Technology, autonomy, and manipulation

Susser, D.; Roessler, B.; Nissenbaum, H.

Technology, autonomy, and manipulation

Daniel Susser
College of Information Sciences and Technology, Pennsylvania State University, United States

Beate Roessler
University of Amsterdam, Netherlands

Helen Nissenbaum
Information Science, Cornell Tech, New York City, United States

Abstract: Since 2016, when the Facebook/Cambridge Analytica scandal began to emerge, public concern has grown around the threat of "online manipulation". While these worries are familiar to privacy researchers, this paper aims to make them more salient to policymakers—first, by defining "online manipulation", thus enabling identification of manipulative practices; and second, by drawing attention to the specific harms online manipulation threatens. We argue that online manipulation is the use of information technology to covertly influence another person's decision-making, by targeting and exploiting their decision-making vulnerabilities. Engaging in such practices can harm individuals by diminishing their economic interests, but its deeper, more insidious harm is its challenge to individual autonomy. We explore this autonomy harm, emphasising its implications for both individuals and society, and we briefly outline some strategies for combating online manipulation and strengthening autonomy in an increasingly digital world.

Keywords: Online manipulation, Behavioural advertising, Privacy

This paper is part of Transnational materialities, a special issue of Internet Policy Review *guest-edited by José van Dijck and Bernhard Rieder.*

Public concern is growing around an issue previously discussed predominantly amongst privacy and surveillance scholars—namely, the ability of data collectors to use information about individuals to manipulate them (e.g., Abramowitz, 2017; Doubek, 2017; Vayena, 2018). Knowing

(or inferring) a person's preferences, interests, and habits, their friends and acquaintances, education and employment, bodily health and financial standing, puts the knower in a position to exercise considerable influence over the known (Richards, 2013).[1] It enables them to better understand what motivates their targets, what their weaknesses and vulnerabilities are, when they are most susceptible to influence and how most effectively to frame pitches and appeals.[2] Because information technology makes generating, collecting, analysing, and leveraging such data about us cheap and easy, and at a scarcely comprehendible scale, the worry is that such technologies render us deeply vulnerable to the whims of those who build, control, and deploy these systems.

Initially, for academics studying this problem, that meant the whims of advertisers, as these technologies were largely developed by firms like Google and Facebook, who identified advertising as a means of monetising the troves of personal information they collect about internet users (Zuboff, 2015). Accordingly, for some time, scholarly worries centred (rightly) on commercial advertising practices, and policy solutions focused on modernising privacy and consumer protection regulations to account for the new capabilities of data-driven advertising technologies (e.g., Calo, 2014; Nadler & McGuigan, 2018; Turow, 2012).[3] As Ryan Calo put it, "the digitization of commerce dramatically alters the capacity of firms to influence consumers at a personal level. A specific set of emerging technologies and techniques will empower corporations to discover and exploit the limits of each individual consumer's ability to pursue his or her own self-interest" (2014, p. 999).

More recently, however, the scope of these worries has expanded. After concerns were raised in 2016 and 2017 about the use of information technology to influence elections around the world, many began to reckon with the fact that the threat of targeted advertising is not limited to the commercial sphere.[4] By harnessing ad targeting platforms, like those offered by Facebook, YouTube, and other social media services, political campaigns can exert meaningful influence over the decision-making and behaviour of voters (Vaidhyanathan, 2018; Yeung, 2017; Zuiderveen Borgesius et al., 2018). Global outrage over the Cambridge Analytica scandal—in which the data analytics firm was accused of profiling voters in the United States, United Kingdom, France, Germany, and elsewhere, and targeting them with advertisements designed to exploit their "inner demons"—brought such worries to the forefront of public consciousness ("Cambridge Analytica and Facebook: The Scandal so Far", 2018; see also, Abramowitz, 2017; Doubek, 2017; Vayena, 2018).

Indeed, there is evidence that the pendulum is swinging well to the other side. Rather than condemning the particular harms wrought in particular contexts by strategies of online influence, scholars are beginning to turn their attention to the big picture. In their recent book *Re-Engineering Humanity*, Brett Frischmann and Evan Selinger describe a vast array of related phenomena, which they collectively term "techno-social engineering"—i.e., "processes where technologies and social forces align and impact how we think, perceive, and act" (2018, p. 4). Operating at a grand scale reminiscent of mid-20th century technology critique (like that of Lewis Mumford or Jacques Ellul), Frischmann and Selinger point to cases of technologies transforming the way we carry out and understand our lives—from "micro-level" to the "meso-level" and "macro-level"— capturing everything from fitness tracking to self-driving cars to viral media (2018, p. 270). Similarly, in her book *The Age of Surveillance Capitalism* (2019)*,* Shoshana Zuboff raises the alarm about the use of information technology to effectuate what she calls "behavior modification", arguing that it has become so pervasive, so central to the functioning of the modern information economy, that we have entered a new epoch in the history of political economy.

These efforts help to highlight the fact that there is something much deeper at stake here than unfair commerce. When information about us is used to influence our decision-making, it does more than diminish our interests—it threatens our autonomy.[5] At the same time, there is value in limiting the scope of the analysis. The notions of "techno-social engineering" and "surveillance capitalism" are too big to wield surgically—the former is intended to reveal a basic truth about the nature of our human relationship with technology, and the latter identifies a broad set of economic imperatives currently structuring technology development and the technology industry.[6] Complementing this work, our intervention aims smaller. For the last several years, public outcry has coalesced against a particular set of abuses effectuated through information technology—what many refer to as "online manipulation" (e.g., Abramowitz, 2017; Doubek, 2017; Vayena, 2018). In what follows, we theorise and vindicate this grievance.[7]

In the first section, we define manipulation, distinguishing it from neighbouring concepts like persuasion, coercion, deception, and nudging, and we explain why information technology is so well-suited to facilitating manipulation. In the second section, we describe the harms of online manipulation—the use of information technology to manipulate—focusing primarily on its threat to individual autonomy. Finally, we suggest directions for future policy efforts aimed at curbing online manipulation and strengthening autonomy in human-technology relations.

1. WHAT IS ONLINE MANIPULATION?

The term "manipulation" is used, colloquially, to designate a wide variety of activities, so before jumping in it is worth narrowing the scope of our intervention further. In the broadest sense, manipulating something simply means steering or controlling it. We talk about doctors manipulating fine instruments during surgery and pilots manipulating cockpit controls during flight. "Manipulation" is also used to describe attempts at steering or controlling institutions and systems. For example, much has been written of late about allegations made (and evidence presented) that internet trolls under the authority of the Russian government attempted to manipulate the US media during the 2016 presidential election.[8] Further, many suspect that the goal of those efforts was, in turn, to manipulate the election itself (by influencing voters). However, at the centre of this story, and at the centre of stories like it, is the worry that *people* are being manipulated, that individual decision-making is being steered or controlled, and that the capacity of individuals to make independent choices is therefore being compromised. It is manipulation in this sense—the attempt to influence individual decision-making and behaviour—that we focus on in what follows.

