Members of the general public may think that terms like ‘Big Data’ are only of relevance to technology geeks and Silicon Valley executives. The reality is that so-called “datafication” marks the beginning of a new human epoch that will have huge implications for all of us – especially generations being born right now. Understanding the ethics of tech has never been more critical than it is today, and any comprehensive analysis should have one of the most apparent challenges right at its core: what Big Data means for our personal autonomy. Some commentators have already expressed nervousness. They are concerned that data-driven technology could lead to the erosion of some of our human capacities as we relinquish more and more of our decision-making to computers.

This paper attempts to frame this emerging concern, before articulating three ways in which an increasing emphasis on Big Data seems to threaten our basic liberty. I identify these as: i) data overload and automation, ii) feedback loops and manipulation, and iii) types and prejudice. I will then argue that, although these factors undoubtedly present a challenge to aspects of our decision-making (and so ethical concerns aren’t entirely misplaced), human autonomy itself is not in danger of being significantly destabilized. Rather, the rapid shift in perspective that characterizes this new era of data, intelligence and mass connectivity simply demands that we reimagine the objects, but not the conditions, of agent autonomy.

I will suggest ways in which we might mitigate some of the more pernicious aspects of these developments, before ultimately concluding that new attitudes and new opportunities for decision-making are actually counteractively extending the domain of the autonomous human agent in positive ways.

Keywords: autonomy, Big Data, decision-making, ethics, technology, society.

---

1 Independent scholar. Email: <fionajmcevoy@gmail.com>.
1. Introduction

Imagine you’re driving through the country. You accelerate through a green light at a quiet, unlit intersection and immediately collide with another car, sending you into a terrifying spin. When the car comes to a rest, you discover that you’ve suffered some minor injuries. What next? Well, according to information experts (IoT Data Panel, 2016) the way in which this story unravels is about to change forever. They say that, not long from now, the smart device in your car will automatically contact a local ambulance. At the same time, it will communicate full details of the incident and key information about your medical history and insurance. It will also – simultaneously – inform your car insurers, summon a drone to take photographs of the scene, and notify whichever city hall is responsible for the faulty traffic light that caused the incident. It may even contact your family, and your lawyer. This incredible, holistic, automated response will issue in just seconds without your having to deliberate, reason, or take a single decision. You will barely even have to think. Pre-programmed AI will either already know, or be able to precisely anticipate, your wishes in such a scenario.

This is not a Jetsons-style future; it is the foreseeable trajectory of automated decision-making. The technology already exists. According to Intel, the connected car is already the third fastest growing technological device after phones and tablets (Floridi 2014, 9). It is only a matter of (short) time before all new cars will be connected to the internet. Critically for this paper, new technology like this is just part of much larger picture within which our data is being used to fuel automated processes that will cater to our every need. If you’re reading this around the time of writing you’re probably part of the last generation to experience a clear distinction between online and offline environments (Floridi 2014, 94).

As we eagerly wade into this brave new data-driven world there are many questions to ask. Here I will focus squarely on one of them:

---

2 Estimates predict that there will be around 50 billion connected devices by 2020 (Floridi 2014, 11).
what does this information revolution mean for us as autonomous human agents? Many have already noted that great technological shifts have significant implications for our self-directed thought and action (Klous and Wieland 2016; Mayer-Schönberger and Cukier 2013; Floridi 2014; Harari 2016).

In what comes next, I will define what I mean by the terms ‘Big Data’ and ‘autonomy’, and provide a brief discussion of how our autonomy has functioned in modernity until now (i.e. what might be ‘under threat’). Then I will describe three examples of how the information revolution, and its embrace of Big Data, presents a challenge to our current understanding of ourselves as autonomous. Lastly, I will argue that Big Data does not represent a threat to autonomy in itself, but rather heralds a sea-change in objects of our autonomous choices. I will conclude that having so many of our actions and decisions wrested from us by this intelligent, data-fueled technology could actually yield much greater human freedoms than we have known until now.

2. Autonomy

Through all but our most recent recorded history, we humans have relied upon gods and other spiritual forces to guide us in times of uncertainty. When our ancestors wanted to know how to act or what to think they consistently looked outside of themselves, and faithfully took instruction from scriptures, sages, priests, and prophets. Then, in the 18th century, humanism came along and largely upended this long-standing convention. For many, this marked the beginning of the modernity in which we now live. An age where we are encouraged to introspect in order to find the solutions to our conundrums.

Unlike religion, humanism informs us that our needs, desires and truths are internal to us, and only through reflection can we access our inner selves and act freely in accordance with its wishes. As Polonius says to Laertes: “This above all: to thine own self be true” (Hamlet, II.III.78).

