

ISSN 1936-5349 (print)
ISSN 1936-5357 (online)

HARVARD

JOHN M. OLIN CENTER FOR LAW, ECONOMICS, AND BUSINESS

MANIPULATION, WELFARE, AND DIGNITY: A REPLY

Cass R. Sunstein

Forthcoming in *Journal of Behavioral Marketing*

Discussion Paper No. 856

02/2016

Harvard Law School
Cambridge, MA 02138

This paper can be downloaded without charge from:

The Harvard John M. Olin Discussion Paper Series:
http://www.law.harvard.edu/programs/olin_center/

The Social Science Research Network Electronic Paper Collection:
<http://ssrn.com/abstract=2661534>

Manipulation, Welfare, and Dignity: A Reply

Cass R. Sunstein*

*This essay responds to six commentaries on my forthcoming essay, *Fifty Shades of Manipulation*. It offers two general points. The first involves the importance of separating three questions: (1) What is manipulation? (2) What is wrong with manipulation? (3) When might manipulation be justified, notwithstanding the answer to (2)? The second involves the relevance of dignity. We might see dignity as a component of welfare, or we might see it as a wholly independent value. But we will not understand manipulation, or what is wrong with it, if we do not see it at all.*

The topic of manipulation raises three separate questions: (1) What is it? (2) What is wrong with it? (3) When might it be justified, notwithstanding the answer to (2)?

In my essay in this issue,¹ I suggest that a statement or action counts as manipulative *when it does not sufficiently engage or appeal to people's capacity for reflective and deliberative choice*. Subliminal advertising is an unambiguous example. Marketing strategies that appeal to people's deepest fears, or that trigger identifiable biases (such as present bias, availability bias, or optimistic bias), might also be counted as manipulative. (To know for sure, we would need to consider the details.) Manipulation is not the same as deception. Manipulators trick you; deceivers lie to you. Your mother or your lover might manipulate you without deceiving you, and the same is true of a marketing campaign for aspirin or for cold medicines – or for a head of state.

Are nudges manipulative²? Most of them are not. Certainly the kinds of nudges that have attracted official support – in the United States, the United Kingdom, Germany, and elsewhere – do not meet the definition.³ Information disclosure, reminders, and warnings are hardly manipulative. Simplification of communications, so that people can more easily understand what is being said, is at the opposite pole from manipulation. When officials ease regulatory requirements,

* Robert Walmsley University Professor, Harvard University.

¹ Cass R. Sunstein, *Fifty Shades of Manipulation*, *J Behav Marketing* (2015, forthcoming).

² On the general idea, see Richard H. Thaler and Cass R. Sunstein, *Nudge* (2008).

³ For a catalogue, see David Halpern, *Inside the Nudge Unit* (2015).

or promote *navigability* (a singularly important goal), they are not manipulating anyone.

Nonetheless, some imaginable nudges can fairly be treated as forms of manipulation. We could easily imagine a public information campaign, designed (for example) to reduce smoking or to promote healthy eating, that would cross the line. Some marketers are engaged in manipulation. For them, manipulation is the coin of their realm. They do not want people to deliberate or reflect. They want to trigger a quick, affective reaction that will get consumers to buy a good or service.

If manipulation is understood in these terms, one reason it is wrong is that it is not respectful of choosers. It offends their autonomy and their dignity. People should be treated as adults, not as children, and if private or public institutions do not sufficiently engage people's capacity for reflective choice, they are treating them without respect. From the standpoint of welfare, the concern is that manipulators might be promoting their own interests, rather than those of the chooser. Why – it might reasonably be asked – is it necessary to resort to manipulation, rather than to provide information and to engage the chooser's reflective capacities?

If we follow Mill and Hayek, and believe that individuals are in the best position to make choices for themselves, then we might fear that any effort to manipulate people will reduce their welfare, because it does not sufficiently allow them to make such choices. The obligation of the adviser, or the choice architect, is to put people in the position to make good choices (by their own lights) -- not to manipulate them. The reason – a follower of Mill and Hayek might claim -- is that choosers have epistemic advantages over outsiders. On this view, manipulators threaten to reduce people's welfare because they find ways to ensure that choosers will choose in the way that manipulators want.

