

Fifty Shades of Manipulation

Cass R. Sunstein*

Harvard University, USA

ABSTRACT

Both marketers and politicians are often accused of “manipulation”, but the term is far from self-defining. A statement or action can be said to be manipulative if it does not sufficiently engage or appeal to people’s capacity for reflective and deliberative choice. One problem with manipulation, thus understood, is that it fails to respect people’s autonomy and is an affront to their dignity. Another problem is that if they are products of manipulation, people’s choices might fail to promote their own welfare, and might instead promote the welfare of the manipulator. To that extent, the central objection to manipulation is rooted in a version of John Stuart Mill’s Harm Principle: People know what is in their best interests and should have a (manipulation-free) opportunity to make that decision. On welfarist grounds, the norm against manipulation can be seen as a kind of heuristic, one that generally works well, but that can also lead to serious errors, at least when the manipulator is both informed and genuinely interested in the welfare of the chooser. For politics and law, a pervasive puzzle is why manipulation is 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. Those who sell products are often engaged in at least arguable forms of manipulation. But as the manipulator’s motives become more self-interested or venal, and as efforts to bypass people’s deliberative capacities become more successful, the ethical objections to manipulation may be very forceful, and the argument for a legal response is fortified. The analysis of manipulation bears on emerging free speech issues raised by compelled disclosure, especially in the context of graphic health warnings. It can also help orient the regulation of financial products, where manipulation of consumer choices is an evident but rarely explicit concern.

Keywords: Manipulation, Nudge, Social influences, Behavioral economics.

*Robert Walmsley University Professor, Harvard University, Cambridge, MA 02138, USA, csunstei@law.harvard.edu. I am grateful to Oren Bar-Gill, Anne Barnhill, Tyler Cowen, Elizabeth Emens, Nir Eyal, George Loewenstein, Martha Minow, Martha Nussbaum,

Introduction

IT ranks among the most spellbinding scenes in the history of American television. Don Draper, an advertising executive in the 1960s, is charged with producing an advertising campaign for Kodak, the film company, which has just invented a new slide projector, one that features continuous viewing. It operates like a wheel. Using the projector to display scenes from a once-happy family (as it happens, his own, which is now broken), Draper tells his potential clients:

In Greek, “nostalgia” literally means, “the pain from an old wound.” It’s a twinge in your heart, far more powerful than memory alone. This device isn’t a spaceship. It’s a time machine. It goes backwards, forwards. It takes us to a place where we ache to go again. It’s not called the Wheel. It’s called a Carousel. It lets us travel the way a child travels. Around and around, and back home again . . . to a place where we know we are loved.¹

— (IMDB 2015)

The Kodak clients are sold; they cancel their meetings with other companies. Now consider the following cases:

1. A parent tries to convince an adult child to visit him in a remote town in Nebraska, saying, “After all, I’m your father, and I raised you for all those years, and it wasn’t always a lot of fun for me — and who knows whether I’m going to live a lot longer?”
2. An automobile company advertises its new vehicle by showing a sleek, attractive couple exiting from it before going to a glamorous party.
3. In an effort to discourage people from smoking, a government requires cigarette packages to contain graphic, frightening health warnings, depicting people with life-threatening illnesses (*R. J. Reynolds Tobacco Co. v. FDA* 2012; *FDA* 2011).
4. In a campaign advertisement, a political candidate displays highly unflattering photographs of his opponent, set against the background of frightening music. An announcer reads quotations that, while accurate and not misleading, are taken out of context to make the opponent look at once ridiculous and scary.

Eric Posner, Lucia Reisch, Mark Tushnet, and Adrian Vermeule for valuable comments on a previous draft. Special thanks to Barnhill for many helpful discussions. Heidi Liu and Lisa Marrone provided excellent research assistance.

¹Revealingly, nostalgia actually means “longing for a return home,” rather than “pain from an old wound.”

5. In an effort to convince consumers to switch to its new, high-cost credit card, a company emphasizes its very low “teaser rate,” by which consumers can enjoy low-cost borrowing for a short period. In its advertisement, it depicts happy, elegant, energized people, displaying their card and their new purchases.
6. To reduce pollution (including greenhouse gas emissions), a city requires public utilities to offer clean energy sources as the default providers, subject to opt-out if customers want to save money (Sunstein and Reisch 2014).

Both public and private laws are pervasively concerned with the problem of coercion, arising from the literal use of force. But what of manipulation, undertaken by either private or public institutions, hoping to succeed in marketing a product, an idea, or a person? In social science and law, there is essentially no sustained analysis of the topic. To be sure, there is a great deal of work on lies and deception (Craswell 1985) and we can identify an overlap among lying, deceiving, and manipulating. We could even see manipulation as a master concept that includes lying and deceiving, or understand the three to be on some kind of continuum. Certainly this is so if our master principle is personal autonomy; if so, the three violate that principle, though for somewhat different reasons. (I shall have something to say about the extent to which this is so.)

But in ordinary usage, it is reasonable to think that the concept of manipulation is distinctive, certainly in the sense that it can occur (as in the mythical Kodak commercial) without lies or deception (at least in their standard forms). What does manipulation entail, and what is wrong with it? How, if at all, should the law respond to it — for example, in the context of consumer protection or in constraining government itself?

It should be clear that an action does not count as manipulative merely because it is an effort to alter people’s behavior. If you are a passenger in a car, and you warn the driver that he is about to get into a crash, you are not engaged in manipulation. The same is true if a company reminds customers that a bill is due. A calorie label and an energy efficiency label are not ordinarily counted as forms of manipulation. So long as a private or public institution is informing people, or “just providing the facts,” it is hard to complain of manipulation.² There is also a large difference between persuading

²A qualification is necessary. If a disclosure requirement focuses on one of many aspects of a situation, and fixes people’s attention on that aspect, a charge of manipulation would not be unreasonable. Consider the controversy over the idea that sellers should have to disclose that food has genetically modified organisms (GMOs) (Noussair *et al.* 2004). For those who object to compulsory labeling about GMOs, there is a plausible claim that labels are a form of manipulation, activating public concern where there is no objective reason for that concern. Of course those in the private sector might engage in similar forms of

people and manipulating them. With (nonmanipulative) persuasion, people are given facts and reasons, presented in a sufficiently fair and neutral way; manipulation is something different.

It is often thought that when people are being manipulated, they are treated as “puppets on a string” (Wilkinson 2013). Almost no one wants to be someone else’s puppet (at least without consent), and in some respects, it is especially bad to be a puppet of government (Bachrach 2009; Westerman 2012). Many of the worst governments in history have attempted to turn their citizens into puppets (Bachrach 2009; Westerman 2012). If we keep the puppet metaphor in mind, the idea of “manipulation” can be applied to many kinds of behavior; but it is not entirely clear that it is a unitary concept, or that we can identify necessary and sufficient conditions (see Coons and Weber 2014 for a number of instructive treatments). Manipulation takes multiple forms. It has at least 50 shades, and some people wonder if they are tightly identified with one another (Wilkinson 2013).

The principal goal of this article is to make progress in understanding what manipulation is and what is wrong with it. If we can make progress on those tasks, we should be better equipped to assess a wide range of problems in ethics, policy, and law. For example, plausible objections might be made to acts of government that can be counted as manipulative; such objections might not treat citizens with respect (Hansen and Jespersen 2013; Waldron 2014). There are also free speech questions. When the government compels speech, is there a difference between mandating a purely factual disclosure and mandating one that has arguably manipulative features? When is advertising manipulative, and when it is, what, if anything, should be done about it (Wood 2014)? How, if at all, should government respond to manipulative behavior by the private sector, for example in the context of financial products (such as credit cards or mortgages) (SEC 2010)? As we shall see, an understanding of manipulation bears directly on the work of the Consumer Financial Protection Bureau (CFPB) in the United States and similar organizations in other nations, and might help to orient some of that work, which is at least implicitly concerned with manipulation (Zibel 2015).

I suggest that an effort to influence people’s choices counts as manipulative *to the extent that it does not sufficiently engage or appeal to their capacity for reflection and deliberation*. The word “sufficiently” leaves a degree of ambiguity and openness, and properly so. It is not possible to know whether manipulation is involved without asking about the sufficiency of people’s capacity to deliberate on the question at hand (Barnhill 2014). We can imagine clear cases of manipulation (such as subliminal advertising) (Bullock 2004), cases that clearly fall outside of the category (such as a warning about

manipulation, drawing people’s attention to a feature of a product that, while real, appears far more important than it actually is.

deer crossings in a remote area), and cases that can be taken as borderline (such as a vivid presentation about the advantages of a particular mortgage or a redesign of a website to attract customers to the most expensive products).³ It is important to emphasize that countless choices are at least partly a product of variables that do not involve reflective deliberation — and choosers tend to be unaware of that fact (Kahneman 2011; Ullmann-Margalit 2006). The problem of manipulation arises when choosers justly complain that because of the actions of a manipulator, they have not, in a sense, had a fair chance to make a decision on their own (Barnhill 2014). Often the distinguishing mark of manipulation is a justified sense of *ex post* betrayal.

