

# 1

## DEFINING MANIPULATION

In 2012, Facebook’s Data Science team decided to conduct an experiment to see whether and how they could affect the emotions of the platform’s users.<sup>1</sup> For a week, they altered the feeds of 689,003 Facebook users. Some of the users were exposed to fewer posts with positive words or phrases; some of the users saw fewer posts with negative words or phrases. The question was simple: Would this intervention change the kinds of emotions that Facebook users displayed online? In other words, would the platform see *emotional contagion*?

The answer was clear. Facebook did indeed have power over the emotions of its users. People who were exposed to fewer negative posts ended up making fewer negative posts. People who were exposed to fewer positive posts made fewer positive posts. As the authors put it, “emotions expressed by others on Facebook influence our own emotions, constituting experimental evidence for massive-scale contagion via social networks.”

The experiment produced widespread alarm, because people immediately saw the implication: If a social media platform wanted to manipulate the feelings of its users, it could do exactly that. A platform could probably make people feel sadder or angrier, or instead happier or more forgiving.

<sup>1</sup> Adam D. I. Kramer et al., Experimental Evidence of Massive-Scale Emotional Contagion Through Social Networks, 111 *PNAS* 8788 (2014), available at [www.pnas.org/doi/10.1073/pnas.1320040111](http://www.pnas.org/doi/10.1073/pnas.1320040111).

4 / Manipulation

\*\*\*

It ranks among the most powerful scenes in the history of television. Don Draper, the charismatic star of *Mad Men*, is charged with producing an advertising campaign for Kodak, which has just invented a new slide projector, with continuous viewing. It operates like a wheel. Using the device to display scenes from a once-happy family (as it happens, his own, which is now broken), Draper tells his potential clients:<sup>2</sup>

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 the 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.<sup>3</sup>

The Kodak clients are sold. They cancel their meetings with other companies.

\*\*\*

In 2018, Cambridge Analytica obtained the personal data of about 87 million Facebook users. It did so by using its app called, “This Is Your Digital Life.” The app asked users a set of questions designed to learn something about their personalities. By using the app, people unwittingly gave Cambridge Analytica not only answers to those questions but also permission to obtain access to millions of independent data points, based on their use of the Internet. In other words, those who answered the relevant questions were taken to have “agreed” to allow Cambridge Analytica to track their online behavior. Far more broadly, they gave the company access to a large number of data points involving the online behavior of all of the users’ friends on Facebook.

<sup>2</sup> See *Mad Men*, Quotes, IMDB, [www.imdb.com/title/tt1105057/quotes](http://www.imdb.com/title/tt1105057/quotes).

<sup>3</sup> Revealingly, nostalgia actually means “longing for a return home,” rather than “pain from an old wound.”

5 / Defining Manipulation

With these data points, Cambridge Analytica believed that it had the capacity to engage in “psychological targeting.” It used people’s online behavior to develop psychological profiles, and then sought to influence their behavior, their attitudes, and their emotions through psychologically informed interventions.

Through those interventions, Cambridge Analytica thought that it could affect people’s political choices. According to a former employee of the firm, the goal of the effort was to use what was known about people to build “models to exploit what we knew about them and target their inner demons.”<sup>4</sup> Pause over that, if you would. We could easily imagine a similar effort to affect people’s consumption choices.

\*\*\*

In 2024, while writing this book, I received an email from “Research Awards.” The email announced that one of my recent publications “has been provisionally selected for the ‘Best Researcher Award.’” It asked me to click on a link to “submit my profile.” It was signed: “Organizing Committee, YST Awards.”

The article, by the way, was on the topic of manipulation. Does the Organizing Committee have a sense of humor?

I did not click on the link.

\*\*\*

To understand manipulation, we need to know something about human psychology – about how people judge and decide, and about how we depart from perfect rationality. Consider the following cases:

1. A social media platform uses artificial intelligence (AI) to target its users. The AI learns quickly what videos its users are most likely to click on – whether the videos involve tennis, shoes, climate change, immigration, the latest conspiracy theory, the newest laptops, or certain kinds of sex. The platform’s goal is to maximize engagement. Its AI is able to create a personalized

<sup>4</sup> Anne Barnhill, How Philosophy Might Contribute to the Practical Ethics of Online Manipulation, in *The Philosophy of Online Manipulation* 49, 50–1 (Fleur Jongepier and Michael Klink eds. 2022).