It is true that the force of the welfarist objections to manipulation depends on the context and the circumstances. Recognizing that fact, welfarists should emphasize the importance of two factors: social roles and market competition. Advertisers and political candidates have particular social roles, and most people know what they are. If they engage in manipulative behavior of certain kinds, the ethical objections might be weakened by virtue of a general understanding that their roles call for it. Perhaps the process of competition works well enough.⁴ With respect to government, a welfarist defender of manipulation might contend that public officials are entitled to meet fire with fire. If cigarette companies or sellers of unhealthy foods are manipulating people to buy their products, perhaps official manipulation might be justified as a form of “debiasing.” And if choosers are making serious or systematic errors, manipulation might be the best way of getting them on track. Perhaps Mill and Hayek are wrong, or at least too categorical, in their claims about the comparative advantages of choosers.

⁴ Or perhaps not. See George Akerlof and Robert Shiller, *Phishing for Phools* (2015).

These points help explain when and how manipulation might be justified. *First*, welfarists would distinguish between beneficial and harmful manipulation. It is fine to manipulate terrorists and kidnappers. In ordinary life, welfarists are not likely to object if friends or family members manipulate people in order to save their lives or to prevent grievous harm, or even to have a little fun. *Second*, manipulation might be part of an overall process or system that is unobjectionable. If parents manipulate their children, perhaps we have no cause for complaint; among friends or spouses, a little manipulation, and possibly a lot, can be fine (though lines can be crossed, depending partly on the consequences). If the system of competition is working well, the same might be said about marketers of many different kinds.⁵ At least relatively mild forms of manipulation might be an inevitable, and fully acceptable, part of competition itself. *Third*, nonwelfarists should acknowledge that some forms of manipulation are trivial or unimportant, and hardly a serious ethical wrong – and that if the consequences of manipulation are good enough, the presumption against it can be overcome.

I am most grateful to the various commentators for their exceptionally valuable essays, which greatly illuminate the topic of manipulation. In this response, I make two general points. The first is the importance of cleanly separating the three questions with which I began. For example, the best definition of manipulation includes a great deal of conduct that can ultimately be justified, and the fact that it can be justified should not deprive us of the privilege of calling it what it is. The second is the importance of dignity. We might see dignity as a component of welfare, or we might see it as a wholly independent value. But we will not understand manipulation, or what is wrong with it, if we do not see it at all.

Eldar Shafir's engaging paper covers a great deal of territory, but it has a unifying theme: Human beings are being manipulated all the time, not least in the marketplace; shouldn't public officials manipulate them for their own good? To motivate that question, he points to a number of contexts in which apparently small and arguably extraneous influences greatly affect behavior. On rainy days, people receive lower ratings in interviews; if a waitress touches people on the shoulder, they tip more; asking people whether they intend to purchase a car makes them more likely to purchase a car; after a lunch break, judges become significantly more lenient. In this light, Shafir thinks that we have a misplaced sense of our own autonomy and dignity. Our decisions "will always be potentially manipulated, by someone or something." Manipulation turns out to be "an integral part of our lives. Why not make the best of it?"

Writing as a welfarist, Shafir urges that we should "do what we can to bring about the best possible outcomes." He is troubled that as of now, the well-intentioned actors respect people's autonomy and dignity, while the less well-intentioned go about their merry way, selling people dangerous products and

⁵ For a sustained argument that the system of competition is not a sufficient safeguard here, see George Akerlof and Robert Shiller, *Phishing For Phools* (2015).

impossible dreams. Those who market unhealthy foods to children are only one prominent example. Shafir favors restrictions on manipulative marketing, but he also wants public officials, experts and nonprofits “to do their own manipulating,” not least with “scary or heart-wrenching stories,” exhibiting their own ability to learn from the advertising industry.

Shafir makes some excellent points about human psychology, but it is important to distinguish between the very broad category of social influence and the narrower one of manipulation. If your doctor gives you purely factual warnings, he is not manipulating you. If public officials disseminate information about healthy eating, or remind you of programs from which you might benefit, they are not manipulating you. Under my definition of manipulation, individuals and institutions have a choice whether to manipulate people. When manipulation occurs, its objects can legitimately ask: *Why did you manipulate me, rather than treating me with dignity?* Perhaps manipulators can answer that question. They might be able to say: *Manipulation saved your life.* But there are good reasons, both welfarist and nonwelfarist, to favor influences that fall short of manipulation – certainly for public officials. A state that felt free to manipulate people, even for ends that should be deemed legitimate and desirable, would not be acting respectfully.