Of course there are degrees of manipulation, as some forms of influence attempt to bypass deliberation altogether (such as subliminal advertising), and other forms merely try to influence it by triggering certain forms of automatic processing (for example, through framing a problem so as to provoke the desired response) (Keren 2010). Some forms of manipulation are modest and relatively benign. In the Kodak commercial, the goal is to connect the product with a set of evocative associations — childhood, a carousel, and a magical ability to recapture, and make permanent, a lost past. Manipulation often occurs through such associations, which are a pervasive feature of both advertising and political campaigns; but the concept is much broader than that (Bread Crust Council 2011). Manipulators often describe choices so as to make certain outcomes vivid and appealing (such as purchases of lottery tickets), or vivid and unappealing (such as failures to buy life insurance) — even though a more neutral frame would present the whole problem in a less tendentious manner, leaving the chooser in a more objective position to weigh the relevant variables (and in that sense more free).

In such cases, a central problem with manipulation is that it can violate people's autonomy (by making them instruments of another's will) and offend their dignity (by failing to treat them with respect). The manipulator is leading the chooser to make a choice, without sufficiently assessing, on the chooser's own terms, its costs and its benefits. For this reason, the most strongly felt moral objections to manipulation are deontological rather than utilitarian in character. The objections reflect a sense that people are not being treated respectfully. Their own capacities — to assess, to weigh, to judge — are not being given appropriate deference. For deontologists, a central question is whether choosers have given appropriate consent to the act of manipulation, or whether the manipulator has properly inferred consent under the circumstances.

³Importantly, the word “sufficiently” applies to the degree of reflection and deliberation that are involved; it does not speak to the issue of justification. For example, would-be kidnappers might be manipulated (in the sense that their deliberative capacities are bypassed) by police officers who are trying to stop a kidnapping, and a terrorist might similarly be subject to (justified) manipulation.

From the welfarist point of view, the objection to manipulation is much less straightforward. Some people can benefit (a great deal) from being manipulated (consider a smoker who desperately wants to quit), and within limits, being manipulated can even be fun. In some forms (including marketing), manipulation is a form of play, undertaken with a smile and a wink. (A speculation: Those who are intensely opposed to manipulation, in all its shades and forms, lack a sense of humor.) In other forms, it is not fun at all, even deadly serious (consider efforts to manipulate kidnappers or terrorists). On welfarist grounds, there is no simple evaluation of manipulation, at least if we embrace the foregoing definition.

The foundation of the welfarist concern, I suggest, is the view, associated with Mill (1859) and Hayek (2014), that the chooser, and not the manipulator, knows what is in his best interest. Of course Mill's principal concern, and Hayek's too, is with coercion, but the welfarist objection to manipulation stems from the same source: A belief that choosers know best. It follows that the anti-manipulation principle is strongly derivative of Mill's Harm Principle; it suggests that choosers ought to be able to make their own decisions, and that the role of others should be restricted to informing them or attempting to persuade them (without manipulation).

If choosers know best, then the welfare-increasing approach is to avoid manipulation and to engage the chooser's deliberative capacities. But the manipulator refuses to do that. The skeptic wonders: Why not? A tempting answer is that the manipulator is promoting his own interests, and not those of the chooser. The use of manipulation, rather than (say) information or persuasion, creates a risk that the manipulator does not have the chooser's interests in mind. For that reason, manipulation undermines the welfare of the chooser. The welfarist analysis of manipulation closely parallels the welfarist analysis of fraud and deceit (Picard 2013). In a sense, the manipulator can even be seen as a kind of thief, taking something from the chooser without real consent. In some cases, that is indeed the right way to assess an act of manipulation; it helps to illuminate recent initiatives in the area of consumer financial protection and in the regulation of commercial advertising.

From the standpoint of the legal system, the problem is that as defined here, manipulation can plausibly be said to be pervasive, even in the freest of free societies. It can be found on television, on the Internet, in every political campaign, in countless markets, in friendships, and in family life. Free markets invite a degree of manipulation, certainly in the area of marketing, even as they impose significant constraints on it. Even if we insist (as we should) that manipulation cannot occur without intentional manipulators,⁴ the scope of

⁴Nature can, in a sense, manipulate people; cold weather and snow, for example, can affect people without sufficiently triggering deliberation. But it seems useful to limit the

the practice is very wide. It would be odd and perhaps pointless to condemn practices that people encounter daily, and with which they live while mounting little or no objection. Indeed, it would be fussy and stern — even a bit inhuman — to try to excise it.

Because of the pervasiveness of manipulation, and because it often does little or no harm, the legal system usually does not attempt to prevent it. At least in general, the costs of regulating manipulation would far exceed the benefits. But as we shall see, the proper evaluation of acts of manipulation depends a great deal on context, including the expectations associated with particular roles (Sunstein 2015; Wilkinson 2013). In some contexts, regulators do aim to prevent manipulation, at least implicitly (CFPB 2013; Smith and Graziano 2014), and from the standpoint of consumer protection, it seems fair to say that there is a mounting interest in constraining manipulation as such, not least in the aftermath of the financial crisis.

Everyone knows that a car company wants to sell cars, and under existing conventions, it is acceptable to produce advertisements that do not exactly target people's deliberative capacities (at least if falsehoods are not involved). Something similar can be said about political campaigns. To be sure, there remains a question whether deliberative capacities are "sufficiently" engaged; in such cases, they are hardly on hold. But in extreme cases, it would be plausible to suggest that the sufficiency requirement is not met. The ethical objection gains strength under two conditions: (1) when the manipulator's goals are self-interested or venal and (2) when the act of manipulation is successful in subverting or bypassing the chooser's deliberative capacities. When both conditions are met, there is good reason for an ethical taboo on manipulation, and perhaps legal constraints, certainly (but not only) when it comes from government.

The remainder of this article is organized as follows: The following section explores conceptual and definitional issues, distinguishing manipulation from deceit, and linking the idea of manipulation to recent psychological and behavioral findings. The section "What is Wrong with Manipulation, So Understood?" turns to normative questions, investigating whether and how manipulation might be unacceptable in light of commitments to autonomy, dignity, and welfare. The section "Consent, Transparency, Democracy" examines the relevance of democratic norms. It concludes that while individual consent justifies manipulation, transparency and democratic authorization do not. The section "Manipulation, Freedom of Speech, and Consumer Protection" explores free speech issues and consumer protection.

category to intentional efforts; in ordinary language, intentionality appears to be a defining characteristic of the concept of manipulation.

What Manipulation Is

Insulting Deliberation

Despite the paucity of relevant work in social science and law, significant effort has been devoted to the definition of manipulation, almost exclusively within the philosophical literature (see Coons and Weber 2014 for an overview). Many of the efforts focus on the effects of manipulation in counteracting or undermining people's ability to engage in rational deliberation. On Wilkinson's (2013) account, for example, manipulation "is a kind of influence that bypasses or subverts the target's rational capacities." Wilkinson urges that manipulation "subverts and insults a person's autonomous decision making," in a way that treats its objects as "tools and fools." He thinks that "manipulation is intentionally and successfully influencing someone using methods that pervert choice" (see also Barnhill 2014).

Recall, for example, efforts to enlist attractive people to sell cars, or to use frightening music and ugly photos to attack a political opponent. We might think that in such cases, customers and voters are being insulted in the sense that the relevant speaker is not giving them anything like a straightforward account of the virtues of the car or the vices of the opponent, but is instead using associations of various kinds to press the chooser in the manipulator's preferred direction. On a plausible view, manipulation is involved to the extent that deliberation is insufficient. Here again, it is important to notice that we should speak of degrees of manipulation, rather than a simple on-off switch.

In a related account, Faden and Beauchamp (1986) define psychological manipulation as "any intentional act that successfully influences a person to belief or behavior by causing changes in mental processes *other than those involved in understanding*." Raz (1986) suggests that "Manipulation, unlike coercion, does not interfere with a person's options. Instead it perverts the way that person reaches decisions, forms preferences or adopts goals."

Of course the idea of "perverting" choice, or people's way of reaching decisions or forming preferences, is not self-defining; it can be understood to refer to methods that do not appeal to, or produce, the right degree or kind of reflective deliberation. If so, an objection to manipulation is that it "infringes upon the autonomy of the victim by subverting and insulting their decision-making powers" (Wilkinson 2013). The objection also offers one account of what is wrong with lies, which attempt to alter behavior not by engaging people on the merits and asking them to decide accordingly, but by enlisting falsehoods, usually in the service of the liar's goals (an idea that also points the way to a welfarist account of what usually makes lies wrong). A lie is disrespectful to its objects, not least if it attempts to exert influence without asking people to make a deliberate choice in light of relevant facts. But when lies are not involved, and when the underlying actions appear to

be manipulative, the challenge is to concretize the ideas of “subverting” and “insulting” (Craswell 1985).

It is tempting to adopt a simple definition to this effect: *A statement or action is manipulative to the extent that it does not engage or appeal to people’s capacity for reflective and deliberative choice.* The problem with this definition is that it is far too broad, sweeping up much action that is a standard part of daily life, and that it is rarely taken as manipulative. Suppose, for example, that a good friend frames an option in the most attractive light and with a cheerful voice; or that the Department of Transportation embarks on a vivid, even graphic public education campaign to reduce texting while driving (Department of Transportation 2011); or that a politician argues in favor of same-sex marriage in a way that points, in an emotionally evocative way, to the lived experience of same-sex couples. In all of these cases, we might have long debates about whether the relevant statements are appealing to people’s capacity for reflective and deliberative choice. And even if we conclude that they are not, we should not therefore be committed to the view that manipulation is involved.

