ONLINE MANIPULATION: HIDDEN INFLUENCES IN A DIGITAL WORLD

Daniel Susser, Beate Roessler, Helen Nissenbaum[*]

CITE AS: 4 GEO. L. TECH. REV. 1 (2019)

TABLE OF CONTENTS

I.   INTRODUCTION

Privacy and surveillance scholars increasingly worry that data collectors can use the information they gather about our behaviors, preferences, interests, incomes, and so on to manipulate us.[1] Consider: investigative journalists recently discovered that Facebook allows advertisers to target vulnerable teenagers at moments when they feel "worthless" and "insecure."[2] "Sharing economy" firms like the ride-hailing company Uber have explored ways to influence not only the behavior of their customers but also that of their drivers, raising concerns about potential manipulation in the workplace.[3] And recent elections in the United States, the United Kingdom, Germany, France, and elsewhere have raised questions about the use of similar techniques to manipulate democratic political processes.[4]

Charges that some practices are manipulative are a strong caution, even a rallying cry to protest. But what it means, exactly, to manipulate someone and how we might systematically distinguish cases of manipulation from other forms of influence—such as persuasion and coercion—has not been thoroughly enough explored in light of the unprecedented capacities that information technologies and digital media enable. This Article endeavors to meet this challenge—to develop a definition of manipulation that addresses these enhanced capacities, to investigate how information technologies facilitate manipulative practices, and to describe the harms to individuals and social institutions that flow from such practices. We use the term "online manipulation" to highlight the particular class of manipulative practices

---

[1] For example, Tal Zarsky offers an early discussion of the problem in *Online Privacy, Tailoring, and Persuasion*, *in* PRIVACY AND TECHNOLOGIES OF IDENTITY: A CROSS-DISCIPLINARY CONVERSATION (Katherine J. Strandburg & Daniela Stan Raicu eds., 2006). Frank Pasquale points to it throughout his book, THE BLACK BOX SOCIETY: THE SECRET ALGORITHMS THAT CONTROL MONEY AND INFORMATION (2015). Frederik Zuiderveen Borgesius provides a helpful treatment of manipulation questions, especially as they relate to European privacy law in IMPROVING PRIVACY PROTECTION IN THE AREA OF BEHAVIOURAL TARGETING (2015). The most far-reaching treatment is Ryan Calo's discussion of "digital market manipulation" and consumer protection law in *Digital Market Manipulation*, 82 GEO. WASH. L. REV. 995 (2014).

[2] Sam Machkovech, *Facebook Helped Advertisers Target Teens Who Feel "Worthless,"* ARS TECHNICA (May 1, 2017), https://arstechnica.com/information-technology/2017/05/facebook-helped-advertisers-target-teens-who-feel-worthless/ [https://perma.cc/BPD9-6NKP].

[3] *See generally* Ryan Calo & Alex Rosenblat, *The Taking Economy: Uber, Information, and Power*, 117 COLUM. L. REV. 1623 (2017).

[4] *See generally* Zeynep Tufekci, *Engineering the Public: Big Data, Surveillance, and Computational Politics*, 19 FIRST MONDAY 7 (July 7, 2014), https://journals.uic.edu/ojs/index.php/fm/article/view/4901/4097 (charting the questions of manipulating democratic political processes) [https://perma.cc/4XVR-KRKL].

enabled by a broad range of information technologies. We aim to contribute to philosophical accounts of manipulation with our conceptual and normative work by clarifying the nature of manipulative practices and drawing attention to the new universe of manipulation cases that information technology raises. Beyond philosophical accounts, however, our work engages with law and policy debates and aims to guide efforts to combat the corrosive impacts of manipulation.

In Part II of the Article, we describe cases that have been cited as instances of manipulation. Then, drawing on the nature of these cases and on discussions about manipulation in moral philosophy, we present our own account of manipulation in Part III, defining manipulation and distinguishing it from neighboring terms. We argue that at its core, manipulation is hidden influence—the covert subversion of another person's decision-making power. In contrast with persuasion, which is the forthright appeal to another person's decision-making power, or coercion, which is the restriction of acceptable options from which another person might choose, manipulation functions by exploiting the manipulee's cognitive (or affective) weaknesses and vulnerabilities in order to steer his or her decision-making process towards the manipulator's ends. Manipulation, therefore, shares certain features with nudging, though we argue that only some nudges are manipulative. In addition, some have argued that manipulation is merely a species of deception.[5] We argue that while deception is often a tool of manipulation, manipulating someone does not necessarily require instilling false beliefs.

In Part IV, we describe the particular forms manipulation takes in a world where digital technologies pervade everyday life. We argue that information technology, for a number of reasons, makes engaging in manipulative practices significantly easier, and it makes the effects of such practices potentially more deeply debilitating. First, widespread digital surveillance makes it easy for data collectors and aggregators to identify our weaknesses. The information we volunteer and shed about our interests, preferences, desires, emotional states, beliefs, habits, and so on, provides everything a would-be manipulator needs to know about how to subvert our decision-making.[6] Second, digital platforms offer a perfect medium through which to leverage those insights. They are dynamic, interactive, intrusive, and incisively personalizable choice architectures—decision-making contexts that can be specifically designed to adapt to and exploit each individual user's

---

[5] For example, see our discussion of Robert Goodin's theory of manipulation, *infra* Section II.B.

[6] Karen Yeung describes the problem in terms of "hypernudging." Karen Yeung, *Hypernudge: Big Data as a Mode of Regulation by Design*, 20 INFO. COMM. & SOC'Y 118 (2017).

particular vulnerabilities.[7] Finally, the reach of digital tools is enormous. Because digital interfaces mediate so much of so many people's lives, they have the potential to affect far more people far more deeply than their analogue counterparts. Social media services, like Facebook, with millions or even billions of users can be leveraged as tools of massive and hyper-targeted manipulation.

In Part V, we turn to the harms of online manipulation. Subverting another person's decision-making power undermines his or her autonomy. Given that respect for individual autonomy is a bedrock principle of liberal democracy, the potential for massive online manipulation is a cause for grave concern. We flesh out the notion of "vulnerability," mapping its various types and the ways they are exploited as well as the places where different kinds of vulnerabilities interact to reinforce or exacerbate one another. We show how altering people's choice architecture to exploit their vulnerabilities affects them. And we look to philosophical accounts of autonomy—specifically to accounts of socially-situated, relational autonomy—to bring these strands together, showing where and how manipulative influences thwart people's capacity to form decisions they can recognize and endorse as their own.

We conclude by considering directions for future research, suggesting that it is especially important to consider the effects of online manipulation in different social contexts. Although we might be willing to tolerate some amount of external influence on our decision-making processes in consumer contexts, we likely want less interference in political contexts. We argue that the role autonomy plays in political decision-making is more fundamental and consequential than it is in consumer decision-making. Threats to the former therefore demand a more vigorous response. As we work to combat the new forms of manipulative practice made possible by information technology, we will need to distinguish carefully among the contexts in which they operate.

II. SOME CASES OF TROUBLING INFLUENCE

To provide concrete vehicles for our analysis, we introduce a few well-known cases to our discussion. First, we consider targeted advertising in the commercial sphere, examining worrying reports that Facebook has the ability to target advertisements at teenagers during moments they are perceived to be especially vulnerable to influence. Second, we consider cases of so-called "algorithmic management"—strategies that gig platforms such as Uber use to influence worker behavior.[8] Some of these strategies may verge on

---

[7] See generally Marjolein Lanzing, "Strongly Recommended" Revisiting Decisional Privacy to Judge Hypernudging in Self-Tracking Technologies, 32 PHIL. & TECH. 549 (2019).

[8] See generally Alex Rosenblat & Luke Stark, Algorithmic Labor and Information Asymmetries: A Case Study of Uber's Drivers, 10 INT'L J. COMM. 3758 (2016).

manipulative, raising difficult questions about appropriate influence in the workplace. Third, we consider Cambridge Analytica and claims that the political advertising firm engaged in voter manipulation. Together, these three cases exemplify concerns about online manipulation across social spheres—in the market, in the workplace, and in the realm of politics.

A. Targeting Advertisements at Vulnerable Teenagers

Consumer advertising has long been contested ethical terrain. Its defenders argue that advertising plays an important role in an ideal competitive free market, informing consumers about products so they may effectively select among the alternatives and thus ensuring supply meets demand and prices adjust accordingly.[9] Critics see the opposite. They identify trends in advertising that are increasingly aimed at provoking action against reason or circumventing reason altogether. Vance Packard famously charged the advertising industry with utilizing "motivation analysis," psychological and psychoanalytical means to exploit "hidden weaknesses and frailties," to appeal to non-rational and subconscious mental processes in service of marketing ends.[10] And, of course, some stake a middle ground by suggesting that advertising can serve a useful function even though certain forms and outcomes are deeply problematic.[11] The answer is neither to abandon nor outlaw but to divine criteria to determine when advertising is performing a useful service (such as informing consumers of and about products, promoting an active marketplace, or even promoting brand loyalty) and when it is not.[12]

With the commercialization of the Internet, it was inevitable that advertising would migrate online, and alongside it, detractors who regretted any such incursions.[13] "Sponsored search," which delivers advertisements tailored to specific search terms, was initially the norm and was developed by Web search services that could easily perform this match.[14] This was quickly followed by advertising targeted at individuals, with DoubleClick (now owned

---

[9] *See, e.g.*, JERRY KIRKPATRICK, IN DEFENSE OF ADVERTISING: ARGUMENTS FROM REASON, ETHICAL EGOISM, AND LAISSEZ-FAIRE CAPITALISM (1994).

[10] *See generally* VANCE PACKARD, THE HIDDEN PERSUADERS (1957).

[11] Some in the advertising industry take this position, recognizing that there is such a thing as unethical advertising but rejecting the notion that all advertising is unethical. *See, e.g.*, Wallace S. Snyder, *Ethics in Advertising: The Players, the Rules, and the Scorecard,* 22 BUS. & PROF. ETHICS J. 1 (2003).

[12] *See, e.g.*, MICHAEL SCHUDSON, ADVERTISING, THE UNEASY PERSUASION: ITS DUBIOUS IMPACT ON AMERICAN SOCIETY (Routledge 2013) (1984).

[13] For a historical account of the emergence of online advertising, see generally TIM WU, THE ATTENTION MERCHANTS: THE EPIC SCRAMBLE TO GET INSIDE OUR HEADS (2016).

[14] *See generally* Bernard J. Jansen & Tracy Mullen, *Sponsored Search: An Overview of the Concept, History, and Technology*, 6 INT'L J. ELEC. BUS. 2 (2008).

by Google) taking advantage of the online advertising industry's victory in the Internet Engineering Task Force (IETF) debate over the Web cookie standard, which allowed third-parties to set and retrieve cookies across numerous sites.[15] Who these third parties are and how they utilize the intelligence they glean from user surveillance is not exactly known, but the dogged research and publication of academics, activists, and the popular press has gone a long way to expose both technical methods and a dizzying array of commercial actors, including those in direct contact with users and those behind the scenes that now populate this space.

A recent case that captured much public attention offers a hint at how these practices are evolving. In May 2017, an Australian newspaper reported that it had obtained a leaked internal Facebook strategy document describing how advertisers could use the company's platform to target advertisements at teenagers as young as fourteen years old at moments when they feel vulnerable. "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,'" the report claims.[16] Facebook responded that the report is misleading, and the features described are simply meant to "help marketers understand how people express themselves on Facebook," not to target ads.[17] They did not deny, however, that their insights into teenagers' emotional states *could* be used to influence the vulnerable, leaving many to wonder if all that stands between us and this kind of purported manipulation is Facebook's company policies.[18]

The marriage of advertising and information technology has thus rendered urgent questions about what Cass Sunstein calls the "ethics of influence."[19] There is a growing sense that the ways of influencing consumers

---

[15] *See generally* David M. Kristol, *HTTP Cookies: Standards, Privacy, and Politics,* 1 ACM TRANSACTIONS ON INTERNET TECH. 151 (2001); John Schwartz, *Giving Web a Memory Cost Its Users Privacy,* N.Y. TIMES (Sept. 4, 2001), https://www.nytimes.com/2001/09/04/business/giving-web-a-memory-cost-its-users-privacy.html [https://perma.cc/8S8L-MPTC].

[16] Darren Davidson, *Facebook Targets 'Insecure' Young People to Sell Ads*, AUSTRALIAN (May 1, 2017), https://www.theaustralian.com.au/business/media/facebook-targets-insecure-young-people-to-sell-ads/news-story/a89949ad016eee7d7a61c3c30c909fa6/ [perma.cc link unavailable].

[17] *Comments on Research and Ad Targeting*, FACEBOOK NEWSROOM (Apr. 30, 2017), https://newsroom.fb.com/news/h/comments-on-research-and-ad-targeting/ [https://perma.cc/K3RM-NBMU].

[18] *See* Nitasha Tiku, *Get Ready for the Next Big Privacy Backlash Against Facebook*, WIRED (May 21, 2017), https://www.wired.com/2017/05/welcome-next-phase-facebook-backlash/ [https://perma.cc/REZ9-QDEA].