## 6 / Manipulation

feed for each user. Some users, including many teenagers, seem to become addicted.

2. A real estate company presents information so as to encourage potential tenants to focus only on monthly price and not on other aspects of the deal, including the costs of electricity service, water, repairs, and monthly upkeep. The latter costs are very high.
3. A parent tries to convince an adult child to visit him in a remote area in California, 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?”
4. A hotel near a beach advertises its rooms as costing “just \$200 a night!” It does not add that guests must also pay a “resort fee,” a “cleaning fee,” and a “beach fee,” which add up to an additional \$90 per night. (There are also taxes.)
5. 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.
6. In a campaign advertisement, a political candidate displays ugly photographs of his opponent, set against the background of terrifying 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.
7. 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.
8. 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.

Which of these are manipulative? And why?

## Coercion

To answer these questions, we need to make some distinctions. Let us begin with coercion, understood to involve the threatened or actual application of force. If the law forbids you to

7 / Defining Manipulation

buy certain medicines without a prescription, and if you would face penalties if you violate that law, coercion is in play. If the police will get involved if you are not wearing your seatbelt, we are speaking of coercion, not manipulation. If you are told that you will be fined or imprisoned if you do not get automobile insurance, you are being coerced.

Note that it is false to say that when coercion is involved, people “have no alternative but to comply.” People do have alternatives. They can pay a fine, or go to jail, or attempt to flee.

It is often said that government has a monopoly on the legitimate use of force, which means that in a certain sense, coercion is the province of government and law. But we might want to understand coercion a bit more broadly. If your employer tells you that you will be fired if you do not work on Saturdays, we can fairly say that in an important respect, you have been coerced. If a thief says, “your money or your life,” we can fairly say that you have been coerced to give up your money. A reference to “coercion,” in cases of this kind, is unobjectionable so long as we know what we are talking about.

Coercion is usually *transparent* and *blunt*. No trickery is involved, and no one is being fooled. Coercion typically depends on transparency.

What is wrong with coercion? That is, of course, a large question. Often coercion is unobjectionable. If you are forbidden from murdering people, or from stealing from them, you have been coerced, but there is no reasonable ground for complaint. People do not have a right to kill or to steal. If you are forbidden to speak freely, or to pray as you think best, we can say that the prohibition offends a fundamental right.

Paternalistic coercion raises its own questions. Suppose that people are forced to buckle their seatbelts, wear motorcycle helmets, or eat healthier foods. Many people object to paternalistic coercion on the ground that choosers know best about what is good for them – about what fits with their preferences, their desires, and their values. Even if they do not know best, they can learn from their own errors. Maybe people have a right to err; maybe they are entitled to be the authors of the narratives of their own lives. These points bear on what manipulation is and why it is wrong; I will return to them.

## Lies

Lies may not coerce anyone. “This product will cure baldness” says a marketer, who knows full well that the product will do nothing of the kind. But what is a lie? According to Arnold Isenberg, summarizing many efforts, “A lie is a statement made by one who does not believe it with the intention that someone else shall be led to believe it.”<sup>5</sup> Isenberg adds: “The essential parts of the lie, according to our definition, are three. (1) A statement – and we may or may not wish to divide this again into two parts, a proposition and an utterance. (2) A disbelief or a lack of belief on the part of the speaker. (3) An intention on the part of the speaker.”

The definition is helpfully specific and narrow. Ordinarily we understand liars to know that what they are saying is false and to be attempting to get others to believe the falsehood. According to a similar account by Thomas Carson, “a lie is a deliberate false statement that the speaker warrants to be true.”<sup>6</sup>

Here is a vivid example from John Rawls, speaking of his experience during World War II:<sup>7</sup>

One day a Lutheran Pastor came up and during his service gave a brief sermon in which he said that God aimed our bullets at the Japanese while God protected us from theirs. I don’t know why this made me so angry, but it certainly did. I upbraided the Pastor (who was a First Lieutenant) for saying what I assumed he knew perfectly well – Lutheran that he was – were simply falsehoods without divine providence. What reason could he possibly have for his trying to comfort the troops? Christian doctrine should not be used for that, though I knew perfectly well it was.