To warrant that conclusion, the word “sufficiently” is required, to add the suggestion that people have been in some sense tricked or fooled, or at least that their deliberative capacities have not been adequately engaged. In this sense, there is a connection between the idea of manipulation and the idea of deceit; we can even see the former as a lighter or softer version of the latter. With an act of deceit, people almost inevitably feel betrayed and outraged once they are informed of the truth. The same is true of manipulation. Once the full context is revealed, those who have been manipulated tend to feel used. They ask: *Why wasn’t I allowed to decide for myself?*

In an illuminating discussion, with implications for policy and law, Barnhill (2014) defines manipulation as “directly influencing someone’s beliefs, desires, or emotions, such that she falls short of ideals for belief, desire, or emotion in ways typically not in her self-interest or likely not in her self-interest in the present context.” Notwithstanding its ambiguity and need for specification, the idea of “falling short of ideals” is helpful, and it should be seen as an effort to capture the same idea as the word “sufficiently.” Note that the standard here is best taken as objective, not subjective. The question is whether someone has, in fact, sufficiently engaged a chooser’s deliberative capacities — not whether the chooser so believes. But there is a problem with Barnhill’s definition, which is that it excludes, from the category of manipulation, influences that are in the self-interest of the chooser. Some acts of manipulation count as such even if they leave the chooser better off. (You might be manipulated to purchase a car that you end up much enjoying.) We might say that such acts are justified — but they are manipulative all the same.

To understand manipulation in this general way, it should not be necessary to make controversial claims about the nature of choice or the role of emotions.

We should agree that many decisions are based on unconscious processing and that people often lack a full sense of the wellsprings of their own choices (Kahneman 2011). Even if this is so, a manipulator might impose some kind of influence that unduly undermines or bypasses reflection and deliberation. It is also possible to acknowledge the view that emotions might themselves be judgments of value (Nussbaum 2003) while emphasizing that manipulators attempt to influence people's choices without promoting much in the way of reflective thinking about the values at stake. In ordinary language, the idea of manipulation is invoked by people who are not committed to controversial views about psychological or philosophical questions, and it is best to understand that idea in a way that brackets the relevant controversies.

Manipulating System 1

We can make progress in understanding some kinds of manipulation with reference to the now-widespread view that the human mind contains not one but two “cognitive systems” (Kahneman 2011; Wilson 2014). In the social science literature, the two systems are described as System 1 and System 2 (Kahneman 2011). System 1 is the automatic, intuitive system, prone to biases and to the use of heuristics, while System 2 is more deliberative, calculative, and reflective. Manipulators often target System 1, and they attempt to bypass or undermine System 2 (Hansen and Jespersen 2013). Savvy marketers know that if they want to influence behavior, they should engage System 1.

System 1 works quickly. Much of the time, it is on automatic pilot. It is driven by habits. When it hears a loud noise, it is inclined to run. When it is offended, it wants to hit back. It certainly eats a delicious brownie. It can procrastinate; it can be impulsive. It is easy to manipulate. It wants what it wants when it wants it. It can be excessively fearful and too complacent. It is a doer, not a planner.

By contrast, System 2 is reflective and deliberative. It calculates. It hears a loud noise, and it assesses whether the noise is a cause for concern. It thinks about probability, carefully though sometimes slowly. It does not really get offended. If it sees reasons for offense, it makes a careful assessment of what, all things considered, ought to be done. It sees a delicious brownie, and it makes a judgment about whether, all things considered, it should eat it. It is harder to manipulate. It insists on the importance of self-control. It is a planner more than a doer.

We need not venture contested claims about the nature of the two systems in order to find it helpful to suggest that many actions count as manipulative because they appeal to System 1, and because System 2 is being subverted, tricked, undermined, insufficiently involved, or not informed. Consider the case of subliminal advertising, which should be deemed manipulative because it operates “behind the back” of the person involved, without appealing to

his conscious awareness. People's decisions are affected in a way that entirely bypasses their own deliberative capacities. If this is the defining problem with subliminal advertising, we can understand why involuntary hypnosis would also count as manipulative. But almost no one favors subliminal advertising, and to say the least, the idea of involuntary hypnosis lacks much appeal. The question is whether admittedly taboo practices can shed light on actions that are more familiar or that might be able to command broader support.

Illustrative (and Not Easy) Cases

Consider some cases that test the boundaries of the concept of manipulation.

(1) Suppose that public officials try to persuade people to engage in certain behavior with the help of relative risk information: "If you do not do X, your chances of death from heart disease will triple!" (Wilkinson 2013). (We could imagine a private institution — say, a pharmaceutical company — doing the same thing in a marketing campaign.) But suppose that for the relevant population, the chance of death from heart disease is very small — say, one in 100,000 — and that people are far more influenced by the idea of "tripling the risk" than they would be if they learned that if they do not do X, they could increase a 1/100,000 risk to a 3/100,000 risk (to say the least, a modest increase). The relative risk frame is far more attention-grabbing than the absolute risk frame; a tripling of a risk sounds alarming, but if the increase is by merely 2/100,000, people might not be much concerned. It is certainly reasonable to take the choice of the relative risk frame (which suggests a large impact on health) as an effort to frighten people and thus to manipulate them (at least in a mild sense).

It is true that any description of a risk requires some choices; people who describe risks cannot avoid some kind of framing. But framing is not the same as manipulation. There is a good argument that this particular choice does not sufficiently engage, or show a great deal of respect for, people's deliberative capacities; it might even be an effort to aim specifically at System 1. As we shall see, that conclusion does not mean that the use of the relative risk frame is necessarily out of bounds (Craswell 1991). This is hardly the most egregious case of manipulation, and if it saves a number of lives across a large population, it might be justified. But it can be counted as manipulative.

(2) Suppose that public officials are alert to the power of loss aversion (Zamir 2014), and hence they use the "loss frame," so as to trigger people's concern about the risks associated with obesity and excessive energy consumption. (Here again we could imagine private marketers doing the same thing.) They might deliberately choose to emphasize, in some kind of information campaign, how much people would *lose from not using* energy conservation techniques, rather than how much people would *gain from using* such techniques (Aronson

1996). Is the use of loss aversion, with its predictably large effects, a form of manipulation?

The answer is not obvious, but there is a good argument that it is not, because deliberative capacities remain sufficiently involved. Even with a loss frame, people remain fully capable of assessing overall effects. But it must be acknowledged that the deliberate use of loss aversion might be an effort to trigger the negative feelings that are distinctly associated with losses.

Here too, it is a separate question whether the use of loss aversion raises serious ethical objections. Within the universe of arguably manipulative statements, those that enlist loss aversion hardly count as the most troublesome, and in the case under discussion, the government's objectives are laudable. If the use of loss aversion produces large gains (in terms of health or economic benefits), we would not have much ground for objection. But we can identify cases in which the use of loss aversion is self-interested, and in which the surrounding context makes it a genuine example of manipulation (Willis 2012). Consider, for example, the efforts of banks, in the aftermath of a new regulation from the Federal Reserve Board, to enlist loss aversion to encourage customers to opt in to costly overdraft protection programs by saying, "*Don't lose your ATM and Debit Card Overdraft Protection*" and "*STAY PROTECTED with [...] ATM and Debit Card Overdraft Coverage*" (Willis 2012). In such cases, there is an evident effort to trigger a degree of alarm, and hence it is reasonable to claim that customers are being manipulated, and to their detriment.

(3) Alert to the behavioral science on social influences, a planner might consider the following approaches:

- (a) Inform people that most people in their community *are engaging in undesirable behavior* (drug use, alcohol abuse, delinquent payment of taxes, environmentally harmful acts).
- (b) Inform people that most/many people in their community *are engaging in desirable behavior*.
- (c) Inform people that most/many people in their community *believe that people should engage in certain behavior*.

The first two approaches rely on "descriptive norms," that is, what people actually do (Cialdini 2003). The third approach relies on "injunctive norms," that is, what people think that people should do. As an empirical matter, it turns out that descriptive norms are ordinarily more powerful (Cialdini 2003). If a change in behavior is what is sought, then it is best to emphasize that most/many people actually do the right thing (Cialdini 2003). But if most/many people do the wrong thing, it can be helpful to invoke injunctive norms (Cialdini 2003; Schultz *et al.* 2007).

Suppose that a public official is keenly aware of these findings and uses them. Is he engaging in manipulation? The word "sufficiently" becomes

relevant here as well. Without doing much violence to ordinary language, some people might think it reasonable to conclude that it is manipulative to choose the formulation that will have the largest impact. At least this is so if social influences work as they do because of their impact on the automatic system, and if they bypass deliberative processing (Charpentier *et al.* 2014). But as an empirical matter, this is far from clear; information about what other people do, or what other people think, can be part of reflective deliberation, and hardly opposed to it. So long as the official is being truthful, it would strain the boundaries of the concept to accuse him of manipulation: When they are informed about what most people do, people's powers of deliberation are sufficiently engaged (Brown *et al.* 2013).

