Manipulation, Deception, the Victim’s Reasoning, and Her Evidence

Abstract: This paper rejects an argument defending the view that the boundary between deception and manipulation is such that some manipulations intended to cause false beliefs count as non-deceptive. On the strongest version of this argument, if a specific behaviour involves compromising the victim’s reasoning, then the behaviour is manipulative but not deceptive and, if it involves exposing the victim to misleading evidence that justifies her false belief, then it is deceptive but not manipulative. This argument has been consistently used as a reason to reject the traditional analysis of human deception, according to which intentionally causing someone to acquire a false belief is sufficient for deception. And because the traditional analysis is also consistent with our most basic intuitions about deception, it does matter whether this argument succeeds.

Keywords: manipulation; deception; reasoning; misinformation.

# Introduction

Understanding manipulation correctly is important not only per sebut also because our analysis of manipulation may affect our analyses of other phenomena, such as deception – and, in this paper, I focus on human deception. And the connection between manipulation and deception is extremely interesting. It has been thought that manipulation is a kind of deception, or that it must be covert, but things have changed. We now have good reasons to think that manipulation need not be covert (e.g., Klenk 2021) and perhaps even to understand deception as a kind of manipulation (Krstić and Saville 2019). However, the predominant voice in philosophy is that these are two distinct phenomena and some important philosophers – e.g., Cohen (2018) and Fallis (2022) – even argue that a sharp distinction can be drawn between them: some kinds of manipulation can never involve deception and vice versa. This view is very interesting and deserves more attention.

The guiding idea behind this distinction is that there is a feature of manipulation, x, that can never be found in deception (call this ‘hypothesis x’), or that there is a feature of deception, y, that can never be found in manipulation (‘hypothesis y’). Accordingly, if we find x in certain behaviour, we will know that it involves manipulation but not deception and, if we find y in some other behaviour, this is deception and not manipulation. And because intentionally causing someone to end up with a false belief may involve x and deception cannot involve x, manipulations that are intended to cause false belief but involve x will be non-deceptive. And because deception must involve y, misleading manipulations that do not involve y will be non-deceptive.

This is a very important argument because it redraws the boundary between deception and not-deception in a rather radical way: it entails that both our most basic intuition and the relevant ‘traditional’ analysis of human deception, which count intentionally causing false beliefs in others as deception, are false.[[1]](#footnote-1) Scholars disagree on a number of issues regarding the nature of human deception, but pretty much everyone will agree that when (a) I believe that *p* is false and *p* is false, (b) I intend to cause you to believe that *p* is true, and (c) I succeed in doing this, I count as deceiving you into believing that *p*. Therefore, the issue of whether there really is a feature x of manipulation or a feature y of deception the presence or absence of which necessarily makes a specific manipulation intended to cause a false belief non-deceptive needs to be carefully scrutinised.

In this paper, I argue that we should not distinguish deception from non-deceptive manipulation in this way and my argument is rather simple. I first identify the strongest versions of hypotheses x and y then argue that they fail. I conclude that we cannot use x or y to distinguish deception from manipulation.

# 2. Manipulation, Evidence, and Reasoning

In order to distinguish manipulation from deception, Shlomo Cohen (2018: 486) writes that manipulation should be understood as intentionally influencing someone in a predictable direction by means that compromise good judgment (feature x); manipulation, that is, interferes with the judgment’s *form*. And this interference should sensibly be understood as including taking advantage of the victim’s suboptimal judgment or some other suboptimal mental state (see below) since this also compromises reasoning. Deception, in contrast, does not involve compromising the victim’s judgement but it must involve feeding it with false premises (feature y); deception interferes with the judgement’s *content*. In short, according to Cohen, the feature x is interfering with the judgement’s form and the feature y is interfering with the judgment’s content. Intentionally causing the victim to acquire a false belief by compromising her good judgement, then, always counts as engaging in non-deceptive manipulation rather than deception – since we have x (‘form’) and deception cannot involve x.

