conducted in-depth studies of the impacts of both descriptive and predictive analytical tools in different working environments.[25] Specifically, she has looked at the impact of real-time analytics in web journalism and algorithmic risk predictions in criminal courts. Although she finds that there is considerable hype and concern about these technologies among workers, the practical impact of the technology is more limited. She finds that workers in both environments frequently ignore or overlook the data and recommendations provided to them by these algorithmic tools. Her findings in relation to risk prediction in criminal justice are particularly striking and important given the focus of this book. As noted in previous chapters, there is much concern about the potential bias and discrimination inherent in these tools. But Christin finds that very little attention is paid to them in practice. Most lawyers, judges, and parole officers either ignore them entirely or actively "game" them in order to reach their own preferred outcomes. These officials also express considerable skepticism about the value of these tools, questioning the methodology underlying them and usually favoring their own expert opinion over that provided to them by the technology.[*]

---

[*] In chapter 5, we saw that AI doesn't induce complacency when the technology is *perceived* to be suboptimal.

Similarly, some of the concerns that have been expressed about ideological conditioning, particularly in the political sphere, appear to be overstated. There is certainly evidence to suggest that fake news and misinformation is spread online by different groups and nation states.[26] They often do this through teams of bots rather than real human beings. But research conducted by the political scientists Andrew Guess, Brendan Nyhan, Benjamin Lyons and Jason Reifler suggests that people don't get trapped in the digital echo chambers that so many of us fear, that the majority of us still rely on traditional mainstream news media for our news, and in fact that we're more likely to fall into echo chambers *off*line than online![27]

Findings like these, combined with the possible autonomy-enhancing effects of AI, provide some justification for cautious optimism. Unless and until we are actively forced to use AI tools against our will, we have more power to resist their ill effects than we might realize. It is important that this is emphasized and that we don't get seduced by a narrative of powerlessness and pessimistic fatalism.

Big Tech or Big Government?

A minor puzzle emerges from the foregoing discussion. We have canvassed the various threats that AI poses to autonomy. We have made the case for nuance and three-dimensional thinking when it comes to assessing those threats. But we have been indiscriminate in how we have thought about the origin of those threats. Is it from big tech or big government that we have most to fear? And does it matter?

You might argue that it does. There's an old trope about staunch libertarians of the pro-business, anti-government type (the kind you often find in Silicon Valley): they are wary of all threats to individual liberty that come from government but seem wholly indifferent to those that come from private businesses.

On the face of it, this looks like a puzzling inconsistency. Surely all threats to autonomy should be taken equally seriously? But there is some logic to the double-standard. Governments typically have far more coercive power than businesses. While you can choose to use the services of a business, and there is often competition among businesses that gives you other options, you *have* to use government services: you can't voluntarily absent yourself from them (short of emigrating or going into exile). If we

accept this reasoning, and if we review the arguments outlined above, we might need to reconsider how seriously we take the threats that AI poses to individual autonomy. Although some of the threats emanate from government, and we should be wary of them, the majority emanate from private enterprises and may provide less cause for concern.

But this is not a particularly plausible stance in the modern day and age. For one thing, there is often a cozy relationship between government and private enterprise with respect to the use of AI. This topic will be taken up in more detail in the next chapter, but for now we can simply note that governments often procure the services of private enterprises in order to carry out functions that impact on citizens' rights. The predictive analytics tools that are now being widely used in policing and criminal sentencing, for example, are ultimately owned and controlled by private enterprises that offer these services to public bodies at a cost. Similarly, the Chinese Social Credit System—which is probably the most invasive and pervasive attempt to use digital surveillance and algorithmic scoring to regulate citizens' behavior—is born of a cozy relationship between the government and the private sector. So it is not easy to disentangle big government from big tech in the fight to protect autonomy.

More controversially, it could be argued that at least some big tech firms have taken on such an outsized role in modern society that they need to be held to the same standards (at least in certain respects) as public bodies. After all, the double-standard argument only really works if there is reasonable competition among private services, and people do in fact have a choice of whether or not to use those services. We may question whether that is true when it comes to the goods and services offered by the tech giants like Google, Amazon, Facebook, and Apple. Even when there is some competition among these firms, they tend to sing from the same "surveillance capitalist" hymnbook.*

In a provocative paper entitled "If Data Is the New Oil, When Is the Extraction of Value from Data Unjust?"[28] the philosopher Michele Loi and the computer scientist Paul Olivier DeHaye have argued that "dominant tech platforms" should be viewed as basic social structures because of their pervasive influence over how we behave and interact with one another.