(4) Default rules often stick, in part because of the force of inertia, in part because of the power of suggestion (Johnson and Goldstein 2013; Sunstein 2014). Suppose that a public official is aware of that fact and decides to reconsider a series of default rules in order to exploit the stickiness of defaults. Seeking to save money, she might decide in favor of a double-sided default for printers (Egebark and Ekström 2013). Seeking to reduce pollution, she might promote, or even require, a default rule in favor of green energy (Sunstein and Reisch 2014). Seeking to increase savings, she might promote, or even require, automatic enrollment in retirement plans (Benartzi and Thaler 2013).

Are these initiatives manipulative? Insofar as default rules carry an element of suggestion — a kind of informational signal — they are not. Such rules appeal to deliberative capacities insofar as they convey information about what planners think people ought to be doing. The analysis is less straightforward insofar as default rules stick because of inertia: Without making a conscious choice, people end up enrolled in some kind of program or plan. In a sense, the official is exploiting System 1, which is prone to inertia and procrastination (Rebonato 2011). The question is whether automatic enrollment fails “sufficiently” to engage reflection and deliberation.

It is surely relevant that an opt-in default is likely to stick as well, in part because of inertia, which means that the question is whether *any* default rule counts as a form of manipulation. The answer to that question is plain: Life cannot be navigated without default rules, and so long as the official is not hiding or suppressing anything, the choice of one or another should not be characterized as manipulative. Note that people do reject default rules that they genuinely dislike, so long as opt-out is easy — an empirical point in favor of the conclusion that such rules should not be counted as manipulative (Beshears *et al.* 2010; Brown *et al.* 2013).

(5) *A potpourri*. There is no question that much of modern advertising is directed at System 1, with attractive people, bold colors, and distinctive aesthetics. (Consider advertisements for Viagra.) Often the goal is to trigger a distinctive affect and more specifically to enlist the “affect heuristic,” which puts the question of manipulation in stark relief (Slovic 2010). Much of

website design is an effort to trigger attention and to put it in the right places (Krug 2014). Cell phone companies, restaurants, and clothing stores use music and colors in a way that is designed to “frame” products in a distinctive manner. Doctors, friends, and family members (including spouses) sometimes do something quite similar. Is romance an exercise in manipulation? Some of the time, the answer is surely yes, though the question of “sufficiently” raises special challenges in that context (Cave 2014).

Acting as advocates, lawyers may be engaged in manipulation; that is part of their job, certainly in front of a jury. The same can be said about some aspects of the provision of medical care, when doctors want patients to choose particular options, and enlist behaviorally informed technique to manipulate them to do so. Or consider certain uses of social media — as, for example, when Facebook attempted to affect (manipulate) the emotions of 689,003 people through the display of positive or negative stories to see how they affected their moods (Kramer *et al.* 2014). A great deal of conduct, however familiar, can be counted as manipulative in the relevant sense.

What is Wrong with Manipulation, So Understood?

Autonomy and Dignity

(1) *Respect.* The most obvious objection to manipulation is that it can insult both autonomy and dignity. From the standpoint of autonomy, the problem is that manipulation can deprive people of agency, resting on a continuum for which coercion is the endpoint. (If people are manipulated into buying a product, they might feel coerced.) From the standpoint of dignity, the problem is that manipulation can be humiliating. Healthy adults, not suffering from a lack of capacity, should not be tricked; they should be treated as fully capable of making their own decisions. Their authority over their own lives should not be undermined by approaches that treat them as children or as puppets. An act of manipulation does not treat people with respect (Baron 2003).

Suppose, for example, that someone thinks, “I want all my friends to do certain things, and I know a number of strategies to get them to do those things. I have read a great deal of psychology and behavioral science, including the best work on social influence, and my project is to use what I know to manipulate my friends.” Such a person would not be respecting her friends’ autonomy. She would be using them as her instruments. Indeed, her actions would be inconsistent with the nature of friendship itself, which entails a relationship that is not strictly or mostly instrumental.

Now turn to the case of government. Suppose that public officials — say, in a governor’s office — similarly learn a great deal about how to influence

people, and suppose that they decide to use what they learn to achieve certain policy goals. Suppose that some of the relevant instruments attempt to subvert or bypass deliberation. We need to know the details — what, exactly, are they doing? But it could well be fair to say that manipulation is involved, and that public officials are not sufficiently respecting their citizens' autonomy (Sunstein 2015). Again we need to know the details, but it could also be fair to say that such officials are not treating citizens with respect; they might be using them as instruments or as puppets for their own ends (or for the perhaps public-spirited ends that they favor).

(2) *Role (and commerce)*. We should be able to see, in this light, that *role* greatly matters to the assessment of manipulation. Suppose that Jones is trying to obtain a job. It is hardly unacceptable for Jones to attempt to get prospective employers to like him, and if Jones learns about social influence and the operations of System 1, it would hardly be illegitimate for him to take advantage of what he learns. To be sure, there are ethical limits on what Jones can do, and even for someone seeking a job, the most egregious forms of manipulation would cross the line. But in interactions or relationships that are instrumental, and that are so understood, deontological constraints are weakened or at least different.

In an advertising campaign, everyone knows the nature of the interaction. In some ways, manipulation is the coin of the realm. The purpose of advertisements is to sell products, and while we can find purely factual presentations, many advertisements do not appeal to reflection or deliberation at all. They try to create certain moods and associations. To the extent that the enterprise is broadly understood, and to the extent that consumers understand it, the ethical objections are weakened; people can and do discount self-interested efforts at manipulation.

Something similar can be said about some aspects of political campaigns. The relationship between a campaign and voters has an instrumental character: Campaigns want votes, and everyone understands that. In the process, both advertisements and speeches will have manipulative features. But it would be extravagant to say that in such cases, people have consented to manipulation in all imaginable forms. Here too, lines can be crossed, above all through the welfare-reducing exploitation of behavioral biases (Akerlof and Shiller 2015; Bar-Gill 2012). But it is important that people are aware of the distinctive nature of the relevant enterprises.

Other roles are accompanied by different norms, and manipulation might not fit with, or might even violate, those norms. When governments deal with their citizens, they face radically different norms from those that apply in campaigns. At least this is so in free and democratic societies, in which it is understood that the public is ultimately sovereign. To be sure, public officials are hardly forbidden from framing options in a way that casts those they prefer in the most favorable light. But as the manipulative characteristics of

their actions become more extreme, the scope for legitimate objection becomes greater (Goodin 1980).

Welfare

Suppose that we are welfarists, and that we believe that what matters is how people's lives are going (on some of the complexities with that concept, see Sen 1979, 2000; Nussbaum 2011; Adler 2011). Suppose too that we care about violations of autonomy and dignity only insofar as such violations affect people's subjective experiences (for example, by making them feel confined or humiliated). If so, how should we think about manipulation?

It should be clear that there is no taboo on the practice. As we shall see, manipulation might promote people's welfare. But there is a distinctive welfarist objection to manipulation, which takes the following form. As a general rule, choosers know what is in their best interest (at least if they are adults, and if they do not suffer from a problem of capacity⁵). They have unique access to their situations, their constraints, and their tastes (Hayek 2014). If they are manipulated, they are deprived of the (full) ability to make choices on their own, simply because they are not given a fair or adequate chance to weigh all variables. If someone wants to help people to make better choices, his obligation is to inform them, so that they can themselves engage in such weighing.

The problem with the manipulator is that he lacks relevant knowledge — about the chooser's situation, tastes, and values (Hayek 2014). Lacking that knowledge, he nonetheless subverts the process by which choosers make their own decisions about what is best for them. Things are even worse if the manipulator is focused on his own interests rather than on those of choosers. It is in this sense that a self-interested manipulator can be said to be stealing from people — both limiting their agency and moving their resources in the preferred direction.

For these reasons, the welfarist objection to paternalism is rooted in the same concerns that underlie Mill's Harm Principle (1859). In Mill's view, the problem with outsiders, including government officials, is that they lack the necessary information. Mill insists that the individual "is the person most interested in his own well-being," and the "ordinary man or woman has means of knowledge immeasurably surpassing those that can be possessed by any one else" (Mill 1859). When society seeks to overrule the individual's judgment, it

⁵ A child, or a person suffering from some form of dementia, has a weaker objection to manipulation. Parents manipulate young children all the time, partly much to promote their welfare. Caretakers manipulate people who are suffering from dementia. These practices are largely taken for granted, but we could imagine situations in which they would raise serious ethical questions. Even if the relevant manipulation is in the interest of those who are being manipulated, the interests in autonomy and dignity impose some constraints even here.

does so on the basis of “general presumptions,” and these “may be altogether wrong, and even if right, are as likely as not to be misapplied to individual cases” (Mill 1859). These points apply to those engaged in manipulation no less than to those engaged in coercion.