Philosophers and political theorists have long struggled to define manipulation. According to Robert Noggle, there are three main proposals (Noggle, 2018b). Some argue that manipulation is *non-rational influence* (Wood, 2014). On that account, manipulating someone means influencing them by circumventing their rational, deliberative decision-making faculties. A classic example of manipulation understood in this way is subliminal messaging, and depending on one's conception of rationality we might also imagine certain kinds of emotional appeals, such as guilt trips, as fitting into this picture. The second approach defines manipulation as a form of *pressure*, as in cases of blackmail (Kligman & Culver, 1992, qtd. in Noggle, 2018b). Here the idea is that manipulation involves some amount of force—a cost is extracted for non-compliance—but not so much force as to rise to the level of coercion. Finally, a third proposal defines manipulation as *trickery*. Although a variety of subtly distinct accounts fall under this umbrella, the main idea is that manipulation, at bottom, means *leading someone along*, inducing them to behave as the manipulator wants, like Iago in Shakespeare's Othello, by

tempting them, insinuating, stoking jealousy, and so on.9

Each of these theories of manipulation has strengths and weaknesses, and our account shares certain features in common with all of them. It hews especially close to the trickery view, but operationalises the notion of trickery more concretely, thus offering more specific tools for diagnosing cases of manipulation. In our view, manipulation is *hidden influence*. Or more fully, manipulating someone means *intentionally and covertly influencing their decision-making, by targeting and exploiting their decision-making vulnerabilities*. Covertly influencing someone—imposing a hidden influence—means influencing them in a way they aren't consciously aware of, and in a way they couldn't easily become aware of were they to try and understand what was impacting their decision-making process.

Understanding manipulation as hidden influence helps to distinguish it from other forms of influence. In what follows, we distinguish it first from persuasion and coercion, and then from deception and nudging. Persuasion—in the sense of rational persuasion—means attempting to influence someone by offering reasons they can think about and evaluate.10 Coercion means influencing someone by constraining their options, such that their only rational course of action is the one the coercer intends (Wood, 2014). Persuasion and coercion carry very different, indeed nearly opposite, normative connotations: persuading someone to do something is almost always acceptable, while coercing them almost always isn't. Yet persuasion and coercion are alike in that they are both *forthright* forms of influence. When someone is trying to persuade us or trying to coerce us we usually know it. Manipulation, by contrast, is hidden—we only learn that someone was trying to steer our decision-making after the fact, if we ever find out at all.

What makes manipulation distinctive, then, is the fact that when we learn we have been manipulated we feel *played*.11 Reflecting back on why we behaved the way we did, we realise that at the time of decision we didn't understand our own motivations. We were like puppets, strung along by a puppet master. Manipulation thus disrupts our capacity for self-authorship—it presumes to decide for us how and why we ought to live. As we discuss in what follows, this gives rise to a specific set of harms. For now, what is important to see is the *kind* of influence at issue here. Unlike persuasion and coercion, which address their targets openly, manipulation is covert. When we are coerced we are usually rightly upset about it, but the object of our indignation is the set of constraints placed upon us. When we are manipulated, by contrast, we are not constrained. Rather, we are directed, outside our conscious awareness, to act for reasons we can't recognise, and toward ends we may wish to avoid.

Given this picture, one can detect a hint of deception. On our view, deception is a special case of manipulation—one way to covertly influence someone is to plant false beliefs. If, for example, a manipulator wanted their partner to clean the house, they could lie and tell them that their mother was coming for a visit, thereby tricking them into doing what they wanted by prompting them to make a rational decision premised on false beliefs. But deception is not the only species of manipulation; there are other ways to exert hidden influence. First, manipulators need not focus on beliefs at all. Instead, they can covertly influence by subtly tempting, guilting, seducing, or otherwise playing upon desires and emotions. As long as the target of manipulation is not conscious of the manipulator's strategy while they are deploying it, it is "hidden" in the relevant sense.

Some argue that even overt temptation, guilting, and so on are manipulative (these arguments are often made by proponents of the "non-rational influence" view of manipulation, described above), though they almost always concede that such strategies are more effective when

concealed.[12] We suspect that what is usually happening in such cases is a manipulator *attempting* to covertly tempt, guilt, etc., but failing to successfully hide their strategy. On our account, it is the attempted covertness that is central to manipulation, rather than the particular strategy, because once one learns that they are the target of another person's influence that knowledge becomes a regular part of their decision-making process. We are all constantly subject to myriad influences; the reason we do not feel constantly manipulated is that we can usually reflect on, understand, and account for those influences in the process of reaching our own decisions about how to act (Raz, 1986, p. 204). The influences become part of how we explain to ourselves why we make the decisions we do. When the influence is hidden, however, that process is undermined. Thus, while we might naturally call a person who frequently engages in overt temptation or seduction *manipulative*—meaning, they frequently attempt to manipulate—strictly speaking we would only say that they have succeeded in manipulating when their target is unaware of their machinations.

Second, behavioural economists have catalogued a long list of "cognitive biases"—unreliable mental shortcuts we use in everyday decision-making—which can be leveraged by would-be manipulators to influence the trajectory of our decision-making by shaping our beliefs, without the need for outright deception.[13] Manipulators can frame information in a way that disposes us to a certain interpretation of the facts; they can strategically "anchor" our frame of reference when evaluating the costs or benefits of some decision; they can indicate to us that others have decided a certain way, in order to cue our intrinsic disposition to social conformity (the so-called "bandwagon effect"); and so on. Indeed, though deception and playing on people's desires and emotions have likely been the most common forms of manipulation in the past—which is to say, the most common strategies for covertly influencing people—as we explain in what follows, there is reason to believe that exploiting cognitive biases and vulnerabilities is the most alarming problem confronting us today.[14]

Talk of exploiting cognitive vulnerabilities inevitably gives rise to questions about nudging, thus finally, we briefly distinguish between nudging and manipulation. The idea of "nudging", as is well known, comes from the work of Richard Thaler and Cass Sunstein, and points to any intentional alteration of another person's decision-making context (their "choice architecture") made in order to influence their decision-making outcome (Thaler & Sunstein, 2008, p. 6). For Thaler and Sunstein, the fact that we suffer from so many decision-making vulnerabilities, that our decision-making processes are inalterably and unavoidably susceptible to even the subtlest cues from the contexts in which they are situated, suggests that when we design other people's choice-making environments—from the apps they use to find a restaurant to the menus they order from after they arrive—we can't help but influence their decisions. As such, on their account, we might as well use that power for good, by steering people's decisions in ways that benefit them individually and all of us collectively. For these reasons, Thaler and Sunstein recommend a variety of nudges, from setting defaults that encourage people to save for retirement to arranging options in a cafeteria in way that encourages people to eat healthier foods.[15]

Given our definition of manipulation as intentionally hidden influence, and our suggestion that influences are frequently hidden precisely by leveraging decision-making vulnerabilities like the cognitive biases nudge advocates reference, the question naturally arises as to whether or not nudges are manipulative. Much has been written on this topic and no consensus has been reached (see, e.g., Bovens, 2009; Hausman & Welch, 2010; Noggle, 2018a; Nys & Engelen, 2017; Reach, 2016; Selinger & Whyte, 2011; Sunstein, 2016). In part, this likely has to do with the fact that a wide and disparate variety of changes to choice architectures are described as nudges. In

our view, some are manipulative and some are not—the distinction hinging on whether or not the nudge is hidden, and whether it exploits vulnerabilities or attempts to rectify them. Many of the nudges Thaler and Sunstein, and others, recommend are not hidden and work to correct cognitive bias. For example, purely informational nudges, such as nutrition labels, do not seem to us to be manipulative. They encourage individuals to slow down, reflect on, and make more informed decisions. By contrast, Thaler and Sunstein's famous cafeteria nudge—placing healthier foods at eye-level and less healthy foods below or above—seems plausibly manipulative, since it attempts to operate outside the individual's conscious awareness, and to leverage a decision-making bias. Of course, just because it's manipulative does not mean it isn't justified. To say that a strategy is manipulative is to draw attention to the fact that it carries a harm, which we discuss in detail below. It is possible, however, that the harm is justified by some greater benefit it brings with it.