Broadly speaking, this is how we still take decisions today: what brand of tomatoes to buy, which school to apply to, what to name
our kids, where to go on vacation. For many, this type of reflective self-determination is one of the central tenets of what it is to be human. It resembles some notions of ‘positive liberty’, described as when I am, “conscious of myself as thinking, willing, active being, bearing responsibility for my choices and able to explain them by references to my own ideas and purposes” (Berlin 1969, 131). We might also call this a (internal) capacity for autonomy.

This idea is often in contrasted with notions of ‘negative liberty’, which is described in terms of the absence of frustration and coercion. I turn to one of the most famous discussions of negative liberty by J. S. Mill, who says:

…the principle requires liberty of tastes and pursuits; of framing the plan of our life to suit our own character; of doing as we like subject to such consequences as may follow: without impediment from our fellow creatures. (Mill 1868, 28)

This is, perhaps, how autonomy is most commonly understood: to live according to our own reasons and motives, unhampered by distorting forces (Christman 2015). We might refer to it as having the opportunity for autonomy. The latter definition could be said to enable the former.

For the purposes of this discussion, I wish to conceive of autonomy (an important facet of liberty) as something that captures both the original capacity for self-determination through reflection and reasoning, and the opportunity to self-determine without external interference.

With that stated, and given that this discussion is concerned with the human experience of Big Data utilization, I will to lay down three rather intuitive conditions for autonomy:

1. First, let us agree that the sort of introspection detailed above is an integral component of most modern conceptions of human autonomy. The ability to access our most truthful personal positions by “looking inside” ourselves. Built into this is the need for contemplation time or the ‘headspace’ to discover what those preferences are. For example, when you ask me if I would like to do activity X, through introspection I might locate the long-term idea or belief y, which is necessary for the enjoyment of X.
2. Next, I want to add the internal capacity to reason from, or to, such favored choices through a kind of internal information processing. When we think of those who lack full autonomy – small children, perhaps – we might note that they do not have sophisticated processing tools which allow them to consistently make appropriate decisions based on available evidence. For example, my decision to $X$ will involve my processing the idea or belief $y$ alongside other relevant ideas and beliefs $l$, $p$, and $h$ (perhaps representing my conflicting preferences or commitments).

3. Lastly, it is clear that to make adequate, life-shaping decisions we must have fair access to information to begin with. We feel sorry for Jim Carey’s character in *The Truman Show* because his world is one of artificially short horizons and external interference, and it means that his choices aren’t made within the full range of possibilities. We wouldn’t consider him to be fully autonomous because he doesn’t have what I will call an unimpeded horizon of choice, the last of our conditions. An example: I am unable to meaningfully process your suggestion to $X$ because key information has been masked or intercepted by external forces in such a way that I have no idea or belief $y$ to begin with (contra the examples in 1 and 2), nor the means to formulate one.

I do not mean to suggest that these conditions alone are sufficient for autonomy, but I am implying that they are necessary. I will also assume that autonomous decisions or choices are the clearest expressions of unimpeded human autonomy, and as such they are required for the ‘summum bonum’. Therefore, any practice that denies agents some aspect of their autonomy needs to be investigated, and either qualified or condemned.

Now let us consider what Big Data is, and how it might challenge these three conditions of autonomy.

3. Big Data

Big Data can be an evasive term that accompanies the tech zeitgeist but rarely receives full or adequate explanation. Depending on whom
you ask, it can alternately refer to the massive amounts of raw data being thrown off by our digital existences, the extraction and analysis of that data (often through complex algorithmic processes), or the utilization or repurposing of data (e.g. in the commercial sense). Most of the raw data collected is about us (Lynch 2016, 90). It can be textual or numerical, past or real-time, and flows from all of our digital interactions – emails, text messages, smart device signals (e.g. Fitbit), social media statuses, Google searches, web-browsing, app usage, geo-location, customer loyalty cards, RFID chips, etc.

Typically, this data is harvested by interested parties who study it – usually by employing algorithms and pattern-recognition techniques – for sentiment, preferences, behavioral indicators, and anything that can usefully predict future action. This in turn allows utilizers to plan, and to categorize the sections of the connected population they wish to monitor or influence. Because Big Data now ‘knows’ so much, it can also allow utilizers to offer us tailored information and alerts, automated services and guidance, and even bespoke medical treatment.

This all sounds laudable, so how do we arrive at this debate about the future of autonomy? Three relevant developments hold the answer.

The first is often termed datafication, and refers to the way in which previously unquantifiable aspects of our lives are being turned into raw, computerized data. This includes aspects that we have never previously thought of as data – like human relationships. Relationships have always existed (and have constituted information), but they were never formerly understood as data until Facebook’s social graph (Mayer-Schönberger and Cukier 2013, 91). Additionally, geospatial data can now be recorded, tallied, analyzed, and communicated in a standardized, numerical format (Mayer-Schönberger and Cukier 2013, 87). Consequently, datafication has led to an explosion in the amount of human-centric data available.