Using the dictionary, Eric Posner begins with this definition of manipulation: “to control or play upon by artful, unfair, or insidious means, especially to one’s advantage.” He narrows that definition to target cases in which “the manipulator perceives that his addressee brings incorrect assumptions to a transaction and does not correct them, or else anticipates and takes advantage of people’s propensity to make incorrect inferences.” When the manipulator is insincere, and benefits himself at the expense of the addressee, Posner contends that we have cases of “exploitation.” And when exploitation is occurring, Posner contends that courts do in fact police it. For example, they will not uphold contractual provisions that disclaim legal responsibility in small print, and thus lead consumers to maintain assumptions that those provisions are expressly negating. Even without fraud or coercion, Posner says that courts will intervene against manipulators who exploit incorrect assumptions that their victims bring to transactions. Here, then, is Posner’s positive claim: Courts police harmful manipulation.

In Posner’s view, the economic analysis of manipulation is essentially the same as the corresponding analysis of fraud and coercion. In the latter cases, the government can deter forced transfers more cheaply than individuals can, so it is efficient for it to do so. The same holds for manipulation: “Sellers should inform buyers of the terms of the contract so that buyers don’t need to hire lawyers to protect them from buying things that they don’t want.” On grounds of efficiency (or welfare), Posner thinks that there is a solid argument for an anti-manipulation principle, which forbids people from obtaining resources “by intentionally saying or doing things that cause the other person to misunderstand the nature of the transaction.”

As a good welfarist, Posner thinks that unless the victim or addressee is harmed, intentional exploitation of behavioral biases is largely unobjectionable. (Note that for those who are committed to autonomy or dignity, this claim misses something of real importance.) If the government uses relative risk information, there is no problem if that method of framing is a better way to communicate the problem, in light of how people think and their antecedent assumptions. Sounding very much like Shafir, Posner urges that “government should feel free to use manipulation so long as the manipulative act advances aggregate welfare or achieves other publicly endorsed goals, whatever they may be.”

Posner is right to say that in some cases, common law courts have policed actions that fall within his definition of manipulation, and that might not be seen as either deceitful or coercive. But his definition seems to sweep up a great deal of conduct, from advertisers and marketers of all kinds, with which the legal system has no problem. Have a look at a random advertisement on television, or a website, and ask whether “the manipulator perceives that his addressee brings incorrect assumptions to a transaction and does not correct them, or else anticipates and takes advantage of people’s propensity to make incorrect inferences.” The answer might well be yes. There is a good argument that advertisements for potato chips, cars, credit cards, beer, chocolate bars, organic food, and exercise clubs run afoul of Posner’s anti-manipulation principle.

Perhaps regulators should do much more to police manipulation, defined as Posner defines it.⁶ But I think that my definition more accurately captures the ordinary meaning of the term. And as for Shafir, so too for Posner: When we are speaking of public officials, there is a good argument for a presumption against manipulation. It is true, of course, that welfare-increasing manipulation will increase welfare, and for welfarists, that would seem to be the end of the matter. But if we do not always trust public officials, and if we think that choosers (often) have epistemic advantages over those who seek to help them, we might insist that such officials will often do better to inform people, and to increase their own agency, rather than to manipulate them. Importantly, however, none of these points is a decisive argument against graphic health warnings or graphic educational campaigns, which may not count as manipulative at all, which might increase people’s understanding, and which might accomplish a great deal of good.

An enthusiastic utilitarian, Baron’s argument overlaps with both Shafir’s and Posner’s. He wants to focus on harmful manipulation, which he defines as “deception with the intent of leading people to choose options that they might regret, or not choose, on the basis of actively open-minded reflection, if they had full information about the nature and intent of the effort to influence their choices.” Baron spotlights manipulation that is hidden, harmful, and intentional.

⁶ See George Akerlof and Robert Shiller, *Phishing for Phools* (2015).

It is useful to emphasize the relevance of “actively open-minded reflection,” and from the welfarist point of view, harmful manipulation is certainly the source of concern. Baron is not at all troubled by some of the cases that I discuss, such as use of relative risk information, loss aversion, descriptive norms, and default rules. He sees these as instances of beneficial manipulation, designed to help choosers.

Of course Baron may ultimately be right on this count, but it is important, I think, to distinguish between deception and manipulation. If a seller of a new asthma drug says during an advertisement, “I know someone who used this drug and soon after, his cancer went into remission,” he is deceiving people. If the seller runs an advertisement for the same drug displaying athletic, glamorous, sexy people playing tennis, deception may not be involved, but we might well be entitled to speak of manipulation. Marketers who deceive are usually worse than marketers who manipulate. Candidates for public office often manipulate people, for example through music and graphics, and we might worry about that, but deception is a more serious matter. It is also useful to separate, more sharply than Baron does, the three issues with which I began: the definition of manipulation, the account of what is wrong with it, the circumstances in which it might be justified.