Having defined manipulation as hidden or covert influence, and having distinguished manipulation from persuasion, coercion, deception, and nudging, it is possible to define "online manipulation" as *the use of information technology to covertly influence another person's decision-making, by targeting and exploiting decision-making vulnerabilities.* Importantly, we have adopted the term "online manipulation" from public discourse and interpret the word "online" expansively, recognising that there is no longer any hard boundary between online and offline life (if there ever was). "Online manipulation", as we understand it, designates manipulation *facilitated by information technology*, and could just as easily be termed "digital manipulation" or "automated manipulation". Since traditionally "offline" spaces are increasingly digitally mediated (because the people occupying them carry smartphones, the spaces themselves are embedded with internet-connected sensors, and so on), we should expect to encounter online manipulation beyond our computer screens.

Given this definition, it is not difficult to see why information technology is uniquely suited to facilitating manipulative influences. First, pervasive digital surveillance puts our decision-making vulnerabilities on permanent display. As privacy scholars have long pointed out, nearly everything we do today leaves a digital trace, and data collectors compile those traces into enormously detailed profiles (Solove, 2004). Such profiles comprise information about our demographics, finances, employment, purchasing behaviour, engagement with public services and institutions, and so on—in total, they often involve thousands of data points about each individual. By analysing patterns latent in this data, advertisers and others engaging in behavioural targeting are able to detect when and how to intervene in order to most effectively influence us (Kaptein & Eckles, 2010).

Moreover, digital surveillance enables detection of increasingly individual- or person-specific vulnerabilities.[16] Beyond the well-known cognitive biases discussed above (e.g., anchoring and framing effects), which condition most people's decision-making to some degree, we are also each subject to particular circumstances that can impact how we choose.[17] We are each prone to specific fears, anxieties, hopes, and desires, as well as physical, material, and economic realities, which—if known—can be used to steer our decision-making. In 2016, the voter micro-targeting firm Cambridge Analytica claimed to construct advertisements appealing to particular voter "psychometric" traits (such as openness, extraversion, etc.) by combining information about social media use with personality profiles culled from online quizzes.[18] And in 2017, an Australian newspaper exposed internal Facebook strategy documents detailing the company's alleged ability to detect when teenage users are feeling insecure. According to the report, "By monitoring posts, pictures, interactions and internet activity in real-time, Facebook can work out when young people feel 'stressed', 'defeated', 'overwhelmed', 'anxious', 'nervous', 'stupid',

'silly', 'useless', and a 'failure'" (Davidson, 2017). Though Facebook claims it never used that information to target advertisements at teenagers, it did not deny that it could. Extrapolating from this example it is easy to imagine others, such as banks targeting advertisements for high-interest loans at the financially desperate or pharmaceutical companies targeting advertisements for drugs at those suspected to be in health crisis.[19]

Second, digital platforms, such as websites and smartphone applications, are the ideal medium for leveraging these insights into our decision-making vulnerabilities. They are dynamic, interactive, intrusive, and adaptive choice architectures (Lanzing, 2018; Susser, 2019b; Yeung, 2017). Which is to say, the digital interfaces we interact with are configured in real time using the information about us described above, and they continue to learn about us as we interact with them. Unlike advertisements of old, they do not wait, passively, for viewers to drive past them on roads or browse over them in magazines; rather, they send text messages and push notifications, demanding our attention, and appear in our social media feeds at the precise moment they are most likely to tempt us. And because all of this is automated, digital platforms are able to adapt to each individual user, creating what Karen Yeung calls "highly personalised choice environment[s]"—decision-making contexts in which the vulnerabilities catalogued through pervasive digital surveillance are put to work in an effort to influence our choices (2017, p. 122).[20]

Third, if manipulation is hidden influence, then digital technologies are ideal vehicles for manipulation because they are already in a real sense hidden. We often think of technologies as objects we attend to and *use* with focus and attention. The language of technology design reflects this: we talk about "users" and "end users," "user interfaces," and "human-computer interaction". In fact, as philosophers (especially phenomenologists) and science and technology studies (STS) scholars have long shown, once we become habituated to a particular technology, the device or interface itself recedes from conscious attention, allowing us to focus on the tasks we are using it to accomplish.[21] Think of a smartphone or computer: we pay little attention to the devices themselves, or even to the way familiar websites or app interfaces are arranged. Instead, after becoming acclimated to them, we attend to the information, entertainment, or conveniences they offer (Rosenberger, 2009). Philosophers refer to this as "technological transparency"—the fact that we see, hear, or otherwise perceive *through* technologies—as though they were clear, transparent—onto the perceptual objects they convey to us (Ihde, 1990; Van Den Eede, 2011; Verbeek, 2005). Because this language of transparency can be confused with the concept of transparency familiar from technology policy discussions, we might more helpfully describe it as "invisibility" (Susser, 2019b). In addition to pervasive digital surveillance making our decision-making vulnerabilities easy to detect, and digital platforms making them easy to exploit, the ease with which our technologies become invisible to us—simply through frequent use and habituation—means the influences they facilitate are often hidden, and thus potentially manipulative.

Finally, although we focus primarily on the example of behavioural advertising to illustrate these dynamics, it is worth emphasising that advertisers are not the only ones engaging in manipulative practices. In the realm of user interface/experience (UI/UX) design, increasing attention is being paid to so-called "dark patterns"—design strategies that exploit users' decision-making vulnerabilities to nudge them into acting against their interests (or, at least, acting in the interests of the website or app), such as requiring automatically-renewing paid subscriptions that begin after an initial free trial period (Brignull, 2013; Gray, Kou, Battles, Hoggatt, & Toombs, 2018; Murgia, 2019; Singer, 2016). Though many of these strategies are as old as the internet and not all rise to the level of manipulation—sometimes overtly

inconveniencing users, rather than hiding their intentions—their growing prevalence has led some to call for legislation banning them (Bartz, 2019).

Worries about online manipulation have also been raised in the context of gig economy services, such as Uber and Lyft (Veen, Goods, Josserand, & Kaine, 2017). While these platforms market themselves as freer, more flexible alternatives to traditional jobs, providing reliable and consistent service to customers requires maintaining some amount of control over workers. However, without access to the traditional managerial controls of the office or factory floor, gig economy firms turn to "algorithmic management" strategies, such as notifications, customer satisfaction ratings, and other forms of soft control enabled through their apps (Rosenblat & Stark, 2016). Uber, for example, rather than requesting (or demanding) that workers put in longer hours, prompts drivers trying to exit the app with a reminder about their progress toward some earnings goal, exploiting the desire to continue making progress toward that goal; Lyft issues game-like "challenges" to drivers and stars and badges for accomplishing them (Mason, 2018; Scheiber, 2017).

In their current form, not all such practices necessarily manipulate—people are savvy, and many likely understand what they are facing. These examples are important, however, because they illustrate our present trajectory. Growing reliance on digital tools in all parts of our lives—tools that constantly record, aggregate, and analyse information about us—means we are revealing more and more about our individual and shared vulnerabilities. The digital platforms we interact with are increasingly capable of exploiting those insights to nudge and shape our choices, at home, in the workplace, and in the public sphere. And the more we become habituated to these systems, the less attention we pay to them.

2. THE HARM(S) OF ONLINE MANIPULATION

With this picture in hand, the question becomes: what exactly is the harm that results from influencing people in this way? Why should we be worried about technological mediation rendering us so susceptible to manipulative influence? In our view, there are several harms, but each flows from the same place—manipulation violates its target's autonomy.