The second development concerns the analytical stage, and what I call intelligence. By this I mean the computerized programs and models used to make sense of this avalanche of raw data; the algorithms and

---

3 Known as the “Internet of Things” (IoT).
4 Governments or private entities.
machine intelligence used to interrogate this information at breakneck speed. Intelligence continues to evolve in terms of the speed and accuracy with which it can spot patterns in Big Data, giving its utilizers a deeper insight into human behaviors, relations, and information structures than we’ve ever had before.

It is important to note that algorithms themselves are inert, meaningless machines until paired with the datasets upon which they function (Gillespie 2014, 167). With the raw materials derived from datafication, however, these encoded procedures are now systematically outstripping expert opinion and forecasting.

The last change concerns our relationship with the giants of the internet kingdom (i.e. the Big Data utilizers) and our connectivity, which I wish to think of not just as a fact, but as a force that bears upon us. In short, we’re losing our ability to “log off”. We are increasingly submerged in what information philosopher Luciano Floridi calls the “Infosphere” or “onlife” (Floridi 2014, 43). As more of our devices are connected and we invest more of ourselves online, walking away is beginning to look inconceivable. A life without Google or Pinterest? For younger generations, it might be an impossibility. As Floridi (2014, 98) predicts: “Generation Z will increasingly feel deprived, excluded, handicapped, or poor to the point of paralysis and psychological trauma whenever it is disconnected from the infosphere”.

What’s more – and as we will see later – being connected (and therefore datafied) needn’t even require that we are signed onto the internet, or hooked-up to a smart device. Anyone, anywhere can be captured within Big Data number-crunching, and become the subject of automated decisions.

Once we realize these developments, we can begin to see how they might present a challenge to the conditions of autonomy outlined above. We are swiftly moving from a position whereby no marketer, government, friend or foe can know what is better for me as an individual agent than myself, to a new era of deference to superior insight and outsourced choices.
4. Three Challenges to Autonomy

Understanding: Data Overload and Automation

There was once a time when a recruitment manager relied upon their own experience when it came to hiring for an important position within their company. Their routine might involve sifting through tens (or even hundreds) of applications, checking qualifications, liaising with previous employers, and devising suitable interview questions. The entire procedure revolved around a relatively small number of relevant ‘data points’, all of which were ultimately considered (somewhat slowly) by the manager’s own in-built human processors. This would involve some pre-existing preferences in terms of the characteristics of good candidates – a sort of introspectively available set of success factors (i.e. interview performance, experience, qualifications, etc.) – which are then used to assist the manager as she internally processes the available data from prospective candidates. The result, she might hope, would be a clear decision about the most suitable candidate for the job.

Now Big Data is being used to revolutionize this process, because datafication and connectivity mean there are thousands of new data points available. Not only can new data be used to find the right candidate, but it can also ascertain the appropriate ‘success factors’ for the role to begin with. Capitalizing on this data overload, companies like Gild are using advanced intelligence to scan millions of websites for strong new prospects, as measured against their own model of the perfect employee. This search might include examining an individual’s social circle (do they have useful contacts?), what they do outside of work (do their hobbies add value?), and even other relevant side interests that algorithms have found to correlate with being a good employee (like the crossover between great coders and admirers of Japanese manga cartoons!) (O’Neil 2017, 120-1). This mass of data is ever-changing and ever-expanding, and many believe the best recruiters need to keep on top of it so that they can make the right offer to the right person at the right time.

The growing popularity of services like those provided by Gild serves to highlight our increasing inability as individuals (or hiring managers) to profitably keep up with, understand, and meaningfully
work through the vast swathes of human data emanating from the internet during every second of every day – and yet, we are loath to ignore it and lose the advantage. Thus, datafication and intelligence are working in combination to usurp our default to personal introspection and our own processing mechanisms (as per autonomy conditions 1 and 2).

There are good reasons for running with this superior insight. Given the flow of data now out-paces human comprehension of the facts, new intelligence is designed to spot useful patterns that are simply unavailable to ordinary human observers. As a natural consequence, we are becoming ever-more dependent upon these algorithmic tools because they seem to ‘understand’ human-centric data better than we do. Some have suggested that Big Data’s “biggest challenge” is convincing people not to trust their judgment (Klous and Wieland 2016, 13) – or to even use it at all. Yet, although decision outsourcing is more expedient, it does also allow us to bypass functions that are central to our self-determination.

If I am correct that autonomy – an essential part of modern conceptions of liberty – is central to humanity, and that autonomy rests (at least partially) on the exercise of our reflection and reason, then these processes are inherently threatened by deferral to automation. Regardless of this, experts are already predicting that in the very near future many aspects of our world that today are the sole purview of human judgment will be augmented or replaced entirely by computer systems (Mayer-Schönberger and Cukier 2013, 12).

