Manipulation and the First Amendment

Helen Norton

MANIPULATION AND THE FIRST AMENDMENT

Helen Norton[*]

INTRODUCTION

Nobody wants to be manipulated. Yet speakers have long sought to manipulate their listeners[1]—in other words, to covertly influence their listeners' decision-making to the speakers' advantage without those listeners' conscious awareness.[2] As one of many examples, think of subliminal advertising, where sellers embed a visual message within an advertisement for a time too brief for the viewer's conscious mind to comprehend.[3]

---

   1   *See, e.g.*, Caroline Mala Corbin, *Compelled Disclosures*, 65 ALA. L. REV. 1277, 1280 (2014) (describing appeals to emotion that rise to the level of manipulation when they "intentionally take[ ] advantage of common cognitive heuristics—cognitive shortcuts that are both helpful and distorting—that we all fall prey to"); Jon D. Hanson & Douglas A. Kysar, *Taking Behavioralism Seriously: Some Evidence of Market Manipulation*, 112 HARV. L. REV. 1420, 1428–29 (1999) (describing various longstanding techniques for manipulative advertising); *see also* Tamara R. Piety, *"Merchants of Discontent": An Exploration of the Psychology of Advertising, Addiction, and the Implications for Commercial Speech*, 25 SEATTLE U. L. REV. 377, 407–22, 432–33, 435, 437–41 (2001) (exploring manipulative technology).

   2   *See* Daniel Susser, Beate Roessler & Helen Nissenbaum, *Online Manipulation: Hidden Influences in a Digital World*, 4 GEO. L. TECH. REV. 1, 26 (2019) (defining manipulation).

   3   *See* Laura R. Salpeter & Jennifer I. Swirsky, *Historical and Legal Implications of Subliminal Messaging in the Multimedia: Unconscious Subjects*, 36 NOVA L. REV. 497, 504 (2012).

Empowered by the ability to collect and aggregate information about users and then to tailor messages designed to shape those users' responses, today's digital technologies can facilitate manipulation unprecedented in its reach and success. "[T]he more information a would-be manipulator has about a person's specific vulnerabilities, the more capably they can exploit them," ethicists Daniel Susser, Beate Roessler and Helen Nissenbaum observe.[4] "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."[5] Contemporary technologies thus enable manipulation different in both degree and in kind from more traditional forms of manipulation.

As one of many illustrations of twenty-first-century manipulation, consider sellers' new ability to monitor changes in the speed and accuracy of your keyboarding to determine when you may be tired or even intoxicated (and thus potentially impaired in your decision-making) and then to craft specific advertisements targeted to exploit that vulnerability.[6] Examples abound in the political context as well: recall, for instance, Russian operatives' use of data collection and algorithms to target African Americans with personalized messages intended to induce them not to vote in the 2016 U.S. elections.[7]

This Article examines new conceptual tools for understanding manipulation and its harms. More specifically, Part I draws from ethicists' insights to explain how manipulation can inflict harms distinct from those imposed by coercion and deception, and to explain why addressing these distinct harms is a government interest sufficiently strong to justify appropriately tailored interventions.[8]

Part II explores how these conceptual tools also help us understand when, how, and why government can regulate manipulation consistent with the First Amendment. As a threshold matter, note that manipulative online interfaces and related design choices may be better understood as conduct, rather than speech protected by the First Amendment.[9] When we recall that the First Amendment fails to cover,

---

[4] Susser, Roessler & Nissenbaum, *supra* note 2, at 41.

[5] *Id.*

[6] *See* Lauren E. Willis, *Deception By Design*, 34 HARV. J. L. & TECH. 115, 143 (2020).

[7] *See* Spencer Overton, *State Power to Regulate Social Media Companies to Prevent Voter Suppression*, 53 U.C. DAVIS L. REV. 1793, 1795–96 (2020).

[8] While I acknowledge that all forms of manipulation can be harmful, I note that the dangers of online manipulation differ in degree (and perhaps in kind) in ways that especially warrant governmental intervention.

[9] *E.g.*, Ryan Calo, *Digital Market Manipulation*, 82 GEO. WASH. L. REV. 995, 1036–37 (2014) (questioning whether the collection and use of data in this context is speech covered by the First Amendment); Julie E. Cohen, *Tailoring Election Regulation: The Platform is the Frame*, 4 GEO. L. TECH. REV. 641, 642 n.1 (2020) ("I disagree that the First Amendment does or should apply to information processing activities regardless of their nature and context, but that is a subject for a different occasion."); Kyle Langvardt, *Regulating Habit-Forming Technology*, 88 FORDHAM L. REV. 129, 133 (2019) ("[C]ourts have hardly begun to address

much less protect, every use of language,[10] one can plausibly understand the First Amendment's coverage to exclude data collection and the use of algorithms (that is, instructions to machines) because they *do* things rather than *say* things.[11] This important possibility deserves attention and consideration. This Article, however, assumes arguendo that courts may characterize the sorts of manipulative practices discussed here as speech covered by the First Amendment, and then explores the constitutional implications of that assumption.

First Amendment law sometimes permits the government to protect comparatively vulnerable listeners from comparatively powerful speakers' false or misleading speech, nondisclosures, or coercion. Think, for example, of the government's requirements that commercial actors provide accurate disclosures about their products—like laws requiring warnings about the dangers of tobacco on cigarette packages and advertisements, and nutritional labels on food packaging.[12]

In other words, differences in power and information sometimes matter to First Amendment law, permitting the government's interventions to protect comparatively vulnerable listeners. The same can and should be true of efforts to regulate manipulative speech. This Article proposes understanding the First Amendment to permit the government to intervene to protect listeners from speakers' manipulative efforts in certain settings.[13]

In commercial settings, more specifically, the Court should refine and extend commercial speech doctrine to add "manipulative" commercial speech to the commercial speech it currently treats as entirely unprotected by the First Amendment because it frustrates listeners' interests.[14] This move tracks the original theoretical justifications of the commercial speech doctrine as steeped in protecting listeners' First Amendment interests. When one recalls that false and misleading commercial speech, as well as commercial speech related to illegal activity, loses its First Amendment protection precisely because it frustrates listeners' First Amendment interests,[15]

---

the First Amendment status of software's technical and nonexpressive components."); Tim Wu, *Machine Speech*, 161 U. PA. L. REV. 1495, 1518–25 (2013) (urging that we be slow to characterize these products as speech for First Amendment purposes).

[10] *See* Frederick Schauer, *The Boundaries of the First Amendment: A Preliminary Exploration of Constitutional Salience*, 117 HARV. L. REV. 1765, 1773–74 (2004).

[11] *See* KENT GREENAWALT, FIGHTING WORDS: INDIVIDUALS, COMMUNITIES, AND LIBERTIES OF SPEECH 6 (Princeton Univ. Press., 1995) (describing "situation-altering" speech as speech that falls outside of the First Amendment's protection because it *does* something rather than just *says* something).

[12] 15 U.S.C. § 1333 (2018).

[13] *See* Susser, Roessler & Nissenbaum, *supra* note 2, at 20 ("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.").

[14] *See* Langvardt, *supra* note 9, at 183.

[15] *See* Schauer, *supra* note 10, at 1776.

one sees that the same can be true of manipulative commercial speech: it frustrates listeners' interests by seeking to covertly influence those listeners' choices without their conscious awareness and by targeting and exploiting their vulnerabilities.

Filling this doctrinal lacuna would also help fill enforcement lacunae within current law. Even though existing consumer protection statutes frequently prohibit "unfair" as well as "deceptive" trade practices, to date enforcement efforts have focused almost entirely on allegedly deceptive practices—largely because of the conceptual difficulty in defining and describing illegally "unfair" practices.[16] Here too ethicists provide conceptual tools that explain why manipulation can be regarded as "unfair" to listeners even when it is hard to characterize as deceptive in traditional terms.[17]

This then requires a workable principle for identifying online commercial speech that is manipulative (and thus unprotected by the First Amendment). To this end, this Article considers two possibilities: a) focusing on evidence of interfaces' manipulative success in changing consumers' choices; and b) targeting interfaces that display key manipulative features that increase the risk of manipulation.