[19] *See* CASS SUNSTEIN, THE ETHICS OF INFLUENCE: GOVERNMENT IN THE AGE OF BEHAVIORAL SCIENCE (2016).

we once (perhaps reluctantly) tolerated have become intolerable—that what some call "persuasive" strategies are in fact manipulative. Ryan Calo claims that "digital market manipulation is a problem, if at all, because it constitutes a form of persuasion that is dangerous to consumers or society."[20] Karen Yeung points to the "persuasive, manipulative qualities" of nudges driven by big data.[21] Anthony Nadler and Lee McGuigan suggest that "persuasive communication can manipulate consumer attitudes and behaviors."[22] Tal Zarsky argues that online surveillance "might facilitate the manipulation of subjects."[23]

We applaud these efforts to bring renewed scrutiny to consumer advertising practices in light of the powerful digital socio-technical systems that now convey them. What is lacking in this emerging literature, however, is conceptual clarity about what exactly online manipulation is, how it differs from other forms of influence, and how, normatively, it ought to be addressed. These are the questions we take up below.

B. Algorithmically-Nudged Labor

On April 21, 2017, Noam Scheiber wrote for *The New York Times*:

> And yet even as Uber talks up its determination to treat drivers more humanely, it is engaged in an extraordinary behind-the-scenes experiment in behavioral science to manipulate them in the service of its corporate growth—an effort whose dimensions become evident in interviews with several dozen current and former Uber officials, drivers and social scientists, as well as a review of behavioral research.[24]

Drawing on academic research, Scheiber cites several practices supporting these assertions. The backdrop is Uber's insistence that drivers are not employees. Instead, they are independent contractors taking advantage of Uber's software platform, which connects drivers with people needing rides and for which Uber charges commission.[25] Although both Uber and its drivers

---

[20] Calo, *supra* note 1, at 1020.

[21] Yeung, *supra* note 6, at 119.

[22] Anthony Nadler & Lee McGuigan, *An Impulse to Exploit: The Behavioral Turn in Data-Driven Marketing*, 35 CRITICAL STUD. MEDIA COMM. 151, 161 (2018).

[23] Tal Z. Zarsky, *Privacy and Manipulation in the Digital Age*, 20 THEORETICAL INQUIRIES L. 157, 158 (2019).

[24] Noam Scheiber, *How Uber Uses Psychological Tricks to Push Its Drivers' Buttons*, N.Y. TIMES (Apr. 21, 2017), https://www.nytimes.com/interactive/2017/04/02/technology/uber-drivers-psychological-tricks.html [https://perma.cc/4UU8-YUV6].

[25] *Id.*

make money through rides, incentives are not neatly aligned. To compete successfully with other companies, generally, Uber must serve riders within the shortest timeframe and at the lowest cost. Purportedly lacking the coercive clout of an employer, Uber has devised a range of features to encourage driver behaviors that are profitable for the company, though, some have argued, not optimal for drivers.

Because conditions of high rider demand and low driver supply yield price surges, they are favored by drivers. Uber's interests, however, are best served by many drivers serving as many riders as possible, with prices that beat those of their competitors. Accordingly, Uber barrages drivers with texts, emails, popups, and carefully designed graphics to keep them behind the wheel and to direct them, ostensibly, to areas of highest demand. One of the techniques cited by critics is the carefully curated graphic representations of predicted needs.[26] For example, the inclusion of sporting events or bar-closing times hint that both demand and the likelihood of surge pricing will be high.[27] In other words, drivers are presented with vague promises and enticed by heat map estimates, which conflate real-time and predictive demand. They are thus nudged toward performing a service for highly uncertain rewards in a manner Ryan Calo and Alex Rosenblat liken to a "bait and switch."[28]

A second feature is to urge drivers to continue working as they reach the end of a shift and try to log out of the system. They may receive push notifications reading, "Are you sure you want to go offline? Demand is very high in your area. Make more money, don't stop now!" and accompanied by a surge icon.[29] Or the app may indicate that they are approaching some arbitrary earnings level for that shift, e.g., "you're $10 away from making $330!"[30] These are attempts to exploit a well-known decision-making vulnerability—"people's preoccupation with goals—to nudge them into driving longer"—evidenced in robust findings from behavior research.[31] Similar effects are imputed to the gamification of Uber's interactive app, which shows work status (hours, earnings, rides, etc.) in game-like formats known for their power to hold players at the game console, and, presumably, drivers at the wheel. And Uber utilizes automatic queuing, a strategy familiar to those who subscribe to streaming services such as Netflix or Amazon Video. Before a ride ends Uber cues up the next ride request, making it just a bit more

---

[26] Calo & Rosenblat, *supra* note 3, at 1662.

[27] Rosenblat & Stark, *supra* note 8, at 3769 ("We also want to remind you that we predict that New Year's Eve will be the busiest time of the year. *With such high demand, it will be a great night to go out and drive!*").

[28] Calo & Rosenblat, *supra* note 3, at 1662.

[29] Rosenblat & Stark, *supra* note 8, at 3768.

[30] Scheiber, *supra* note, 24.

[31] *Id.*

difficult to refuse than to take on another ride.[32] Although Uber responded to concerns by allowing automatic queuing to be disengaged, the feature is engaged by default and reloads after breaks.[33] Uber is not alone in employing these practices.[34]

In describing Uber's practices, critics regularly cite to behavioral science research that reveals human vulnerability to forms of irrational, biased, and bounded thinking, as well as tendencies toward compulsive and even addictive ruts.[35] These cognitive vulnerabilities range from well-studied phenomena,[36] such as loss aversion and preoccupation with goals, to what Natasha Schüll, in her study of addictive gambling machines, calls "ludic loop," a compulsion to keep playing.[37] It even extends to the practice of Uber employees taking on female personas based on experimental findings that suggest the majority male driver population is more likely to engage with messages emanating from female rather than male communicators.[38] While these critics have raised a number of ethical concerns, including coercion, excessive control through punitive measures, exploitation, and power and information asymmetries, importantly for this discussion they also claim these practices can be manipulative, raising deep ethical questions about which forms of influence are appropriate in the workplace.[39]

C. Psychographic Profiling and Election Influence

In March 2018, sudden public interest arose in whether or not the political services firm, Cambridge Analytica, had improperly used Facebook's advertising platform to exert influence in the 2016 U.S. presidential election. While many questions about the case remain unanswered, many commentators have pointed to the possibility that the purported influences were manipulative.

Piecing together a story from public media, and conceding that it includes contradictions and denials, it appears that Cambridge Analytica

---

[32] *Id.* ("It requires very little effort to binge on Netflix; in fact, it takes more effort to stop than to keep going.")

[33] *Id.*

[34] *Id.*; *see also* Tae Wan Kim & Kevin Werbach, *More Than Just a Game: Ethical Issues in Gamification*, 18 ETHICS & INFO. TECH. 157, 157–73 (2016).

[35] *See, e.g.*, Calo & Rosenblat, *supra* note 3; Scheiber *supra* note 24; Luke Stark, *Algorithmic Psychometrics and the Scalable Subject,* 48 SOC. STUD. SCI. 204 (2018).

[36] *See* RICHARD H. THALER & CASS R. SUNSTEIN, NUDGE: IMPROVING DECISIONS ABOUT HEALTH, WEALTH, AND HAPPINESS (2008).

[37] NATASHA DOW SCHÜLL, ADDICTION BY DESIGN: MACHINE GAMBLING IN LAS VEGAS (2014).

[38] Scheiber, *supra* note 24.

[39] *Id.*

employed a company called Global Science Research (GSR) to generate vast repositories of digital user profiles. These repositories were initially seeded by the results of a personality quiz, "thisisyourdigitallife" (2014), administered by Aleksandr Kogan, a lecturer in the Department of Psychology at the University of Cambridge and the head of GSR.[40] Kogan distributed the quiz through a Facebook app after being refused access to a dataset Michael Kosinski had assembled with colleagues at Cambridge and Microsoft Research for their widely cited study published in the *Proceedings of the National Academy of Science*.[41] Subjects, limited to U.S. voters with Facebook accounts, each received a few dollars for taking the quiz and providing access to their Facebook accounts. From the few hundred thousand quiz-takers, Cambridge Analytica accumulated tens of millions of Facebook user accounts through a feature (no longer active) that allowed developers to gain access to the accounts of "friends" of quiz-takers—a number possibly as high as 87 million.[42]

We can only surmise that Kogan's work with Cambridge Analytica was inspired by the Kosinski, et al. study, which demonstrated that a tremendous amount can be inferred about individuals from Facebook "likes" alone. Such inferences include gender, sexual orientation, race, religion, political views, relationship status, substance use, and size and density of friendship networks.[43] The study further claimed to uncover correlations between "like" patterns, psychological traits (such as intelligence as measured by Raven's Standard Progressive Matrices), and personality profiles (such as openness, agreeableness, emotional stability, and conscientiousness as measured by the International Personality Item Pool).

This provides context for the claims made by Christopher Wylie, one of Cambridge Analytica's co-founders-turned-whistleblower: "We exploited Facebook to harvest millions of people's profiles and built models to exploit what we knew about them and target their inner demons. That was the basis

---

[40] Carole Cadwalladr & Emma Graham-Harrison, *Revealed: 50 Million Facebook Profiles Harvested for Cambridge Analytica in Major Data Breach*, GUARDIAN (Mar. 17, 2018), https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election [https://perma.cc/X58U-SJ8Z].

[41] Michael Kosinski et al., *Private Traits and Attributes are Predictable from Digital Records of Human Behavior*, 110 PROC. NAT'L ACAD. SCI. 5802 (2013).

[42] Cecilia Kang & Sheera Frankel, *Facebook Says Cambridge Analytica Harvested Data of Up to 87 Million Users*, N.Y. TIMES (Apr. 4, 2018), https://www.nytimes.com/2018/04/04/technology/mark-zuckerberg-testify-congress.html [https://perma.cc/VFZ8-3W2Q].

[43] Kosinski et al., *supra* note 41.

the entire company was built on."[44] According to news reports, Cambridge Analytica claims to have built "psychographic" profiles on approximately 220 million U.S. voters based on 5,000 pieces of data.[45] In a 2016 speech, then-CEO Alexander Nix boasted about Cambridge Analytica's ability to personalize messages in a range of media from direct mail, to online cookie-driven ad targeting, to social media banners, and even to set-top televisions.[46] Cambridge Analytica's personalized approach was different from past advertising, Nix claimed, because "we don't need to guess at what creative solution may or may not work. We can use hundreds or thousands of individual data points on our target audiences to understand exactly which messages are going to appeal to which audiences."[47] Cambridge Analytica's strategies also differed from traditional mass advertising—even advertising that is targeted at defined demographic groups, such as age, race, socioeconomic class, etc. "[W]e've rolled out a long form quantitative instrument to probe the underlying traits that inform personality," Nix claimed, which measures:

> openness—how open you are to new experiences; conscientiousness—whether you prefer order and habits and planning in your life; extraversion—how social you are; agreeableness—whether you put other people's needs and society and community ahead of yourself; and finally neuroticism—a measurement of how much you tend to worry. By having hundreds and hundreds of thousands of Americans undertake this survey, we were able to form a model to predict the personality of every single adult in the United States of America.[48]

Compiling profiles is one thing. But even assuming one accepts the premises upon which claims about these profiles are based—that personality tests, augmented with inferences based on online measures, such as Facebook "likes," produce sound profiles—it is another thing to go from psychographic

---

[44] *Cambridge Analytica and Facebook: The Scandal So Far*, AL JAZEERA (Mar. 28, 2018), https://www.aljazeera.com/news/2018/03/cambridge-analytica-facebook-scandal-180327172353667.html [https://perma.cc/PU4C-KNSN].

[45] Carole Cadwalladr, *Google, Democracy and the Truth About Internet Search*, GUARDIAN (Dec. 4, 2016), https://www.theguardian.com/technology/2016/dec/04/google-democracy-truth-internet-search-facebook [https://perma.cc/5ZJ4-GAWV].

[46] Alexander Nix, Cambridge Analytica–The Power of Big Data and Psychographics, Presentation at the 2016 Concordia Summit (Sept. 27, 2016), *in* YOUTUBE, https://www.youtube.com/watch?v=n8Dd5aVXLCc [https://perma.cc/N9C6-HTMJ].

[47] *Id.*

[48] *Id.*

profiles to targeted messaging. This step requires tailoring messages to correspond to the specific personality traits of their recipients. Matz et al. claim to have shown experimentally that "targeting people with persuasive appeals tailored to their psychological profiles can be used to influence their behavior as measured by clicks and conversions."[49] The business literature is filled with studies that claim to "prove" the effectiveness of conventional ad targeting, and there are possibly industry studies, not revealed publicly, showing that targeting works. As to pinpointing advertisement recipients, we learned that Facebook, while it does not allow marketers to target advertisements based on psychological traits directly, "it does so indirectly by offering the possibility to target users based on their Facebook Likes."[50] In a 2018 paper, computer scientists Irfan Faizullabhoy and Aleksandra Korolova demonstrated that one can get around technical enforcement of policies disallowing overly narrow targeting and successfully target particular messages down to the individual.[51]

In any event, the popular embrace of claims about the efficacy of these methods in the broadest terms very likely stirs its continuing appeal even in the absence of hard evidence. For the purposes of this article, we do not need to demonstrate that Cambridge Analytica's efforts had a deep and meaningful impact on the U.S. presidential election. While it is unlikely that its effects were dispositive, that they worked, at some level, seems incontrovertible. More importantly, there is every reason to believe efforts like these will continue to evolve, and worries about online manipulation in political contexts will continue to grow.[52] Our goal, then, is to reveal the dimensions of these practices that make them manipulative, and in so doing, expose why they are deeply disturbing.