I think I know why the Pastor’s sermon made Rawls so angry. The Pastor was not treating the troops respectfully. Actually he was treating them with contempt. Even worse, he was using Christian doctrine in bad faith. He was using what he said to be

<sup>5</sup> See Arnold Isenberg, Deontology and the Ethics of Lying, 24 *Philosophy and Phenomenological Research* 463, 466 (1964).

<sup>6</sup> See Thomas L. Carson, *Lying and Deception* 15 (2010).

<sup>7</sup> John Rawls, *A Brief Inquiry into the Meaning of Sin and Faith* (2009).

9 / Defining Manipulation

God's will, but did not believe to be God's will, in order to make people feel better. That is a form of fraud. It is horrific. It is a desecration.

Inadvertent falsehoods belong in an altogether different category. If you say, "climate change is not real," and if that is what you think, you are not lying, even if what you say is false. You might be reckless or you might be negligent, but you are not lying.

It is noteworthy that the standard definition of lying does not include false statements from people with various cognitive and emotional problems, who may sincerely believe what they are saying. In the 1960s, my father had a little construction company and he worked near a mental hospital. He was often visited by someone who lived there, who would give my father a large check and say, "Mr. Sunstein, here's a check for you. I have enough funds to cover it. Will this be enough for the day?"

It is fair to say that while my father's visitor was not telling the truth, he was not exactly lying. We would not respond to my father's benefactor with the same anger that Rawls felt toward the Pastor.

Or consider the case of confabulators, defined as people with memory disorders who fill in gaps with falsehoods, not knowing that they are false. Nor does the definition include people who believe what they say because of motivated reasoning. People often believe what they want to believe, even if it is untrue, and people whose beliefs are motivated may not be liars. They might be motivated to say that they are spectacularly successful, and while that might not be true, they might not be lying, because they believe what they say. They might say, "My company is bound to make huge profits in the next year," not because there is any evidence to support their optimistic prediction, but because they wish the statement to be true. Such people might be spreading falsehoods, but if they do not know that what they are spreading is false, it does not seem right to describe them as "lying."

## Deception

Now turn to deception. It is often said that the term is broader than "lying." On a standard definition, it refers to *intentionally causing other people to hold false beliefs*. You might deceive

people into holding a false belief without making a false statement. You might sell a new “natural pain medication” to willing buyers, referring truthfully to the testimony of “dozens of satisfied customers,” even though there is no evidence that the pain medication does anyone any good. You might sell an airline ticket to a beautiful location, stating that the ticket costs “just \$199,” without mentioning that the ticket comes with an assortment of extra fees, leading to a total cost of \$299. You might sell your house to a naïve buyer, emphasizing that it has not been necessary to make any repairs over the last ten years, even though it is clear that the house will need plenty of repairs over the next six months. In all of these cases, you have deceived people without lying.<sup>8</sup>

What is wrong with lying and deception? (For present purposes, we can group them together.) They are often taken to be illicit forms of influence. Return to Rawls’ Pastor, who treated the soldiers disrespectfully. Much of the time, liars and deceivers are thieves. They try to take things from people (their money, their time, their vote, their sympathy) without their consent.

Still, we have to be careful here. It is widely though not universally<sup>9</sup> agreed that some lies are acceptable or perhaps even mandatory. A few decades ago, my father started to stumble on the tennis court. My mother and I brought him to the hospital for various tests. After conferring with the doctor, my mother came to my father’s hospital room to announce, “I have great news. They didn’t discover anything serious. They will keep you here for another day, out of an excess of caution, but basically, you’re fine!”

When my mother took me downstairs, to drive me back to law school, her face turned ashen. She said, “Your father has a brain tumor, and there’s nothing they can do about it. He has about a year. But I’m not going to tell him, and you’re not, either.”

<sup>8</sup> See Schlomo Cohen, Manipulation and Deception, 96 *Australasian Journal of Philosophy* 483 (2018).