Trying to refine Cohen’s view, Don Fallis offers his *evidence-based* definition of deception, according to which A deceivesB if and only if A intentionally causes B to have a false belief that *p* by exposingB to misleading evidence that justifiesB in believing that *p* (feature y). The justification is such that, if *p* were true, B would count as knowing that *p* (Fallis 2022: 11, 12). On this analysis, x (‘form’) and y (‘content’) seem to be such that one cannot simultaneously interfere with both the content and the form of the judgement of one’s victim. If x is present, then y is absent and vice versa: if the evidence justifies the false belief, then the reasoning from this evidence is not flawed (no x) and, if the reasoning is flawed, the evidence does not justify the belief (no y). Intentionally causing the victim to acquire a false belief by compromising her good judgement (x) involves exposing her to evidence that does *not* justify believing the relevant proposition (y does not obtain) and, thus, counts as engaging in non-deceptive manipulation rather than deception (e.g., Fallis 2022: 6).

Arguably, neither Cohen nor Fallis thinks that *x* is necessary for manipulation and the former also seems to think that misleading non-deceptive manipulation could involve interfering with both the form (x)and the content (y) of the judgement(Cohen 2018: 490–492, esp. 486). However, one might argue that some behaviour designed to cause a false belief is not manipulative because it does not involve interfering with the judgement’s form (x) and this is necessary for manipulation, and I would like to pre-empt this argument as well. Therefore, allow me to first argue that interfering with the judgement’s form (x) is not necessary for manipulation by discussing two cases. After that, I will argue that it (x) is not exclusive to manipulation either.

In *Friends* (Season 5, [Episode 13](https://www.youtube.com/watch?v=UdN8sM2BtG4)), during a 1999 New Year’s celebration, Chandler and Monica want to kiss at midnight but they cannot because nobody else is going to and they are still keeping their romance a secret from their friends Ross (Monica’s brother), Rachel, and Phoebe. But Joey knows about them and he has a solution. Joey approaches Ross and asks ‘Who are you kissing at midnight, Rachel [his ex-girlfriend] or Phoebe? You gotta kiss someone and you can’t kiss your sister’. ‘Well, who’s gonna kiss my sister?’ asks Ross, and the answer is ‘Chandler’. Ross is unhappy with this but Joey responds convincingly: ‘Dude! Who would you rather kiss your sister? Me [a known playboy], or Chandler [had only one girlfriend]?’ ‘Good point’ replies Ross and, since he has ‘this whole history with Rachel’ [a devastating breakup], he goes for Phoebe. Joey then approaches Phoebe: ‘Listen, Ross is gonna kiss you at midnight since he cannot kiss Monica or Rachel and he has to kiss someone’. Phoebe agrees, since this makes perfect sense, so Joey goes to Rachel and says ‘I’m gonna kiss you at midnight’. Rachel is surprised but, since everyone’s gotta kiss someone, she has all this history with Ross, and she would rather kiss Joey than Chandler, she agrees. Joe’s plan worked perfectly.

Joe manipulates Ross, Phoebe, and Rachel but he does not bypass their autonomy or capacity to reason, he does not compromise their good judgment or block some possible options, and he does not directly influence their beliefs, desires, or emotions such that they fall short of ideals for belief, desire, or emotion in ways typically not in their self-interest. In fact, Joe causes no faulty reasoning or mental states in them; he exploits their sound reasoning. Everyone wants to avoid the awkwardness of Ross and Rachel kissing each other, Ross should not kiss his sister, and Rachel would rather kiss Joe than Chandler.

The case, thus, supports the idea that manipulation may just utilize the way victim’s judgement normally or even ideally functions (Krstić and Saville 2019), it sits uneasily with some theories of manipulation (e.g., Blumenthal-Barby and Burroughs 2012: 5; Barnhill 2014; Noggle 2018; 2020), and it suggests that compromising the form of the victim’s judgement (x) is not necessary for manipulation. What makes Joe’s behaviour manipulative is that he gives everyone good reasons not because they are good reasons, he does not care about that (see, Gorin 2014a, 2014b; Klenk 2021), but rather because they promote his goals.

Even if Ross, Phoebe, and Rachel falsely believe that Chandler and Monica kissed because this is the only reasonable option, we should not think that the manipulation compromises their reasoning. Notwithstanding this belief, Ross should not kiss Rachel or Monica, Rachel still prefers kissing Joe over Chandler, and everybody relevant still prefers Chandler rather than Joe kissing Monica. Also, it seems more reasonable to think that Ross, Phoebe and Rachel do not have false beliefs regarding Chandler and Monica’s kiss but rather merely do not know all relevant facts (they are kept in the dark). So, this manipulation does not cause the victims to end up epistemically worse off.