Finally, this Article briefly examines how online manipulation in the political setting poses harms of its own that may also justify appropriately tailored regulatory intervention (even while recognizing that the First Amendment barriers to such regulation are significantly greater in this context than in the commercial setting) and closes by highlighting some possible interventions that deserve further consideration.

I. UNDERSTANDING MANIPULATION AND ITS HARMS

Thinkers have long struggled to define the concept of manipulation with precision.[18] This Part draws from thoughtful recent work co-authored by ethicists Daniel Susser, Beate Roessler, and Helen Nissenbaum, who describe manipulation as "imposing a hidden or covert influence on another person's decision-making."[19] More specifically, they define manipulation as covertly influencing listeners' decision-making by targeting and exploiting their decision-making vulnerabilities: "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."[20] The key

---

[16]   *See* Piety, *supra* note 1, at 442.

[17]   *See generally id.*

[18]   *See* Shaun B. Spencer, *The Problem of Online Manipulation*, 2020 U. ILL. L. REV. 959, 984–88 (2020) (canvassing various efforts to define manipulation); Susser, Roessler & Nissenbaum, *supra* note 2, at 12 ("Manipulation is a tricky term, much like the behavior it describes.").

[19]   *See* Susser, Roessler & Nissenbaum, *supra* note 2, at 26 (emphasis omitted).

[20]   *Id.* (emphasis omitted); *see also id.* ("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.").

features of manipulation—as they define it, and as discussed here—are thus a speaker's *hidden* efforts to shape listeners' decision-making that *target* and *exploit* those listeners' vulnerabilities in ways that the targets are not consciously aware of, and in ways that those targets could not easily become aware of if they were to try.[21]

To further understand the concept of manipulation, consider how it differs from other efforts (some more ethical than others) to influence targets' decision-making.

First, manipulation differs from both persuasion as well as coercion in that the latter two efforts are apparent while manipulation is not. Susser, Roessler, and Nissenbaum define persuasion as "the forthright appeal to another person's decision-making power."[22] And since coercion "is blunt and forthright: one almost always *knows* one is being coerced."[23] In other words, both persuasion and coercion are transparent efforts to influence the target—although to be sure, persuasion (as the term is used here) is generally respectful of the target's autonomy while coercion is not. But neither persuasion nor coercion is sneaky.

Manipulation, in contrast, *is* sneaky: "[r]ather 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."[24] So, for purposes of this Article, the terms "manipulation," "persuasion," and "coercion" describe mutually exclusive concepts.[25]

Turn next to the relationship between manipulation and "nudges," that is, "interventions that steer people in particular directions but that also allow them to go

---

[21] *Id.* at 2–4.

[22] *Id.* at 3.

[23] *Id.* at 17.

[24] *Id.*; *see also id.* at 15 ("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.").

[25] Of course, speakers can and do use twenty-first-century expressive technologies not only to manipulate but also to discriminate, or to coerce, hector and deluge unwilling listeners, among other harms. *See, e.g.*, Pauline T. Kim, *Manipulating Opportunity*, 106 VA. L. REV. 867, 870–71 (2020) ("[O]nline intermediaries have the ability to precisely target an audience, selecting some users to receive information and others to be excluded in ways that are not at all transparent. . . . [W]hen predictive algorithms are used to allocate access to opportunities, there is a significant risk that they will do so in a way that reproduces existing patterns of inequality and disadvantage."); Tony Romm, *Robocalls Are Overwhelming Hospitals and Patients, Threatening a New Kind of Health Crisis*, WASH. POST (June 17, 2019), https://www.wash ingtonpost.com/technology/2019/06/17/robocalls-are-overwhelming-hospitals-patients -threatening-new-kind-health-crisis/ [https://perma.cc/7A4L-DQ2U] (describing how spam robocalls that flood hospitals' phone lines and squeeze out communications to patients "amount to a literal life-or-death challenge, one that increasingly is threatening doctors and patients in a setting where every second can count. . . . [Hospital administrators] fear that robocallers could eventually outmatch their best efforts to keep hospital phone lines free during emergencies, creating the conditions for a potential health crisis.").

their own way."[26] Nudges may be or may not be manipulative, depending on whether the nudger hides the intentions underlying, and the effects of, the nudge. As Cass Sunstein explains, manipulative nudges are those that attempt "to influence people in a way that does not sufficiently engage or appeal to their capacities for reflective and deliberative choice."[27] Subliminal advertising is the classic example of the non-transparent, and thus manipulative, nudge.[28]

In contrast, nudges are *not* manipulative when they are transparent and when they "have the goal of increasing navigability—of making it easier for people to get to their preferred destination. Such nudges stem from an understanding that life can be simple or hard to navigate, and a goal of helpful choice architecture is to promote simpler navigation."[29] Illustrations of non-manipulative nudges include nutrition labels with calorie information, or utilities' notices to us about how our home energy use compares with that of our neighbors.[30]

Now consider the relationship between manipulation and deception. As Susser, Roessler, and Nissenbaum explain, deception is a subset of the broader concept of manipulation: deception is a particular type of covert effort to influence listeners' decision-making through false or misleading representations about objectively verifiable facts.[31] In the commercial context, for example, deception includes false representations about the quality or hazards of goods and services, or about the actual terms and conditions of a transaction. But manipulation is not limited to deception: manipulation also includes a variety of hidden efforts to influence listeners' decision-making that don't involve factual misrepresentations but instead exploit listeners' emotional, cognitive, or other vulnerabilities. This Article sets deception aside and focuses instead on these other forms of manipulation.[32]

---

[26] Cass R. Sunstein, *The Ethics of Nudging*, 32 YALE J. REGUL. 413, 417 (2015). Sunstein thus defines nudging and coercion to be mutually exclusive, in that the former "avoids coercion or material incentives and thus fully maintains freedom of choice." *Id.* at 433.

[27] *Id.* at 443 (emphasis omitted); *see also* Susser, Roessler & Nissenbaum, *supra* note 2, at 25–26 ("Another distinction between nudges and manipulation . . . stems from the fact that manipulation is usually targeted. In order to exploit someone's vulnerabilities, one must know something about what those vulnerabilities are and how precisely to leverage them. Most nudges, by contrast, are not targeted to particular individuals.") (emphasis omitted).

[28] Sunstein, *supra* note 26, at 446–47.

[29] *Id.* at 426 (emphasis omitted).

[30] *Id.* at 425.

[31] *See* Susser, Roessler & Nissenbaum, *supra* note 2, at 21–22.

[32] Elsewhere I have explored the harms of lies and other forms of deception, along with the First Amendment implications of the government's efforts to address those harms. *See, e.g.*, Helen Norton, *Lies and the Constitution*, 2012 SUP. CT. REV. 161; Helen Norton, *Truth and Lies in the Workplace: Employer Speech and the First Amendment*, 101 MINN. L. REV. 31 (2016) [hereinafter Norton, *Truth and Lies in the Workplace*]; Helen Norton, *Lies to Manipulate, Misappropriate, and Acquire Governmental Power*, *in* LAW AND LIES: DECEPTION AND TRUTH-TELLING IN THE AMERICAN LEGAL SYSTEM (Austin Sarat ed., Cambridge Univ. Press 2015); Helen Norton, *(At Least) Thirteen Ways of Looking at Election Lies*, 71 OKLA. L. REV. 117 (2018).