In truth, even if our human processing power could keep pace with datafication, our own levels of connectivity mean as would-be decision-makers we are losing the personal space we need in order to properly introspect and survey the options without interference from yet more data. Intelligence doesn’t need contemplative ‘headspace’ which in many respects makes it more effective. Yuval Harari pushes this idea of outsourcing, and imagines a near future in which we defer to Google for everything, including our dating prospects:

Google will answer: “Well, I know you from the day you were born. I have read all your emails, recorded all your phone calls, and know your favorite films, your DNA and the entire biometric history of your heart. I have exact data about each
date you went on, and I can show you second-by-second graphs of your heart rate, blood pressure and sugar levels whenever you went on a date with John or Paul. And, naturally enough, I know them as well as I know you. Based on all this information, on my superb algorithms and on decades’ worth of statistics about millions of relationships — I advise you to go with John, with an 87 per cent probability of being more satisfied with him in the long run. (Harari 2016, 398)

Is this an absurd prediction? We already farm-out routine tasks for convenience and rely on intelligence to understand us based on our data. We expect search engines to know what we want, mail programs to prevent embarrassing mistakes, and music services to order concert tickets before others know our favorite artist is in town (Klous and Wielaard 2016, 35). In short, the locus of knowledge and agency is already shifting and becoming external to us. Harari believes that once this is fully realized, “the belief in individualism will collapse and authority will shift from individual humans to networked algorithms” (Harari 2016, 329).

Raw data and its analysis were once the subjects of our own processing and introspection, yielding our autonomous choices. Now prospective partners are intelligently suggested to us, families are beginning to let their cupboards use their data to order groceries, and entrepreneurs are letting algorithms set prices in real-time (e.g. Uber). Though, for the most part, we still operate the fuel dispensers at gas stations, translate GPS’ instructions into driving maneuvers, and make the supermarket interact with our refrigerator (Floridi 2014, 31), how long will it be before we let data-driven intelligence edit us out of the picture entirely? After all, once I have a system that knows me better, it seems foolhardy to trust myself (Harari 2016, 338).

In sum, if processing and introspection are necessary conditions of autonomy, our relinquishment of decisions to data-driven AI arguably represents a weakening of autonomy, leaving us less space to establish our sovereign character, as expressed through choice.

**Being understood: Feedback Loops and Manipulation**

Even if we were to retain much of our day-to-day decision-making, there are other ways in which datafication, intelligence and connectivity can
collude to limit our *horizon of choice*. I refer to environments where Big Data is employed to ‘nudge’ and shepherd our decision-making behaviors. Karen Yeung calls the mechanism “digital guidance processes”:

…designed so that it is not the machine, but the targeted individual, who makes the relevant decision. These technologies seek to direct or guide the individual’s decision-making processes in ways identified by the underlying software algorithm as ‘optimal’, by offering ‘suggestions’ intended to prompt the user to make decisions preferred by the choice architect. (Yeung 2016, 5-6)

These subtle prompts are extremely powerful. Outwardly catering to our tastes, they favor the success factors specified by the Big Data utilizer. This can mean pushing us toward some things and masking others in a way that constitutes interference, causing problems for autonomy condition 3. I will explain this further.

The advent of the internet was something of a dream scenario for governments, big business, and anyone else with an interest in ‘listening’ intensively to our behaviors. As the web evolved from static webpages to the interactive ‘Web 2.0’, our raw data output increased exponentially.⁵ Social platforms like Facebook and Twitter, alongside search engines like Google, could employ *intelligence* to scan this data for deep, commercially valuable insights about our individual preferences. For the first time, the web started to use our information to tailor itself to our needs.

Nowadays, if you and I google the same terms we won’t always be delivered the same results. That’s because Google knows our likes, dislikes, and what we’ve asked before: it shapes our world accordingly. Meanwhile, other companies track our browsing data to adapt the style of their websites to appeal to our tastes⁶. Enjoy the *New York Times*? The next site you hit might employ the same kind of layout and vocabulary in a bid to charm you.

Back in 2012, Google’s Ad Preference settings caused a splash in the news (Kessler 2012). Their search engine was using cookies to follow users around the internet and make guesses about who we were. For the

---

⁵ Stored data grows four times faster than the world economy (Mayer-Schoenberger and Cukier 2013, 9).

⁶ Big Willow/DemandBase.
large part, these were unnervingly precise diagnoses about factors like age, gender, and residence, as well as preferences for snowboarding or rock music or motherhood. Google continues to compile and share this information with other companies so that they can cater to – and anticipate – our preferences. This involves adapting your online experience to increase exposure to goods or services that are preferred by people that it deems to be “just like you”.