Gatignon and Le Nagard are concerned that for purposes of thinking about marketing, my definition of manipulation sweeps up far too much. They want to confine manipulation to cases of indirect, covert influence, through the use of deception or abusive tactics. Subliminal advertising would fall within the category. The same is true for lies and deception, which “go directly against the objective of customer long-term satisfaction.” And of course “abusive” is “an extremely strong qualifier” -- which means that selective information, appeals to emotions, uses of music and smells, and loyalty programs cannot be counted as manipulative. Gatignon and Le Nagard think that to avoid treating these standard marketing practices as such, “a more precise definition” than my own is required.

It is not clear that their preferred definition is more precise, but it is certainly narrower. For example, Gatignon and Le Nagard would never, or almost never, consider emotional appeals to be manipulative. But why is that a good idea? At first glance, we should define manipulation in a way that fits with ordinary language, which counts many statements or actions as manipulative even if they do not involve deception or abuse. (A mother, elderly but in good health, asks an adult son to visit, saying, “You know, I might not live much longer.”) Many marketers manipulate people, even if they do not deceive or abuse them. Doctors and teachers do the same. If a nation were to propose a new law, forbidding manipulation, my definition would of course be too vague and too broad. But if the goal is to make progress in understanding the concept of manipulation, even within marketing, Gatignon and Le Nagard seem to exclude too much.

Gatignon and Le Nagard are also concerned about the definition of marketing. In ordinary language, marketing seem to refer to efforts to persuade people to purchase good or services. But Gatignon and Le Nagard are not happy

with that definition, or for that matter with the (hilariously self-serving) 1935 definition of the American Marketing Association: “the performance of business activities that direct the flow of goods and services from producers to consumers.” Gatignon and Le Nagard prefer what the same organization offered in 2013 (less hilariously but also amusingly): “the activity, set of institutions, and processes for creating, communicating, delivering, and exchanging offerings that have value for customers, clients, partners, and society at large.” But their own definition is different and even more upbeat: “The marketing concept is a view of exchanges realized through practices intended to create sustained value equitably for a set of involved parties (individuals, firms or organizations.” For “marketing management,” Gatignon and Le Nagard would go further to emphasize “equitable relationships” and consideration of “long-term consequences” both for involved parties and for society.

Is it unfair to wonder whether some definitions of marketing are themselves forms of marketing? Suppose that a tobacco company tries to sell cigarettes, with the unambiguous goal of persuading millions of people to buy its product. Suppose that it is not at all concerned with creating “sustained value equitably.” More simply, it wants to sell cigarettes. Is it not engaged in marketing? That would be a surprising conclusion. It is laudable to attempt to introduce ethical standards into marketing, but there is a question whether a definition of marketing is the best way to achieve that goal. Gatignon and Le Nagard seem to me to offer unduly narrow definitions of both manipulation (to exclude much marketing) and marketing (to incorporate ethical principles).

Anne Barnhill’s superb essay urges that my definition of manipulation is not too broad but too narrow. Many of our attitudes and actions, she urges, are *properly nondeliberative*. You might simply like a drink (it tastes great), or enjoy an apple (it’s so red!), or want to spend time with a friend (she’s a lot of fun). As Barnhill notes, social influences often occur without engaging people in reflection and deliberation. Influencing people in that way need not be manipulative if reflection and deliberation are not called for. But if someone provides bad inputs in such settings, manipulation is nonetheless involved. Thus Barnhill urges, contrary to my argument, that people can be manipulated in situations in which responses are properly nondeliberative. On Barnhill’s view, “sometimes we manipulate people by providing ‘bad inputs’ into non-deliberative processes.”

Consider commercial advertising. There is no problem if you just like a product – say, a soft drink, because it tastes great. But if you like it because of subliminal advertising, you have been manipulated; your properly nondeliberative response has been distorted. Barnhill offers a more general point, which is that when advertising is manipulative, it is sometimes because it is providing “distorting inputs” into a properly nondeliberative process. Her broadest claim is that we should find manipulation whenever people are caused “to fall short of ideals for belief, desire and emotion.” That can happen even if deliberation is not involved.