Notwithstanding these points, it should be clear that from the welfarist standpoint, there should be no ban on manipulation.⁶ Everything depends on whether manipulation improves people’s welfare. To see the point, imagine a benign, all-knowing, welfare-promoting manipulator — a kind of idealized parent — who is concerned only with the welfare of those who are being manipulated, who has all the knowledge he needs, and who simply does not make mistakes. By hypothesis, the welfare-promoting manipulator should be celebrated on welfarist grounds. The major qualification is that if people know that they are being manipulated, and do not like it, there will be a welfare loss, and that loss will have to be counted in the overall assessment. The simple point is that if people hate manipulators, manipulation is less likely to be supportable on welfare grounds (unless it is hidden, which raises problems of its own.⁷)

But put that point to one side. The main problem with the thought experiment is that manipulators are unlikely to be either benign or all-knowing. Often they have their own agendas, and the fact that they engage in manipulation attests to that fact. If they are genuinely concerned about the welfare of the chooser, why not try to persuade them? Why cross the line to manipulation? To be sure, the manipulator might be able to answer this question if, for example, time is of the essence, or if the chooser lacks capacity (because, for example, he is a child or very ill). Or suppose that graphic health warnings, aimed directly at System 1, save numerous lives; suppose too that numerous lives cannot be saved with a merely factual presentation unaccompanied by graphic health warnings. On welfarist grounds, a great deal might be said on behalf of graphic health warnings (Jolls 2013).

The example shows that from the standpoint of welfare, everything depends on the context; the fact that manipulation is involved does not *necessarily* impeach the manipulator’s welfare calculus. But in many situations, suspicion about manipulators’ goals is perfectly justified. To this point it must be added that even when those goals are admirable, manipulators may not know enough to justify their actions. Consider Hayek’s (2014) remarkable suggestion that

⁶Note that the question here is whether manipulation increases or decreases welfare; it is not whether the law, or some regulator, should ban manipulation. The latter question raises issues about institutional competence and decision costs. It also requires attention to the effect of manipulation on large populations with heterogeneous understandings. In response to an advertising campaign, for example, some people might be manipulated (in the sense that System 1 is essentially all that is affected) while others are not (because the campaign triggers a significant amount of deliberation).

⁷Hidden manipulation is risky, because it might be disclosed, and people will not be happy to learn that it has been hidden.

“the awareness of our irremediable ignorance of most of what is known to somebody [who is a planner] is *the chief basis of the argument for liberty*” (emphasis added). I will return to this point, and the question of regulation, in the section “Manipulation, freedom of speech and consumer protection.

We should emphasize that insofar as people who sell products are engaged in manipulation, the welfarist objection runs into even more complicated territory. Recall the importance of role. Insofar as consumers understand that sellers are in the business of selling, and do not necessarily have consumers’ interests in mind, we might think that efforts at manipulation are less harmful, or perhaps not harmful at all. Efforts at manipulation might not work; they might be self-defeating. In any case, market forces impose discipline on what sellers can do or say, and under optimistic assumptions, welfare-reducing manipulation will be sharply limited. For standard economic reasons, involving incomplete and asymmetric information, those assumptions are unrealistically optimistic, and an understanding of behavioral biases compounds the problem (Akerlof and Shiller 2015; Bar-Gill 2012). But acknowledging those points, we should accept a degree of manipulation on the part of sellers. The question is when the appropriate line is crossed — a question that takes us into the domain of cost–benefit analysis.

Consent, Transparency, Democracy

Manipulation With Consent: “I Welcome It!”

Suppose that people consent to manipulation.⁸ An alcoholic might tell his wife: “I am trying so hard to quit. Please use whatever techniques you can think of to help me. Manipulation is very much on the table. I welcome it!” Or suppose that the overwhelming majority of smokers tell their government: “I want to stop! If you can find a way to help me to overcome my addiction, I would be grateful.” Wilkinson (2013) notes that it is too crude to say that manipulation infringes upon autonomy, because “manipulation could be consented to. If it were consented to, in the right kind of way, then the manipulation would at least be consistent with autonomy and might count as enhancing it.”

The conclusion has a great deal of force. We can understand consent as suggesting support from System 2, which might welcome a little manipulation (or possibly a lot) as a way of cabining the adverse effects of System 1. The tale of Ulysses and the Sirens is instructive here, whether Ulysses was requesting manipulation or something else (Elster 1984). Nor is there an objection,

⁸On one view, the concept of manipulation presupposes a lack of consent (Goodin 1980). But the examples given in text suggest that manipulation can be a product of consent and even invitation.

in the case of consent, from the standpoint of welfare. The chooser has decided that he will be better off if he is manipulated. If we see his choice as presumptively promoting his welfare, we should respect it, even if what he chose is manipulation.

In the easiest cases, consent is explicit. In harder cases, it is only implicit, in the sense that the manipulator infers it from the circumstances or believes, with good reason, that the chooser would consent if asked. If the inference is less than reliable, the consent justification is correspondingly weakened. If the belief is reasonable but potentially wrong, it might make sense to insist on obtaining explicit consent in order to avoid the risk of error. It is important to see that consensual manipulation is an unusual case; those who need help do not ordinarily ask, "Please manipulate me." But such cases certainly do exist, at least when people face serious problems of self-control.

A qualification: I have emphasized the importance of role and noted that in the domain of commerce, consumers generally understand that sellers want them to buy their products. But we should not take that general understanding as implicit consent to the most egregious forms of manipulation. In free markets, people are hardly asking to be manipulated.

Transparency and Manipulation

The idea of manipulation is sometimes taken to imply a lack of transparency, as if something important is being hidden or not being disclosed (Goodin 1980). If a manipulator is acting as a puppeteer, he might be failing to reveal his own role; that can be an important feature of manipulation. With respect to manipulation, however, it is not entirely clear what transparency even means (Hansen and Jespersen 2013).

Sometimes manipulation itself *consists* in a lack of transparency about a relevant feature of a situation; that is the very manipulation involved. (A parent tells a small child: If you are very good, Santa Claus will bring you a toy giraffe.) Once the relevant feature is revealed, the manipulation is gone. But the manipulative character of some acts does not consist in their hidden quality. (Recall the use of relative risk reduction.) Some acts can be both manipulative and fully revealed to those who are being manipulated (Felsen *et al.* 2013; Loewenstein *et al.* 2014).⁹ A graphic health warning, for example, is perfectly transparent (and if it is required by regulation, it is even likely to be preceded by a period for public comment, as was the case for the FDA regulation invalidated by *R. J. Reynolds Tobacco Co. v. FDA* (2012)). Subliminal advertising

⁹On the fact that manipulation can be transparent rather than covert, Barnhill (2014) urges: "I think that deceptiveness or covertness is a favorite technique of manipulators — manipulation is more likely to succeed if its target doesn't realize what's happening. But manipulation needn't be covert. Covertness isn't what's definitive of manipulation."

could be preceded by an explicit warning: “This movie contains subliminal advertising.” If so, it would still count as a form of manipulation.

In the pivotal scene in *The Wizard of Oz*, the Wizard says, “Pay no attention to the man behind the curtain.” The man behind the curtain is of course a mere human being who is masquerading as the great Wizard — and who is both claiming far more authority than he deserves and designing social situations in a way that hides features that, if revealed, would greatly alter people’s judgments and choices. Or consider a less celebrated movie, *The Truman Show*, in which the life course of Truman, the movie’s hero-protagonist, is affected by multiple decisions from a master manipulator, who conceals the facts that Truman is the unwitting star of a television show and that his friends and acquaintances consist of actors. Covert and hiding are common features of manipulation. Whenever people who are imposing influence conceal their own role, it seems reasonable to object. A lack of transparency offends both autonomy and dignity.

From the standpoint of welfare, we might ask why, exactly, someone has failed to be upfront with the chooser, who ought to be able to make his own decisions, armed with relevant information. As before, however, the analysis of welfare is more complex, for we could imagine cases in which transparency is not necessary and may in fact be a problem. Suppose that someone suffers from a serious self-control problem and that his life is in danger (from, say, alcohol or drug addiction). Suppose too that a manipulator has come across a life-saving strategy and that transparency would render the manipulation less effective. By hypothesis, welfarist considerations argue against transparency. Points of this kind have the strongest intuitive force when applied to people who lack capacity (young children, the mentally ill), but we can imagine contexts in which adults with full capacity would benefit from being manipulated (Only 2012).

There might also be a welfarist justification for hidden manipulation in other extreme circumstances — as, for example, when people are trying to stop a kidnapping or to save a kidnapping victim. If the goal is to stop a wrongdoer, or someone who threatens to do real harm, it may be perfectly acceptable or even obligatory to manipulate them and to hide that fact. They have forfeited their right to be treated with respect, and their welfare, as choosers, is not a matter of much concern.

In standard cases, however, this argument will be unavailable. It follows that most of the time, manipulation should not be hidden or covert, even when it is justified; return to the case of graphic health warnings. Transparency is a necessary condition. Note, however, that it is not sufficient (Hansen and Jespersen 2013). Subliminal advertising would not become acceptable merely because people were informed about it. If a movie chain announced that its previews would be filled with subliminal advertisements, people could fairly object.