The notion of autonomy points to an individual's capacity to make meaningfully independent decisions. As Joseph Raz puts it: "(t)he ruling idea behind the ideal of personal autonomy is that people should make their own lives" (Raz, 1986, p. 369). Making one's own life means freely facing both existential choices, like whom to spend one's life with or whether to have children, and pedestrian, everyday ones. And facing them freely means having the opportunity to think about and deliberate over one's options, considering them against the backdrop of one's beliefs, desires, and commitments, and ultimately deciding for reasons one recognises and endorses as one's own, absent unwelcome influence (J. P. Christman, 2009; Oshana, 2015; Veltman & Piper, 2014). Autonomy is in many ways the guiding normative principle of liberal democratic societies. It is because we think individuals can and should govern themselves that we value our capacity to collectively and democratically self-govern.

Philosophers sometimes operationalise the notion of autonomy by distinguishing between its competency and authenticity conditions (J. P. Christman, 2009, p. 155f). In the first place, being autonomous means having the cognitive, psychological, social, and emotional *competencies* to think through one's choices, form intentions about them, and act on the basis of those intentions. Second, it means that upon critical reflection one identifies with one's values,

desires, and goals, and endorses them *authentically* as one's own. Of course, many have criticised such conceptions of autonomy as overly rationalistic and implausibly demanding, arguing that we rarely decide in this way. We are emotional actors and creatures of habit, they argue, socialised and encultured into specific ways of choosing that we almost never reflect upon or endorse. But we understand autonomy broadly—our conception of deliberation includes not only beliefs and desires, but also emotions, convictions, and experiences, and critical reflection can be counterfactual (we must in principle be able to critically reflect on and endorse our motivations for acting, but we need not actually reflect on each and every move we make).

In addition to rejecting overly demanding and rationalistic conceptions of autonomy, we also reject overly atomistic ones. In our view, autonomous persons are socially, culturally, historically, and politically situated. Which is to say, we acknowledge the "intersubjective and social dimensions of selfhood and identity for individual autonomy and moral and political agency" (Mackenzie & Stoljar, 2000, p. 4).[22] Though social contexts can constrain our choices, by conditioning us to believe and behave in stereotypical ways (as, for example, in the case of gendered social expectations), it is also our social contexts that bestow value on autonomy, teaching us what it means to make independent decisions, and providing us with rich sets of options from which to choose. Moreover, it is crucial for present purposes that we emphasise our understanding of autonomy as more than an individual good—it is an essential social and political good too. Individuals express their autonomy across a variety of social contexts, from the home to the marketplace to the political sphere. Democratic institutions are meant to register and reflect the autonomous political decisions individuals make. Disrupting individual autonomy is thus more than an ethical concern; it has social and political import.

Against this picture of autonomy and its value, we can more carefully explain why online manipulation poses such a grave threat. To manipulate someone is, again, to covertly influence them, to intentionally alter their decision-making process without their conscious awareness. Doing so undermines the target's autonomy in two ways: first, it can lead them to act toward ends they haven't chosen, and second, it can lead them to act for reasons not authentically their own.

To see the first problem, consider examples of targeted advertising in the commercial sphere. Here, the aim of manipulators is fairly straightforward: they want people to buy things. Rather than simply put products on display, however, advertisers can construct decision-making environments—*choice architectures*—that subtly tempt or seduce shoppers to purchase their wares, and at the highest possible price (Calo, 2014). A variety of strategies might be deployed, from pointing out that one's friends have purchased the item to countdown clocks that pressure one to act before some offer expires, the goal being to hurry, evade, or undermine deliberation, and thus to encourage decisions that may or may not align with an individual's deeper, reflective, self-chosen ends and values.

Of course, these strategies are familiar from non-digital contexts; all commercial advertising (digital or otherwise) functions in part to induce consumers to buy things, and worries about manipulative ads emerged long before advertising moved online.[23] Equally, not all advertising—perhaps not even all targeted advertising—involves manipulation. Purely informational ads displayed to audiences actively seeking out related products and services (e.g., online banner ads displaying a doctor's contact information shown to visitors to a health-related website) are unlikely to covertly influence their targets. Worries about manipulation arise in cases where advertisements are *sneaky*—which is to say, where their effects are achieved covertly. If, for example, the doctor was a psychiatrist, his advertisements were shown to people

suspected of suffering from depression, and only at the specific times of day they were thought to be most afflicted, our account would offer grounds for condemning such tactics as manipulative.

It might also be the case that manipulation is not a binary phenomenon. We are the objects of countless influence campaigns and we understand some of them more than others; perhaps we ought to say that they are more or less manipulative in equal measure. On such a view, online targeted (or "behavioural") advertising could be understood as *exacerbating* manipulative dynamics common to other forms of advertising, by making the tweaks to individual choice architectures more subtle, and the seductions and temptations that result from them more difficult to resist (Yeung, 2017). Worse still, the fluidity and porousness of online environments makes it easy for marketers to conflate other distinct contexts with shopping, further blurring a person's reasoning about whether they truly want to make some purchase. For example, while chatting with friends over social media or searching for some place to eat, an ad may appear, thus requiring the target to juggle several tasks—in this case, communication and information retrieval—along with deliberation over whether or not to respond to the marketing ploy, thus diminishing the target's ability to sustain focus on any of the them. This problem is especially clearly illustrated by so-called "native advertising" (advertisements designed to look like user-generated, non-commercial content). Such advertisements are a kind of Trojan horse, intentionally conflating commercial and non-commercial activities in an attempt to undermine our capacity for focused, careful deliberation.

In the philosophical language introduced above, these strategies challenge both autonomy's competency and authenticity conditions. By deliberately and covertly engineering our choice environments to steer our decision-making, online manipulation threatens our competency to deliberate about our options, form intentions about them, and act on the basis of those intentions. And since, as we've seen, manipulative practices often work by targeting and exploiting our decision-making vulnerabilities—concealing their effects, leaving us unaware of the influence on our decision-making process—they also challenge our capacity to reflect on and endorse our reasons for acting as authentically on our own. Online manipulation thus harms us both by inducing us to act *toward ends* not of our choosing and *for reasons* we haven't endorsed.

Importantly, undermining personal autonomy in the ways just described can lead to further harms. First, since autonomous individuals are wont to protect (or at least to try and protect) their own interests, we can reasonably expect that undermining people's autonomy will lead, in many cases, to a diminishment of those interests. Losing the ability to look out for ourselves is unlikely to leave us better off in the long run. This harm—e.g., being tricked into buying things we don't need or paying more for them than we otherwise would—is well described by those who have analysed the problem of online manipulation in the commercial sphere (Calo, 2014; Nadler & McGuigan, 2018; Zarsky, 2006; Zarsky, 2019). And it is a serious harm, which we would do well to take seriously, especially given the fact that law and policy around information and internet practices (at least in the US) assume that individuals are for the most part capable of safeguarding their interests (Solove, 2013). However, it is equally important to see that this harm to welfare is derivative of the deeper harm to autonomy. Attempting to "protect consumers" from threats to their economic or other interests, without addressing the more fundamental threat to their autonomy, is thus to treat the symptoms without addressing the cause.

To bring this into sharper relief, it is worth pointing out that even purely beneficent

manipulation is harmful. Indeed, it is harmful to manipulate someone even in an effort to lead them more effectively toward *their own self-chosen ends*. That is because the fundamental harm of manipulation is to the process of decision-making, not its outcome. A well-meaning, paternalistic manipulator, who subtly induces his target to eat better food, exercise, and work hard, makes his target better off in one sense—he is healthier and perhaps more materially well-off—but it harms him as well by rendering him opaque to himself. Imagine if some bad habit, which someone had spent their whole life attempting to overcome, one day, all of a sudden, disappeared. They would be happy, of course, to be rid of the habit, but they might also be deeply confused and suspicious about the source of the change. As T.M. Scanlon writes, "I want to choose the furniture for my own apartment, pick out the pictures for the walls, and even write my own lectures despite the fact that these things might be done better by a decorator, art expert, or talented graduate student. For better or worse, I want these things to be produced by and reflect my own taste, imagination, and powers of discrimination and analysis. I feel the same way, even more strongly, about important decisions affecting my life in larger terms: what career to follow, where to work, how to live" (Scanlon, 1988).