Twentieth-century exemplars of this sort of manipulation include subliminal advertising,[33] and sellers' infusion of sweet scents throughout their stores that cause consumers to linger longer and more happily.[34] Law professors Jon Hanson and Douglas Kysar have examined a wide range of additional twentieth-century examples of "the possibility of market manipulation—that is, the possibility that market outcomes can be influenced, if not determined, by the ability of one actor to control the format of information, the framing and presentation of choices, and, more generally, the setting within which market transactions occur."[35]

The many twenty-first-century examples of manipulation now include sellers' ability to target online advertisements to consumers when surveillance of social media posts shows those consumers to be sad or lonely and thus especially vulnerable to buying certain goods and services they would not normally buy—or to paying higher prices than they would normally be willing to pay.[36] And using webcams and smartphone cameras "to analyze consumers' facial expressions as they looked at a sales website and instantaneously deliver offers personalized to those consumers' nonverbal responses to the websites."[37] And controlling the content of individuals' newsfeeds to steer their emotions to anger or fear—emotional states associated with barriers to careful decision-making.[38]

Twenty-first-century technologies—including the use of predictive algorithms informed by the collection and analysis of huge amounts of data—thus create opportunities for manipulation different in both degree and in kind from more traditional forms of manipulation. "By running tens of thousands of consumers through interfaces that were identical in every respect but one, firms can determine exactly which interface, which text, which juxtapositions, and which graphics maximize revenues. What was once an art is now a science," legal scholars Jamie Luguri and Lior Strahilevitz explain.[39] "As a result, consumers' ability to defend themselves has degraded."[40] Law professor Julie Cohen finds the same to be true in

---

[33] *See* Sunstein, *supra* note 26, at 444.

[34] *See* Shmuel I. Becher & Yuval Feldman, *Manipulating, Fast and Slow: The Law of Non-Verbal Market Manipulation*, 38 CARDOZO L. REV. 459, 475 (2016).

[35] Jon D. Hanson & Douglas A. Kysar, *Taking Behavioralism Seriously: The Problem of Market Manipulation*, 74 N.Y.U. L. REV. 630, 630 (1999); *see also id.* at 748 (detailing how manufacturers rely on behavioral research to shape consumers' behavior).

[36] *See* Langvardt, *supra* note 9, at 149; Willis, *supra* note 6, at 122–23.

[37] Willis, *supra* note 6, at 126.

[38] *Id.* at 144–47; *see also* Spencer, *supra* note 18, at 979 (describing technological developments that allow tech companies to detect and analyze users' moods and then fashion and deliver targeted responses).

[39] Jamie Luguri & Lior Jacob Strahilevitz, *Shining a Light on Dark Patterns*, 13 J. LEG. ANALYSIS 43, 103 (2021).

[40] *Id.; see also* Langvardt, *supra* note 9, at 184–85 ("[T]hese firms' actual 'product' consists in the ability to raise the odds that a targeted consumer will perform a desired action following a behavioral cue. At some point, through pervasive surveillance and conditioning and

political settings: "Manipulation in platform-based information environments is neither occasional nor accidental; it is endemic and results from capabilities that platforms systematically design, continually reoptimize, and deliberately offer up to third parties for exploitation."[41]

These technological changes inspired law professor Ryan Calo to extend Hanson's and Kysar's work to "digital market manipulation" that "stands to generate dramatic asymmetries of information and control between firms and consumers."[42] In other words, contemporary online manipulation not only exploits vulnerabilities but can even create them. "[D]igital market manipulation combines, for the first time, a certain kind of personalization with the intense systemization made possible by mediated consumption," Calo observes.[43] "A firm with the resources and inclination will be in a position to surface and exploit how consumers tend to deviate from rational decisionmaking on a previously unimaginable scale. Thus, firms will increasingly be in the position to create suckers, rather than waiting for one to be born."[44]

Online manipulation is far from harmless. Manipulation injures listeners' autonomy and welfare when it shapes those listeners' choices to their economic and other detriment. This is the case, for example, "when a firm uses personal information to extract as much rent as possible from the consumer," Calo explains.[45]

> [T]he consumer is shedding information that, without her knowledge or against her wishes, will be used to charge her as much as possible, to sell her a product or service she does not need or needs less of, or to convince her in a way that she would find objectionable were she aware of the practice.[46]

The case for regulation becomes even stronger when we recall that manipulation, by definition, covertly targets and exploits users' vulnerabilities, thus inflicting harm that its targets cannot avoid through the traditional self-help remedies of

---

visual stimuli embedded in users' contact lenses, some tech developers may become so good at raising the odds of a purchase that probability approaches certainty . . . ."); Willis, *supra* note 6, at 128 ("A 2014 experiment showed targeting marketing texts based on a model derived by machine learning from past data resulted in thirteen times more conversions than targeting texts based on variables selected by human marketers.").

[41] Cohen, *supra* note 9, at 658.
[42] Calo, *supra* note 9, at 999–1000.
[43] *Id.* at 1021 (emphasis omitted).
[44] *Id.* at 1018 (emphasis omitted); *see also id.* at 1034 ("The advancement in this Article is to observe that intervention may be justified not only where a consumer is already vulnerable, and firms are taking advantage, but also [ ] and indeed a fortiori [ ] where the firm is leveraging what it knows about the consumer in order to purposefully render that specific consumer vulnerable.").
[45] *Id.* at 1029.
[46] *Id.* at 1030.

avoidance and counter-speech. "For politics and law, a pervasive puzzle is why manipulation is so rarely policed. The simplest answer is that manipulation has so many shades, and in a social order that values free markets and consumer sovereignty, it is exceptionally difficult to regulate manipulation as such," Cass Sunstein notes.[47] "But as the manipulator's motives become more self-interested or venal, and as efforts to bypass people's deliberative capacities becomes more successful, the ethical objections to manipulation may be very forceful, and the argument for a legal response is fortified."[48]

Online manipulation is also far from inevitable. Instead, it is the product of conscious design choices, carefully studied and tested to maximize their effectiveness in shaping targets' choices without those targets' conscious awareness, and deliberately unleashed to advantage some at the expense of others. Government and private actors alike have made legal, policy, and design choices that have enabled increases in online manipulation and its attendant harms.[49] So too can we choose instead to make legal, policy, and design decisions that deter and prevent these practices and their harms.

Ethicists like Susser, Roessler, and Nissenbaum provide powerful conceptual tools for thinking about these contemporary problems of manipulation—and how they relate to (and sometimes differ from) the problems of coercion and deception (and the non-problem of persuasion). The next Part explores how these conceptual tools can also explain when, why, and how the First Amendment permits the government to regulate manipulation.[50]

---

[47]   Cass R. Sunstein, *Fifty Shades of Manipulation*, 1 J. MKTG. BEHAV. 213, 213 (2015).

[48]   *Id.*; *see also* Susser, Roessler & Nissenbaum, *supra* note 2, at 44 ("[B]ecause 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.") (emphasis omitted).

[49]   *See* Julie E. Cohen, *Law for the Platform Economy*, 51 U.C. DAVIS L. REV. 133, 161 (2017) ("As the networked information environment has redistributed control over reputational development, powerful economic actors have worked to craft narratives that make unaccountability for certain types of information harms seem logical, inevitable, and right. They have relied heavily on the U.S. [F]irst [A]mendment tradition, which characterizes the public sphere as a marketplace of ideas—an arena for neutral truth production through deliberate, reasoned exchange, where the goods on offer can be evaluated on their merits, where the volume and quality of information are regulated by the laws of supply and demand, and where those making decisions about the quality of information function as separate, individual nodes of rationality.").

[50]   This Article focuses on constitutional constraints on the government's regulation of manipulative speech rather than statutory restraints like Section 230 of the Communications Decency Act, which currently immunizes internet service providers from liability for harm caused by content published by others. For thoughtful discussion of those statutory barriers and possibilities for their application and reform, see Danielle Keats Citron & Mary Anne Franks, *The Internet as a Speech Machine and Other Myths Confounding Section 230 Speech Reform*, 2020 U. CHI. L.F. 45 (2020); Overton, *supra* note 7; Olivier Sylvain, *Recovering Tech's Humanity*, 119 COLUM. L. REV. F. 252 (2019).