III. DEFINING MANIPULATION

Manipulation is a tricky term, much like the behavior it describes. Colloquially, to manipulate something is to steer or control it. One often speaks of manipulating complex technical instruments, devices that would do nothing at all without human direction. For example, cars are manipulated via steering wheels and pedals, and computers are manipulated via keyboards and

---

[49] S.C. Matz et al., *Psychological Targeting as an Effective Approach to Digital Mass Persuasion*, 114 PROC. NAT'L ACAD. SCI. U.S. 12714, 12715 (2017).

[50] *Id.*

[51] Irfan Faizullabhoy & Aleksandra Korolova, *Facebook's Advertising Platform: New Attack Vectors and the Need for Interventions*, (Workshop on Technology & Consumer Protection, arXiv:1803.10099, 2018), https://arxiv.org/abs/1803.10099 [https://perma.cc/39Z8-7P45].

[52] For a sense of the broad contours of debates about voter microtargeting, see Frederick J. Zuiderveen Borgesius et al., *Online Political Microtargeting: Promises and Threats for Democracy*, 14 UTRECHT L. REV. 82 (2018).

mouse devices. Similarly, one can manipulate living, animate things by changing the way they would behave absent any interventions. Gardeners manipulate tomato plants by fastening them to a trellis, and trainers manipulate dogs by enticing them with treats. To manipulate a *person* is likewise to steer or control them, as though they were a car or a tomato plant.[53] Manipulators are often described as "puppet masters" who pull their targets' strings.

It is worth distinguishing at the outset between manipulation understood in this way—as steering or controlling a person—and the manipulation of institutions or systems. Given the recent rise of online disinformation campaigns, especially targeted at voters, there is growing concern about what some have termed "media manipulation" and its effects on election outcomes.[54] Obviously, the ultimate goal of influencing the media is to influence the people consuming it. Likewise, influencing people can be a means of altering the institutions they participate in—for example, when they vote. At the heart of these worries, though, are concerns about individuals and the independence of their decision-making processes. It is this sense of manipulation—as influence over individuals—that occupies us here.

Manipulation, then, is a kind of influence—an attempt to change the way someone would behave absent the manipulator's interventions. The question is what kind of influence it is. In the following section, we show how manipulation differs from related concepts. We consider what distinguishes manipulation from persuasion, as well as other familiar forms of influence— coercion and deception. In general, persuasion is thought to be perfectly acceptable, while deception and coercion are not. If manipulation is like persuasion in relevant ways, then we might decide it is not worth worrying about. If it is like deception or coercion, we might worry indeed. Finally, we consider how manipulation differs from nudging, a form of influence about which there is considerably less agreement.

A. Persuasion, Coercion, and Manipulation

"Persuasion" has both broad and narrow meanings. In the broad sense, to persuade simply means to change someone's mind—it is an umbrella term

---

[53] As Allen Wood argues, "The manipulative person 'steers' the other as a driver steers an automobile. The automobile is already moving through its own internal combustion engine and momentum, but its direction is influenced by the one who steers it." Allen Wood, *Coercion, Manipulation, Exploitation, in* MANIPULATION: THEORY AND PRACTICE 17, 33–34 (Christian Coons & Michael Weber eds., 2014).

[54] *See generally* Yochai Benkler et al., *Network Propaganda: Manipulation, Disinformation, and Radicalization in American Politics* (2018); Alice Marwick & Rebecca Lewis, *Media Manipulation and Disinformation Online*, DATA & SOC'Y RES. INST. (May 15, 2017), https://datasociety.net/pubs/oh/DataAndSociety_MediaManipulationAndDisinformationOnli ne.pdf [https://perma.cc/H4TY-446Q].

covering nearly all forms of influence, from argumentation and rhetoric to pointing a gun at someone's head. In this sense, persuasion includes both making someone an offer and making someone *an offer they can't refuse.* By contrast, persuasion in the narrow sense (i.e., rational persuasion) means changing someone's mind by giving reasons he or she can reflect on and evaluate. Distinguishing between these two meanings of the term is important because they carry significantly different normative connotations. Persuasion in the narrow sense is generally thought to be perfectly morally acceptable, whereas persuasion in the broad sense denotes both behaviors we accept and behaviors we do not accept.

Many authors writing about these issues use the term "persuasion" in the broad sense. Conceptually, there is nothing wrong with that. Normatively, however, in discussions about manipulation, it muddies the waters, raising questions about how and why a good thing (persuasion) becomes a bad thing (manipulation). For this reason, throughout this Article we use "persuasion" in the narrow sense, meaning rational persuasion, and the generic term "influence"—a term without normative baggage—to describe the category of which rational persuasion and manipulation are both members.

At an abstract level, there are two ways of influencing another person's decision-making: change the options available to the other person (their decision space) or change how they understand their options (their internal decision-making process).[55] Persuasion, the paradigmatic form of respectful influence, works by operating either of these two levers. When we persuade someone to do something (or to refrain from doing it), we appeal openly to their capacity for conscious deliberation and choice. We offer arguments or incentives. We assume, in other words, that they pursue their own ends, and we try to demonstrate that doing things our way would advance them. Take, for instance, the manager of a car dealership. To persuade her sales team to work longer hours, she might argue that because shoppers stay out later during long summer days, working late during the summer would increase their odds of earning extra commissions. In this way, she appeals to her team members' own ends—earning an income—and tries to show them that behaving the way she wants is a means to achieving those ends. She attempts to influence her team by changing the way they understand their options—i.e., by appealing to their internal decision-making process.

If her argument fails, the manager can shift strategies. Instead of offering an argument, she might offer an incentive, such as overtime pay. Rather than change how her team members understand their situation, the manager changes the situation itself. She alters their decision-space by reconfiguring the options available to them. This is equally a form of

---

[55] We will complicate this dichotomy later. *See infra* Section V.B.

persuasion. In either case, the attempted influence is forthright and transparent, and the decision is ultimately left entirely up to the staff. The manager successfully influences her team's behavior only when she *convinces* them to behave the way she wants, motivating them with abstract arguments or concrete incentives.

This example highlights what Joel Rudinow would call "resistible incentives."[56] Although the incentives change the terms being deliberated over, the change is not determinative; it is easy to imagine some of the team members deciding, on balance, that the overtime pay is not worth it. In that case, the manager has yet another option: she can coerce them. To coerce someone is to offer "*irresistible* incentives."[57] Or, to put the same point slightly differently, as Allen Wood does, coercing someone means eliminating all of the "acceptable alternatives."[58] The manager could threaten her team members' jobs, or she could put a gun to their heads. She could, in other words, restrict her team members' decision-spaces, arranging them such that they only have one acceptable option—behaving the way the manager wants.[59] If we face irresistible incentives (or lack acceptable alternatives) we are *forced* to act as our coercers would have us act; we are deprived of choice.

Persuading someone leaves the choice of the matter entirely up to them, while coercing someone robs them of choice. At the same time, although coercing someone deprives them of choice, in an important sense, it leaves their capacity for conscious decision-making intact. After all, recognizing that some incentive is irresistible, or that an alternative is unacceptable, requires having our wits about us. If one did not understand that the only acceptable option available to them was to do as their coercer instructed, or if they could not act on that understanding, then they would have no motivation or no means to go along with the coercer's plan. Coercing someone forces them to act the

---

[56] Joel Rudinow, *Manipulation*, 88 ETHICS 338, 342 (1978).

[57] An irresistible incentive, Rudinow argues, is one which could only be avoided through "heroism, madness, or something similarly extraordinary." *Id*. at 341.

[58] Wood, *supra* note 53, at 21–23.

[59] What counts as irresistible or unacceptable might differ somewhat from person to person (the threat of being fired might be resistible/acceptable for someone with other job prospects, but irresistible/unacceptable for someone without them). Which is not to say that they are determined by the agent's own beliefs or feelings about them. Rather, it is to say that what counts as an acceptable or unacceptable option can be determined, in part, by context. *Id*. As Wood writes, "there is sometimes an objective fact of the matter that certain alternatives are (or are not) acceptable to a given agent *under specific circumstances*." *Id.* (emphasis added).

way the coercer wants, not by undermining or circumventing their decision-making faculties, but by making the coercer's way the only acceptable one.[60]

In one respect, persuasion and coercion are, therefore, opposites, since persuading someone leaves the choice of the matter entirely up to the target, while coercing someone deprives them of choice. But what persuasion and coercion have in common is that they attempt to influence without undermining the target's decision-making powers. In cases of persuasion and coercion, the agent is steering the ship. When coerced, a person is forced to abandon their self-chosen ends (the destination, say), but it is still the coerced person who does the abandoning. They understand what is happening; they recognize it as the only acceptable option. They direct themselves to do as their coercer demands. If asked later what happened, why they acted the way that they did, the coerced person has no trouble explaining it. "My hands were tied," they say, regretfully, "It was the only available option. I was forced to act against my will." Or perhaps not: rare though they may be, there are occasions when a person with a gun to their head defies their would-be coercer anyway. There are heroic stories of German gentiles, under threat of death, refusing to give up their Jewish neighbors to the Nazis. That such a thing is even possible demonstrates that no matter how forceful the attempted coercion, it is ultimately the coerced person who makes the decision to act.

Manipulation, by contrast, means taking hold of the controls. To manipulate people is to displace them as the decider, to "subvert," as Wood puts it, their capacity for self-government, to "undermine or disrupt the ways of choosing that they themselves would critically endorse if they considered the matter in a way that is lucid and free of error."[61] It is to deprive them of

---

[60] Christian Coons and Michael Weber write: "the instruments of coercion (threats, incarceration, and other penalties) are attempts to alter the context of choice, making it rational for you to comply. In this way, the coercer typically treats the coerced as rational. In fact, coercion depends on the target's being rational." Christian Coons & Michael Weber, *Introduction: Investigating the Core Concept and Its Moral Status*, in MANIPULATION: THEORY AND PRACTICE 1, 15 (Christian Coons & Michael Weber eds., 2014). We agree. Note in the next sections, though, that ultimately, we do not argue that manipulation subverts *rationality*; rather, we argue that it subverts conscious, self-aware decision-making.

[61] We should note that we only follow Wood part of the way. For Wood, self-government is coextensive with *rational* decision-making—to self-govern is to decide rationally. Wood, *supra* note 53, at 35 ("What is characteristic of manipulative behavior is that it influences people's choices in ways that circumvent or subvert their rational decision-making processes, and that undermine or disrupt the ways of choosing that they themselves would critically endorse if they considered the matter in a way that is lucid and free of error."). We do not tie self-government or its disruption so closely to rationality. It is possible, on our account, to be irrationally influenced without being manipulated, just as it is possible to be manipulated and still decide rationally. The salient issue, to our minds, is not whether the influence appeals to the target's rational faculties, but, rather, whether it appeals to their *conscious* decision-making process. Which is to say, the issue for us is not rationality but *awareness*.

authorship over their actions.[62] That is what it means to feel like someone else's puppet: when a person is coerced that person feels used, when a person is manipulated that person feels *played.*

Of course, manipulation is rarely so thorough as to *totally* deprive its target of self-control or self-government. That is why many people intuitively believe that manipulation involves less control than coercion (and consequently, that people should almost always be excused for doing things they were coerced to do, but only sometimes be excused for things they were manipulated into doing). Rather than entirely displacing the target as the decision-maker, the manipulator insinuates himself in his target's decision-making process. In Sarah Buss's words, he interferes with "the self-governed (and self-governing) activity we call 'making up one's own mind about how to act.'"[63] To say that you "feel manipulated" is to say that you do not fully understand why you acted the way that you did, or whether your actions served your own or someone else's ends.

Whereas persuasion and coercion work by appealing to the target's capacity for conscious decision-making, manipulation attempts to subvert that capacity. It neither convinces the target (leaving all options open) nor compels the target (eliminating all options but one). Instead, it interferes with the target's decision-making process in order to steer them toward the manipulator's ends. Importantly, this is not meant to suggest that coercion is acceptable or less bad than manipulation, but that coercion and manipulation threaten an individual's capacity to choose and pursue their own ends in different ways. Coercion is blunt and forthright: one almost always *knows* one is being coerced. Manipulation is subtle and sneaky. Rather than simply depriving a person of options as the coercer does, the manipulator infiltrates their decision-making process, disposing it to the manipulator's ends, which may or may not match their own.