Having said that, we have not demonstrated that manipulation is necessarily wrong in every case—only that it always carries a harm. One can imagine cases where the harm to autonomy is outweighed by the benefit to welfare. (For example, a case where someone's life is in immediate danger, and the only way to save them is by manipulating them.) But such cases are likely few and far between. What is so worrying about online manipulation is precisely its banality—the fact that it threatens to become a regular part of the fabric of everyday experience. As Jeremy Waldron argues, if we allow that to happen, our lives will be drained of something deeply important: "What becomes of the self-respect we invest in our own willed actions, flawed and misguided though they often are, when so many of our choices are manipulated to promote what someone else sees (perhaps rightly) as our best interest?" (Waldron, 2014) That we also lack reason to believe online manipulators really do have our best interests at heart is only more reason to resist them.

Finally, beyond the harm to individuals, manipulation promises a collective harm. By threatening our autonomy it threatens democracy as well. For autonomy is writ small what democracy is writ large—the capacity to self-govern. It is only because we believe individuals can make meaningfully independent decisions that we value institutions designed to register and reflect them. As the Cambridge Analytica case—and the public outcry in response to it—demonstrates, online manipulation in the political sphere threatens to undermine these core collective values. The problem of online manipulation is, therefore, not simply an ethical problem; it is a social and political one too.

3. TECHNOLOGY AND AUTONOMY

If one accepts the arguments advanced thus far, an obvious response is that we need to devise law and policy capable of preventing and mitigating manipulative online practices. We agree that we do. But that response is not sufficient—the question for policymakers is not simply how to mitigate online manipulation, but how to strengthen autonomy in the digital age. In making this claim, we join our voices with a growing chorus of scholars and activists—like Frischmann, Selinger, and Zuboff—working to highlight the corrosive effects of digital technologies on autonomy. Meeting these challenges requires more than consumer protection—it requires creating the positive conditions necessary for supporting individual and collective self-determination.

We don't pretend to have a comprehensive solution to these deep and complex problems, but some suggestions follow from our brief discussion. It should be noted that these suggestions—like the discussion, above, that prompted them—are situated firmly in the terrain of contemporary liberal political discourse, and those convinced that online manipulation poses a significant threat (especially some European readers) may be struck by how moderate our responses are. While we are not opposed to more radical interventions, we formulate our analysis using the conceptual and normative frameworks familiar to existing policy discussions in hopes of having an impact on them.

CURTAIL DIGITAL SURVEILLANCE

Data, as Tal Zarsky writes, is the "fuel" powering online manipulation (2019, p. 186). Without the detailed profiles cataloguing our preferences, interests, habits, and so on, the ability of would-be manipulators to identify our weaknesses and vulnerabilities would be vastly diminished, and so too their capacity to leverage them to their ends. Of course, the call to curtail digital surveillance is nothing new. Privacy scholars and advocates have been raising alarms about the ills of surveillance for half a century or more. Yet, as Zarsky argues, manipulation arguments could add to the "analytic and doctrinal arsenal of measures which enable legal intervention in the new digital environment" (2019, p. 185). Furthermore, outcry over apparent online manipulation in both the commercial and political spheres appears to be generating momentum behind new policy interventions to combat such strategies. In the US, a number of states have recently passed or are considering passing new privacy legislation, and the U.S. Congress appears to be weighing new federal privacy legislation as well. ("Congress Is Trying to Create a Federal Privacy Law", 2019; Merken, 2019). And, of course, all of that takes place on the heels of the new General Data Protection Regulation (GDPR) taking effect in Europe, which places new limits on when and what kinds of data can be collected about European citizens and by firms operating on European soil.[24] To curb manipulation and strengthen autonomy online, efforts to curtail digital surveillance ought to be redoubled.

PROBLEMATISE PERSONALISATION

When asked to justify collecting so much data about us, data collectors routinely argue that the information is needed in order to personalise their services to the needs and interests of individual users. Mark Zuckerberg, for example, attempted recently to explain Facebook's business model in the pages of the *Wall Street Journal*: "People consistently tell us that if they're going to see ads, they want them to be relevant," he wrote. "That means we need to understand their interests" (2019).[25] Personalisation seems, on the face of it, like an unalloyed good. Who wouldn't prefer a personalised experience to a generic one? Yet research into different forms of personalisation suggests that individualising—personalising—our experiences can carry with it significant risks.

These worries came to popular attention with Eli Pariser's book *Filter Bubble* (2011), which argued forcefully (though not without challenge) that the construction of increasingly singular, individualised experiences, means at the same time the loss of common, shared ones, and describes the detriments of that transformation to both individual and collective decision-making.[26] In addition to personalised information environments—Pariser's focus—technological advances enable things like personalised pricing - sometimes called "dynamic pricing" or "price discrimination" (Calo, 2014) and personalised work scheduling - or "just-in-time" scheduling (De Stefano, 2015). For the reasons discussed above, many such strategies may well be manipulative. The targeting and exploiting of individual decision-making vulnerabilities enabled by digital technologies—the potential for online manipulation they create—gives us reason to question whether the benefits of personalisation really outweigh the costs. At the very

least, we ought not to uncritically accept personalisation as a rationale for increased data collection, and we ought to approach with care (if not skepticism) the promise of an increasingly personalised digital environment.

PROMOTE AWARENESS AND UNDERSTANDING

If the central problem of online manipulation is its hiddenness, then any response must involve a drive toward increased awareness. The question is what form such awareness should take. Yeung argues that the predominant vehicle for notifying individuals about information flows and data practices—the privacy notice, or what is often called "notice-and-consent"—is insufficient (2017). Indeed, merely notifying someone that they are the target of manipulation is not enough to neutralise its effects. Doing so would require understanding not only *that* one is the target of manipulation, but also who the manipulator is, what strategies they are deploying, and why. Given the well-known "transparency paradox", according to which we are bound to either deprive users of relevant information (in an attempt to be succinct) or overwhelm them with it (in an attempt to be thorough), there is little reason to believe standard forms of notice alone can equip users to face the challenges of online manipulation.[27]

Furthermore, the problem of online manipulation runs deeper than any particular manipulative practice. What worries many people is the fact that manipulative strategies, like targeted advertising, are becoming basic features of the digital world—so commonplace as to escape notice or mention.[28] In the same way that machine learning and artificial intelligence tools have quickly and quietly been delegated vast decision-making authorities in a variety of contemporary contexts and institutions, and in response, scholars and activists have mounted calls to make their decision-making processes more explainable, transparent, and accountable, so too must we give people tools to understand and manage a digital environment designed to shape and influence them.[29]

ATTEND TO CONTEXT

Finally, it is important to recognise that moral intuitions about manipulation are indexed to social context. Which is to say, we are willing to tolerate different levels of outside influence on our decision-making in different decision-making spheres. As relatively lax commercial advertising regulations indicate, we are—at least in the US—willing to accept a fair amount of interference in the commercial sphere. By contrast, somewhat more stringent regulations around elections and campaign advertising suggest that we are less willing to accept such interference in the realm of politics.[30] Responding to the threats of online manipulation therefore requires sensitivity to where—in which spheres of life—we encounter them.