Democratically Authorized Manipulation

What if manipulation is democratically authorized? Suppose that a national legislature expressly votes for it, perhaps in order to improve public health (as, for example, by discouraging smoking or unhealthy eating), or perhaps to promote other goals (such as enlistment in the military or adoption of the currently preferred ideology). In relatively benign cases, involving little or no manipulation, a legislature might support an educational campaign that is designed to reduce illnesses and deaths and that enlists a series of behaviorally informed strategies, targeting System 1, to accomplish its goals.

It should be clear that democratic authorization ought not by itself to dissolve otherwise reasonable objections to manipulation. The most obvious problems arise if the national legislature has illegitimate ends (say, the inculcation of racial prejudice or self-entrenchment of some kind). But the familiar objections — involving autonomy, dignity, and welfare — apply even if the ends are legitimate. If a national legislature authorizes subliminal advertising, it remains fully possible to object on grounds of both autonomy and dignity. An objection from the standpoint of welfare is also possible: Why did the democratic process authorize manipulation, rather than some other kind of communication?

To be sure, we could understand democratic authorization as a form of majority or collective consent, suggesting support from System 2, which might welcome a little manipulation (or possibly a lot) as a way of cabining the adverse effects of System 1. In general, however, there are evident risks in authorizing public officials to pursue this line of argument. The objection to manipulation comes from individuals, who do not want to be manipulated; the fact that a majority wants to manipulate them is no defense.

But in certain contexts, the argument on behalf of at least a modest degree of manipulation might not be implausible, and it is strengthened if the democratic process has supported it. Imagine, for example, a public education campaign that is designed to reduce the risks associated with texting while driving, or an effort to combat the use of dangerous drugs or to convince people to stay in school. Many such campaigns are vivid and have an emotional component; they can be understood as efforts to combat self-control problems and to focus people on the long term. (To be sure, educational campaigns of this kind might not be counted as manipulation at all, because deliberative capacities are sufficiently engaged.)

If government is targeting System 1 — perhaps through framing, perhaps through emotionally evocative appeals — it might be responding to the fact that System 1 has already been targeted, and to people's detriment. In the context of cigarettes, for example, it is plausible to say that previous manipulations — including advertising and social norms — have influenced people to become smokers. If this is so, perhaps we can say that public officials

are permitted to meet fire with fire. But some people might insist that two wrongs do not make a right — and that if the government seeks to lead people to quit, it must treat them as adults, and appeal to their deliberative capacities.

Recall that there are degrees of manipulation, and there is a large difference between a lie and an effort to frame an alternative in an appealing, unappealing, or ugly light. In ordinary life, we would not be likely to accuse our friends or loved ones of manipulation if they characterized one approach as favored by most members of our peer group, or if they emphasized the losses that might accompany an alternative that they abhor, or if they accompanied a description of one option with a grave look and a frown. These are at most mild forms of manipulation, to be sure, and it is important to see that mild forms might well be acceptable and benign (and a bit fun) if they promote the interests of those people at whom they are aimed. No legal system has a general tort called “exploitation of cognitive biases.”

Unifying Strands

If the various arguments are put together, we might be able to evaluate acts of manipulation with the help of the matrix shown in Table 1.

The matrix helps to orient the appropriate response to manipulation from the standpoint of ethics, politics, and law, and indeed it captures widespread intuitions. In the bottom right cell, we can find actions by self-interested or venal public officials both in undemocratic systems and occasionally in democratic ones as well (Goodin 1980). In the top right, we can find foolish or venal statements or actions by private and public actors that do not entirely bypass people’s deliberative capacities, but that hardly do justice to them. Many government agencies, and many ordinary companies, act in accordance with the top left cell; they portray their behavior in an appealing light, and they try to attract favorable attention, but the particular form of manipulation

Table 1: Matrix for evaluating acts of manipulation.

	Benign and informed	Malign or uninformed
Not highly manipulative	Acceptable on welfare grounds; might be acceptable by reference to autonomy or dignity	Unacceptable
Highly manipulative	Acceptable on welfare grounds; objectionable on grounds of autonomy and dignity	Highly unacceptable

is hardly egregious. Some governments, some of the time, act in a way that fits in the bottom left cell, perhaps especially with graphic campaigns.

The matrix also provides a start toward an analysis of how the legal system should respond to manipulation. From the welfarist standpoint, the central question is whether the benefits of restricting the manipulative action or statement justify the costs. To answer that question, we need to know what would happen if the action or statement were not made (or were transformed into a nonmanipulative version). In this respect, the analysis of manipulation closely parallels the analysis of deception (Craswell 1985). The costs of manipulation depend, in large part, on whether the manipulator is malign or uninformed. To the extent that it is, there is a risk of serious welfare losses. But suppose that an advertiser is part of a well-functioning competitive process, and that its advertisement includes a degree of manipulation in order to sell a product. If the competitive process is genuinely well-functioning, consumers are not likely to lose much, and market pressures may discipline the use and the effectiveness of manipulation (Glaeser 2006).

The question, then, is whether some kind of market failure exists, so that manipulative behavior can persist or be rewarded. In light of information problems and behavioral biases, the answer is likely to be affirmative in some markets, especially those that involve complex financial products (Akerlof and Shiller 2015; Bar-Gill 2012). Of course consumers will have diverse understandings of, and reactions to, statements and actions that plausibly fall within the category of manipulation (Craswell 1985). Empirical testing of representative populations could provide highly informative here. The fact of heterogeneous understandings will create serious challenges for regulators seeking to prevent arguably harmful forms of manipulation. As we shall see, however, some manipulative acts are so plainly welfare-reducing that it makes sense to restrict them.

Manipulation, Freedom of Speech, and Consumer Protection

Under established doctrine in the United States, government can regulate threats (Watts, v. United States 1969); it can also regulate false or misleading commercial speech (Virginia State Pharmacy Board v. Virginia Citizens Consumer Council 1976) and certain forms of coercive speech. May it also regulate manipulation? It is also clear that government can compel certain kinds of speech (Nunes 2014). May it compel speech that is arguably manipulative?

Compelled Speech

As a testing case, consider the efforts, in the United States, of the Food and Drug Administration (FDA) to require cigarette packages to contain graphic

health warnings. The court of appeals invalidated the requirement on free speech grounds, concluding that the FDA lacked sufficient evidence to justify the compelled speech (*R. J. Reynolds Tobacco Co. v. FDA* 2012). In so ruling, the court did not emphasize the arguably manipulative nature of the graphic warnings. But the lower court opinion did exactly that (*R. J. Reynolds Tobacco Co. v. FDA* 2011).

The court found it relevant that “the Rule’s graphic-image requirements are *not* the type of purely factual and uncontroversial disclosures that are reviewable under this less stringent standard” (*R. J. Reynolds Tobacco Co. v. FDA* 2011). It added, plausibly, that “it is abundantly clear from viewing these images that the emotional response they were crafted to induce is calculated to provoke the viewer to quit, or never to start, smoking: An objective wholly apart from disseminating purely factual and uncontroversial information” (*R. J. Reynolds Tobacco Co. v. FDA* 2011). The court concluded that when the government compels speech that does not involve the “purely factual and uncontroversial,” it has to meet a higher burden of justification.

As a matter of law, the central idea here lacks support in Supreme Court decisions, but it has some appeal: The free speech principle imposes particular barriers to government efforts to require speech that does not merely appeal to deliberative or reflective capacities, but that engages and attempts to activate System 1. On this view, there is no firm rule against compelling manipulative speech of that kind (so long as it is not false or deceptive), but if government is engaging in such compulsion, it must have a strong justification for doing so.

This analysis raises an assortment of issues. Do the graphic warnings count as manipulative? They are certainly designed to create a visceral response (and they do exactly that). But the question is whether they do not sufficiently engage or appeal to people’s capacity for reflective and deliberative choice. Here the answer has to come from specifying the idea of “sufficiently.” There is an empirical component to the specification: What, exactly, do people understand after they see the warnings? Suppose that for a large part of the population, understanding is actually improved. If so, there is a good argument that manipulation is not involved (Jolls 2013). But suppose that understanding is not improved. Can the warnings nonetheless be justified?

We could imagine two such justifications. The first is welfarist: Graphic health warnings will save a significant number of lives, and purely factual information will have a far weaker effect. If this is so, then the graphic health warnings do have a sufficient justification. The second is rooted in autonomy: Smokers, and prospective smokers, do not sufficiently appreciate the health risks of smoking, and graphic warnings can promote a kind of “debiasing” that statistical information fails to provide (Jolls 2013). To this point, it might be added that government regulation is hardly being imposed on a blank slate. Recall that efforts to promote smoking involve a high degree of manipulation — portraying happy, attractive smokers — and the government can legitimately

respond. In light of the number of lives at risk and the underlying evidence, these kinds of justifications do seem sufficient in the particular context of smoking.

Regulating Manipulation

Should government regulate manipulation? In the context of political speech? Commercial advertising? We could imagine, or find, egregious cases in which it is tempting to say that it should. But the free speech barriers are severe.

(1) *Political speech and public figures.* In the context of political speech, the leading case in the United States is *Hustler Magazine v. Falwell* (1988), where the Court ruled that the first amendment protects a parody, depicting Protestant minister Jerry Falwell as engaged in an incestuous act with his mother at an outhouse. The parody was satirical; it can also be seen as a form of manipulation, designed to lead readers to see Falwell as a ridiculous figure and also a hypocrite. In the terms used here, the parody was an effort to appeal directly to System 1, so that people would not be able to regard Falwell in the same light in the future.