Some manipulations may even *improve* the victim’s epistemic state by exploiting their correct reasoning. Consider the events from *How I Met Your Mother* (season 4, episode 17 – ‘The Front Porch’).

Ted and Robin are Lily’s closest friends who have been dating for a while. However, they want completely different things in life but refuse to deal with this. Ted wants children; Robin hates children and she wants to travel and focus on her career. Because of their inconsistent desires, their relationship is making them miserable. Lily will make them break up for their own good. If she gets them to discuss their future, they’ll figure out that they are not right for each other. Lily knows that she cannot confront them directly; they’ll immediately shut down. And so, during her casual hanging out with Ted, she says ‘Hey here’s a fun question: Where do you see yourself in five years?’ ‘I don’t know,’ replies Ted. ‘Not that I’m putting words in your mouth, but you probably want to be married’. Lily is correct. Ted is repressing his desires. Lily uses the same trick on Robin, except that she now says ‘You probably want to be in Tokyo … or Paris’. And, again, Lily is right. Robin wants to travel the world but she is too afraid to admit this to herself. Not long after, Ted and Robin meet for dinner. Prompted by their conversations with Lily, they start discussing their future in light of their real desires and, as a result, they realize that they need to break up.

This is another example of a manipulation that does not interfere with the victim’s good reasoning, it does not interfere with its proper form (x), but rather exploits it. The difference with the existing examples is that this manipulation improves the victim’s epistemic position and it brings her doxastic states closer to their ideals. Lily manipulated Ted and Robyn into realizing that they are not good for each other; they learned something new. And she did not hide any information relevant to the question of whether they should break up (she only hid the fact that she was manipulating them). Therefore, compromising the victim’s reasoning or her mental states, x, is not necessary for manipulation.

However, while not necessary, it could be that x is exclusive to manipulation and can never be found in deception; it could be that deceivers never interfere with the judgement’s form or exploit their victim’s bad reasoning. Fallis (2022: 5–6) discusses a case put forward by Fuller (1975) in a way that supports this version of hypothesis x. Fuller (1975: 25) writes:

Suppose that I wanted you to believe that my neighbour is Norwegian, and I know that you falsely believe that only Norwegians are blondes. Remarkably, Pilar is both Spanish and blonde. So I send you next door. I do not think that … we would want to say that I deceived you … about Pilar’s nationality.

The explanation is that this is not deception because, ‘unlike with cases of deception, in the Blonde Norwegian case …, the victim acquires the false belief as a result of bad reasoning. … [And] when the perpetrator takes advantage of the victim’s “suboptimal judgment” in this way [hypothesis x], it is merely a case of non-deceptive manipulation’ (Fallis 2022: 6). In short, this is not deception because it involves x (incorrect form) and x is exclusive to manipulation. Fuller (1975: 27) explains his ‘powerful car’ example in a similar way and Cohen (2018: 489–490) uses similar reasoning in discussing his cases of misleading by communicating the whole truth and nothing but the truth.

While I recognize its appeal, I think that this approach is problematic. Fallis and Cohen use cases like *Blond Norwegian* to support their hypothesis that interfering with the victim’s reasoning (x) is exclusive to manipulation but then they use the same hypothesis to support the interpretation that these are not cases of deception, which strikes me as viciously circular. They also appeal to intuitions when interpreting their examples, but my intuitions differ from theirs. However, I will not defend my position here since my intuitions could easily be tainted by my preferred theories. Rather, I propose the following approach.

Because our intuitions are conflicting when applied to cases like *Blond Norwegian*, let us not discuss these cases in assessing the plausibility of x and y. Rather, let us apply x and y to cases where our intuitions will not vary. With that in mind, I proceed to examine cases that unequivocally show that x and y are both problematic.

# 3. Deception, Evidence, and Reasoning

In Shakespeare’s *Othello*, Desdemona drops her handkerchief and her maid, Emilia, finds it. Before it can be returned to Desdemona, the handkerchief is taken from Emilia by Iago, the villain of the piece. Iago takes advantage of this situation and exploits Othello’s profound jealousy to cause him to falsely believe that Desdemona is having an affair with Cassio: he lies to Othello by telling him that he saw Cassio wiping his beard with Desdemona’s handkerchief, the exact one Iago is holding in his hand.