CONCLUSION

The idea that technological advancements bring with them new arrangements of power is, of course, nothing new. That online manipulation threatens to subordinate the interests of individuals to those of data collectors and their clients is thus, in one respect, a familiar (if nonetheless troubling) problem. What we hope to have shown, however, is that the threat of online manipulation is deeper, more insidious, than that. Being steered or controlled, outside our conscious awareness, violates our autonomy, our capacity to understand and author our own lives. If the tools that facilitate such control are left unchecked, it will be to our individual and collective detriment. As we've seen, information technology is in many ways an ideal vehicle for these forms of control, but that does not mean that they are inevitable. Combating online

manipulation requires both depriving it of personal data—the oxygen enabling it—and empowering its targets with awareness, understanding, and savvy about the forces attempting to influence them.

REFERENCES

Abramowitz, M. J. (2017, December 11). Stop the Manipulation of Democracy Online. *The New York Times*. Retrieved from https://www.nytimes.com/2017/12/11/opinion/fake-news-russia-kenya.html

Anderson, J., & Honneth, A. (2005). Autonomy, Vulnerability, Recognition, and Justice. In J. Christman & J. Anderson (Eds.), *Autonomy and the Challenges to Liberalism* (pp. 127–149). doi:10.1017/CBO9780511610325.008

Bartz, D. (2019, April 13). U.S. senators introduce social media bill to ban "dark patterns" tricks. *Reuters*. Retrieved from https://www.reuters.com/article/us-usa-tech-idUSKCN1RL25Q

Benkler, Y., Faris, R., & Roberts, H. (2018). *Network Propaganda: Manipulation, Disinformation, and Radicalization in American Politics*. New York: Oxford University Press.

Blumenthal, J. A. (2005). Does Mood Influence Moral Judgment? An Empirical Test with Legal and Policy Implications. *Law & Psychology Review*, *29*, 1–28.

Boerman, S. C., Kruikemeier, S., & Zuiderveen Borgesius, F. J. (2017). Online Behavioral Advertising: A Literature Review and Research Agenda. *Journal of Advertising*, *46*(3), 363–376. doi:10.1080/00913367.2017.1339368

Bovens, L. (2009). The Ethics of Nudge. In T. Grüne-Yanoff & S. O. Hansson (Eds.), *Preference Change: Approaches from Philosophy, Economics and Psychology* (pp. 207–219). Dordrecht: Springer Netherlands.

Brignull, H. (2013, August 29). Dark Patterns: inside the interfaces designed to trick you. Retrieved June 17, 2019, from The Verge website: https://www.theverge.com/2013/8/29/4640308/dark-patterns-inside-the-interfaces-designed-to-trick-you

Calo, M. R. (2014). Digital Market Manipulation. *The George Washington Law Review*, *82*(4). Retrieved from https://www.gwlr.org/wp-content/uploads/2018/01/82-Geo.-Wash.-L.-Rev.-995.pdf

Cambridge Analytica and Facebook: The Scandal so Far. (2018, March 28). *Al Jazeera News*. Retrieved from https://www.aljazeera.com/news/2018/03/cambridge-analytica-facebook-scandal-180327172353667.html

Christman, J. P. (2009). *The Politics of Persons: Individual Autonomy and Socio-Historical Selves*. Cambridge; New York: Cambridge University Press.

Congress Is Trying to Create a Federal Privacy Law. (2019, February 28). *The Economist*. Retrieved from https://www.economist.com/united-states/2019/02/28/congress-is-trying-to-create-a-federal-privacy-law

Davidson, D. (2017, May 1). Facebook targets "insecure" to sell ads. *The Australian*.

De Stefano, V. (2015). The Rise of the "Just-in-Time Workforce": On-Demand Work, Crowd Work and Labour Protection in the "Gig-Economy." *SSRN Electronic Journal*. doi:10.2139/ssrn.2682602

Doubek, J. (2017, November 16). How Disinformation And Distortions On Social Media Affected Elections Worldwide. Retrieved March 24, 2019, from NPR.org website: https://www.npr.org/sections/alltechconsidered/2017/11/16/564542100/how-disinformation-and-distortions-on-social-media-affected-elections-worldwide

Dubois, E., & Blank, G. (2018). The echo chamber is overstated: The moderating effect of political interest and diverse media. *Information, Communication & Society*, *21*(5), 729–745. doi:10.1080/1369118X.2018.1428656

Franken, I. H. A., & Muris, P. (2005). Individual Differences in Decision-Making. *Personality and Individual Differences*, *39*(5), 991–998. doi:10.1016/j.paid.2005.04.004

Frischmann, B., & Selinger, E. (2018). *Re-Engineering Humanity* (1st ed.). doi:10.1017/9781316544846

Gray, C. M., Kou, Y., Battles, B., Hoggatt, J., & Toombs, A. L. (2018). The Dark (Patterns) Side of UX Design. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18*, 1–14. doi:10.1145/3173574.3174108

Hausman, D. M., & Welch, B. (2010). Debate: To Nudge or Not to Nudge. *Journal of Political Philosophy*, *18*(1), 123–136. doi:10.1111/j.1467-9760.2009.00351.x

Ihde, D. (1990). *Technology and the Lifeworld: From Garden to Earth.* Bloomington: Indiana University Press.

Kahneman, D. (2013). *Thinking, Fast and Slow* (1st pbk. ed). New York: Farrar, Straus and Giroux.

Kaptein, M., & Eckles, D. (2010). Selecting Effective Means to Any End: Futures and Ethics of Persuasion Profiling. In T. Ploug, P. Hasle, & H. Oinas-Kukkonen (Eds.), *Persuasive Technology* (Vol. 6137, pp. 82–93). doi:10.1007/978-3-642-13226-1_10

Kligman, M., & Culver, C. M. (1992). An Analysis of Interpersonal Manipulation. *Journal of Medicine and Philosophy*, *17*(2), 173–197. doi:10.1093/jmp/17.2.173

Lanzing, M. (2018). "Strongly Recommended" Revisiting Decisional Privacy to Judge Hypernudging in Self-Tracking Technologies. *Philosophy & Technology*. doi:10.1007/s13347-018-0316-4

Levinson, J. D., & Peng, K. (2007). Valuing Cultural Differences in Behavioral Economics. *ICFAI Journal of Behavioral Finance*, *4*(1).

Mackenzie, C., & Stoljar, N. (Eds.). (2000). *Relational Autonomy: Feminist Perspectives on Automomy, Agency, and the Social Self*. New York: Oxford University Press.

Mason, S. (2018, November 20). High score, low pay: Why the gig economy loves gamification. *The Guardian*. Retrieved from https://www.theguardian.com/business/2018/nov/20/high-score-low-pay-gamification-lyft-uber-drivers-ride-hailing-gig-economy

Matz, S. C., Kosinski, M., Nave, G., & Stillwell, D. J. (2017). Psychological Targeting as an Effective Approach to Digital Mass Persuasion. *Proceedings of the National Academy of Sciences*, *114*(48), 12714–12719. doi:10.1073/pnas.1710966114

Merken, S. (2019, February 6). States Follow EU, California in Push for Consumer Privacy Laws. Retrieved March 25, 2019, from Bloomberg Law website: https://news.bloomberglaw.com/privacy-and-data-security/states-follow-eu-california-in-push-for-consumer-privacy-laws-1

Murgia, M. (2019, May 4). When manipulation is the business model. *Financial Times*.