---

[62] Or more precisely, it deprives them of *part* authorship. In thinking about how we influence others, it is tempting to use as one's baseline an imaginary *un*influenced person—a perfectly unencumbered decision-maker. This person is the imagined subject of much liberal social and political discourse. She is rational, deliberative, and fully autonomous, an individual in every sense of the word. She is the author of her own life story. It is important to recognize that this subject is a fiction. Although we are individuals, we are also social beings who exist inexorably with others, and the people we live and interact with affect us. Thus, as Joseph Raz argues, "an autonomous person is *part* author of his own life. His life is, *in part,* of his own making." JOSEPH RAZ, THE MORALITY OF FREEDOM 204 (1986) (emphasis added). Autonomy is not the absence of influence, but the presence of self-government. To be autonomous—to be part-author of one's own life—is to know that one's decision-making is conditioned, and yet, still, to take one's own reasons for acting as authoritative. It is this authority of one's own reasons that manipulation subverts, as we discuss in Part IV, below.

[63] Sarah Buss, *Valuing Autonomy and Respecting Persons: Manipulation, Seduction, and the Basis of Moral Constraints*, 115 ETHICS 195, 195 (2005).

B. The Means of Manipulation

The question, then, is how this is done—how are manipulators able to alienate their targets from their own decision-making powers and interfere with the way they make up their minds about how to act? Philosophers and political theorists have described a wide variety of purportedly manipulative techniques. Marcia Baron points to lies, false promises, applying pressure, and playing on people's emotions.[64] Wood adds encouraging false assumptions, fostering self-deception, and appealing to "character flaws."[65] Kate Manne discusses guilt trips.[66] Buss considers certain kinds of seduction.[67]

The most systematic account is Robert Noggle's influential theory that one manipulates another by causing their beliefs, desires, or emotions to deviate from certain ideals.[68] Beliefs, for instance, are ideally true. Desires, according to Noggle, are ideally directed toward things we have reason to want. Emotions are ideally appropriate to the situation at hand. When we make decisions according to right-functioning rational decision-making processes, our beliefs, desires, and emotions approximate these ideals.[69] Manipulating someone, Noggle argues, means corrupting rational decision-making by "adjusting" these "psychological levers"—belief, desire, and emotion—away from their ideal settings. One might deceive their target (causing them to have false beliefs), tempt them (creating a desire for what they lack reason to want), incite them (causing an inappropriate emotional response), and so on.

While there is much to like about this story, it fails to capture what is distinctive about manipulation—that it undermines our sense of authorship over our decisions. Causing someone to make non-ideal decisions is sometimes manipulative, but it is not manipulative in every case. Suppose you are a great lover of martinis but really ought not to drink. And knowing this, but not wanting to feel like lushes themselves, your friends parade the finest martinis before you, tempting you to have one. In this case your friends have

---

[64] *See* Marcia Baron, *Manipulativeness*, 77 PROC. & ADDRESSES AM. PHIL. ASS'N 37 (2003).

[65] *See* Wood, *supra* note 53, at 35.

[66] *See generally* Kate Manne, *Non-Machiavellian Manipulation and the Opacity of Motive*, *in* MANIPULATION: THEORY AND PRACTICE 221 (Christian Coons & Michael Weber eds., 2014).

[67] *See* Buss, *supra* note 63, at 195.

[68] Which ideals exactly are relevant here is up for some debate. For Noggle, manipulation involves inducing someone to deviate from what the manipulator believes are the ideals for beliefs, desires, and emotions. *See* Robert Noggle, *Manipulative Actions: A Conceptual and Moral Analysis*, 33 AM. PHIL. Q. 43, 47–48 (1996). Anne Barnhill endorses all of Noggle's account save this. In Barnhill's view, the relevant ideals are those objectively in the target's self-interest. Anne Barnhill, *What is Manipulation?*, *in* MANIPULATION: THEORY AND PRACTICE 51, 65–72 (Christian Coons & Michael Weber eds., 2014).

[69] Noggle, *supra* note 68, at 44–47.

caused you to desire what you ought not to have (or to feel particularly drawn to it). But have they manipulated you?

Surely not. If you really cannot resist the temptation, then your friends have offered you an irresistible incentive—they have coerced you. If you *can* resist it, and yet you do not, then you have simply been persuaded by bad reasons (e.g., "If a thing you love is paraded before you then you really ought to have it, regardless of the consequences"). Either way, you have not been deprived of authorship over your decision. You have chosen to drink; you have willed it. Or, at the very least, you were unable to will yourself not to drink. Indeed, this last possibility—that you suffered from weakness of will— illustrates the point most clearly. If parading martinis in front of you caused an internal struggle, one you were aware of and engaged in with intention, there can be no doubt that it was you (in all your complexity) who made the final choice.[70]

To sharpen the point even further, manipulation cannot simply mean causing someone to make less-ideal decisions than they otherwise would, since it is possible to manipulate someone into making more ideal decisions. Imagine your friends (different ones) know you love martinis, but they also know you worry about gaining weight. Having heard that you intend to spend the evening drinking—which you agree you really should not do—your friends share articles on Facebook and Twitter about the high calorie content of alcohol and the correlation between drinking and weight gain, knowing you are likely to see them. You are unaware that your friends have plotted to influence you, but as a result of seeing the articles you decide not to drink. In this case, you have been manipulated since your friends have insinuated themselves covertly into your decision-making process and redirected it to their own ends. This is despite the fact that their ends (and the decision-making process that led you to them) are—by your own admission—more ideal.

Apart from the direction of the influence (toward or away from ideals), the salient difference between these two cases is that the influence in the first case is overt, while the influence in the second case is hidden. You would feel manipulated in the second case—if you later found out why your friends had posted those articles—because you would realize that you had been motivated by someone else's reasons. You made what seemed at the time like a decision that was yours, only to find out that it was infected by external machinations. In the first case, by contrast, you knew while you were deciding that your friends were trying to influence your decision. Because you were aware of their influence, you could treat it as you would any other decision—you could

---

[70] Alan Ware makes a similar argument to this, but he frames his analysis in terms of opportunity for consent, rather than awareness of and authorship over the influence. *See* Alan Ware, *The Concept of Manipulation: Its Relation to Democracy and Power*, 11 BRITISH J. POL. SCI. 163, 169–70 (1981).

contemplate it, weigh it against other considerations, attempt to mount defenses against it, and so on. The decision you reached was therefore your own (to whatever extent our decisions are ever our own). It was no different, in this respect, from just happening to see a perfect martini at the next table over at dinner. You would be tempted by it, sure, but not manipulated.

The hiddenness of manipulative influences explains how it is possible to alienate someone from their own decision-making powers. In order to get someone to act the way you want without realizing *why* they are acting that way, they must be unaware of the influence. As soon as we become conscious of outside influence, of someone else's plans and how we are implicated in them, we incorporate that influence into our own decision-making. Once you know someone else is trying to get you to do something, that fact becomes a regular part of how you make up your mind. It becomes one of the reasons that helps you explain your actions to yourself. Since we are never totally free of outside influence, what gives us (part) authorship over our own actions is that we regard our own reasons for acting as authoritative.[71] Manipulation thwarts that.

Robert Goodin places the notion of hidden influence (or what he calls "deceptive influence") at the center of his theory of manipulation. "One person manipulates another," he writes, "when he deceptively influences him, causing the other to act contrary to his putative will."[72] Alan Ware agrees, arguing that for A to manipulate B it must be true that (among other things) "B either has no knowledge of, or does not understand, the ways in which A affects his choices."[73] Although nearly everyone else concedes that hidden influence is an effective means of manipulation, not everyone agrees that hiddenness is necessary. To demonstrate this point, a number of theorists have proposed cases of supposedly overt manipulation.[74] Noggle invokes the image of Satan tempting Christ as he fasts in the wilderness.[75] Anne Barnhill claims that blatant guilt trips are manipulative.[76] Moti Gorin describes an employee who tries openly to get his co-worker—a recovering alcoholic—to relapse in order

---

[71] *See* RAZ, *supra* note 62, at 204; Buss, *supra* note 63, at 195.

[72] ROBERT E. GOODIN, MANIPULATORY POLITICS 19 (1980). We distinguish between manipulation and deception, below, and use the term "hidden influence" rather than "deceptive influence," in order to keep the two concepts apart. Furthermore, as we argue in Part IV, in our view, it is not strictly necessary for manipulation to result in the target acting contrary to their putative will. It is possible, in other words, to be manipulated and still behave as one would have, absent the manipulation.

[73] Ware, *supra* note 70, at 165.

[74] Wood, *supra* note 53, at 38. Wood leaves the door open to the possibility that there might be cases of overt manipulation, but he does not develop the point.

[75] Noggle, *supra* note 68, at 44.

[76] Barnhill, *supra* note 68, at 60.

to beat her out for a promotion.[77] In our view, these are not cases of manipulation, since they share all the relevant features of the martini case above. If you know you are being guilted (in Barnhill's case) or tempted (in Noggle's and Gorin's cases), then you have not been deprived of authorship over your decision. Either you cannot resist the influence and have therefore been coerced, or you can resist it and do not, in which case you have simply been moved by bad reasons.

These arguments give us little reason to think manipulation can operate out in the open. It is only because the target of influence is unaware of or does not understand how they are being influenced that the manipulator's intervention can infiltrate and disrupt the target's decision-making process, steering them without force.

C. Deception and Manipulation

With this in mind, one can begin to see how manipulation is related to deception. To deceive someone is to cause them to hold false beliefs. This, of course, can be a powerful tool of manipulation. Imagine, once again, the manager of our car dealership. If she told her team they would be fired for refusing to work late but in fact lacked the authority to carry out the threat, or if she held a convincing, but fake, gun to their heads, then she would be deceiving them. She would be inducing them to act under false pretenses. Instead of changing or restricting the options open to her team members (i.e., persuading or coercing them), she would be disposing them to act the way she wanted by undermining their ability to understand their options. Here we have an example of deception in the service of manipulation. Although the team members know their manager is influencing them, they are misled about *how* she is influencing them. They are fed false beliefs about their options, which induces them to decide in ways they likely would not endorse if they had access to all the facts.

Deception is thus an important tool in the manipulator's toolkit, but it is not the only one. While instilling false beliefs is a blunt way of controlling another person's decision-making process, there are subtler means of shaping a person's beliefs. Social scientists have shown that human reasoning is "bounded" in significant ways. When evaluating information, people use a variety of unreliable shortcuts and heuristics, which researchers describe as "cognitive biases." Famously, Daniel Kahneman and Amos Tverksy demonstrated that people are often influenced by irrelevant information (so-called "anchoring effects") and give more weight to evidence they can easily

---

[77] Moti Gorin, *Towards a Theory of Interpersonal Manipulation*, in MANIPULATION: THEORY AND PRACTICE 73, 80–81 (Christian Coons & Michael Weber eds., 2014).

recall (the "availability heuristic").[78] People draw different conclusions from the same information depending on how it is presented ("framing effects"), and so on.[79] Because these cognitive biases are widespread and predictable, manipulators can easily treat them as vulnerabilities to exploit. Manipulators can remind targets of unimportant facts so that they give them undue weight. They can point out that their targets' friends believe certain things in hopes that they will believe them too. They can frame information in misleading ways. Manipulation, therefore, need not involve outright deception; the truth can also be used to control our decision-making.

In fact, manipulators need not influence beliefs at all. To use Noggle's metaphor, there are other "psychological levers" a manipulator can "adjust."[80] Some of the most common examples of manipulation involve leveraging emotions and desires. If you and your partner are having a disagreement and you know they still feel guilty about something unrelated from the week before, you can subtly mention the source of guilt in conversation to wear down their resolve. Rather than create straightforward advertisements, where products are center-stage, marketers try to subliminally connect their products to our fantasies and aspirations by paying celebrities to appear with them or having them casually turn up on television and in films.

Though people like to imagine themselves as reliable and independent decision-makers, in fact they have many vulnerabilities to exploit. Beliefs, desires, and emotions form in response to a wide variety of factors, many of which operate outside conscious awareness. If a manipulator wants to influence someone without their knowledge, they need only intervene in a way that flies under the metaphorical radar. Deception exploits one kind of vulnerability—lack of perfect information—and can therefore be understood as a species of manipulation. But exploiting vulnerabilities in the way people form and manage desires and emotions is manipulative too.

D. Nudges and Manipulation

We have so far distinguished manipulation from persuasion and coercion, demonstrating that the former is hidden and works by exploiting vulnerabilities in the way we make decisions, while the latter operate out in the open, appealing directly to our capacity for conscious decision-making. We have also posited the relationship between manipulation and deception,

---

[78] Amos Tversky & Daniel Kahneman, *Judgment under Uncertainty: Heuristics and Biases*, 185 SCIENCE 1124, 1128 (1974). For Kahneman's more recent, popular treatment, see DANIEL KAHNEMAN, THINKING, FAST AND SLOW (2011).

[79] *See* Amos Tversky & Daniel Kahneman, *Rational Choice and the Framing of Decisions*, 59 J. BUS. 251, 257 (1986).

[80] Noggle, *supra* note 68, at 44–47.

arguing that people may deceive in order to manipulate, but that manipulation does not require instilling false beliefs. It remains to distinguish manipulation from nudging.

Now well-known, the term "nudge" was coined by the behavioral economist Richard Thaler and legal scholar Cass Sunstein, and refers to "any aspect of the choice architecture that alters people's behavior in a predictable way without forbidding any options or significantly changing their economic incentives."[81] By "choice architecture," Thaler and Sunstein mean the context in which people make decisions. To take one of their familiar examples, when someone is deciding what to buy at a cafeteria, the arrangement of options—some at eye-level, others slightly out of reach—influences the choices they make as people are more likely to reach for what's closest.[82] Thus, in order to shape other people's decision-making, one can alter their choice architecture to nudge them in a preferred direction.