The Court unanimously ruled that the first amendment protected the parody. The Court acknowledged that to prevent genuine harm, states could regulate false statements of fact, which “are particularly valueless; they interfere with the truth-seeking function of the marketplace of ideas, and they cause damage to an individual’s reputation that cannot easily be repaired by counterspeech, however persuasive or effective” (*Hustler Magazine v. Falwell* 1988). But satire must be treated differently. “Were we to hold otherwise, there can be little doubt that political cartoonists and satirists would be subjected to damages awards without any showing that their work falsely defamed its subject” (*Hustler Magazine v. Falwell* 1988). Even the most outrageous forms of satire are protected, because the idea of outrageousness has “an inherent subjectiveness about it which would allow a jury to impose liability on the basis of the jurors’ tastes or views, or perhaps on the basis of their dislike of a particular expression” (*Hustler Magazine v. Falwell* 1988).

To be sure, the Court’s reasoning was not unbounded. It acknowledged that the “actual malice” standard — allowing recovery of damages for statements known to be false or made with reckless indifference to the question of truth or falsity — would apply if readers had taken the parody as depicting actual facts (*New York Times v. Sullivan* 1964). But readers could not so take the parody here. This point leaves open the possibility that even in the political domain, certain forms of manipulation could be regulated if readers or viewers were affirmatively misled. But in view of the fact that the Court has pointedly declined to create a *general* exception to the free speech principle even for false statements of fact (*United States v. Alvarez* 2012), any effort to regulate

manipulative speech in the political context would run into severe trouble (*Whitney v. California* 1927).

With the design of any restrictions on such speech, there are independent questions of vagueness and overbreadth. If a government wants to prohibit the most egregious forms of manipulation in the political context, what, exactly, would it say? I have ventured a definition of manipulation here, but it is not exactly easy to adapt that definition to fit a provision of civil or criminal law. Manipulation has too many shades — which means that any effort to restrict it would likely be too vague and too broad.

(2) *Commercial speech*. The context of commercial advertising is different, because the burden on regulators is lighter. But at least in the United States, the free speech obstacles are formidable in the United States as well. So long as the relevant speech is neither false nor deceptive, the government would need a powerful justification for imposing regulation (*Central Hudson Gas & Electric Corp. v. Public Service Commission* 1980). The definitional issues remain severe, and even if they could be resolved, it would be plausible to say, in the general spirit of *Hustler Magazine*, that the marketplace of ideas is full of efforts to appeal to System 1, and to downplay or bypass deliberation and reflection.

Even if the commercial sphere is less immune from speech regulation, it is emphatically a place where manipulation is pervasive. The hope is that consumers will understand that advertisements are generally self-serving and that the process of competition will provide a sufficient corrective. To be sure, behavioral economics has raised serious questions about the realism of that hope (Bar-Gill 2012). But it is highly doubtful that those questions provide a sufficient basis for a *general* “manipulation exception” to the existing protection accorded to commercial speech.

Consumer Protection

None of these conclusions mean that narrower forms of regulation could not be imagined. In the context of consumer financial products, various forms of manipulation are a widespread problem. Indeed, manipulation can be seen as a defining motivation for recent regulatory initiatives (Bar-Gill 2012; Bubb and Pildes 2014). The Dodd-Frank Wall Street Reform and Consumer Protection Act states that the CFPB should ensure that “markets for consumer financial products and services are fair, transparent, and competitive” (12 U.S.C. §5511 2010). It calls for attention not only to “unfair and deceptive” acts and practices but also to “abusive” ones, which can be seen as a reference to the worst forms of manipulation. In monitoring the relevant markets, the CFPB must consider the “understanding by consumers of the risks of a type of consumer financial product or service” — a phrase that can easily be taken to reflect a concern about manipulation (12 U.S.C. §5512 2010).

Implementing these requirements, the CFPB has adopted as its slogan, “know before you owe,” and its various efforts to ensure informed choices can be understood as an attack on manipulation as I have understood it here (CFPB 2014). In consumer markets, of course, one problem is complexity, which can defeat understanding (Bar-Gill 2012). But another problem falls in the general category of manipulation, as in the form of “teaser rates” and various inducements that fall short of deceit, but that emphatically prey on System 1 (Bubb and Pildes 2014).

A short, simple credit card agreement, of the sort provided by the CFPB (CFPB 2014), can be seen as a direct response to the risk of manipulation — and as an effort to ensure that System 2 is firmly in charge. Proposals to ban or restrict teaser rates can be understood in similar terms (Bar-Gill and Bubb 2012). In cases of this kind, there is ample room for considering the problem of manipulation in deciding how best to regulate financial products. It is important to see that in such cases, government is regulating commercial practices, not advertising, and that its real concern is with practices that do not sufficiently trigger reflective deliberation on the part of consumers. We have seen that this is far from a self-defining category, but the CFPB’s initiatives can be taken as initial efforts to specify it.

Conclusion

A statement or action can be counted manipulative to the extent that it does not sufficiently engage or appeal to people’s capacity for reflective and deliberative choice. Some forms of manipulation are egregious, as where a vivid, graphic description of an outcome (winning the lottery, dying in an airplane crash, losing a child) is invoked in order to convince people to engage in certain conduct (to buy a lottery ticket, to take a train, to buy extra life insurance). Some arguable forms of manipulation are mild, as when a politician, an employer, or a waiter uses loss aversion, tone of voice, and facial expressions to encourage certain decisions. Thus defined, manipulation is a pervasive feature of human life. It is for this reason that while the legal system is generally able to handle lies and deception, it has a much harder time in targeting manipulation.

In their troublesome forms, manipulative acts fail to respect choosers; they undermine people’s autonomy and do not respect their dignity. The welfarist objection, rooted in the idea that choosers know what is in their best interests, is that when they are products of manipulation, people’s choices may not promote their own welfare, precisely because choosers have not been put in a position to deliberate about relevant variables and values. This is likely to be true if the manipulator is ill-motivated, but it might also be true because the manipulator lacks relevant information.

From the welfarist point of view, manipulation is only presumptively disfavored. A benign, knowledgeable manipulator could make people's lives go better and possibly much better. But under realistic assumptions, the presumption against manipulation is justifiably strong, because manipulators are unlikely to be either benign or knowledgeable. It is important to acknowledge that in the commercial realm, manipulation is widespread; it is part of the basic enterprise. For that reason, the ethical taboo on manipulation is substantially weakened, in part on the theory that competitive markets impose appropriate constraints against undue harm. But in some cases, those constraints are too weak, and it is appropriate to invoke social norms or even the law to discipline welfare-reducing acts of manipulation.

References

- Adler, Matthew (2011), *Well-Being and Fair Distribution*, Oxford University Press.
- Akerlof, George and Robert Shiller (2015), *Phishing for Phools*, Princeton University Press.
- Aronson, Elliot (1996), *The Social Animal*, 6th Edition, Worth Publishers.
- Bachrach, Susan (2009), *State of Deception: The Power of Nazi Propaganda*, U.S. Holocaust Memorial Museum.
- Bar-Gill, Oren (2012), *Seduction By Contract*, Oxford University Press.
- Bar-Gill, Oren and Ryan Bubbs (2012), "Credit Card Pricing: The Card Act and Beyond," *Cornell Law Review*, 97, 967–1018.
- Barnhill, Anne (2014), "What is Manipulation?" In *Manipulation: Theory and Practice*, ed. Christian Coons and Michael Weber, Oxford University Press, 51–72.
- Baron, Marcia (2003), "Manipulativeness," *Proceedings and Addresses of the American Philosophical Association*, 77(2), 37–54.
- Benartzi, Shlomo and Richard H. Thaler (2013), "Behavioral Economics and the Retirement Savings Crisis," *Science*, 339(6124), 1152–3.
- Beshears, John, James J. Choi, David Laibson, and Brigitte C. Madrian (2010), "The Limitations of Defaults," <http://www.nber.org/programs/ag/rrc/NB10-02%20Beshears,%20Choi%20et%20al%20SUMMARY.pdf>.
- Bread Crust Council (2011), "Daisy, 1964 Lyndon Johnson ad," <https://www.youtube.com/watch?v=uO0R4k1tVMs>.
- Brown, Zack, Nick Johnstone, Ivan Haščič, Laura Vong, and Francis Barascud (2013), "Testing the Effects of Defaults on the Thermostat Settings of OECD Employees," *Energy Economics*, 39, 128–34.
- Bubbs, Ryan and Richard H. Pildes (2014), "How Behavioral Economics Trims Its Sails and Why," *Harvard Law Review*, 127, 1593–677.