Nadler, A., & McGuigan, L. (2018). An Impulse to Exploit: The Behavioral Turn in Data-Driven Marketing. *Critical Studies in Media Communication*, *35*(2), 151–165. doi:10.1080/15295036.2017.1387279

Nissenbaum, H. (2011). A Contextual Approach to Privacy Online. *Daedalus*, *140*(4), 32–48. doi:10.1162/DAED_a_00113

Noggle, R. (2018a). Manipulation, Salience, and Nudges. *Bioethics*, *32*(3), 164–170. doi:10.1111/bioe.12421

Noggle, R. (2018b). The Ethics of Manipulation. In E. N. Zalta (Ed.), *Stanford Encyclopedia of Philosophy* (p. 24). Retrieved from https://plato.stanford.edu/entries/ethics-manipulation/

Nys, T. R., & Engelen, B. (2017). Judging Nudging: Answering the Manipulation Objection. *Political Studies*, *65*(1), 199–214. doi:10.1177/0032321716629487

Oshana, M. (Ed.). (2015). *Personal Autonomy and Social Oppression: Philosophical Perspectives* (First edition). New York: Routledge, Taylor & Francis Group.

Pariser, E. (2011). *The Filter Bubble: What the Internet Is Hiding from You*. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&db=nlabk&AN=1118322

Rachlinski, J. J. (R). Cognitive Errors, Individual Differences, and Paternalism. *University of Chicago Law Review*, *73*(1), 207–229. Available at https://chicagounbound.uchicago.edu/uclrev/vol73/iss1/11/

Raz, J. (1986). *The Morality of Freedom* (Reprinted). Oxford: Clarendon Press.

Reach, G. (2016). Patient education, nudge, and manipulation: Defining the ethical conditions of the person-centered model of care. *Patient Preference and Adherence*, *10*, 459–468. doi:10.2147/PPA.S99627

Richards, N. M. (2013). The Dangers of Surveillance. *Harvard Law Review*, *126*(7), 1934–1965. Available at https://harvardlawreview.org/2013/05/the-dangers-of-surveillance/

Rosenberger, R. (2009). The Sudden Experience of the Computer. *AI & Society*, *24*(2), 173–180. doi:10.1007/s00146-009-0190-9

Rosenblat, A., & Stark, L. (2016). Algorithmic Labor and Information Asymmetries: A Case Study of Uber's Drivers. *International Journal of Communication*, *10*, 3758–3784. Retrieved from https://ijoc.org/index.php/ijoc/article/view/4892

Rudinow, J. (1978). Manipulation. *Ethics*, *88*(4), 338–347. doi:10.1086/292086

Scanlon, T. M. (1988). The Significance of Choice. In A. Sen & S. M. McMurrin (Eds.), *The*

Tanner Lectures on Human Values (Vol. 8, p. 68).

Scheiber, N. (2017, April 2). How Uber Uses Psychological Tricks to Push Its Drivers' Buttons. *The New York Times*. Retrieved from https://www.nytimes.com/interactive/2017/04/02/technology/uber-drivers-psychological-tricks.html

Selbst, A. D., & Barocas, S. (2018). The Intuitive Appeal of Explainable Machines. *Fordham Law Review*, *87*(3), 1085-1139. Retrieved from https://ir.lawnet.fordham.edu/flr/vol87/iss3/11/

Selinger, E., & Whyte, K. (2011). Is There a Right Way to Nudge? The Practice and Ethics of Choice Architecture. *Sociology Compass*, *5*(10), 923–935. doi:10.1111/j.1751-9020.2011.00413.x

Singer, N. (2016, May 14). When Websites Won't Take No for an Answer. *The New York Times*. Retrieved from https://www.nytimes.com/2016/05/15/technology/personaltech/when-websites-wont-take-no-for-an-answer.html

Solove, D. J. (2004). *The Digital Person: Technology and Privacy In The Information Age*. New York: New York University Press.

Solove, D. J. (2013). Privacy Self-Management and the Consent Dilemma. *Harvard Law Review*, *126*(7), 1880–1903. Retrieved from https://harvardlawreview.org/2013/05/introduction-privacy-self-management-and-the-consent-dilemma/

Stanovich, K. E., & West, R. F. (1998). Individual Differences in Rational Thought. *Journal of Experimental Psychology: General*, *127*(2), 161–188. doi:10.1037/0096-3445.127.2.161

Stole, I. L. (2014). Persistent Pursuit of Personal Information: A Historical Perspective on Digital Advertising Strategies. *Critical Studies in Media Communication*, *31*(2), 129–133. doi:10.1080/15295036.2014.921319

Sunstein, C. R. (2016). *The Ethics of Influence: Government in the Age of Behavioral Science*. Cambridge: Cambridge University Press.

Susser. (2019a). Notice After Notice-and-Consent: Why Privacy Disclosures Are Valuable Even If Consent Frameworks Aren't. *Journal of Information Policy*, *9*, 37–62. doi:10.5325/jinfopoli.9.2019.0037

Susser, D. (2019b). *Invisible Influence: Artificial Intelligence and the Ethics of Adaptive Choice Architectures*. Presented at the AAAI/ACM Conference on AI, Ethics, and Society (AIES '19), Honolulu. Available at http://www.aies-conference.com/wp-content/papers/main/AIES-19_paper_54.pdf

Susser, D., Roessler, B., & Nissenbaum, H. (2018). Online Manipulation: Hidden Influences in a Digital World. *SSRN Electronic Journal*. Retrieved from https://papers.ssrn.com/abstract=3306006

Thaler, R. H., & Sunstein, C. R. (2008). *Nudge: Improving Decisions About Health, Wealth, and Happiness*. New Haven: Yale University Press.

Tufekci, Z. (2014). Engineering the Public: Big Data, Surveillance and Computational Politics. *First Monday*, *19*(7). doi:10.5210/fm.v19i7.4901

Turow, J. (2012). *The Daily You: How the New Advertising Industry Is Defining Your Identity and Your Worth*. Retrieved from https://books.google.com/books?id=rK7JSFudXA8C

Vaidhyanathan, S. (2018). *Antisocial Media: How Facebook Disconnects Us and Undermines Democracy*. New York; Oxford: Oxford University Press.

Van Den Eede, Y. (2011). In Between Us: On the Transparency and Opacity of Technological Mediation. *Foundations of Science*, *16*(2/3), 139–159. doi:10.1007/s10699-010-9190-y

Vayena, M. I., Effy. (2018, March 30). Cambridge Analytica and Online Manipulation. Retrieved March 24, 2019, from Scientific American Blog Network website: https://blogs.scientificamerican.com/observations/cambridge-analytica-and-online-manipulati on/

Veen, A., Goods, C., Josserand, E., & Kaine, S. (2017, June 18). "The way they manipulate people is really saddening": Study shows the trade-offs in gig work. Retrieved June 16, 2019, from The Conversation website: http://theconversation.com/the-way-they-manipulate-people-is-really-saddening-study-shows-the-trade-offs-in-gig-work-79042

Veltman, A., & Piper, M. (Eds.). (2014). *Autonomy, Oppression, and Gender*. Oxford; New York: Oxford University Press.

Verbeek, P.-P. (2005). *What Things Do: Philosophical Reflections on Technology, Agency, and Design*. University Park: Pennsylvania State University Press.

Waldron, J. (2014, October 9). It's All for Your Own Good. *The New York Review of Books*. Retrieved from https://www.nybooks.com/articles/2014/10/09/cass-sunstein-its-all-your-own-good/

Westin, A. F. (2015). *Privacy and Freedom*. New York: IG Publishing.

Wood, A. (2014). Coercion, Manipulation, Exploitation. In C. Coons & M. Weber (Eds.), *Manipulation: Theory and Practice*. Oxford□; New York: Oxford University Press.