Iago lies to Othello because he intends to thereby cause him to falsely believe that Desdemona is having an affair with Cassio. Not all lies seem to be intended to deceive but deceptive lying is a prototypical and the most common kind of human deception nonetheless. Thus, *Othello* involves deception and his judgement has been fed with a false premise (the lie); the example involves the supposed feature of deception, y (interfering with the content of a judgement). However, because Iago lies in order to exploit Othello’s profound jealousy, *Othello* also involves the supposed feature of manipulation, x (interfering with the form). By lying, that is, Iago manipulates Othello in a way that compromises Othello’s good judgment: the jealousy triggered by the lie prevents Othello from assessing the situation correctly.

A case that involves both x and y is a problem for Cohen’s view since he says that deception cannot involve x (Cohen 2018: 486). Fallis’s (2022) refinement of y (the misleading evidence must be sufficient) might seem as avoiding this problem. Iago’s testimony (lie) is sufficient evidence that Cassio wiped his beard with Desdemona’s handkerchief, but it is only weak evidence that Cassio and Desdemona are having an affair. Cassio could have wiped his beard with her handkerchief for many innocent reasons. On Fallis’s view, then, this could be a non-deceptive misleading manipulation – y is not present.

However, it seems wrong to think that a lie (a) intended to mislead, that (b) genuinely misleads, and that misleads (c) in the exact same way as intended does not involve deception. Intuitively, *Othello* involves deception to a great extent because the judgement’s form has been compromised (x). Fallis’s idea that deception requires exposing the victim to misleading evidence that justifies her false belief (y) seems to raise the bar for deception too high. But, let us suppose that I am wrong and consider a variation of *Othello*.

Say that everything is the same as in *Othello* except that Iago says to Othello ‘Desdemona is having an affair with Cassio’ and that Othello has no reason not to trust Iago. Let us call this case *Othello Obvious* because it is an obvious case of deceiving by lying. *Othello Obvious* satisfies Fallis’s version of hypothesis y (the testimony is sufficient evidence), but it also involves a manipulation that compromises the form of Othello’s good judgment (x) – Othello is still uncontrollably jealous.

What I mean is the following. Indeed, the testimony is sufficient evidence that Desdemona is having an affair with Cassio but Othello does not form this belief *exclusively* because of the nature or the content of the testimony; rather, he principally forms it because he is uncontrollably jealous. Even though he is given sufficient evidence, Othello does not need it to form this belief. Moreover, we should not forget that Othello’s behaviour is pathological: while the lie may justify the false belief, it does not justify Othello’s reaction. Othello was manipulated into killing Cassio just as a parent may be manipulated by their crying child into caving to the child’s demands. Thus, we have a clear case of deception that involves both x and y. The case involves y (interfering with the content) because the testimony is sufficient evidence of the affair and the victim takes it to be sufficient evidence, and it involves x (interfering with the form) because it triggers a pathological behaviour on the part of the victim. The case, thus, shows that x and y are not mutually exclusive.

Cohen’s theoretical framework also struggles with *Othello Obvious*. Cohen (2018: 490) explicitly writes that ‘intentionally causing false beliefs by means of triggering the victims’ lapses of judgment [form] is the sign of manipulation’ and that, in deception, the misleading information (content) must be the cause which ‘the listener takes to justify her belief’. *Othello Obvious* satisfies both requirements: the misleading information is both the cause that Othello takes to justify his (false) belief and a means of triggering a lapse in his judgement on how to act on this belief.

Therefore, while *Friends* and *How I Met Your Mother* show that interfering with the judgement’s form (x) is not necessary for manipulation, *Othello Obvious* shows that it is not exclusive to it either. *Othello Obvious* also shows that interfering with the judgement’s content (y) is not exclusive to deception: Othello is both exposed to evidence that justifies his false belief and is being manipulated. Finally, unlike in *Blond Norwegian* or Cohen’s cases, there is nothing unclear in *Othello Obvious* and the intuitions that it involves both deception and manipulation should not vary. What remains is the issue of whether feeding the victim’s judgement with false premises (y) is necessary for deception.

The importance of the *Othello* cases lies in the fact that Othello’s jealousy can turn any information into misleading evidence. Say that Cassio, rather than the maid, found the handkerchief. Say that Cassio did wipe his beard with the handkerchief without knowing that it belongs to Desdemona and that Iago saw this. Iago can still exploit Othello’s jealousy to generate the same false belief by asserting that he saw Cassio wiping his beard with her handkerchief. Even though Iago does not lie in this variation, call it *Othello Modified*, Iago *acts with the same intention* (to mislead) and he causes Othello to acquire *the same false belief* by giving him *the same piece of evidence* as in the original *Othello*.