II. THE FIRST AMENDMENT IMPLICATIONS OF THE GOVERNMENT'S
INTERVENTIONS TO PROTECT LISTENERS FROM MANIPULATION[51]

Challenging constitutional problems are often difficult because they force us to choose between important constitutional values—for example, between liberty and security,[52] or among speech, religion, and equality.[53] With respect to the First Amendment, the choice is often between speakers' and listeners' First Amendment interests—interests that include autonomy, enlightenment, and democratic self-governance.[54] This is the case, for instance, when speakers wish to tell lies while their listeners hunger for the truth; when listeners pine for speakers to reveal information that speakers would prefer to conceal; and when listeners hope for respite from speakers resolved to address them.[55]

Bedrock First Amendment law often privileges speakers' interests over listeners' because it presumes that listeners can usually protect themselves from unwelcome or harmful speech through avoidance and rebuttal.[56] At the same time, however, First Amendment law sometimes permits the government's intervention where asymmetries of power and information between speakers and listeners not only increase the

---

[51] As explained, *supra* notes 9–11 and accompanying text, one can plausibly understand the First Amendment's coverage to exclude manipulative practices of the sort discussed here. This Part assumes arguendo that courts may characterize the sorts of manipulative practices discussed here as speech covered by the First Amendment and explores the constitutional implications of that assumption.

[52] *E.g.*, Youngstown Sheet & Tube Co. v. Sawyer, 343 U.S. 579, 589 (1952) (exploring the trade-offs between security and liberty when interpreting the scope of the President's inherent Article II authority).

[53] *See* Masterpiece Cakeshop, Ltd. v. Colo. C.R. Comm'n, 138 S. Ct. 1719 (2018) (considering Free Exercise Clause and Free Speech Clause challenges to state antidiscrimination law).

[54] *See* Toni M. Massaro & Helen Norton, *Free Speech and Democracy: A Primer for Twenty-First Century Reformers*, 54 U.C. DAVIS L. REV. 1631, 1658–60 (2021) (describing First Amendment values).

[55] I have explored a variety of these tensions elsewhere. *E.g.*, Helen Norton, *Powerful Speakers and Their Listeners*, 90 U. COLO. L. REV. 441 (2019) (describing asymmetries of information and power between speakers and interests in several contexts); Helen Norton, *Pregnancy and the First Amendment*, 87 FORDHAM L. REV. 2417 (2019) (examining tensions between the First Amendment interests of pregnant women seeking reproductive health care and the speakers who seek to influence them); Norton, *Truth and Lies in the Workplace*, *supra* note 32 (describing conflicts between employers' and workers' First Amendment interests); Helen Norton, *Secrets, Lies, and Disclosure*, 27 J.L. & POL. 641 (2012) (describing tensions between speakers' and listeners' interests in political campaign settings).

[56] *See* Cohen v. California, 403 U.S. 15, 21 (1971) ("Those in the Los Angeles courthouse [offended by a jacket bearing the words "Fuck the Draft"] could effectively avoid further bombardment of their sensibilities simply by averting their eyes."); Whitney v. California, 274 U.S. 357, 377 (1927) (Brandeis, J., concurring) ("If there be time to expose through discussion the falsehood and fallacies, to avert the evil by the processes of education, the remedy to be applied is more speech, not enforced silence.").

likelihood and severity of harm to listeners, but also limit the effectiveness of listeners' traditional self-help remedies.[57]

In these settings, courts sometimes interpret the First Amendment to permit the government to intervene on listeners' behalf by prohibiting false and misleading speech, requiring speakers to stay away from listeners who prefer to be left alone, or requiring speakers to make accurate disclosures of material matters.[58] In the commercial setting, think of governmental requirements that manufacturers and sellers affirmatively disclose the costs of, or dangers posed by, their products even when those speakers would prefer not to reveal that information.[59] And in the context of public discourse, recall the government's campaign disclosure and disclaimer requirements that serve listeners' informational interests in knowing the source of campaign advertisements and contributions—even though some of those campaign speakers and contributors would prefer not to disclose their identities.[60]

In short, inequalities of information and power sometimes matter to First Amendment doctrine. And online manipulative speech, as defined in Part I, inherently involves such inequalities: those who deploy manipulative interfaces enjoy informational advantages because their ability to collect, aggregate, and analyze data about their listeners means that they often know more about listeners and their vulnerabilities than the listeners themselves know.[61] These informational advantages also often draw from, or exacerbate, power advantages.[62]

---

[57] *See* Cent. Hudson Gas & Elec. Corp. v. Pub. Serv. Comm'n, 447 U.S. 557, 563 (1980) ("The government may ban forms of communication more likely to deceive the public than to inform it."); *see also* Friedman v. Rogers, 440 U.S. 1 (1979); Ohralik v. Ohio State Bar Assn., 436 U.S. 447 (1978).

[58] But not always. *See* Nat'l Inst. Fam. and Life Advocs. v. Becerra, 138 S. Ct. 2361 (2018) (privileging pregnancy service centers' autonomy interests as speakers in not disclosing factual information about the availability and quality of services over pregnant women's interests as listeners in receiving accurate information that is material to their reproductive decisions).

[59] *E.g.*, Zauderer v. Off. of Disciplinary Couns. of Sup. Ct., 471 U.S. 626, 651 (1985) (upholding commercial disclosure requirements that serve consumers' informational interests as listeners).

[60] *E.g.*, Citizens United v. Fed. Election Comm'n, 558 U.S. 310, 371 (2010) (upholding federal campaign disclosure and disclaimer requirements and noting that "transparency enables the electorate to make informed decisions and give proper weight to different speakers and messages").

[61] *See* Susser, Roessler, & Nissenbaum, *supra* note 2, at 44 ("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. 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.").

[62] Jack M. Balkin, *Free Speech in the Algorithmic Society: Big Data, Private Governance, and New School Speech Regulation*, 51 U.C. DAVIS L. REV. 1149, 1156 (2018) ("Big Data

The First Amendment thus can be understood to permit the government to protect listeners from manipulation and its harms in settings where listeners cannot protect themselves because they are unaware of their manipulation. The remainder of this Part discusses more specifically what this means for manipulative speech in commercial and in political settings.

A. Regulating Manipulative Speech in Commercial Settings

The Supreme Court's longstanding commercial speech doctrine takes a transparently listener-centered approach by treating commercial speech as protected or unprotected depending on whether it provides value to consumers as listeners. As Felix Wu explains, "Commercial speech protection [] originated in and is justified by protecting consumers' rights to receive commercial information, not in protecting merchants' rights to frame that information."[63] Under the Court's framework, the government's restriction of accurate commercial speech about legal activity (like accurate speech about prescription drug prices[64]) triggers courts' suspicion in the form of intermediate scrutiny because such speech is generally valuable to listeners.[65] At the same time, the Court treats commercial speech that is false, misleading, or related to illegal activity as entirely unprotected by the First Amendment because such speech frustrates listeners' interests.[66]

---

collects and analyzes information about people—their locations, actions, characteristics, and behaviors. But the people whose information is collected are not necessarily the people who control the information. Quite the contrary: information about the world's populations serves as grist for the mill of computation, analysis, and decisionmaking by governments and large corporations. Big Data enables new ways of classifying people, making decisions about them, and exercising power over them."); *see also* NEIL POSTMAN, TECHNOPOLY: THE SURRENDER OF CULTURE TO TECHNOLOGY 11 (1992) (advising that we ask "to whom will the technology give greater power and freedom? And whose power and freedom will be reduced by it?"); *id.* at 13 ("[E]mbedded in every tool is an ideological bias, a predisposition to construct the world as one thing rather than another, to value one thing over another . . . .").

[63]   Felix T. Wu, *Commercial Speech Protection as Consumer Protection*, 90 U. COLO. L. REV. 631, 637 (2019); *see also* Felix T. Wu, *The Commercial Difference*, 58 WM. & MARY L. REV. 2005, 2060 (2017) ("[T]he First Amendment imposes no impediment to the government regulating marketing techniques that unduly take advantage of consumer weaknesses. . . . [B]ecause the First Amendment value of the solicitation speech lies in its value to the recipient, rather than the speaker, there can be no infringement on First Amendment rights in a regulation that protects the recipient in that encounter.").