Yeung, K. (2017). Hypernudge: Big Data as a Mode of Regulation by Design. *Information, Communication & Society*, *20*(1), 118–136. doi:10.1080/1369118X.2016.1186713

Zarsky, T. (2006). Online Privacy, Tailoring, and Persuasion. In K. J. Strandburg & D. S. Raicu (Eds.), *Privacy and Technologies of Identity: A Cross-Disciplinary Conversation* (pp. 209–224). doi:10.1007/0-387-28222-X_12

Zarsky, T. Z. (2019). Privacy and Manipulation in the Digital Age. *Theoretical Inquiries in Law*, *20(1)*, 157–188. http://www7.tau.ac.il/ojs/index.php/til/article/view/1612

Zittrain, J. (2014). Engineering an Election. *Harvard Law Review Forum*, *127*(8), 335–341. Retrieved from https://harvardlawreview.org/2014/06/engineering-an-election/

Zuboff, S. (2015). Big Other: Surveillance Capitalism and the Prospects of an Information Civilization. *Journal of Information Technology*, *30*(1), 75–89. doi:10.1057/jit.2015.5

Zuboff, S. (2019). *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power* (First edition). New York: Public Affairs.

Zuckerberg, M. (2019, January 25). The Facts About Facebook. *Wall Street Journal*. Retrieved from

http://ezaccess.libraries.psu.edu/login?url=https://search-proquest-com.ezaccess.libraries.psu.edu/docview/2170828623?accountid=13158

Zuiderveen Borgesius, F. J., Möller, J., Kruikemeier, S., Ó Fathaigh, R., Irion, K., Dobber, T., … De Vreese, C. (2018). Online Political Microtargeting: Promises and Threats for Democracy. *Utrecht Law Review*, *14*(1), 82–96. doi:**10.18352/ulr.420**

Zuiderveen Borgesius, F. J., Trilling, D., Möller, J., Bodó, B., De Vreese, C. H., & Helberger, N. (2016). Should We Worry About Filter Bubbles? *Internet Policy Review*, *5*(1). doi:**10.14763/2016.1.401**

FOOTNOTES

1. Richards describes this influence as "persuasion" and "subtle forms of control". In our view, for reasons discussed below, the subtler forms of influence ought really to be called "manipulation".

2. For a wide-ranging review of the scholarly literature on targeted advertising, see (Boerman, Kruikemeier, & Zuiderveen Borgesius, 2017).

3. See, for example, Zarsky gestures at there being more at stake than consumer interests, but he explicitly declines to develop the point, framing the problem instead as one of consumer protection. See (2006; 2019)

4. Which is not to say that no one saw this coming. As far back as 1967, Alan Westin warned about "the entire range of forthcoming devices, techniques, and substances that enter the mind to implant influences or extract data" and their application "in commerce or politics" (Westin, 2015, p. 331). See also (Tufekci, 2014; Zittrain, 2014).

5. Frischmann and Selinger write: "Across cultures and generations, humans have engineered themselves and their built social environments to sustain capacities for thinking, the ability to socialize and relate to each other, free will, autonomy, and agency, as well as other core capabilities. […T]hey are at risk of being whittled away through modern forms of techno-social engineering." (2018, p. 271). And Zuboff argues that the behaviour modifications characteristic of surveillance capitalism "sacrifice our right to the future tense, which comprises our will to will, our autonomy, our decision rights, our privacy, and, indeed, our human natures" (2019, p. 347).

6. As Frischmann and Selinger write, "We are fundamentally techno-social animals" (2018, p. 271).

7. For a more fully developed and defended version of our account, see Susser, Roessler, and Nissenbaum (2018).

8. (Benkler, Faris, & Roberts, 2018). See also the many excellent reports from the Data & Society Research Institute's "Media Manipulation" project: https://datasociety.net/research/media-manipulation/

9. Examples from Noggle (2018b).

10. The term "persuasion" is sometimes used in a broader sense, as a synonym for "influence".

Here we use it in the narrower sense of *rational persuasion,* since our goal is precisely to distinguish between different forms of influence.

11. Assuming we ever do learn that we have been manipulated. Presumably we often do not.

12. As Luc Bovens writes about nudges (discussed below), such strategies "typically work better in the dark" (2009, p. 209).

13. The classic formulation of these ideas comes from Daniel Kahneman and Amos Tversky, summarised in (Kahneman, 2013). See also (Thaler & Sunstein, 2008).

14. Writing about manipulation in 1978, Joel Rudinow observed: "Weaknesses are rarely displayed; they are betrayed. Since our weaknesses, in addition to making us vulnerable, are generally repugnant to us, we generally do our best to conceal them, not least from ourselves. Consequently too few people are insightful enough into or familiar enough with enough other people to make the use of resistible incentives a statistically common form of manipulation. In addition we are not always so situated as to be able genuinely to offer someone the incentive which we believe will best suit our manipulative aims. Just as often it becomes necessary to deceive someone in order to play on his weakness. Thus it is only to be expected that deception plays a role in the great majority of cases of manipulation." (Rudinow, 1978, p. 347) As we'll see below, it is precisely the limitations confronting the would-be manipulator in 1978, which Rudinow identifies, that thanks to technology have since been overcome.

15. Thaler and Sunstein refer to this as "libertarian paternalism" (2008).

16. Our thanks to a reviewer of this essay for the term "person-specific vulnerability."

17. In fact, while we are all susceptible to the kinds of cognitive biases discussed by behavioral economists to some degree, we are not all susceptible to each bias to the same degree (Rachlinski, R; Stanovich & West, 1998). Empirical evidence suggests that individual differences in personality (Franken & Muris, 2005), cultural background (Levinson & Peng, 2007), and mood (Blumenthal, 2005), among others, can modulate how individuals are impacted by particular biases. It is not difficult to imagine digital tools detecting these differences and leveraging them to structure particular interventions.

18. Cambridge Analytica's then-CEO Alexander Nix discusses these tactics here: https://www.youtube.com/watch?v=n8Dd5aVXLCc. Research suggests such tactics are plausible, see Matz, Kosinski, Nave, and Stillwell (2017).

19. For a deeper discussion about vulnerability, its varieties, and the ways vulnerabilities can be leveraged by digital tools, see Susser, Roessler, and Nissenbaum (2018).

20. See also Susser (2019b).

21. For an excellent discussion of the different ways this idea has been elaborated by a variety of philosophers and STS scholars, see Van Den Eede (2011).

22. It is worth noting, however: just because individuals and their capacities are inextricably social, that does not mean autonomy is only possible in egalitarian social contexts. See Anderson and Honneth (2005).

23. "While much about digital advertising appears revolutionary, it would be wrong to accept

the notion of customer surveillance as a modern phenomenon. Although the internet's technological advances have taken advertising in new directions and the practice of 'data-mining' to almost incomprehensible extremes, nearly all of what is transpiring reflects some of the basic methods developed by marketers beginning a hundred years ago" (Stole, 2014).

24. See https://eugdpr.org

25. Zuckerberg also cited needing user information for "security and operating our services".

26. Some empirical researchers have expressed skepticism about the alleged harms of filter bubbles, some even suggesting that they are beneficial (Dubois & Blank, 2018; Zuiderveen Borgesius et al., 2016). Their findings, however, are far from conclusive.

27. On the "transparency paradox," see Nissenbaum (2011). Though privacy notices are, in themselves, insufficient for shielding individuals from the effects of online manipulation, that does not mean that they are entirely without value. They might support individual autonomy, even if they can't guarantee it: see Susser (2019a).

28. For example, Marcella Vayena writes: "[N]ot just Cambridge Analytica, but most of the current online ecosystem, is an arm's race to the unconscious mind: notifications, microtargeted ads, autoplay plugins, are all strategies designed to induce addictive behavior, hence to manipulate" (Vayena, 2018).

29. For a helpful discussion about the calls for—and limits of—explainable artificial intelligence, see (Selbst & Barocas, 2018)

30. In a longer version of this paper, we also consider online manipulation in the context of the workplace. See Susser et al. (2018).