---

*A potential exception might be Apple, which is trying to position itself as the privacy-protecting big tech firm.

The best examples of this are the large social media platforms such as Facebook and Twitter. They affect how we communicate and interact with one another on a daily basis. Loi and DeHaye argue that, when viewed as basic social structures, these dominant tech platforms must be required to uphold the basic principles of social and political justice, which include protecting our fundamental freedoms such as freedom of speech and freedom of association. Platforms like Facebook even seem to be waking up to their responsibilities in this regard (although more cynical interpretations are available).[29] In any case the distinction between big tech and big government is, at least in some cases, a spurious one.

Finally, even if you reject this and think there are good grounds for distinguishing between the threats from big tech and big government, it is important to realize that just because it might be appropriate to take threats from government more seriously than threats from private enterprise, it doesn't mean we can discount or ignore the latter. Citizens still have an interest in ensuring that private enterprise does not unduly compromise their autonomy, and governments have a responsibility to prevent that from happening.

What Should We Do about It?

Now that we have a clearer sense of the nature and value of autonomy as well as the potential threats to which AI exposes it, we can turn to the question of what to do about these threats. We can start by noting that how you approach this question will depend on how you value autonomy. If you think autonomy is not particularly valuable or that it is only valuable to the extent that it enhances individual well-being, then you might be relatively sanguine about our current predicament. You might, for example, embrace the paternalistic ethos of hypernudging and hope that, through AI assistance, we can combat our biases and irrationalities and lead longer, healthier, and happier lives. In other words, you might be willing to countenance a post-freedom society because of the benefits it will bring.

But if you reject this and think that autonomy is an important, perhaps central value, you will want to do something to promote and protect it. There is an old saying widely but probably falsely attributed to Thomas Jefferson that states that "eternal vigilance is the price we pay for liberty." Whatever its provenance, this seems like a good principle to take with us

as we march into the brave new world that is made possible by pervasive AI. Comprehensive legal regulations on the protection of personal data and privacy—such as the EU's General Data Protection Regulation—are consequently a good start in the fight to protect autonomy because they give citizens control over the fuel (data) that feeds the fire of AI. Having robust consent protocols that prevent hidden or unknown uses of data, along with transparency requirements and rights to control and delete data, are all valuable if we want to protect autonomy.

But they may not be enough by themselves. We may, ultimately, require specific legal rights and protections against the autonomy-undermining powers of AI. In this respect, the proposal from Brett Frischmann and Evan Selinger, for the recognition of two new fundamental rights and freedoms is worth taking seriously.[30] Frischmann and Selinger argue that in order to prevent AI and algorithmic decision-making systems from turning us into simple, unfree stimulus-response machines, we need to recognize both a freedom to be off (i.e., to not use technology and not be programmed or conditioned by smart machines) and a freedom from engineered determinism. Frischmann and Selinger recognize that we cannot be completely free from the influence and interference of others. We are necessarily dependent on each other and on our environments. But the dependence on AI is, they argue, different and should be treated differently from a legal and political perspective. Too much dependence on AI and we will corrode our capacity for reflective rationality and independent thought. Indeed, at its extreme, dependence on AI will obviate the need for autonomous choice. The AI will just do the work for us. The two new freedoms are designed to stop us from sliding down that slippery slope.

As they envisage it, these freedoms would entail a bundle of positive and negative rights. This bundle could include rights to be let alone and reject the use of dominant tech platforms—essential if we are to retain our capacity to resist manipulation and interference—as well as positive obligations on tech platforms and governments to bolster those capacities for reflective rationality and independent thought. The full charter of relevant rights and duties remains to be worked out, but one thing is obvious: public education about the risks that the tech poses to autonomy is essential. As Jamie Susskind notes in *Future Politics*, an active and informed citizenry is ultimately the best bulwark against the loss of liberty.[31]

Summing Up

So what can be said by way of conclusion? Nothing definitive, unfortunately. Autonomy is a relatively modern ideal.[32] It is cherished and protected in liberal democratic states but is constantly under siege. The rise of AI introduces new potential threats that are maybe not categorically different from the old kind but are different in their scope and scale. At the same time, AI could, if deployed wisely, be a boon to autonomy, enhancing our capacity for rational reflective choice. It is important to be vigilant and perhaps introduce new rights and legal protections against the misuse of AI to guarantee autonomy in the future. However we think about it, we must remember that autonomy is a complex, multidimensional ideal. It can be promoted and attacked in many different ways. If we wish to preserve it, it is important to think about it in a multidimensional way.