The notion of choice architecture is useful for understanding manipulation in general, and online manipulation in particular; we carry this notion with us through the discussion that follows. Like manipulation, intentionally shaping someone's choice architecture to nudge them in a particular direction involves influencing their decision-making without force. What's more, nudges often work by leveraging cognitive biases. Those familiar with the growing literature on nudges will recognize many of their key features in our characterization of manipulation as hidden influence and our description of exploiting vulnerabilities as the primary means of manipulation. If the way options are selected, arranged, and presented (i.e. choice architecture) affects the way people understand and respond to them, and if such effects operate outside conscious awareness, the question arises as to where nudging ends and manipulation begins.[83]

Recognizing an overlap, Sunstein argues that while some nudges might be manipulative, most are not. To support this claim he offers a definition of manipulation very much in the spirit of our discussion. An influence is manipulative, he argues, "to the extent that it does not sufficiently engage or

---

[81] THALER & SUNSTEIN, *supra* note 36, at 6.

[82] *Id.* at 1–4.

[83] *See generally, e.g.*, Robert Noggle, *Manipulation, Salience, and Nudges*, 32 BIOETHICS 164, 166–68 (2018); Thomas RV Nys & Bart Engelen, *Judging Nudging: Answering the Manipulation Objection*, 65 POL. STUD. 199, 203–08 (2016); Gérard Reach, *Patient Education, Nudge, and Manipulation: Defining the Ethical Conditions of the Person-Centered Model of Care*, 10 PATIENT PREFERENCE & ADHERENCE 459, 460–61 (2016); Luc Bovens, *The Ethics of Nudge*, *in* PREFERENCE CHANGE: APPROACHES FROM PHILOSOPHY, ECONOMICS AND PSYCHOLOGY 207 (Till Grüne-Yanoff & Sven Ove Hansson eds., 2009); Evan Selinger & Kyle Whyte, *Is There a Right Way to Nudge? The Practice and Ethics of Choice Architecture*, 5 SOC. COMPASS 923, 928–930 (2011); Daniel M. Hausman & Brynn Welch, *Debate: To Nudge or Not to Nudge*, 18 J. POL. PHIL. 123, 130–35 (2010).

appeal to [the target's] capacity for reflection and deliberation."[84] He disagrees, however, that manipulative influences must be hidden. Unlike the challenges to the hiddenness condition that we have already considered, Sunstein argues that while manipulation is usually hidden, we ought not to consider hiddenness its defining criterion because the notion of "hiddenness" is too difficult to spell out.[85] What about an influence needs to be exposed, he asks, in order for it not to be hidden? When the government orders that graphic health warnings be presented on cigarette packs, must it also—in order not to be manipulative—disclose information about the particular psychological mechanisms that make such warnings effective?[86]

Although the question Sunstein raises is a good one, the answer is not as difficult as he suggests. Sunstein is surely right that when people make everyday decisions, much of what conditions their outcomes is not wholly apparent to them. People are not entirely self-transparent. They do not—and perhaps cannot—reflect on every belief, desire, emotion, and habit that impacts the decisions they reach. Moreover, they cannot know everything about how the world of choices they confront—whether in the natural world or the human-built one—came to be precisely the way it is, i.e., how the decision-making contexts they face were formed. Yet the reason this does not, generally, thwart a person's capacity to act competently is that they are able to act on the basis of reliable assumptions. Upon encountering graphic health warnings, for example, the average person has no trouble intuiting why they are there or how—roughly—they work: the health bureaucracy is encouraging people not to smoke, both by providing information and by appealing to a visceral sense of fear and disgust. Hidden influences thwart such assumptions. If people learned after some time that the *real* reason graphic health warnings were placed on cigarette packages is that the alcohol lobby paid off government officials, in an attempt to drive people away from smoking and toward drinking, the influence would be hidden in the relevant sense and the public would rightly feel manipulated.

Interestingly, Sunstein's definition raises an even more difficult question. Namely, what is "sufficient" engagement of a target's capacity for reflection and deliberation? The notion of sufficiency here is a normative one—what constitutes sufficient engagement is a question about values. When attempting to influence someone, how much *ought* the influencer to engage their target's capacity for reflection and deliberation? This normative component of Sunstein's definition creates internal tensions in his account. He argues that "from the standpoint of welfare, there might . . . be a justification

---

[84] SUNSTEIN, *supra* note 19, at 82.
[85] *Id.* at 102–05.
[86] *Id.*

for hidden manipulation in . . . extreme circumstances—as, for example, when people are trying to stop a kidnapping or to save a kidnapping victim."[87] We agree that manipulation might be justified in such circumstances, but it is not clear why that situation is manipulative under Sunstein's theory. If the hidden influence is justified, then it would seem that the level of engagement with the target's (i.e., the kidnapper's) capacity for reflection and deliberation was decidedly sufficient. Yet it is important to recognize that even in situations where manipulation might, ultimately, be justified, it is still manipulation; and, as a result, it carries with it a real harm, which must be weighed against potential benefits.[88]

Thus, while we agree with Sunstein's conclusion that some nudges are manipulative and others are not, we believe that our account of manipulation as hidden influence better delineates between them. Sunstein suggests, for example, that disclosures (purely informational nudges) are not manipulative since they represent "an effort to appeal to [people's] deliberative capacities" rather than bypass them.[89] But there are different types of disclosures. Nutrition labels are straightforward—they simply provide consumers with some of the information they need to make an informed choice about what foods to buy. They are not manipulative because they are not designed to exploit a cognitive bias, and they instead encourage and strengthen the consumer's capacity for conscious deliberation. And there is nothing hidden about them—people understand why they are there and what effects they are meant to have. Another case that Thaler and Sunstein cite approvingly is less clear. Deploying a so-called "social nudge," the state of Minnesota informed its citizens that 90% of Minnesotans pay their taxes in order to encourage the 10% who do not pay to comply.[90] Depending on how this information was conveyed, this strategy is plausibly manipulative because, although the information it conveyed was true, its purpose may have been hidden.

In our view, nudges are manipulative if they are hidden and exploit vulnerabilities. Many nudges—indeed, most of the nudges that Thaler and Sunstein recommend—are transparent and work to *correct* cognitive vulnerabilities rather than exploit them. As such, they are not manipulative at all. Another distinction between nudges and manipulation, which we discuss at length in the sections that follow, stems from the fact that manipulation is usually *targeted*.[91] In order to exploit someone's vulnerabilities, one must know something about what those vulnerabilities are and how precisely to

---

[87] *Id.* at 104.

[88] We discuss the precise nature of that harm in Part IV, below.

[89] SUNSTEIN, *supra* note 19, at 82.

[90] THALER & SUNSTEIN, *supra* note 36, at 66–65.

[91] *See* Patrick Todd, *Manipulation*, *in* 5 INT'L ENCYCLOPEDIA ETHICS 3139, 3140 (Hugh LaFollette ed., 2013).

leverage them. Most nudges, by contrast, are not targeted to particular individuals. Like the cafeteria example, above, the sorts of nudges Thaler and Sunstein advocate for are meant to be applied in the same way to everyone.

As we discuss below, part of what makes information technology particularly well-suited to facilitating manipulation is that it allows for fine-grained microtargeting, making it possible for potential manipulators to engage in what Karen Yeung calls "hypernudging."[92] Hypernudges are not only hidden, they precisely target and exploit individual vulnerabilities, making them much more difficult to resist. Although we can imagine mass manipulation, which attempts to influence large groups of people all in exactly the same way, the more targeted manipulation is the more we ought to worry about it.

E. Manipulation and Manipulative Practices Defined

At this point, we can pull these different strands together and define manipulation as *imposing a hidden or covert influence on another person's decision-making.* That means influencing someone's beliefs, desires, emotions, habits, or behaviors without their conscious awareness, or in ways that would thwart their capacity to *become* consciously aware of it by undermining usually reliable assumptions. This definition captures what is essential about manipulation—namely, that it disrupts the target's capacity for self-authorship. Which is to say, it explains why, upon learning they have been manipulated, people feel like puppets. It differentiates manipulation from persuasion and coercion, which are forthright efforts to alter a person's decision-making process. It reveals that nearly all instances of deception are also instances of manipulation, but not all manipulation is deceptive. And it helps distinguish between nudges that are manipulative and those that are not. Furthermore, we have considered *how* manipulators can impose hidden influences on their targets: by targeting and exploiting their cognitive, emotional, or other decision-making vulnerabilities.

An issue we have not considered in depth is whether or not a hidden influence must be *intentionally* imposed to be manipulation. We believe that it does. The notion of manipulation naturally evokes the image of a manipulator, and influences hidden by natural or accidental means would likely not undermine a person's sense of self-authorship. It is known and expected that some of the underlying causes of experience lay outside conscious awareness. What is not expected is intentional meddling in other people's decision-making. People assume—and ought to be able to assume—that they do not live in the company of Descartes' evil demon or on the set of

---

[92] Yeung, *supra* note 6.

the *Truman Show*. Thus, when we discuss manipulation as "hidden or covert influence," we mean such cases where hiddenness is intended.

Even with this definition in hand, however, identifying cases of manipulation is difficult because manipulation is a "success concept."[93] Which is to say, the claim that someone has been manipulated refers not only to the strategies employed by the influencer but also to the effects of those strategies on the influenced. Analogously, consider the distinction between lying and deceiving. To lie is to assert a falsehood. To deceive, by contrast, is to instill false beliefs. In order to determine whether or not someone has lied, one only needs to investigate the actions of the alleged liar. In order to determine whether or not someone has deceived, one needs to investigate the actions of the deceiver *as well as the resulting beliefs of the deceiver's target.* The same is true of manipulation. Manipulation only exists where the attempt to manipulate succeeds and the manipulee's decision-making is affected by the influence. Investigating cases of manipulation, therefore, requires a focus on the status of two parties—the intentions of the suspected manipulator and the beliefs and actions of potential manipulees. In many cases, establishing impacts on targets of manipulation may require far-flung empirical findings that are difficult, if not impossible, to access.

For this reason, we focus in what follows on the concept of *manipulative practices*—strategies that a reasonable person should expect to result in manipulation—and not on the success concept of manipulation, in toto. The preceding discussion points to three key characteristics of manipulative practices: they involve influences that (1) are hidden, (2) exploit cognitive, emotional, or other decision-making vulnerabilities, and (3) are targeted. Strictly speaking, the only necessary condition of manipulation is that the influence is hidden; targeting and exploiting vulnerabilities are the means through which a hidden influence is imposed. Indeed, targeting is best understood as an exacerbating condition: the more closely targeted a strategy is to the specific vulnerabilities of a particular manipulee, the more effective one can expect that strategy to be. Hidden influences not targeted at all still count, therefore, as manipulative, but we ought to worry more about manipulative practices the more targeted they are. As we will see in the next section, this is especially important when considering manipulation online.

Drawing on this analysis, let us briefly return to the cases we discussed earlier in this article.[94] Did Facebook, Uber, or Cambridge Analytica engage in manipulative practices? The Facebook case is the clearest. Its strategy memo described identifying moments when teenagers felt emotionally vulnerable and deploying advertisements to leverage that vulnerability. Such

---

[93] *See* Wood, *supra* note 53, at 11.
[94] *See supra* Part II.

practices are targeted to the individual level, exploit vulnerabilities by design, and are hidden—i.e., there is reason to believe few people targeted would know why they were seeing that particular advertisement at that particular moment. On our account, this is manipulative beyond any doubt. The only question, in this case, is whether Facebook engaged in the practice; the company denies that it did.

The Cambridge Analytica case has much in common with the Facebook case. By its own account, Cambridge Analytica attempted to "exploit what we knew about [voters]" and "target their inner demons" [95]—i.e., Cambridge Analytica aimed to frame political messages in ways each targeted individual was most inclined to accept and internalize. This practice is targeted and exploits individual vulnerabilities by design, and as in the case of Facebook above, there is little reason to believe that individuals seeing Cambridge Analytica's advertisements would know either that the ads were tailored to them specifically or how the tailoring was accomplished. These practices are manipulative under our framework. Unlike with Facebook, we know that Cambridge Analytica actually engaged in these practices; it remains a mystery however, the extent to which they were effective. Even if Cambridge Analytica's manipulative practices only had minor impacts on recent elections, we ought to treat this case as a cautionary tale. As we discuss in the next Part, these practices are likely to increase in power and sophistication.

Finally, Uber: this case is complicated, not least because there are so many potentially manipulative practices in which the company is known to have engaged. Some strategies seem clearly manipulative, such as intentionally misleading drivers with surge pricing "heat maps" that conflate real-time and predictive demand. But other strategies may not necessarily be manipulative. For example, automatically queuing the next ride request before the driver has time to decide whether or not to continue working might *feel* manipulative because it exploits the fact that we need time to make decisions. However, there is nothing hidden about this strategy. Drivers likely understand perfectly well what is happening, and though they may feel pressured, they likely do not feel tricked. Thus, on our account, automatic queueing is not an obviously manipulative practice, though there may be other reasons for questioning its ethical standing.