[64]   *See* Va. State Bd. of Pharmacy v. Va. Citizens Consumer Council, Inc., 425 U.S. 748, 756–63 (1976) (explaining the value of information about prescription drug prices and other commercial speech to consumer decision-making).

[65]   Cent. Hudson Gas & Elec. Corp. v. Pub. Serv. Comm'n, 447 U.S. 557, 563–64 (1980).

[66]   *Id.* Despite the contemporary antiregulatory turn in its Free Speech Clause doctrine, the Court continues to apply this framework to the regulation of commercial speech. *E.g.*, Matal v. Tam, 137 S. Ct. 1744, 1763–65 (2017) (discussing commercial speech doctrine with respect to the government's regulation of trademarks); Expressions Hair Design v. Schneiderman,

Writing in 2014, Ryan Calo proposed that courts categorize online manipulation as commercial speech that is entirely unprotected under this precedent because it is misleading.[67] Although I agree that manipulation can be considered false or misleading (that is, when it seeks to cause the listener to believe a false factual representation to be true), many manipulative practices instead target and exploit cognitive, emotional, and other vulnerabilities, rather than make false or misleading representations of fact.[68]

For this reason, the Court's current commercial speech framework is incomplete in its failure to address a large volume of manipulative (yet not deceptive) commercial speech, especially in online environments.[69] To fill this doctrinal lacuna, I propose that the Court should refine and extend current commercial speech doctrine to add "manipulative" commercial speech to the commercial speech that it already treats as unprotected by the First Amendment. This move tracks the theoretical justifications of the commercial speech doctrine as steeped in protecting listeners' First Amendment interests.[70] When one recalls that false and misleading commercial speech, as well as commercial speech related to illegal activity, loses its First Amendment protection

---

137 S. Ct. 1144, 1151 (2017) (discussing commercial speech doctrine with respect to the government's regulation of retailers' communication about prices).

[67] Calo, *supra* note 9, at 1038–39. For a related but distinct take on manipulative marketing outside of the digital context, see Micah L. Berman, *Manipulative Marketing and the First Amendment*, 103 GEO. L.J. 497, 501 (2015) (contending that the *Central Hudson* commercial speech framework addresses only informational advertising and asserting that "noninformational marketing practices that manipulate consumers—and particularly marketing practices that seek to subconsciously influence consumer decisionmaking—should be entitled to limited, if any," First Amendment protection). Berman focuses on "noninformational marketing" that takes the form of product placement and package coloring. *See id.* at 526–33.

[68] *See* Luguri & Strahilevitz, *supra* note 39, at 90 (describing techniques "like nagging, price comparison prevention, intermediate currency, toying with emotion, or confirmshaming").

[69] Calo suggested that online manipulation might also be characterized as coercive, and thus unprotected by the First Amendment for that reason as well. Calo, *supra* note 9, at 1038–39. But recall that manipulation and its harms are conceptually distinct from coercion: manipulation is sneaky, coercion is not. *See supra* notes 8–10 and accompanying text; David C. Vladeck, *Digital Marketing, Consumer Protection, and the First Amendment: A Brief Reply to Professor Ryan Calo*, 82 GEO. WASH. L. REV. ARGUENDO 156, 166–67 (2014) (doubting that the effects of online marketing can accurately be described as coercive).

[70] If courts decline the invitation to treat manipulative commercial speech as entirely unprotected, then *Central Hudson* intermediate scrutiny would apply to the government's regulation of such manipulative commercial speech. Even so, appropriately designed regulations may still survive this scrutiny. *See* Greater Phila. Chamber of Com. v. City of Philadelphia, 949 F.3d 116, 154–57 (3d Cir. 2020) (holding that the city's law prohibiting employers' inquiries about applicants' salary history survived *Central Hudson* intermediate scrutiny); King v. Gen. Info. Servs., Inc., 903 F. Supp. 2d 303, 304–05, 307–08, 310–11, 313 (E.D. Pa. 2012) (holding that the Fair Credit Reporting Act requirement that credit reports exclude outdated arrest record information survived *Central Hudson* intermediate scrutiny); Berman, *supra* note 67, at 541–46 (describing how the government's regulation of noninformational marketing, like product placement and package coloring, to protect consumers from manipulation could satisfy intermediate scrutiny).

precisely because that speech frustrates listeners' First Amendment interests, one sees that the same can be true of manipulative commercial speech: it frustrates listeners' interests by seeking to covertly influence those listeners' choices to the speaker's advantage without their conscious awareness and by targeting and exploiting their vulnerabilities.[71]

Filling this doctrinal lacuna would also help fill enforcement lacunae within current law. Even though existing consumer protection statutes often prohibit "unfair" as well as "deceptive" trade practices,[72] to date "[u]nfairness has been the basis for decision in only a handful of litigated cases"[73]—largely because of the conceptual difficulty in defining and describing illegally "unfair" practices.[74] Here too Susser, Roessler, and Nissenbaum give us the conceptual tools to help understand why manipulation—that is, speakers' *hidden* efforts to shape listeners' decision-making that *target* and *exploit* those listeners' vulnerabilities in ways that the targets are not consciously aware of—is "unfair" to consumers even when it is hard to characterize as "deceptive."[75]

This doctrinal move would also support the constitutionality of new legislation if it were to be enacted. A variety of recent legislative proposals seek to regulate manipulative online practices of the sort described here. Examples include the Deceptive Experiences To Online Users Reduction Act, which proposes to prohibit interfaces designed "with the purpose or substantial effect of obscuring, subverting, or impairing user autonomy, decision-making, or choice to obtain consent or user data;" practices that "subdivide or segment consumers of online services into groups for the purposes of behavioral or psychological experiments or studies, except with the informed consent of each user involved;" and practices that cultivate children's

---

[71] Note that workers as well as consumers are vulnerable to manipulative commercial speech. *See* Susser, Roessler & Nissenbaum, *supra* note 2, at 7–9 (observing how algorithms deployed in the gig economy can manipulate workers' choices).

[72] 15 U.S.C. § 45(n) (defining a trade practice to be illegally unfair if it "causes or is likely to cause substantial injury to consumers which is not reasonably avoidable by consumers themselves and not outweighed by countervailing benefits to consumers or to competition").

[73] Michael M. Greenfield, *Unfairness Under Section 5 of the FTC Act and Its Impact on State Law*, 46 WAYNE L. REV. 1869, 1877 (2000); *see also* Coburn Keegan & Calli Schroeder, *Unpacking Unfairness: The FTC's Evolving Measures of Privacy Harms*, 15 J.L. ECON. & POL'Y 19, 19–20 (2019) (explaining the FTC's primary reliance on deception rather than unfairness).

[74] *See* Willis, *supra* note 6, at 120 (noting "thorny questions about when marketing crosses from fair persuasion to unfair or abusive manipulation" and the lack of "societal consensus on where to draw the line"). Given the lack of enforcement activity, "[t]here is virtually no legal authority addressing the question of whether commercial speech that satisfies the FTC's test for unfairness, but is neither misleading nor deceptive, is protected by the First Amendment." Luguri & Strahilevitz, *supra* note 39, at 100. Treating manipulative commercial speech as unprotected—akin to false and misleading commercial speech—would, of course, resolve this question.

[75] Note that "unfairness" may be an even more capacious term than manipulation; for example, coercive sales tactics can be unfair even if not manipulative as the term is used in this Article.

compulsive platform usage.[76] As examples of such interfaces or practices, the bill's cosponsors pointed to:

> [A] sudden interruption during the middle of a task repeating until the user agrees to consent; a deliberate obscuring of alternative choices or settings through design or other means; or the use of privacy settings that push users to 'agree' as the default option, while users looking for more privacy-friendly options often must click through a much longer process, detouring through multiple screens.[77]

The bill proposes to treat such practices as violations of the Federal Trade Commission's "rule defining an unfair or deceptive act or practice under . . . the Federal Trade Commission Act."[78]

Treating manipulative commercial speech as unprotected by the First Amendment then requires that we have a workable principle for identifying such manipulative commercial speech. The remainder of this subpart explores two possibilities.