Two things here are relevant to our autonomy question. Firstly, this implies that whatever we see online, from advertising to top search results, is not a straight representation of what there is, but an altered reality designed especially for the type of person it suspects you are. Secondly, in reacting to these stimuli – reading the news article, buying the boots, trying the product – we are often letting the internet giant pre-select the very things we use to define our “me-ness”. Clearly, there could come a point where data-driven algorithms stop simply mapping our preferences, and start actively defining them. These are both in conflict with autonomy condition 3.

Furthermore, this self-perpetuating feedback loop isn’t just limited to the internet. Given that it is considerably more potent type of consumer-targeting than old-fashioned market research surveys, large tech companies at the fore of Big Data want to fully leverage it. Companies like Google, Apple and Samsung want us to buy their products – cameras, phones, televisions – not only because they make money selling us their hardware, but also so they can plant another ‘sensor’ on us. These sensors increase our connectivity and provide utilizers with more information about our behaviors and where their future business might lie (Klous and Wielaard 2016, 4-5). As Steve Jobs once said, “it isn’t the consumers’ job to know what they want!” Indeed, increasingly we just behave and our world is structured to marshal us towards the option an algorithm has decided we will like. The problem is, if we aren’t careful (and I shall address later), very soon we won’t even get to see or come to know of the things that the intelligence supposes we aren’t interested in, based on proxies and past decisions.

Once our preferences have been refined, utilizers ideally want to lead us into the sort of material decision-outsourcing that guarantees
them a long-term income. Programmatic commerce\(^7\) pioneers Salmon claim 57% of UK shoppers are ready for automatic purchasing without any need for approval within two years (Salmon 2016). Controlling the horizon of choice (effectively the ‘input’) means predicting consumer predilections (ideas or ‘output’) with greater accuracy, making us less likely to feel the need to fully introspect or process variables ourselves. We can handover and “let the data decide.”

Deployed in this way, Big Data could make us forever prisoners of our previous actions and choices (Mayer-Schönberger and Cukier 2013, 195). What’s more, elevating the status of prior preferences also has the knock-on effect of decreasing the autonomy we enjoy in redefining ourselves, given that forgetting is part of the process of self-reconstruction (Lynch 2016, 72). This undermines our capacity for autonomy, as represented by conditions 1 and 2.

Some have speculated that new technology is less about listening to our inner voices, and more about controlling them. Using predictive proxies and subtle choice ‘nudges’, Big Data utilizers are making decisions for us in cunning and elusive ways, and challenging all three of our autonomy conditions. In essence, masking the information that allows me to formulate idea or belief \(y\) (or non-\(y\)) means that: a. I cannot process other ideas in light of \(y\) (or non-\(y\)), and b. I cannot have \(y\) or non-\(y\) as part of my self-definition in the way that I might otherwise.

**Being Misunderstood: Types and Prejudice**

Lastly, it would be remiss not to briefly address the ways in which intelligence and connectivity can limit the horizon of choice for certain individuals in even more damaging ways.

Big Data algorithms ‘personalize’ our feedback by assigning us a type and predicting our choices based both on our data and the data of others with similar profiles. Thus, we are ultimately reduced to a type of customer, a type of driver, a type of citizen, a type of person who lives at

---

\(^7\) Whereby smart devices are enabled to make automated purchases, e.g. a coffee machine which automatically reorders supplies.
that postal code, who drives that type of car, who goes to that type of restaurant, etc. (Lynch 2016, 58). This division might be innocuous in most cases, but we should be concerned when the ‘type’ we’re assigned negatively affects the opportunities we’re offered by utilizers.

Intelligence might appear neutral, but we shouldn’t assume that this makes its output objective. Algorithms are still built by (inherently subjective) human-beings with a subjectively driven definition of success. Additionally, the historical data they process can be laden with socio-political salience. If models are assigning type, and blindly predicting that the future will mirror the past without recourse to causality, we end up with scenarios where, for example, a job advert is not displayed to a qualified woman because the career type has been historically masculine, which biases the algorithm (Oram 2016). This is profiling, and once any personal feature – our name, our address, our associates – puts us on a poor track, algorithms can automatically limit our horizons. We may never even know.

Take the smartphone game Pokémon GO, which uses a phone’s camera and GPS to augment the real world with Pokémon. Players catch Pokémon using Pokéballs, which can be picked up at Pokéstops located near local landmarks (Kooragayala and Srini 2016). Soon after the game was launched, controversy broke out as users found that Pokéstops were hard to find in poorer and minority neighborhoods. With relevance to autonomy condition 3, a direct consequence of reducing the potential for gameplay amongst minorities is that they are omitted from the pool of data about players that is inevitably networked and tapped by various utilizers. In turn, we might assume that these minorities don’t receive the associated offers, opportunities, suggestions, benefits etc., that are made available to white high school kids, college students, and people with desk jobs (Huffaker 2016). They are on the outside of the feedback loop. Without even these stimuli to react to, they are not captured in data that is used to find patterns and assign attributes to certain types of citizens or consumers.