One last Uber practice is worth considering—namely, the notifications designed to nudge drivers to stay on the road. Much like automatic queuing, drivers may experience this pressure as manipulative, but as it is neither hidden nor significantly targeted, we ultimately judge it not to be so. As a nudge, it is designed to exploit drivers' desire to accomplish goals, even if the goals are

---

[95] *Cambridge Analytica and Facebook: The Scandal So Far*, *supra* note 44.

functionally meaningless, but it is not a manipulative practice. Tweak some aspects of this scenario, however, and one can imagine future versions of this practice that are indeed manipulative. Consider if notifications were timed to appear right when drivers were desperate—say, if the earnings goals were not arbitrary, but instead were indexed to bills coming due. If the apparatus of such influence were hidden from the driver, manipulation would be a worry.

That possibility is important as we move into the next section, where we discuss more systematically how information technology can facilitate manipulation. Sometimes, as we have seen, technologically-mediated practices are manipulative; sometimes they are not. Our goal in the next section is to make clear how information technology and manipulation intersect.

IV.   MANIPULATION THROUGH INFORMATION TECHNOLOGY

Having defined manipulation and developed the concept of manipulative practices, we are now in a position to address this article's next question: what is it about information technology that, together with manipulation, makes such a worrying combination? Building on our definition of manipulation, generally, we define online manipulation as *the use of information technology to covertly influence another person's decision-making.* And, accordingly, we define online manipulative practices as *applications of information technology that impose hidden influences on users, by targeting and exploiting decision-making vulnerabilities.*

We argue, below, that as digital technologies are incorporated into all aspects of people's everyday lives, they become increasingly susceptible to this kind of manipulation. Widespread digital surveillance means it takes little effort to identify people's vulnerabilities. Digital platforms are the perfect medium through which to leverage those insights. And because information technology mediates so much of so many people's lives, there is virtually limitless opportunity to invisibly influence.

A. Surveillance

Living in an "Information Age" means that nearly everything people do is tracked.[96] Both the information individuals knowingly disseminate about themselves (e.g., when they visit websites, make online purchases, and post

---

[96] *See, e.g.*, SHOSHANA ZUBOFF, SURVEILLANCE CAPITALISM: THE FIGHT FOR A HUMAN FUTURE AT THE NEW FRONTIER OF POWER (2019); Julia Angwin, WHAT THEY KNOW (2019), http://juliaangwin.com/the-what-they-know-series/ [https://perma.cc/88QA-4LTU]. There is, of course, an enormous academic literature on surveillance. For a helpful introduction, see DAVID LYON, SURVEILLANCE STUDIES: AN OVERVIEW (2007).

photographs and videos on social media) and the information they unwittingly provide (e.g., when those websites record data about how long they spend browsing them, where they are when they access them, and which advertisements they click on) reveals a great deal about who each individual is, what interests them, and what they find amusing, tempting, and off-putting.[97] Moreover, one need not "go online" in the traditional sense to be digitally tracked. Credit card purchases log what people buy in brick-and-mortar stores,[98] law enforcement license plate readers track where they drive,[99] and facial recognition software identifies them as they move through public spaces.[100] The workplace has also become a site of intense digital surveillance. Private firms carefully monitor and analyze how their employees conduct themselves throughout the workday and beyond, leading to what some have called a new "digital Taylorism."[101]

Tal Zarsky has warned that information technology could potentially be used to manipulate people by harnessing this wealth of information to precisely tailor advertisements that exploit their vulnerabilities.[102] Ryan Calo has deepened Zarsky's account by describing these practices in terms of what Jon Hanson and Douglas Kysar call "market manipulation"—exploiting cognitive biases in order to influence consumer behavior.[103] Calo updates

---

[97] As Julie Cohen argues, this information is a source of tremendous value and is treated by private firms as a repository of resources to be mined. Understood through the lens of political economy, Cohen suggests that we think of this information as a "biopolitical public domain": "Personal information processing has become the newest form of bioprospecting, as entities of all sizes compete to discover new patterns and extract their marketplace value." Julie E. Cohen, *The Biopolitical Public Domain: The Legal Construction of the Surveillance Economy*, 31 PHIL. & TECH. 213 (2017).

[98] Michael Reilly, *Google Now Tracks Your Credit Card Purchases and Connects Them to Its Online Profile of You*, MIT TECH. REV. (May 25, 2017), https://www.technologyreview.com/s/607938/google-now-tracks-your-credit-card-purchases-and-connects-them-to-its-online-profile-of-you/ [https://perma.cc/K33S-UBDB].

[99] American Civil Liberties Union, YOU ARE BEING TRACKED: HOW LICENSE PLATE READERS ARE BEING USED TO RECORD AMERICANS' MOVEMENTS (2013), https://www.aclu.org/sites/default/files/field_document/071613-aclu-alprreport-opt-v05.pdf [https://perma.cc/2QKK-4UWZ].

[100] Jennifer Lynch, FACE OFF: LAW ENFORCEMENT USE OF FACE RECOGNITION TECHNOLOGY (2018), https://www.eff.org/files/2018/02/15/face-off-report-1b.pdf [https://perma.cc/MS78-QTR8].

[101] *See, e.g.*, Christian Parenti, *Big Brother's Corporate Cousin: High-Tech Workplace Surveillance is the Hallmark of a New Digital Taylorism*, NATION (July 27, 2001), https://www.thenation.com/article/big-brothers-corporate-cousin/ [https://perma.cc/7X7B-E8X8]; Richard Salame, *The New Taylorism*, JACOBIN (Feb. 20, 2018), https://www.jacobinmag.com/2018/02/amazon-wristband-surveillance-scientific-management [https://perma.cc/FR49-GT5V].

[102] *See* Zarsky, *supra* note 1, at 219–20.

[103] *See* Calo, *supra* note 1, at 1000.

Hanson and Kysar's theory to reflect the increased capacities for market manipulation that information technology furnishes (what Calo calls "digital market manipulation").[104] For example, he points to the use of "disclosure ratcheting"—using behavioral nudges to dispose individuals to reveal more information about themselves than they intend (information which is then used to optimize further nudges)—and "means-based targeting"—using online experiments like A/B testing[105] to tailor the presentation of each advertisement.[106]

On the whole, information that was once difficult and expensive to uncover is now available easily and cheaply. People reveal their vulnerabilities constantly, and as the Facebook case suggests, many parties are eager to accrue information in order to leverage it for manipulative ends. Importantly, ubiquitous digital surveillance means that people reveal information about their vulnerabilities not only through what some might judge frivolous or unnecessary disclosures—e.g., on social media. Rather, as the Uber case suggests, people are rendered susceptible to this kind of tracking merely in the routine, day-to-day activities of everyday life: going to work, commuting, and communicating, or even as the Cambridge Analytica case highlights, by simply being associated with others who make disclosures about themselves.

B. Digital Platforms

In addition to providing insight into vulnerabilities that advertisers, employers, and political campaigns may want to exploit, information technology also makes it increasingly easy to leverage those insights. As Karen Yeung argues, digital platforms facilitate "Big Data-driven decision-guidance techniques," which constitute a kind of "hypernudging."[107] Unlike traditional advertisements, which were static and disseminated en masse, digitally-mediated platforms, such as websites and social media applications, constitute dynamic, interactive, intrusive, and personalized choice

---

[104] *Id.*

[105] In website and app design, A/B testing is a method for determining the relative effectiveness of messaging strategies or user interface elements by presenting different variations to different groups of users and measuring user response.

[106] Calo, *supra* note 1, at 1012–14; *see also* Alessandro Acquisti et al., *Privacy and Human Behavior in the Age of Information*, 347 SCI. 509 (2015).

[107] Yeung, *supra* note 6, at 121; *see also* Lanzing, *supra* note 7.

architectures.[108] Websites can alter the presentation of information depending on the specific things they know about each individual visitor. And if the websites do not know much, they can learn as users interact with them. If an advertiser knows that someone is more likely to buy things at a particular time of the day or a particular day of the week—as the Facebook case demonstrated—the advertiser does not have to wait, passively, for that person to browse past their campaigns; the advertiser can send emails or text notifications, or present themselves in the target's social media feeds at optimal moments. And because this tailoring process is automated, it is fully personalized for each individual target. As Calo puts it, "firms will increasingly be in the position to *create* suckers, rather than waiting for one to be born."[109]

Crucially, as Yeung points out, the insights into people's vulnerabilities that digital platforms utilize in shaping their choice architectures are not limited to what can be gleaned from information collected about each individual. Platforms also enable "monitoring and refinement of the individual's choice environment in light of *population-wide* trends identified via population-wide Big Data surveillance analysis."[110] In other words, the digital systems people interact with study both their individual idiosyncrasies *and* the patterns that emerge amongst demographic groups to which they belong, potentially revealing weaknesses and dispositions that individuals themselves cannot see. As Frank Pasquale argues, "the real basis of commercial success in Big Data-driven industries is likely the quantity of

---

[108] By digital platforms, we mean the built online environments, such as websites and smartphone applications, that facilitate online interaction, and through which individuals communicate, access information, and engage in a wide range of other activities, from reading and writing, conducting research, and playing games to finding work, shopping, and arranging services. For a critical analysis of the concept of platforms, see Tarleton Gillespie, *The Politics of "Platforms*,*"* 12 NEW MEDIA & SOC'Y. 347 (2010).

[109] Calo, *supra* note 1, at 1018. These worries are connected to growing discussions around so-called "dark patterns"—user interface design strategies that drive users to behave in ways beneficial to online services and potentially harmful to users themselves. A number of scholars have suggested these design strategies can be manipulative. *See, e.g.*, Arunesh Mathur et al., *Dark Patterns at Scale: Findings from a Crawl of 11K Shopping Websites*, *in* 22ND ACM CONFERENCE ON COMPUTER-SUPPORTED COOPERATIVE WORK AND SOCIAL ENGINEERING 81 (2019); Jamie Luguri & Lior Strahilevitz, *Shining a Light on Dark Patterns* (U. Chicago, Poverty Law Working Paper No. 719, Aug. 7, 2019). While a thorough analysis of dark patterns is outside the scope of this paper, the characteristics of manipulative practices highlighted above (i.e., that they are hidden, exploit decision-making vulnerabilities, and are targeted) might be applied productively to debates in this area, helping to distinguish between manipulative and non-manipulative patterns. We discuss this briefly in related work, see Susser et al., *Technology, Autonomy, and Manipulation*, 8 INTERNET POL'Y REV. 7 (2019).

[110] Yeung, *supra* note 6, at 122 (emphasis added).

relevant data collected *in the aggregate*—something not necessarily revealed or shared via person-by-person disclosure."[111]

C. Mediation

Finally, there is a tendency to treat information technology as the kind of thing people approach, direct their attention toward, and engage with intention. But it is more appropriate to understand information technology as a set of tools that increasingly mediates everyday experience. Digital platforms are more like eyeglasses than magnifying glasses—technologies one wears and forgets about rather than those one picks up and puts to use. As Don Ihde argues, when people use such technologies, the technologies recede from view.[112] They attend not to the eyeglasses themselves, but to what they can see *through* them.[113] It is only when something goes wrong—when one *cannot* see through the glasses because they get dirty or break—that we notice the technology itself. Likewise, people pay very little attention to smartphones and computers themselves, focusing instead on the information they can access through them, the pictures they can see and videos they can watch, the things they can buy, and directions they can follow.[114] Even the very tech-savvy, who understand what is going on behind-the-scenes of the technologies they use— for example, the way data is collected and stored, analyzed and transmitted— likely spend little time considering all that takes place under the hood. For the average person, information technology sits almost entirely outside conscious awareness; few ponder its nature and inner-functioning while they are actively engaged in using it.[115]

---

[111] PASQUALE, *supra* note 1, at 153; *see also*, Solon Barocas & Helen Nissenbaum, *On Notice: The Trouble with Notice and Consent*, *in* PROCEEDINGS OF THE ENGAGING DATA FORUM: THE FIRST INTERNATIONAL FORUM ON THE APPLICATION AND MANAGEMENT OF PERSONAL ELECTRONIC INFORMATION (2009) ("Aggregated user data can be subject to computerized analyses to produce predictive models that can be used to estimate other like users' propensity to respond to certain ads.").

[112] *See generally* Don Ihde, Technology and the Lifeworld: From Garden to Earth (1990).

[113] *Id.*

[114] Robert Rosenberger & Peter-Paul Verbeek, *A Field Guide to Postphenomenology*, *in* POSTPHENOMENOLOGICAL INVESTIGATIONS: ESSAYS ON HUMAN-TECHNOLOGY RELATIONS 9, 37–38 (Robert Rosenberger & Peter-Paul Verbeek eds., 2015).

[115] For more on what postphenomenologists call "transparency" (not to be confused with the notion of transparency as it is used in policy contexts) and its relation to using information technology, see Daniel Susser, *Transparent Media and the Development of Digital Habits*, *in* POSTPHENOMENOLOGY AND MEDIA: ESSAYS ON HUMAN-TECHNOLOGY-WORLD RELATIONS 27 (Yoni Van Den Eede et al. eds., 2017); *see also*, Diane Michelfelder, *Postphenomenology with an Eye to the Future*, *in* POSTPHENOMENOLOGICAL INVESTIGATIONS: ESSAYS ON HUMAN-TECHNOLOGY RELATIONS 237, 237–46 (Robert Rosenberger & Peter-Paul Verbeek eds., 2015).