<sup>9</sup> See Carson, *supra* note 6, for a good discussion of why absolutist or near-absolutist positions do not work; Sissela Bok, *Lying: Moral Choice in Public and Private Life* (1999); for a discussion of absolutist or near-absolutist positions with respect to lies and lying; see also, Christine M. Korsgaard, What’s Wrong With Lying?, in *Philosophical Inquiry: Classic and Contemporary Readings* (Jonathan E. Adler and Catherine Z. Elgin eds. 2007), for what is easily taken as a near-absolute ban on lying, on Kantian grounds.



## 11 / Defining Manipulation

Was she wrong to lie to him? I don't know. Was Bill Clinton wrong to lie to the American people about his relationship with Monica Lewinsky? I believe so, but not everyone agrees.<sup>10</sup>

Consider the following propositions:

1. If an armed thief comes to your door and asks you where you keep your money, you are entitled to lie.
2. If a terrorist captures a spy and asks her to give up official secrets, she is under no obligation to tell the truth.
3. If you tell your children that Santa Claus is coming on the night before Christmas, you have not done anything wrong.
4. If you compliment your spouse on his appearance, even though he is not looking especially good, it would be pretty rigid to say that you have violated some ethical stricture.

We should conclude that lies and deception are generally wrong, and I will have something to say about why. But they are not always wrong.

## Meaning and Morality

A great deal of effort has been devoted to the definition of manipulation, almost exclusively within the philosophical literature.<sup>11</sup> Most of those efforts are both instructive and honorable, and I will build on them here. In the end, it might be doubted that a single definition will exhaust the territory. I will be offering two quite different accounts (one in this chapter and one in Chapter 6), and there are others.

Let us begin with some methodological remarks. On one account, defining manipulation is an altogether different enterprise from explaining why and when manipulation is wrong. We might say that manipulation occurs in certain situations and then make an independent judgment about when or whether it is wrong. We might conclude that it is presumptively wrong, or almost always wrong, because of what it means or does. Or we might think that it is often

<sup>10</sup> See Thomas Nagel, *Concealment and Exposure*, 27 *Phil & Pub Affairs* 3 (1998).

<sup>11</sup> An excellent overview is Christian Coons and Michael Webster eds. *Manipulation: Theory and Practice* (2014). Also superb is Robert Noggle, *The Ethics of Manipulation* (2018); *Stanford Encyclopedia of Philosophy*, available at <https://plato.stanford.edu/entries/ethics-manipulation/>.

12 / Manipulation

justified – if, for example, it is necessary to increase public safety or improve public health.

On another account, manipulation is a “thick” or “moralized” concept, in the sense that it always carries with it a normative evaluation. Compare the words “generous,” “brave,” “kind,” and “cruel”; any definition of these words must be accompanied with a positive or negative evaluation. To say that an act is “generous” is to say that it is good. You might think that to say that an act is “manipulative” is to say that it is bad.

In my view, the first account is right, and it is the right way to proceed. A manipulative act might be good. You might manipulate your spouse in order to reduce serious health risks that she faces. A doctor might manipulate a patient for the same reason. True, you might say, in such cases, that manipulation is *pro tanto* wrong, or presumptively wrong, and that like coercion, deception, or lying, a *pro tanto* wrong or a presumptive wrong might ultimately be right (and perhaps morally mandatory). But we can, I think, define manipulation before offering a judgment about when and whether it is right or wrong. In any case, that is the strategy that I will be pursuing here. And my own preferred definitions in this chapter, intended not to be exhaustive but to capture a set of important cases of manipulation, will turn out to embed something like an account of why manipulation is presumptively wrong.

The task of definition presents other puzzles. When we insist that “deception” means this, and that “manipulation” means that, what exactly are we saying? Are we trying to capture people’s intuitions? Are we trying to capture ordinary usage? What if intuitions or usages diverge? Are we seeking to capture the views of the majority? If so, are we asking an *empirical* question, to be resolved through empirical methods? What would those methods look like? Can artificial intelligence help? More broadly: When people (including philosophers) disagree about what manipulation means, what exactly are they disagreeing about?

I suggest that when we try to define a term like manipulation, we are engaging in something like interpretation in Ronald Dworkin’s sense.<sup>12</sup> In Dworkin’s account, interpretation involves

<sup>12</sup> See Ronald Dworkin, *Law’s Empire* (1985).