I think that Barnhill is convincing here, and that a full account of manipulation must take account of the category of properly nondeliberative responses. Her point has particular relevance to marketing, where “distorting inputs” are a legitimate source of concern. One question is the precise domain of the properly nondeliberative. It is fine to say that you just like Coke Zero, Taylor Swift’s 1983, or a day on the beach in Marblehead, Massachusetts. To like those things, you need not reflect or deliberate. Moreover, there is no question that you can be manipulated into liking those things. But suppose the question is whether you should buy a can of Coke Zero, buy 1983, or book a room at Hotel Preston. For such actions, a degree of deliberation can be expected, and an advertiser might engage in manipulation, in my sense rather than Barnhill’s, in subverting that process. Nonetheless, Barnhill is right to say that the “liking” can itself be a product of manipulation.

Joseph Alba and Yanmei Zheng are interested in the relationship between deliberation on the one hand and welfare on the other. They produce an illuminating typology of relationships between the two, resulting in four sets of cases that deserve attention here. The first is straightforward: Information, or some other kind of educative nudge, improves deliberation and increases welfare (while also promoting dignity). The second is also straightforward: Someone exploits System 1 and reduces welfare, for example by leading people to make decisions that they would not make if their decisions were more deliberative. The first category includes some of the most attractive nudges, and the second includes some of the most egregious forms of manipulation (and essentially what I have in mind).

But Alba and Zheng’s third and fourth categories are more unsettling. The third category includes interventions that suppress deliberation, but in a way that increases people’s welfare. Suppose that in deliberating, people give insufficient weight to certain attributes, or give a lot of weight to irrelevant information. In such cases, *less deliberation means more welfare*. The fourth category is the mirror-image: interventions that reduce welfare but increase deliberation. With additional options or new information, people might think harder, and have more with which to think, but they might be more likely to blunder. Clever manipulators might well promote their own ends by increasing people’s thinking, not by reducing it. Self-interested choice architects might increase complexity, or introduce irrelevant options, to skew people’s thinking in their preferred directions. Alba and Zheng argue that in light of “the uncertain relationships among information, deliberation, and optimal decision-making,” we should not support “any universal effort to promote deliberation.”

This is a convincing argument. Most simply, Alba and Zheng show that manipulation, understood as an effort to undermine deliberative processes, can promote people’s welfare. Shafir and Baron would certainly agree, and almost all doctors, lawyers, and parents are aware of that fact. But there are serious cautionary notes. Humility is a virtue, and doctors, lawyers, and even parents ought to consider whether non-manipulative methods might be better (as well as more respectful).

Alba and Zheng's more complicated point is that efforts to increase deliberation might end up reducing people's welfare by impairing the quality of their decisions. Suppose that people (marketers, for example) make such efforts, not because they want to help choosers, but in order to produce decisions that promote their own self-interested goals. We can easily imagine insurance and mortgage companies, or those who sell cars, engaging in an effort to increase complexity and deliberation, not to improve consumer welfare but to add to profits. Should we characterize that effort as manipulative? I think that we should. In cases of this kind, someone is trying to trick choosers, by altering the context of choice in a way that increases the burdens of decision so as to produce an inferior judgment.

It would be possible, and perhaps correct, to suggest that my definition needs to be altered or expanded to pick up cases of this kind. But perhaps we can include such cases within that definition, as a distinctive and unusual subcategory of situations in which an agent is not "sufficiently" engaging people's capacity for reflective and deliberative choice. Overload of the deliberative process does not count as sufficient engagement (though it may be more natural to describe it as excessive engagement). In other words, the capacity to deliberate and reflect is not sufficiently engaged simply because people are devoting a lot of time to thinking, or struggling hard with a problem. If the goal of the manipulator (and let's call him that) is to increase such struggles in order to skew people's decisions, we can fairly say that manipulation is involved.

Let me conclude with some words about the relationship between manipulation and dignity. If we care about welfare, we will pay a great deal of attention to dignity: When people feel that they have been treated disrespectfully, or as mere children, they will rebel. Objects of manipulation often have a sense of ex post betrayal. And within political philosophy, it is standard to see a serious offense to dignity as an independent wrong, one that requires, at a minimum, a compelling justification. If public officials have manipulated their citizens, there will be hell to pay, and rightly so.

It is not customary to introduce considerations of human dignity into the analysis of marketing, where manipulation is hardly uncommon, and which is more standardly approached in welfarist terms. But if marketers are known to treat people disrespectfully, their own efforts may be counterproductive. And though manipulation has at least fifty shades, and is therefore difficult to police, there are strong arguments for a regulatory response to the most egregious forms of manipulation, which involve something very close to theft. The worst manipulators are thieves. Among other things, theft is an affront to dignity.