As a result, there’s potential for a significant minority to be ‘edited out’ in a way that narrows their horizon of choice. For example, a Hispanic youth might not be assigned as the ‘type’ of person who would buy an Apple product. Far worse still, having limited representation in certain
pools of data might lead learning algorithms to believe that you don’t fit the ‘type’ who can be relied upon to pay back a loan or drive a car safely.

Landing up in certain datasets, or being omitted from them, could lead to a stark contrast in the kinds of lifestyles different communities identify – and are identified – with. This is likely to have a determining effect on their autonomous self-definition (given that their ‘choosing’ of a particular product actually arises from their being denied access to the full range of other options). Although we might have seen similarly divisive segmentation tactics in years gone by (e.g. commercial advertisers bypassing certain communities), the nature of connectivity and the blending of off-life with on-life means this new incarnation serves to amplify its troubling effects.

Furthermore, if – as ideas of relational autonomy would have it – our individual self-understanding and self-development are pervasively shaped by our surrounding environments, this ubiquitous categorization will shape the world in which people perceive and understand themselves (Yeung 2016, 18-19). Such theories would agree that where categories of information exist in social and economic institutions and are negatively-valenced, this also erodes an agent’s autonomy (perhaps regardless of what is or isn’t included within their horizon of choice).

Both views of autonomy understand intelligence as shaping the way reality really is by a different route. If we concede that environmental factors are critical to introspection and processing, it seems that we can say that these two conditions for autonomy are undermined, not only as a by-product of a limited horizon of choice, but also independently, by pervasive and negative societal factors that minimize the status of certain individuals.

5. An Objection

Of course, there’s a clear objection to all of this. We don’t need to partake in the digital “Infosphere” if we are concerned about our free will. We don’t have to maintain a social media profile, use loyalty cards or bedeck ourselves and our homes with sensor-ridden smart devices. This, in itself, is a choice and we can easily go ‘off-grid’.
I believe that this objection overlooks three points: the first is that such an ‘opt out’ is just a misnomer, especially for upcoming generations. It is extremely unlikely that children born now will understand a world bereft of Facebook and iGadgets. This point is made beautifully by Jean-Luis Constanza, a top telecoms executive who uploaded a YouTube video of his baby daughter prodding at images in a magazine. He comments: “In the eyes of my one-year-old daughter, a magazine is a broken iPad. That will remain the case throughout her entire life. Steve Jobs programmed part of her operating system” (Matyszczycyk 2011).

The rest of us are equally unlikely to avoid playing a role in a Big Data society. As Sarah Kessler (2016) notes, even something as simple as using electricity generates data about our habits – and not just about whether or not the lights are on, but also about how many people are in the house and when they are usually around. If we think we can avoid the collection and use of our data, we are mistaken.

Second, even if we imagine that it is possible to delete ourselves from databases and withdraw from the digital society, can we really argue that this increases our autonomy? We can live without smart devices and social media but coming off-grid is the equivalent of modern day hermitry. Most of us know someone who cannot or will not ‘connect’: perhaps an older relative. Would we argue that their autonomy is greater than ours? I think not. We access the internet for them because of the options and information available for us to choose from is more, not less. Their horizon of choice is impeded by the fact that utilizers now demand our connectivity in order to acquire any key information at all. Though I may not wish to be ‘nudged’ when browsing online for real estate, by opting to not to connect my options are fewer still because of how this market is now structured. I cannot fulfill my objectives without assenting to the possibility of subtle influences.

Lastly, whether or not we interact with the Infosphere or Big Data, we are still controlled by it (as per the previous section). Even if we block data-driven technology from taking decisions on our behalf, or believe we can resist letting it guide us in our selections online, it is still the fuel of a great many decisions that are taken about us – some of which could interfere with our horizon of choice. These range from how
our communities are policed, to what products and designs appear in the stores we visit, and even our city infrastructure.

In short, the internet and data can cause problems for autonomy, regardless of our active participation.

### 6. Re-claiming Autonomy

Although this discussion has looked to demonstrate that ethical concerns about autonomy have good grounds, in-line with other commentators (O’Neil 2016; Mayer-Schönberger and Cukier 2013; Klous and Wielaaard 2016) I believe that there are measures society can take to mitigate against potential harms. I now wish to address the concerns laid out above with recommendations that should allow us to maintain our technological progress without abandoning our idea of autonomy as central to liberty. I address each of the points in reverse order with reference to regulation, awareness, and acceptance.