1. Focusing on Manipulative Effects

One approach to identifying manipulative (and thus unprotected) commercial speech is to require evidence of its manipulative success.

Recall that manipulative interfaces and other online practices are the products of conscious design choices, carefully studied and tested to maximize their effectiveness in shaping targets' choices without those targets' conscious awareness.[79] This means that academics and regulators can also measure their effects. Testing by legal

---

[76] S. 1084, 116th Cong. § 3(a)(1) (2019).

[77] Press Release, The Office of United States Senator Deb Fischer, Senators Introduce Bipartisan Legislation to Ban Manipulative 'Dark Patterns' (Apr. 9, 2019), https://www.fischer .senate.gov/public/index.cfm/2019/4/senators-introduce-bipartisan-legislation-to-ban-manipu lative-dark-patterns [https://perma.cc/GC9B-99KM]. For a related proposal under consideration in the European Union, see *Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonized Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts*, at 12–13, COM (2021) 206 final (Apr. 21, 2021) (proposing to ban artificial intelligence systems with "significant potential to manipulate persons through subliminal techniques beyond their consciousness or exploit vulnerabilities of specific vulnerable groups such as children or persons with disabilities in order to materially distort their [behavior] in a manner that is likely to cause them or another person psychological or physical harm").

[78] S. 1084, 116th Cong. § 3(d) (2019).

[79] *See* Langvardt, *supra* note 9, at 142 ("Aza Raskin, inventor of the infinite scroll at Mozilla, told interviewers that behind every phone screen, there are about a thousand engineers who work on increasing addictiveness.").

scholars Luguri and Strahilevitz, for instance, found certain interfaces to double, triple, sometimes even quadruple consumers' willingness to accept sellers' offers and requests compared to neutral choice architecture: "[o]ur bottom line is that dark patterns are strikingly effective in getting consumers to do what they would not do when confronted with more neutral user interfaces."[80] These design choices include interfaces that obscure options that are popular to consumers but less lucrative for sellers, minimize material information with smaller print in less prominent locations, or require users to jump through numerous hoops to reject or withdraw from a service or product (so-called "roach motels").[81] Luguri and Strahilevitz thus recommend that regulators engage in testing of their own to identify manipulative interfaces[82]—that is, design choices that should trigger regulatory interventions because of their measurably stark effects on consumers' decisions.

Indeed, algorithmic manipulation at times may be easier to identify and measure—and thus responsibly regulate—than that by manipulative humans. "[D]ark patterns are different from other forms of dodgy business practices because of the scale of e-commerce. There may be poetic justice in the fact that this very scale presents an opportunity for creative legal regulators," Luguri and Strahilevitz note.[83] "Now that scholars can test dark patterns, we can isolate causation in a way that's heretofore been impossible in the brick-and-mortar world. Unlike brick-and-mortar manipulation, dark patterns are hiding in plain sight, operate on a massive scale, and are relatively easy to detect."[84]

Of course, we can imagine challenges in identifying baselines for neutral choice architecture. (But, as Luguri and Strahilevitz note, that is not always the case: "It should not be hard to generate consensus around the idea that a simple Yes/No or Accept/Decline prompt is neutral, provided the choices are presented with identical fonts, colors, font sizes, and placement.")[85] Effects-based approaches to identifying illegally manipulative practices may also invite objections that comparisons to so-called "neutral" baselines rely on a contested liberal understanding of the autonomous self that assumes that our preferences are stable: in other words, our preferences are

---

[80] Luguri & Strahilevitz, *supra* note 39, at 46 (emphasis omitted); *see also* Becher & Feldman, *supra* note 34, at 497–99 (proposing an "evidence-based approach" to regulating market manipulations that involves regulators' testing in laboratory settings to assess designs' manipulative success as well as users' awareness (or lack thereof) of those effects).

[81] Luguri & Strahilevitz, *supra* note 39, at 49–50, 52.

[82] *Id.* at 44–45.

[83] *Id.* at 48; *see also id.* ("It is exceedingly difficult to figure out whether a door-to-door salesperson's least savory tactics significantly affected the chances of a purchase—was the verbal sleight of hand material or incidental? Who knows? But with e-commerce, firms run thousands of consumers through identical interfaces at a reasonable cost and see how small software tweaks might alter user behavior. Social scientists working in academia or for the government can do this too; we just haven't done so before today.").

[84] *Id.*

[85] *Id.* at 98.

not always consistent and, in fact, can be shaped by algorithms and other forces, both technological and otherwise.[86]

For these reasons, rather than (or in addition to) focusing on manipulative success, we might identify manipulative (and thus unprotected) commercial speech as that involving certain manipulative characteristics that signal the intent and tendency to interfere with our choices, an effort that we might objectively view as harmful in and of itself. The next section explores this possibility.

2. Focusing on Manipulative Features

A second approach to identifying commercial speech that is manipulative (and thus unprotected by the First Amendment) focuses on design choices that display the three characteristics of manipulation as defined by Susser, Roessler, and Nissenbaum—*hidden* efforts that *target* and *exploit* users' vulnerabilities:

> [W]e 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.[87]

To be sure, much more work remains to be done in defining each of these three characteristics more precisely. But to start I will suggest "hidden" to mean not apparent to the user; "target" to mean identifying specific users with certain vulnerabilities; and "exploit" to mean deliberately deploying knowledge of those vulnerabilities in settings where there's reason to believe it will shape users' choices to the speaker's advantage. In other words, we can understand a speaker's choice to deploy designs with features that display these manipulative characteristics as a proxy for that choice's tendency to manipulate those listeners.[88] Such a regulatory focus on risk of

---

[86]   *See* Ian Kerr, *The Devil Is in the Defaults*, 4 CRITICAL ANALYSIS L. 91, 92 (2017); Susser, Roessler & Nissenbaum, *supra* note 2, at 4 (relying on philosophical "accounts of socially-situated, relational autonomy"); *see also* Julie Cohen, *What Privacy is For*, 126 HARV. L. REV. 1904, 1907, 1909 (2013) (describing conceptions of the idealized "liberal self" that "possesses both abstract liberty rights and the capacity for rational deliberation and choice and is capable of exercising its capacities in ways uninfluenced by cultural context" and identifying "the need for a postliberal theory of selfhood—one capacious enough to accommodate the full spectrum of relational, emergent subjectivity").

[87]   *See* Susser, Roessler & Nissenbaum, *supra* note 2, at 41.

[88]   *See* Cohen, *supra* note 9, at 655 ("In the context of platform-based, massively interme-diated environments, the legal system should be less concerned with intentionality as to specific pieces of content—a lens that inevitably implicates the state in choice of political

harm (rather than causation) is by no means unprecedented in the consumer protection context. The Federal Trade Commission, for instance, identifies commercial advertisements as deceptive by focusing not on a showing that they actually *cause* consumer deception, but instead on a showing that they are *likely to* mislead a reasonable consumer and that those representations are material (and thus likely to cause injury to the consumer).[89]

B. Regulating Manipulative Speech in Political Settings

While manipulation in commercial settings threatens harm to individual consumers, online manipulation in public discourse additionally threatens collective harm to our democratic self-governance. "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," Susser, Roessler, and Nissenbaum observe.[90] "Add to this the fact that the tools of online manipulation are concentrated in only a few hands, and it is easy to see how the nexus of influence and information technology stands to make already problematic power dynamics far worse."[91]

Recall, for instance, Cambridge Analytica's use of big data to microtarget messages to specific voters in efforts to influence those voters' choices in the 2016 U.S. elections.[92] Authoritarians and others similarly seek to exploit the manipulative possibilities enabled by twenty-first-century platforms in ways that threaten democracy, as Julie Cohen explains:

> Authoritarian information systems have developed sophisticated information strategies that leverage platform-based environments to undermine common knowledge about how democratic institutions function and, by extension, to destabilize the behavioral norms that lend such institutions continuing legitimacy. Such attacks, which are now well-documented, exploit platform capabilities for microtargeting, automaticity, and cascading, socially-networked information spread to stoke conspiracy theories and foster distrust—of government, of the 'mainstream media,' of scientific consensus around topics such as climate change and the efficacy of vaccines, and so on. Powerful domestic factions that

preferences—and more concerned with a deliberate design orientation that privileges automatic, habitual response and reflexive amplification.").