- Bullock, August (2004), *The Secret Sales Pitch: An Overview of Subliminal Advertising*, Norwich Publications.
- Cave, Eric M. (2014), "Unsavory Seduction and Manipulation," in *Manipulation: Theory and Practice*, ed. Christian Coons and Michael Weber, Oxford University Press, 176–200.
- Central Hudson Gas & Electric Corp. v. Public Service Commission (1980), 447 US 557.
- Charpentier, Caroline J., Christina Moutsiana, Neil Garrett, and Tari Sharot (2014), "The Brain's Temporal Dynamics from a Collective Decision to Individual Action," *Journal of Neuroscience*, 34(17), 5816–23.
- Cialdini, Robert B. (2003), "Crafting Normative Messages to Protect the Environment," *Current Directions in Psychological Science*, 12(4), 105–9.
- Conly, Sarah (2012), *Against Autonomy*, Cambridge University Press.
- Consumer Financial Protection Bureau (CFPB) (2013), "Know Before You Owe: Credit Cards," <http://www.consumerfinance.gov/credit-cards/knowbeforeyouowe/>.
- Consumer Financial Protection Bureau (CFPB) (2014), "12 CFR Parts 1024 and 1026: Integrated Mortgage Disclosures under the Real Estate Settlement Procedures Act (Regulation X) and the Truth in Lending Act (Regulation Z)," RIN 3170-AA19, http://files.consumerfinance.gov/f/201311_cfpb_final-rule_integrated-mortgage-disclosures.pdf.
- Craswell, Richard (1985), "Interpreting Deceptive Advertising," *Boston University Law Review*, 65, 657–732.
- Craswell, Richard (1991), "Regulating Deceptive Advertising: The Role of Cost-Benefit Analysis," *Southern California Law Review*, 64, 549–604.
- Department of Transportation (2011), "U.S. Department of Transportation Releases New 'Faces of Distracted Driving' Video," Press Release, <https://www.transportation.gov/briefing-room/us-department-transportation-releases-new-%C3%A2%C2%80%C2%9Cfaces-distracted-driving%C3%A2%C2%80%C2%9D-video-1>.
- Egebark, Johan and Mathias Ekström (2013), "Can Indifference Make the World Greener?" <http://blogg.nhh.no/thechoicelab/wp-content/uploads/2014/03/EgebarkEkstr%C3%83%C2%B6m2014.pdf>.
- Elster, Jon (1984), *Ulysses and the Sirens*, Cambridge University Press.
- Faden, Ruth and Tom Beauchamp (1986), *A History and Theory of Informed Consent*, Oxford University Press.
- Felsen, Gidon, Noah Castelo, and Peter B. Reiner (2013), "Decisional Enhancement and Autonomy: Public Attitudes Toward Overt and Covert Nudges," *Judgment and Decision Making*, 8(3), 202–13.
- Glaeser, Edward L. (2006), "Paternalism and Psychology," *University of Chicago Law Review*, 73(1), 133–56.
- Goodin, Robert E. (1980), *Manipulatory Politics*, Yale University Press.

- Hansen, Pelle Guldberg and Andreas Maaløe Jespersen (2013), "Nudge and the Manipulation of Choice," *The European Journal of Risk Regulation*, 4(1), 3–28.
- Hayek, Friedrich A. (2014), *The Market and Other Orders*, ed. Bruce Caldwell, University of Chicago Press.
- Hustler Magazine v. Falwell (1988), 485 US 45.
- IMDB (2015), "Mad Men: The Wheel Quotes," <http://www.imdb.com/title/tt1105057/quotes>.
- Johnson, Eric J. and Daniel G. Goldstein (2013), "Decisions by Default," in *The Behavioral Foundations of Public Policy*, ed. Eldaer Shafir, Princeton University Press.
- Jolls, Christine (2013), "Product Warnings, Debiasing, and Free Speech: The Case of Tobacco Regulation," *Journal of Institutional and Theoretical Economics*, 169(1), 53–78.
- Kahneman, Daniel (2011), *Thinking, Fast and Slow*, Farrar, Straus and Giroux.
- Keren, Gideon, ed. (2010), *Perspectives on Framing*, Psychology Press.
- Kramer, Adam D. I., Jamie E. Guillory, and Jeffrey T. Hancock (2014), "Experimental Evidence of Massive-Scale Emotional Contagion Through Social Networks," *Proceedings of the National Academy of Sciences*, 111(24), 8788–90.
- Krug, Steve (2014), *Don't Make Me Think Revisited: A Common Sense Approach to Web and Mobile Usability*, Newriders.
- Loewenstein, George, Cindy Bryce, David Hagmann, and Sachin Rajpal (2014), "Warning: You Are About To Be Nudged," http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2417383.
- Mill, John Stuart (1859), On Liberty.
- New York Times v. Sullivan (1964), 376 US 254.
- Noussair, Charles, Stéphane Robin, and Bernard Ruffieux (2004), "Do Consumers Really Refuse to Buy Genetically Modified Food?" *Economic Journal*, 114(492), 102–20.
- Nunes, Bianca (2014), "Note: The Future of Government-Mandated Health Warnings," *University of Pennsylvania Law Review*, 163, 177–213.
- Nussbaum, Martha (2003), *Upheavals of Thought*, Cambridge University Press.
- Nussbaum, Martha. (2011), *Creating Capabilities*, Harvard University Press.
- Picard, Pierre (2013), "Economic Analysis of Insurance Fraud," in *Handbook of Insurance*, ed. Georges Dionne, 349–95.
- R. J. Reynolds Tobacco Co. v. FDA (2011), 823 F.Supp.2d 36 (D.D.C. 2011).
- R. J. Reynolds Tobacco Co. v. FDA (2012), 696 F.3d 1205 (DC Cir. 2012).
- Raz, Joseph (1986), *The Morality of Freedom*, Oxford University Press.
- Rebonato, Riccardo (2011), *Taking Liberties: A Critique of Libertarian Paternalism*, Palgrave.

- Schultz, P. Wesley, Jenny M. Nolan, Robert B. Cialdini, Noah J. Goldstein, and Vidas Griskevicius (2007), "The Constructive, Destructive, and Reconstructive Power of Social Norms," *Psychological Science*, 18(5), 429–34.
- Securities and Exchange Commission (SEC) (2010), "17 CFR Part 240: Prohibition Against Fraud, Manipulation and Deception in Connection with Security-Based Swaps Proposed Rule," RIN 3235-AK77, <http://www.sec.gov/rules/proposed/2010/34-63236.pdf>.
- Sen, Amartya (1979), "Utilitarianism and Welfarism," *Journal of Philosophy*, 76(9), 463–89.
- Sen, Amartya (2000), *Development As Freedom*, Anchor.
- Slovic, Paul, ed. (2010), *The Feeling of Risk: New Perspectives on Risk Perception*, Routledge.
- Smith, Adam C. and Todd Graziano (2014), "Behavior, Paternalism, and Policy: Evaluating Consumer Financial Protection," http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2408083.
- Sunstein, Cass R. (2014), "Choosing Not To Choose," *Duke Law Journal*, 64, 1–52.
- Sunstein, Cass R. (2015), "Nudges and Choice Architecture: Ethical Considerations," *Yale Journal of Regulation*, forthcoming.
- Sunstein, Cass R. and Lucia Reisch (2014), "Automatically Green: Behavioral Economics and Environmental Protection," *Harvard Environmental Law Review*, 38, 127–58.
- Ullmann-Margalit, Edna (2006), "Big Decisions: Opting, Converting, Drifting," *Royal Institution of Philosophy Supplement*, 58, 157–72.
- United States Code (U.S.C.) (2010a), "12 U.S.C. §5511 Purpose, Objectives, and Functions," <http://uscode.house.gov/view.xhtml?req=granuleid:USC-prelim-title12-section5511&num=0&edition=prelim>.
- United States Code (U.S.C.) (2010b), "12 U.S.C. §5512 Rulemaking Authority," <http://uscode.house.gov/view.xhtml?req=granuleid:USC-prelim-title12-section5512&num=0&edition=prelim>.
- United States v. Alvarez (2012), 567 US.
- Virginia State Pharmacy Board v. Virginia Citizens Consumer Council (1976), 425 US 748.
- Waldron, Jeremy (2014), "It's All For Your Own Good," *New York Review of Books*, <http://www.nybooks.com/articles/archives/2014/oct/09/cass-sunstein-its-all-your-own-good/>.
- Watts, v. United States (1969), 394 US 705.
- Westerman, Frank (2012), *Engineers of the Soul: The Grandiose Propaganda of Stalin's Russia*, Overlook Books.
- Whitney v. California (1927), 274 US 357.
- Wilkinson, T. M. (2013), "Nudging and Manipulation," *Political Studies*, 61(2), 341–55.

- Willis, Lauren E. (2012), "When Defaults Fail: Slippery Defaults," *University of Chicago Law Review*, 80(3), 1155–229.
- Wilson, Timothy D. (2014), *Strangers to Ourselves*, Harvard University Press.
- Wood, Allen W. (2014), "Coercion, Manipulation, Exploitation," in *Manipulation: Theory and Practice*, ed. Christian Coons and Michael Weber, Oxford University Press, 17–50.
- Zamir, Eyal (2014), *Law, Psychology, and Morality: The Role of Loss Aversion*, Oxford University Press.
- Zibel, Alan (2015), "CFPB Sets Sights on Payday Loans," *Wall Street Journal*, <http://www.wsj.com/articles/cfpb-sets-sights-on-payday-loans-1420410479>.