**Regulation in Response to “Types and Prejudice”**

Prejudice cannot be left unchecked. It’s clear to most that data-driven intelligence should be objectively interrogated to prevent it from instantiating bias or participating in any other type of societal warping. Attitudinally, this should be accompanied by a general dismissal of the idea that any of this information is value-neutral or truthfully representative to begin with.

Some commentators (O’Neil 2016) have suggested the introduction of a kind of “Hippocratic Oath” for data scientists, compelling them to acknowledge the potential for destructive downstream effects. Others want to enforce more rigorous data hygiene amongst utilizers, proposing a patient-focused model of distributed moral responsibility which lays ethical liability for data bias at the door of all parties in the

---

8 As with ‘smart cities’ that leverage Big Data to make efficiency decisions with regards to infrastructure and planning.
supply chain, regardless of their knowledge or intent (Floridi 2016). At present, work is also being done to develop de-biasing algorithms that can neutralize biased machine learning datasets (Bolukbasi et al. 2016, 4349-4357). Nevertheless, at the time of writing Big Data continues to be mined by companies and governments whose models derive tenuous proxies, whilst the nature of the system allows wild inaccuracies to remain concealed.

Wherever the eventual solution lies in terms of corrective efforts, it is clear that expecting individual users to understand and assess the risks of divulging their personal information is inadequate. However, here is not the place for fleshed-out policy recommendations, and considerable energy is being invested by others.

**Awareness in Response to “Feedback Loops and Manipulation”**

Only an elite few have any detailed understanding of how our data is used to shape our worldview. Back in 2014 Facebook ran an experiment where it tweaked our news-feeds to see if it could change our moods, as judged by what we wrote as a consequence (Tsukayama and Fung 2014). We patently need to ensure that there is broader awareness of the types of manipulation at play.

In so many ways the internet provides us with a window on the world, but if data-driven models have a tendency to narrow that privileged vision then we need to invent new ways of forcefully broadening the scope of our experiences and understanding. Here I suggest just a few ways by which society and individuals might enlighten themselves.

The first is by calling for greater algorithmic transparency. The forthcoming EU General Data Protection Regulation (GDPR) will allow data subjects to access limited information regarding automated decisions, but citizens worldwide also need to better understand how their data is employed. Though there may be difficulty – particularly for machine learning – explaining how algorithms arrive at actionable conclusions, suggestions have been made with regards to improving the transparency and accountability of automated decision-making
(Wachter, Mittelstadt and Floridi 2016). This includes legally mandated right of explanation about specific decisions.

Secondly, we need to encourage Big Data utilizers to supplement their predictive output with qualitative, so-called “thick data” (Wang 2016) where available. In 2009, an ethnographer living and working amongst poorer populations in China made a discovery that challenged the model of her then employers, one-time telecommunications giant, Nokia. Tricia Wang’s indicators showed that these low-income consumers were ready to pay for more expensive smartphones. Her bosses insisted that the findings from her small study of just 100 subjects did not show up in their existing datasets. Wang says: “I told them that it made sense that they haven’t seen any of my data show up in their quantitative datasets because their notion of demand was a fixed quantitative model that didn’t map to how demand worked as a cultural model in China.” Nokia neglected her research.

By reminding corporate and government utilizers that small sample, qualitative information about people and their communities can shed useful light on Big Data predictions, we can also protect against their shaping the world in a way that warps our reality. This is not about improving commercial insight: it is in all of our interests for utilizers to solicit feedback and adapt their models so as to avoid entrenching wrong assumptions and guiding us around in narrower and narrower circles, irrespective of our needs and preferences.

Lastly, we must ensure that ‘Generation Z’ (or the ‘Post-Millennials’) advance as critical thinkers, questioning all information and its sources. This is not just my opinion. In September 2016 philosophy was introduced into Irish schools as: “one of the most powerful tools we have at our disposal to empower children in an ever more complex, interconnected, and uncertain world” (Blease 2017). With a stark information asymmetry between utilizers and citizen users, it is vital that more of us are equipped with tools that allow us to look beyond our tailored bubbles and question the horizons that are constructed for us.

In sum, upping awareness involves demanding more information about algorithmic mechanisms, imploring utilizers to supplement their
Big Data findings, and ensuring that citizens have the critical tools to look beyond the artificial perimeters of their connected lives.

Outsourcing decision-making is only bad if the outcome is bad. It’s hard to argue that an algorithm that suggests I buy a specific brand of shampoo is bad in itself. Where it goes wrong is if this shampoo is somehow very wrong for me, or if it masks my access to other shampoo brands. It is because of these potentialities that we need to make sure we are knowledgeable about, and engaged with, the intelligence process – even if 9 times out of ten the result is right.