[89] *See* FED. TRADE COMM'N, ENFORCEMENT POLICY STATEMENT ON DECEPTIVELY FORMATTED ADVERTISEMENTS 1–2 (2015).

[90] Susser, Roessler & Nissenbaum, *supra* note 2, at 43.

[91] *Id.* at 43–44.

[92] *See* Ido Kilovaty, *Legally Cognizable Manipulation*, 34 BERKELEY TECH. L.J. 449, 467–68 (2019).

> should have mobilized to defend these assaults on our founda-
> tional institutions instead have adopted weaponization techniques
> to further their own ends. As such strategies become more pow-
> erful, they produce and amplify modes of public discourse about
> institutional actors that are incompatible with the knowledge
> structure of a stable democracy.[93]

The existential democratic threats posed by today's manipulative online practices thus may now outweigh the traditional dangers of the government's regulation of speech in public discourse.[94]

Of course, speakers' efforts to manipulate listeners' decisions about voting and other core political activity occur in the realm not of commercial speech but instead in public discourse, an area where courts' suspicion of governments' regulatory interventions is considerably greater (and understandably so) than in the commercial context.[95] Under the Court's longstanding First Amendment doctrine, the government's content-neutral regulation of speech (that is, its regulation of expression's time, place, or manner rather than its content) in public discourse triggers a form of intermediate scrutiny,[96] while the government's content-based regulation triggers strict scrutiny.[97] (To be sure, some suggest that the threats to free speech and democracy posed in the twenty-first-century speech environment should inspire tweaks, and perhaps even

---

[93] Cohen, *supra* note 9, at 659.

[94] *Id.* at 661–62.

[95] *See* Jack M. Balkin, Keynote Address at the Association for Computing Machinery Symposium on Computer Science and Law: How to Regulate (and Not Regulate) Social Media (Oct. 28, 2019) (transcript available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id =3484114 [https://perma.cc/5N9V-FEA6]) (defining "public sphere" as "the space in which people express opinions and exchange views that judge what is going on in society").

[96] *See* Clark v. Cmty. for Creative Non-Violence, 468 U.S. 288, 293 (1984) (explaining that the government's content-neutral restrictions of speech survive First Amendment review so long as they "are justified without reference to the content of the regulated speech, that they are narrowly tailored to serve a significant governmental interest, and that they leave open ample alternative channels for communication of the information").

[97] *See* Adam Winkler, *Fatal in Theory and Strict in Fact: An Empirical Analysis of Strict Scrutiny in the Federal Courts*, 59 VAND. L. REV. 793, 844–55 (2006) (explaining the application of strict scrutiny, which requires the government to show that its intervention is narrowly tailored to achieve a compelling interest). For an example of the government's content-based regulation of public discourse to serve listeners' interests that survived strict scrutiny, see Burson v. Freeman, 504 U.S. 191, 211 (1992) (upholding a town's ban on the distribution of campaign literature within 100 feet of polling places to prevent voters' coercion and harassment). And for thoughtful discussion of how existing doctrine can be, and sometimes is, applied to sensibly balance competing constitutional values, see Ashutosh Bhagwat, *The Test That Ate Everything: Intermediate Scrutiny in First Amendment Jurisprudence*, 2007 U. ILL. L. REV. 783 (2007); Ashutosh Bhagwat, *In Defense of Content Regulation*, 102 IOWA L. REV. 1427 (2017).

topples, to this doctrine.)[98] In any event, the challenge is to cabin the government's considerable potential for overreach through appropriately tailored regulation of manipulative speech.

To this end, the government can tailor its intervention through its choice of regulatory target (i.e., *what* it targets for regulation), through its choice of regulatory tool (i.e., *how* it regulates its target), or both.[99]

1. Tailoring Through Targeting

As explained in the preceding subpart, appropriately tailored interventions could target interfaces with certain manipulative effects, certain manipulative features, or both.

First, interventions could target interfaces with particularly stark manipulative effects through A–B testing (where an "A/B test randomly distributes an 'A' version of something to some people and a 'B' version to others and measures differences between the responses of the two groups").[100] Indeed, those who design and deploy these features engage in this sort of testing themselves to identify effective means for influencing users' choices to click on, read, or forward specific content in political as well as commercial settings[101]: online actors deploy these manipulative interfaces precisely because they've tested them extensively and *know* they are effective in changing their target's choices.[102] (In political settings, however, the causal

---

[98] *E.g.*, Massaro & Norton, *supra* note 54, at 1638 (proposing "a framework for assessing when and why the First Amendment permits carefully designed regulations of speech to further free speech and democracy values"); Tim Wu, *Is the First Amendment Obsolete?*, EMERGING THREATS 2, 17 (2017), https://s3.amazonaws.com/kfai-documents/documents/5d8a0f848d/Is-the -First-Amendment-Obsolete-.pdf [https://perma.cc/5L8T-HLCY] (suggesting that we "imagine how First Amendment doctrine might adapt" to the speech pathologies in the twenty-first-century environment). Justice Breyer, for instance, has long urged courts and policymakers to more frankly balance the harms of the government's regulation of speech against the harms that speech sometimes threatens. *See, e.g.*, Iancu v. Brunetti, 139 S. Ct. 2294, 2304 (2019) (Breyer, J., concurring in part and dissenting in part) ("I believe we would do better to treat this Court's speech-related categories not as outcome-determinative rules, but instead as rules of thumb."); United States v. Alvarez, 567 U.S. 709, 731 (2012) (Breyer, J., concurring) ("[S]ome such approach is necessary if the First Amendment is to offer proper protection in the many instances in which a statute adversely affects constitutionally protected interests but warrants neither near-automatic condemnation (as 'strict scrutiny' implies) nor near near-automatic approval (as is implicit in 'rational basis' review).").

[99] *See* Langvardt, *supra* note 9, at 153 ("[T]he regulatory response should escalate with the sense of urgency. Light-touch responses would try to help consumers make good decisions. More dramatic responses would simply disable products' most dangerous features.").

[100] Willis, *supra* note 6, at 127 n.52.

[101] *Id.* at 127–28.

[102] *See* Cohen, *supra* note 49, at 165 ("Networked environments configured to optimize data harvesting and surplus extraction operate—and are systematically designed to operate— in ways that preclude even the most perceptive and reasonable consumer from evaluating the

relationship between design and outcome may be considerably more contested, as the decision whether or for whom to vote can be more causally complex than the decision to decide whether to buy a product or agree to a transactional condition.)

Second, interventions could instead target "[m]anipulative practices—characterized, as we have argued, by concealment, exploitation of vulnerabilities, and targeting—[as] cause for concern, regardless of whether they succeed in every instance."[103] Along these lines, Julie Cohen emphasizes that "[t]he First Amendment does not require legislators or judges to privilege design for automaticity and reflexive amplification";[104] she identifies particularly manipulative features in public discourse to include "predictive profiling and microtargeting based on behavioral and psychographic data; interface design to elicit automatic, precognitive responses; and algorithmic optimization to amplify patterns of cascading, socially-networked spread."[105]

2. Tailoring Through Tools

Moving from the question of *what* to regulate (target) to the question of *how* to regulate (tool), this section highlights a few suggestions offered by thoughtful commentators; my objective here is to introduce a range of possibilities for further consideration and exploration rather than to detail (much less exhaust) them. In thinking through these and other available options, recall that the choice of tool drives the level of suspicion under the Court's First Amendment doctrine: tools characterized as content-neutral receive "only" intermediate scrutiny, as compared to the strict scrutiny generally applied to the government's content-based regulation of protected speech.[106]

Kyle Langvardt, for instance, has described a variety of friction-creating restrictions of manipulative interfaces, like restrictions on infinite scrolls (that unceasingly feed users with new posts); restrictions on autoplay features that seek to manipulate users into remaining online for longer periods of time (during which time users spend more money and shed more data); and restrictions that limit or delay users' ability to mass-forward (and thus amplify) content.[107]

---

goods or services on offer. Predictive profiling seeks to minimize the need to persuade by targeting directly those potential customers most strongly predisposed to buy and appealing to everything that is known about those customers' habits and predilections.").