This is important to recognize in the context of worries about online manipulation because (as we argued in the previous section) lack of awareness is the defining feature of manipulative strategies. Beyond the fact that information technology provides insight into people's vulnerabilities and furnishes platforms that can leverage those insights, it is designed to be *seen through* and is thus already, in a real sense, hidden. A world increasingly structured by information technology is a world increasingly removed from view—a world of screens people look through, cameras they walk past, and sensors they unknowingly impress upon. A determined manipulator could not dream up a better infrastructure through which to carry out his plans.

Furthermore, because information technology mediates so much of so many people's lives, the reach of online manipulation is virtually limitless. Just as information technology increases the scale of other activities, it increases the potential scale of manipulative influence. Facebook, for instance, currently has more than 2 billion monthly active users.[116] If its platform were leveraged to manipulate users—as it allegedly was in the Cambridge Analytica case—the impact could be massive.[117] Uber has 75 million monthly active drivers.[118] The strategies it devises to influence its drivers can affect what happens in millions of cars around the world. Unlike "offline manipulation," which is constrained by the manipulator's ability to understand and influence a finite number of other people, online manipulation is practically unbounded.

On the whole, information technology has made people's vulnerabilities both easy to detect and easy to exploit. We have become more manipulable, and—there is reason to believe—more manipulated.

V.   HARMS OF MANIPULATION: THE SIGNIFICANCE OF AUTONOMY

One might agree with everything we have claimed thus far but not be concerned. What harm does manipulation cause and why should anyone care that the insurgence of digital technologies into everyday life exacerbates it? A plausible answer might point to the motivations of manipulators and suggest that the reason for resorting to manipulation is generally to induce targets to

---

[116] *Facebook Company Information*, FACEBOOK, (accessed Nov. 13, 2019), https://newsroom.fb.com/company-info/ [https://perma.cc/4Z3B-Q4HH].

[117] In 2012, Facebook conducted its infamous "emotional contagion" experiment to gauge whether or not altering the content it showed users affected their emotional states, which involved more than 689,000 (unwitting) user-subjects. *See* Adam D.I. Kramer et al., *Experimental Evidence of Massive-Scale Emotional Contagion Through Social Networks*, 111 PROC. NAT'L ACAD. SCI. 8788, 8788 (2014).

[118] Johana Bhuiyan, *Uber Powered Four Billion Rides in 2017. It Wants To Do More—And Cheaper—in 2018*, RECODE (Jan. 5, 2018), https://www.recode.net/2018/1/5/16854714/uber-four-billion-rides-coo-barney-harford-2018-cut-costs-customer-service [https://perma.cc/X9JJ-XYR8].

act against their interests. If, for example, Facebook targeted advertisements at vulnerable teenagers, the suspicion would be that they did so to exploit moments of weakness, to incline teenagers to buy something they did not need or to pay more for it than they otherwise would. As Calo argues, it is not these forms of online influence alone that are cause for concern; rather, it is the fact that they are "*coupled with divergent interests* that should raise a red flag."[119] According to such accounts, the incorporation of digital technologies into all aspects of people's lives, and the attendant threat of manipulation, is worrying because the designers and operators of these technologies are often incentivized to subordinate others' interests to theirs.

The subordination of interests, in certain cases, may be sufficient reason to condemn manipulative practices. But it fails to reach the heart of the matter. First, as we saw in the martini examples above, it is easy to imagine manipulative practices that advance, rather than diminish, a target's interests. More importantly, beyond the direct, material harms that result from manipulation, such as exploitation, impoverishment, unfairness, and the deprivation of benefits, the deeper harm is infringement of individual autonomy. Since autonomy lies at the normative core of liberal democracies, the harm to autonomy rendered by manipulative practices extends beyond personal lives and relationships, reaching public institutions at a fundamental level.

In the following Section, we first clarify the conception of autonomy that motivates our analysis. Second, we look closer at the key characteristics of manipulation to expose how manipulative practices undermine autonomy. Our main focus is the hidden changes to choice architectures that exploit individual vulnerabilities. Finally, we consider possible objections to our account.

A. Autonomy at a Glance

Personal autonomy is the capacity to make one's own choices, with respect to both existential and everyday decisions. As Joseph Raz, writes, "[t]he ruling idea behind the ideal of personal autonomy is that people should make their own lives."[120] This notion of autonomy is rooted in the view that people can (mostly) rationally deliberate on the different options they are faced

---

[119] Calo, *supra* note 1, at 1023.

[120] RAZ, *supra* note 62, at 369; *see also, e.g.*, Marilyn Friedman, *Autonomy, Social Disruption, and Women*, *in* RELATIONAL AUTONOMY: FEMINIST PERSPECTIVES ON AUTONOMY, AGENCY, AND THE SOCIAL SELF 35, 37 (Catriona Mackenzie & Natalie Stoljar eds., 2000) ("Autonomy involves choosing and living according to standards or values that are, in some plausible sense, one's own.").

with, that they know (mostly and roughly) what they believe and desire, and that they can act on the reasons they think best.[121]

Autonomy theorists often distinguish between competency conditions and authenticity conditions.[122] Autonomous persons have the cognitive, psychological, social, and emotional *competencies* to deliberate, to form intentions, and to act on the basis of that process. And autonomous persons can (at least in principle) critically reflect on their values, desires, and goals, and act for their *own* reasons—i.e., endorse them *authentically* as their own. Importantly, this conception of autonomy is not overly rationalistic— autonomous persons deliberate on the basis of their beliefs and desires, as well as on the basis of their emotions, convictions, experiences, and commitments. While broad, this understanding of autonomy and its value is indispensable to a normative theory of manipulation: without recourse to some notion of autonomy there is no basis for holding people responsible for their actions, no way to ascribe authorship (or part-authorship) of behavior, to oneself or to others.[123]

But let us steer clear of two possible misunderstandings: first, autonomy is sometimes still understood entirely in terms of solipsistic "rational choosers," without taking social contexts (and social others) into account.[124] We agree with criticisms of this conception of autonomy and develop our own conception from theories of relational autonomy. Such accounts understand autonomy as a capacity of socially-situated persons, whose decisions issue from deliberative processes rich with emotion, feeling, imagination, and reason, and which are conditioned in part by their contexts.[125]

Second, in our view, autonomy is not a capacity of ideal choosers, but rather of ordinary people making decisions in their everyday lives. When we say that to act autonomously people have to be able to act *for their own reasons*, we do not mean that everyone must always and necessarily critically reflect on their every step. We mean that when questioned (by others or by themselves) people are generally able to identify some reasons for their

---

[121] *See* JOHN CHRISTMAN, THE POLITICS OF PERSONS: INDIVIDUAL AUTONOMY AND SOCIO-HISTORICAL SELVES 149–56 (2009); *see generally* AUTONOMY, OPPRESSION, AND GENDER (Andrea Veltman & Mark Piper eds., 2014); PERSONAL AUTONOMY AND SOCIAL OPPRESSION: PHILOSOPHICAL PERSPECTIVES (Marina A.L. Oshana ed., 2015).

[122] *See* CHRISTMAN, *supra* note 121, at 155.

[123] *See* RAZ, *supra* note 62.

[124] *See* JULIE E. COHEN, CONFIGURING THE NETWORKED SELF: LAW, CODE, AND THE PLAY OF EVERYDAY PRACTICE 107–26 (2012).

[125] *See* Catriona Mackenzie & Natalie Stoljar, *Introduction: Autonomy Refigured, in* RELATIONAL AUTONOMY: FEMINIST PERSPECTIVES ON AUTONOMY, AGENCY, AND THE SOCIAL SELF 3, 4 (Catriona Mackenzie & Natalie Stoljar eds., 2000).

actions, reasons they themselves can endorse.[126] In other words, we are describing a hypothetically valid condition: if someone were to reflect on an action, that person could give a rough account of the reasons that motivated it.

On this account, autonomy is exercised by persons with particular identities and who always already live in social, cultural, political, and other contexts. Idealized conceptions of autonomy forget the "intersubjective and social dimensions of selfhood and identity for conceptions of individual autonomy and moral and political agency."[127] Social contexts and relations both *enable* and *constrain* autonomous choices.[128] On one hand, autonomy requires living with others who teach one to act with reflection and deliberation, and it requires social contexts that enable autonomous choices by providing horizons of options. On the other hand, social contexts also *condition* choices because (often stereotypical) societal expectations influence the ways people choose.

Finally, autonomy is not only an individual good; it is also a social and political good. After all, people are not merely consumers, and markets are not the only institutions meant to respond to the preferences their decisions express; people also act as citizens, and democratic institutions are designed (ideally) to reflect autonomous decisions reached in the political sphere. When manipulators aim to elicit votes rather than purchases, it is the integrity of this realm—the political realm—that is called into question. As we saw in the Cambridge Analytica case, political advertisers have attempted to use psychographic profiling to create campaigns that exploit the decision-making vulnerabilities of individual voters. Such practices threaten the autonomy of citizens, and in doing so, they threaten democracy. Interfering with people's autonomy is thus both an ethical concern and a political one.

---

[126] There is an interesting debate about which sorts of reasons should be acceptable here—any reasons, even seemingly delusional or irrational ones? Such questions are outside the scope of this Article. *See, e.g.*, THOMAS NAGEL, THE VIEW FROM NOWHERE 164–88 (1986) (discussing difference between subjective and objective reasons); SUSAN WOLF, MEANING IN LIFE AND WHY IT MATTERS 37 (2010) (discussing delusionary and irrational activities-cum-reasons).

[127] Mackenzie & Stoljar, *supra* note 125, at 4. But one has to be careful here: although persons are always socially situated and have to realize their autonomy in social contexts, this does not imply that autonomy is only possible in (social and political) egalitarian relations supporting self-respect, self-esteem, and self-trust—and thereby autonomy. *See* Joel Anderson & Axel Honneth, *Autonomy, Vulnerability, Recognition, and Justice*, *in* AUTONOMY AND THE CHALLENGES TO LIBERALISM: NEW ESSAYS 127 (John Christman & Joel Anderson eds., 2005).

[128] *See* Friedman, *supra* note 120, at 35.

B. Autonomy and Manipulation

Throughout this article, we have argued that manipulative practices are characterized by the hiddenness of their influence on decision-making. By fine-tuning the contexts in which individuals choose—their choice architectures—a manipulator can exploit each person's decision-making vulnerabilities, and in doing so interfere with their beliefs, desires, emotions, and behavior. But why is it that these interferences undermine autonomy? How do they differ from, say, overt guilt trips or transparent forms of pressure—influences we deem non-manipulative?

1. *Autonomy, Manipulation and Choice Architecture*

As we have seen, manipulators intentionally alter the contexts in which their targets make decisions. But, unlike those who persuade or coerce, manipulators do so without the target's conscious awareness. The hiddenness of manipulation challenges both conditions of autonomy—competency and authenticity. Because manipulees are unaware that features of their choice environments have been intentionally designed to influence them, their capacity to (competently) deliberate is undermined, yielding decisions they cannot endorse (authentically) as their own.

In many cases of behavioral advertising that is obviously the point. Without realizing that their choice architectures were constructed to influence them, subtle cues are used to tempt or seduce people to buy things. So-called "native advertising"—advertisements designed to mimic the appearance of non-advertising content—provides an especially clear example of a manipulative online practice, since it intentionally conflates information and advertising. While reading the news or browsing social media, people contextualize and understand the information they absorb within that frame. It requires special vigilance to identify and re-contextualize native advertisements, in order to competently evaluate them.

The vast stores of information and detailed behavioral profiles compiled about each individual greatly facilitate manipulative practices. The more that is known about each person's personality, preferences, habits, and vulnerabilities, the easier it is to construct choice environments that will guide their decision-making in the desired direction. In addition, as we argued in the previous section, the less users know about the inner-workings of the technologies that facilitate online manipulation, the easier it is for these mechanisms to function, making online manipulation especially worrying and its potential threat to autonomy so grave.

The relationship between manipulation, autonomy, and the construction of choice architectures raises another issue worth mentioning. We

pointed out in Part II that, at an abstract level, there are two ways to influence a person's decision-making: changing the options available to them and changing how they understand their options. Some argue that the term "manipulation" has two distinct senses, each corresponding to one of these strategies.[129] Manipulating a *person* involves interfering directly with their psychological processes, whereas manipulating a *situation* involves interfering with the options available to them. In our view, the preceding discussion reveals the limits of this distinction. Manipulators often change the choices available to someone—their situation—precisely to interfere with their decision-making processes. The construction of online choice contexts makes this clear. While it can be helpful in certain contexts to draw attention to the differences between situations and persons, explaining manipulation requires a holistic understanding. In autonomous actions, the person and the situation are always interconnected; one cannot be analyzed without the other.

Crucial to the self-understanding of autonomous actors is the sense that the choices they make are *their* choices, that they *know* what they are doing, and that they can, in principle, endorse the terms of their decisions. Hidden adjustments to people's choice environments fundamentally interfere with their autonomy, threatening their ability to act for reasons of their own.[130] As we have seen, this works so well online because online environments are especially well-suited to identifying the adjustments that most effectively shape people's choices.