**Acceptance in Response to “Data Overload and Automation”**

We cannot play King Cnut and attempt to hold back the tide of inevitable change. Allowing technology to play a bigger role in our decision-making is not necessarily equivalent to an erosion of autonomy. The latter assumes that we are preparing to customarily abandon any recourse to introspection and processing – or in other words, willingly give up the use of our self-determining rationality, and submit to the intelligence of data-driven machines.

I believe, however, that even in a world equipped with widespread automation, we will continue to make decisions and display character. It is just the case that, as Floridi writes: “Whatever defines us uniquely, it can no longer be playing chess, checking the spelling of a document or translating it into another language, calculating the orbit of a satellite, parking a car, or landing an aircraft” (Floridi 2014, 96). But this doesn’t mean we are less autonomous, less free or less human. It simply means that our self-definition appears to lie somewhere outside of the many humdrum practical decisions we make on a day-to-day basis. Yet we don’t exist in a vacuum: where we lose the need to take decisions about new hires and the best travel directions, their space is taken up with other decision-making opportunities.

I am contesting that – so long as we’re alert to feedback loops and prejudice – datafication, intelligence and connectivity do not disenfranchise us, but furnish us with a dazzling array of choices (and, thus, opportunities to introspect) that are superior in both quantity and quality.
Let’s consider what Klous and Wielaard call the “DIY society” forming as a result of Big Data. This gives people the freedom to set-up collectives and abolish old institutions – like residents who buy a windmill and no longer need the energy company, or entrepreneurs who fund each other’s ideas through crowdfunding and therefore no longer need banks (Klous and Wielaard 2016, 65). In our personal lives people are finding love on an unprecedented scale with websites like eHarmony and Match.com that use data to bring people together, fundraisers and charities can find philanthropists and donors, and because we are now ‘patchwork people’ who come together based on shared passions rather than territories, those in developing countries are finding their voice on the internet and consequently being identified by Big Data in statistics that previously ignored them. The benefits of all this are endless and (ironically) unquantifiable. Big Data actually enriches human comprehension in these cases (Mayer-Schönberger and Cukier 2013, 96). They deliver a whole new brand of human autonomy, one which displaces dictatorial institution and convention, and that privileges all three of our original conditions.

Though privacy concerns often surface in these discussions, as Lynch points out, one of the most fascinating facts about the amount of tracking going on in the United States is that hardly anyone seems to care (Lynch 2016, 91). Some find this a difficult exchange. Harari comments:

> In the high days of European imperialism, conquistadors and merchants bought entire islands and countries in exchange for colored beads. In the twenty-first century our personal data is probably the most valuable resource most humans still have to offer, and we are giving it to the tech giants in exchange for email and funny cat videos. (Harari 2016, 341)

Contrary to Harari’s cynicism, what we get in exchange for our data is infinitely more valuable than colored beads. To offer just one example, if we allow various sensors and gadgets to monitor our health – from our pulse rate to our blood sugar – it could give us lifesaving advice. Moreover, our information can be aggregated with the information of millions of other people and enable data scientists to identify all sorts of patterns that allow us to prevent problems or improve outcomes for health, wealth and politics. As such, rather than
undermining freedom, Big Data can help release people from debilitating situations that inhibit their natural agency, like disease and poverty.

If our concern about Big Data is an ethical one, then we must measure these benefits against the potential negatives outlined earlier in this paper. It seems to me that if the latter can be managed to preserve the former (which, I have suggested, can be done), we have a moral obligation to do so.

7. Conclusion

I have considered the threat that Big Data poses to autonomy in some detail and have concluded that, although there are concerns that need to be addressed, our relationship with the data should emerge as a symbiotic one, rather than the sort of dystopian master-slave scenarios imagined by some (typically the popular media). An increasing number of tasks will inevitably be given over to Big Data-driven algorithms, but this is just a new chapter in the history of technological innovation. We can choose to be Luddites, or we can choose to adapt.

Ultimately, we are not losing our capacity or opportunity for autonomy. Rather, where we lose deliberations about soap powder, we gain time that we’re being invited to fill in a number of new, liberty-enhancing ways. Raw data is too large and overwhelming for us to crunch, so we have intelligence do it for us and still exercise our right to question, interrogate and choose another option. This is so long as we are aware of the algorithmic tendency to limit our horizons if we allow ourselves not to look beyond them.

Finally, we need to keep our heads when we consider the project at hand. The hysteria of some might be off-set by the ignorance and passivity of others, but if we are truly to remain autonomous (i.e. neither withdraw from nor give into a Big Data society) we need some attitudinal leadership, and that might start with this: what we are able to collect in terms of data is still only a tiny fraction of the information that exists in the world. A small, two-dimensional smattering of human emissions that allows us to test some assumptions and change the world for the better. However, it can only ever be a simulacrum of reality, like
the shadows on the wall of Plato’s cave (Mayer-Schoenberger and Cukier 2013, 197). Where we are able, we should treat it as such.

REFERENCES