[103]   Susser, Roessler & Nissenbaum, *supra* note 2, at 41; *see also id.* at 27 (recognizing the difficulties of proving causation and thus focusing 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").

[104]   Cohen, *supra* note 9, at 655.

[105]   *See id.* at 660.

[106]   *See supra* notes 96–97 and accompanying text.

[107]   Kyle Langvardt, *Platform Speech Governance and the First Amendment: A User-Centered Approach*, *in* THE DIGITAL SOCIAL CONTRACT: A LAWFARE PAPER SERIES (2020), at 13–15; *see also id.* at 14 ("Facebook, as discussed further below, has already experimented with broad-based content-neutral restrictions in its WhatsApp and Messenger products on users' ability to mass-forward content. Twitter has experimented with a feature that would ask for

Lauren Willis urges that law "compel businesses to engage in fair marketing by design" by, for example, requiring that platforms or retailers refrain from directing materials to consumers "whose demographics or behaviors indicate persistent or transitory impairment."[108] Ian Kerr suggests that "technological defaults ought to be regulated in a manner similar to contractual defaults," urging that "the strictest privacy settings automatically apply once a customer acquires a new product or service," such that users need make no manual change to their privacy settings to protect themselves to the fullest.[109] And Woody Hartzog, who has written extensively on how law should "better reflect how design influences our perceptions and actions,"[110] urges regulators to "discourage design that tricks us, lies to us, exploits us . . . ." and advises courts and regulators to ferret out abusive design that seeks to exploit users' biases and vulnerabilities.[111]

In sum, recognizing the harm of online manipulation in political settings does not tell us how to address that harm. Although much work remains to be done in developing appropriate responses, this Part has identified some possibilities for consideration.

CONCLUSION

Some may fear the government's restriction of manipulative interfaces, viewing such restriction as unacceptably paternalistic. But paternalism describes interference with autonomous choices that others think unwise, rather than interference with practices that themselves frustrate autonomous choice.[112] Precisely because manipulation's targets are unaware of the ways in which online actors gather, aggregate, and exploit their data to influence their decision-making, online manipulation occurs in a setting that defies the traditional First Amendment model of fully informed rational listeners freely choosing among available options. In other words, nobody consciously chooses to be manipulated (as the term is used in this Article) because, by definition, targets are unaware of their manipulation, and thus can't take steps to protect themselves. As Langvardt observes, "[i]f the regulatory goal is simply to make product design less manipulative, then regulation in principle exists to enhance rather than diminish tech users' freedom of choice."[113]

---

confirmation—suggesting that '[y]ou can read the article on Twitter before Retweeting'—when a user attempts to retweet a news article seconds after seeing it.").

[108]   Willis, *supra* note 6, at 119–20, 184.

[109]   Kerr, *supra* note 86, at 101–02.

[110]   WOODROW HARTZOG, PRIVACY'S BLUEPRINT: THE BATTLE TO CONTROL THE DESIGN OF NEW TECHNOLOGIES 7 (2018).

[111]   *Id.* at 126.

[112]   *See Paternalism*, OXFORD ENG. DICTIONARY (3d ed. 2005) (defining "paternalism" as "[t]he policy or practice of restricting the freedoms and responsibilities of subordinates or dependants in what is considered or claimed to be their best interests").

[113]   Langvardt, *supra* note 9, at 148.

Others may worry that any effort to regulate manipulation is folly because manipulation is endemic to the human condition. But so too are violence, discrimination, and falsehoods—and yet government restricts each of those choices at times precisely because of the harms they inflict.[114] Rather than ask whether to regulate these all-too-human behaviors, the better question is when, why, and how to regulate them.

To this end, this Article suggests that ethicists' understanding of manipulation and its harms offers helpful tools for thinking more carefully about legal and policy responses.[115] Relying upon these tools, I propose that courts should add manipulation to the list of harms to listeners' interests that sometimes justify the government's intervention in certain settings consistent with the First Amendment.

More specifically, one can understand manipulative commercial speech to be entirely unprotected by the First Amendment, with the government's regulation subject only to rational-basis scrutiny. Like commercial speech that is false, misleading, or related to illegal activity—and thus treated as unprotected by the First Amendment[116]—manipulative commercial speech frustrates rather than furthers listeners' First Amendment interests. This leaves open possibilities for identifying commercial speech as manipulative (and thus unprotected by the First Amendment) by focusing on its effects (i.e., its manipulative success), its design (i.e., its deployment of certain manipulative features), or both. And although First Amendment doctrine appropriately poses a considerably larger barrier to the regulation of manipulative speech in political settings, there too we can understand the First Amendment to permit certain interventions that are carefully tailored in terms of regulatory target, regulatory tool, or both.[117]

Many may understandably wonder whether the contemporary Court will be receptive to these ideas, given the antiregulatory turn in its First Amendment doctrine,[118] along with its inconsistent attention to the ways in which twenty-first-century technologies inflict harms that are different in degree and sometimes in kind from

---

[114] *See, e.g.*, 18 U.S.C. § 1111 (criminalizing murder); *id.* § 242 (criminalizing deprivation of rights, privileges, and immunities "on account of such person being an alien, or by reason of his color, or race" under color of law); *id.* § 1621 (criminalizing perjury).

[115] The many remaining areas for further study include the legal and policy implications of the government's own manipulative practices, especially (but not only) in the digital age. *See* Aziz Z. Huq, *Constitutional Rights in the Machine-Learning State*, 105 CORNELL L. REV. 1875 (2020) (considering the constitutional implications of the government's use of algorithms that involve machine learning); Ryan Calo & Danielle Keats Citron, *The Automated Administrative State: A Crisis of Legitimacy*, 70 EMORY L.J. 797 (2021) (proposing a positive vision of administrative agencies' use of automation that permits agencies to adopt such tools only when they enhance agency legitimacy).

[116] Cent. Hudson Gas & Elec. Corp. v. Pub. Serv. Comm'n of N.Y., 447 U.S. 557, 563–64 (1980).

[117] *See* Burson v. Freeman, 504 U.S. 191, 196–97 (1992).

[118] *See* Norton, *Truth and Lies in the Workplace*, *supra* note 32, at 34 (describing this turn).

earlier counterparts.[119] Even so, as constitutional law scholar Mark Graber reminded us in another context: "Advocating doctrines unlikely to be accepted immediately is still a worthwhile activity."[120] Indeed, the challenges posed by twenty-first-century expressive technologies "may inspire more careful reflection about how to define and mitigate the harmful effects of covered speech, while preserving its manifold benefits."[121]

---

[119] *Compare* Carpenter v. United States, 138 S. Ct. 2206, 2223 (2018) (interpreting the Fourth Amendment to protect certain cell phone data from the government's intrusion in light of changes in technology and thus changes in our expectations of privacy), *with* Rucho v. Common Cause, 139 S. Ct. 2484, 2491, 2508 (2019) (discounting how contemporary technology enables partisan gerrymandering to be effective in ways never before imagined).

[120] MARK A. GRABER, TRANSFORMING FREE SPEECH: THE AMBIGUOUS LEGACY OF CIVIL LIBERTARIANISM 223 (1991).

[121] *See* Toni M. Massaro, Helen Norton & Margot E. Kaminski, *SIRI-OUSLY 2.0: What Artificial Intelligence Reveals About the First Amendment*, 101 MINN. L. REV. 2481, 2514 (2017).