---

[129] *See* Claudia Mills, *Politics and Manipulation*, 21 SOC. THEORY & PRAC. 97 (1995).

[130] This point can also be expressed through a traditional Kantian vocabulary. When manipulated, people are treated purely as means, and not, at the same time, as ends in themselves. In this way, manipulation can be understood as reification: people are being made into things; or, as Kant would put it, they are treated as not having dignity, but a price. Examining the problem through this lens is useful as well, as it reveals another worrying effect of manipulation: when people are treated solely as means, and not as ends in themselves, it likely has an impact on their self-understanding. *See* Jeremy Waldron, *It's All for Your Own Good*, N.Y. REV. BOOKS (Oct. 9, 2014), http://www.nybooks.com/articles/2014/10/09/cass-sunstein-its-all-your-own-good/ ("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?") [perma.cc link unavailable]. Here, Waldron criticizes nudges exercised for paternalist aims, but it is easy to see how much more damaging nudges are when people are manipulated not for the sake of paternalist aims but for the sake of sheer acquisitiveness.

2. *Manipulation and Vulnerabilities*

The concept of vulnerability has recently received a great deal of philosophical attention.[131] Scholars have focused on different aspects of vulnerability, with some exploring the distinction between ontological and contingent vulnerabilities, others questioning whether vulnerability is a special source of moral obligations, and so on.[132] Following Catriona Mackenzie and others' example,[133] we distinguish ontological vulnerabilities—vulnerabilities all human beings share in virtue of their embodied condition—from situated, socially constructed, or contingent vulnerabilities, which will be our primary focus.

Contingent vulnerabilities can result from structural conditions or individual differences.[134] Structural vulnerabilities derive from the fact that individuals are also members of groups, which enjoy varying degrees of privilege and experience varying levels of discrimination. For example, one might be vulnerable because of economic disadvantage, or on account of one's gender or sexual identity. Individual vulnerabilities, on the other hand, are those that exist irrespective of group memberships and may be the result of a person's particular history, circumstances, personality, habits, or practices. For instance, one person might be more vulnerable to online manipulation than another person who is similarly socially situated if the latter is more technologically savvy. Moreover, contingent vulnerabilities can overlap and exacerbate each other. Structural conditions can make a person more susceptible to having their individual weaknesses targeted and exploited. For example, many communities of color in America are more closely surveilled than white communities.[135] As a result, there is more information available

---

[131] *See generally* VULNERABILITY, AUTONOMY, AND APPLIED ETHICS (Christine Straehle ed., 2016); VULNERABILITY: NEW ESSAYS IN ETHICS AND FEMINIST PHILOSOPHY (Catriona Mackenzie et al. eds., 2013).

[132] *See* VULNERABILITY, AUTONOMY, AND APPLIED ETHICS, *supra* note 131, at 1–10. Some authors see an incompatibility between autonomy and vulnerability; we do not share this view but cannot go into detail here. The form of vulnerability that we analyze in the following does not follow precisely the taxonomies that these authors suggest. *See, e.g.*, *id.*; VULNERABILITY: NEW ESSAYS IN ETHICS AND FEMINIST PHILOSOPHY, *supra* note 131.

[133] Catriona Mackenzie et al., *Introduction*, *in* VULNERABILITY: NEW ESSAYS IN ETHICS AND FEMINIST PHILOSOPHY 1 (Catriona Mackenzie et al. eds., 2013).

[134] *Id.*

[135] Alvaro M. Bedoya, *The Color of Surveillance: What an Infamous Abuse of Power Teaches Us About the Modern Spy Era*, SLATE (Jan. 18, 2016), https://slate.com/technology/2016/01/what-the-fbis-surveillance-of-martin-luther-king-says-about-modern-spying.html [https://perma.cc/U27Z-H5YR].

about the individual vulnerabilities of people of color living in those communities, potentially rendering them more vulnerable to manipulation.[136]

Thus, vulnerabilities are not monolithic—they represent a variety of pressure-points, which can be exploited in a number of ways. Manipulative practices can reinforce and be reinforced by structural vulnerabilities, but they take advantage of the individual vulnerabilities of particular persons. Consequently, and as we have seen throughout, the more information a would-be manipulator has about a person's specific vulnerabilities, the more capably they can exploit them.

Distinguishing between the various kinds of vulnerabilities helps emphasize a significant difference between online manipulation and the old-fashioned manipulative practices characteristic of some pre-digital advertising. Rather than aiming only to exploit vulnerabilities almost all of us share, as television advertisements and static billboards often attempt to do, online manipulation targets individuals, exploiting vulnerabilities specific to them.

C. Objections and Responses

There are at least two potential objections to our account of the autonomy harms manipulation threatens. First, one might object that it is difficult, if not impossible, to determine whether a person's choices are made autonomously. As such, how could one determine in any particular case whether or not someone else has been manipulated?

In response, we emphasize our focus on manipulative *practices*, rather than manipulation per se. While it is indeed difficult to determine if someone has been manipulated in a particular instance (requiring information about the target's beliefs, the intentions of the alleged manipulator, and so on), much can be surmised by investigating the influence strategy itself. Rather than attempting to determine whether the target of influence was moved or whether the influence was successfully hidden, we should attempt to determine whether the influencer was trying to conceal their efforts, whether the influence was intended to exploit the manipulee's vulnerabilities, and to what extent the influence was targeted. Manipulative practices—characterized, as we have argued, by concealment, exploitation of vulnerabilities, and targeting—are cause for concern, regardless of whether they succeed in every instance. Indeed, adopting this perspective allows one to reframe the issue at

---

[136] The exploitation of structural vulnerabilities thus has discriminatory consequences. This problem has been discussed in the literature several times, but not with a focus on manipulation. *See, e.g.*, PASQUALE, *supra* note 1, at 38; *see generally* CATHY O'NEIL, WEAPONS OF MATH DESTRUCTION: HOW BIG DATA INCREASES INEQUALITY AND THREATENS DEMOCRACY 70 (2016).

a macro level: as we have argued throughout, information technology increasingly pervades and mediates nearly all aspects of our lives. If online manipulation becomes standard practice and thus an integral and indistinguishable part of our lives, it will become ever more difficult to identify specific cases of manipulation, and thus to think and talk about autonomous individuals who are responsible for their actions at all.

A second objection asks whether some alleged manipulees might not have behaved in the same way even if they hadn't been the subjects of manipulation. In other words, it points to a counterfactual component of claims to manipulation—that absent the manipulative intervention, the target would have decided differently.[137] Although some argue that this condition must be fulfilled for an influence to count as manipulative, i.e., the influence must change the way the target would have acted absent the influence, on our account manipulation interferes with the decision-making *process*, irrespective of its outcome, and one can therefore be manipulated even without such a change. While this may seem counterintuitive, it points to an issue of deep significance. The distinguishing feature of autonomy is that people want *to make choices for themselves*, and they would feel estranged from their choices if they knew they were being secretly driven toward a certain result. Autonomous choices and the decision procedures that yield those choices express who the decider is. As T.M. Scanlon argued:

> 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.[138]

Scanlon shows that it is not just the result that matters when someone acts autonomously, but rather the choice-making, the processes of making up one's own mind. Autonomous choosers want to be in a position, at least in principle, to stand behind their choices with reasons of their own.

Finally, a third objection concerns a problem mentioned earlier: since autonomy is always socially situated, choices are always framed and conditioned by features of their social contexts. Indeed, social, cultural,

---

[137] *See generally* GOODIN, *supra* note 72.

[138] Thomas Scanlon, *The Significance of Choice*, *in* 7 THE TANNER LECTURES ON HUMAN VALUES 149, 180 (1986).

economic, and political contexts determine the sets of options from which one chooses in the first place. Thus, it can be difficult to draw a line between contextual influences and the influences of manipulators.

Imagine deciding to pursue a university education. In addition to determining whether or not someone is even given the option to pursue a university education, their social, cultural, political, and economic contexts may frame the decision in a particular way. For instance, girls may be discouraged from pursuing a technical education or dissuaded from training to become doctors because they are told that women are "better suited" to other careers. While this kind of treatment is sometimes called manipulative, it is not the kind of manipulation we have in mind.[139] And there are decisive differences between this kind of influence and online manipulation. Most important to our analysis, online manipulation is aimed precisely at *individual* choosers, and it is the specific information about each target that enables online manipulators to exploit that target's vulnerabilities. As discussed in the previous section, information about groups can reveal individual vulnerabilities, and individual vulnerabilities can be exacerbated by structural conditions, which discriminate against groups. However, for the purposes of this analysis, these disparate sources of vulnerability can and must be distinguished from one another.

Furthermore, while it is sometimes difficult to define the specific beneficiaries of structural conditions (such as gender-stereotyping), the parties benefiting from manipulative online practices are easy to find. Critically analyzing structural conditions thus requires different terms and different theoretical tools than the forms of manipulation at issue here.[140]

On the whole, we have seen that using autonomy as the normative lens through which to understand the harms of online manipulation helps foreground its individual, social, and political effects. The consequences are manifold. On one hand, manipulative practices undermine individual autonomy—people's capacity for self-government, their ability to pursue their own goals. This is troubling in itself. But perhaps more worrying are the threats to *collective* self-government. When citizens are targets of online manipulation and voter decisions rather than purchase decisions are swayed by hidden influence, democracy itself is called into question. Add to this the fact that the tools of online manipulation are concentrated in only a few hands,

---

[139] *See, e.g.*, Ann E. Cudd, *Adaptations to Oppression: Preference, Autonomy and Resistance*, *in* PERSONAL AUTONOMY AND SOCIAL OPPRESSION: PHILOSOPHICAL PERSPECTIVES 142 (Marina A.L. Oshana ed., 1st ed. 2015); *see generally* SERENE KHADER, ADAPTIVE PREFERENCES AND WOMEN'S EMPOWERMENT (2011).

[140] For literature on analyzing discrimination in liberal democracies, see, for instance, SALLY HASLANGER, RESISTING REALITY: SOCIAL CONSTRUCTION AND SOCIAL CRITIQUE (2012), and ANN CUDD, ANALYZING OPPRESSION (2006).

and it is easy to see how the nexus of influence and information technology stands to make already problematic power dynamics far worse.

VI.   CONCLUSION

Our aim in this Article has been to put forward a systematic account of the nature of online manipulation, how it differs from other forms of influence, and what harms it threatens, to both individuals and society at large. The question that remains unanswered is, of course, how these threats can be addressed. Our Article offers a starting point and guide for answering that question.

First, online manipulation poses a threat across a wide variety of social contexts—from commercial contexts, to the workplace, to the political realm. Importantly, normative commitments regarding the types and degree of influence vary with each sphere. In the United States, for example, consumer decisions are generally considered less worthy of protection from outside influence than political decisions. With the rise of algorithmically mediated labor, protection for decision-making in the workplace will be the subject of much future debate. Therefore, we must pay close attention to where—in which spheres of life—influence is exerted, and what the effects of that influence mean for the individuals inhabiting them.

Second, because manipulation is, by our definition, *hidden*, combating it requires extra vigilance. The effects will often only become apparent after the harm has already been done. Further, the threat of online manipulation presents additional challenges to the predominant model of data regulation in the United States, which places the full burden of managing information flows and data practices on individuals.[141] This model assumes that people are aware of the ways data about them is flowing and the risks and benefits associated with the data practices that implicate them. The emergence and proliferation of hidden, manipulative online practices pushes beyond the outermost limits of this approach. Individuals, unaware of the ways data is collected, aggregated, and used to influence them, simply cannot be left alone to fend off these incursions into their everyday decision-making. Bringing meaningful regulation to the digital sphere is no easy task. However, given that the tools

---

[141] *See, e.g.*, Daniel Solove, *Privacy Self-Management and the Consent Dilemma*, 126 HARV. L. REV. 1880 (2013). In Europe, data protection law works differently and follows slightly different principles, certainly since the implementation of the GDPR in May 2018. *See* Natali Helberger et al., *The Perfect Match? A Closer Look at the Relationship Between EU Consumer Law and Data Protection Law*, 54 COMMON MKT. L. REV. 1427, 1431 (2017) ("Data protection law operates on the basis of a number of central principles: lawfulness, fairness and transparency; purpose limitation; data minimization; accuracy; storage limitation; integrity and confidentiality; and accountability.") (internal citations omitted). They also point out the potentially fruitful connection between the GDPR and European consumer law.

of online manipulation are largely controlled by a few powerful actors, we should begin our efforts by holding them to account.

Finally, many of the manipulative practices causing concern today evolved from advertising practices that have long been tolerated. If we are right about the nature of harm wrought by manipulative practices, we should no longer concede this ground. Although some forms of consumer targeting might be acceptable, we have suggested that even within the advertising sphere, there is a line that ought not to be crossed. Moreover, advertising may have received disproportionate attention from the research community because it is the visible tip of the iceberg. Understanding manipulative advertising provides insight into what might lie beneath. The use of information technology to facilitate manipulative practices greatly enhances their ability to shape our decision-making, raising anew questions about their ethical and political legitimacy.