The only difference between *Othello* and *Othello Modified* is that the evidence is not objectively misleading in the latter case (it is misleading only subjectively). This difference suggests that what is important in this strategy of human deception is not that it involves communicating misinformation but rather that the deceiver covertly manipulates the dupe’s dispositions to cause them epistemic harm; the function of Iago’s manipulation is to mislead (see, Krstić forthcoming 1). *Othello Modified* not only shows that x is not necessary for manipulation, but also suggests that y is not necessary for deception (also, Krstić forthcoming). We should, thus, conclude that deception need not involve communicating a *misleading* message intended to mislead; rather, deception must involve communicating *a* message intended to mislead.

# 4. Instead of the Conclusion

According to my argument, interfering with the form of the victim’s judgement (x) is neither necessary nor exclusive to manipulation andinterfering with the content of her judgement (y) is neither necessary nor exclusive to deception. The existence of non-deceptive misleading manipulation does make prima facie sense and it is based on the common view that deception must involve communicating misinformation. This view is standard not only in our analyses of human deception but also in pretty much all philosophical analyses of biological deception. However, philosophical analyses of deception seem to be rather narrow. *Othello Modified* is not an isolated case. Many organisms, such as fungi or plants, do appear to be victims of deception even though they are insufficiently sophisticated to engage in communication (e.g., Waterman and Bidartondo 2008) and some strategies of biological deception appear to involve messages that do not carry misinformation (Krstić and Saville 2019; Krstić forthcoming). This has important consequences for our discussion and it must be taken into account when considering whether some supposed examples of non-deceptive misleading manipulation are genuinely non-deceptive.

Nazarbayev University,

History, Philosophy, and Religious Sciences

Kazakhstan

[drpop1@yahoo.com](mailto:drpop1@yahoo.com)

[vladimir.krstic@nu.edu.kz](mailto:vladimir.krstic@nu.edu.kz)

# Acknowledgement:

I thank the two anonymous reviewers and the Area Editor of this journal for their very helpful comments.

# References

Barnhill, A. 2014. What Is Manipulation? In Christian Coons and Michael Weber (eds.), *Manipulation: Theory and Practice*, 51–72. Oxford: Oxford University Press.

Blumenthal-Barby, J.S., and H. Burroughs. 2012. Seeking better health care outcomes: The ethics of using the ‘nudge’. *American Journal of Bioethics* 12: 1–10.

Cohen, S. 2018. Manipulation and Deception. *Australasian Journal of Philosophy* 96/3: 483–97.

Fallis, D. 2022. Deceiving versus manipulating: An evidence-based definition of deception. *Analytic Philosophy* (Online First): 1–18.

Fuller, G. 1976. Other-deception. *Southwestern Journal of Philosophy* 7: 21–31.

Gorin, M. 2014a. Do Manipulators Always Threaten Rationality? *American Philosophical Quarterly* 51/1: 51–61.

———. 2014b. Towards a Theory of Interpersonal Manipulation. In Christian Coons and Michael Weber (eds.), *Manipulation: Theory and Practice*, 73–97. Oxford: Oxford University Press.

Klenk, M. 2021. (Online) manipulation: sometimes hidden, always careless. Review of Social Economy (Online First): 1–21.

Krstić, V. and C. Saville. 2019. Deception (Under Uncertainty) as a Kind of Manipulation. *Australasian Journal of Philosophy* 97: 830–35.

Krstić, V. Forthcoming. We Should Move on From Signalling-Based Analyses of Biological Deception. *Erkenntnis*. <https://doi.org/10.1007/s10670-023-00719-x>

———. Forthcoming 1. A Functional Analysis of Human Deception. *Journal of American Philosophical Association*.

Mahon, J.E. 2007. A definition of deceiving. *International Journal of Applied Philosophy* 21: 181–94.

Noggle, R. 2018. Manipulation, salience, and nudges. *Bioethics* 32/3:164–70.

———. 2020. Pressure, Trickery, and a Unified Account of Manipulation. *American Philosophical Quarterly* 57: 241–52.

Waterman, R.J. and M.I. Bidartondo. 2008. Deception above, deception below: linking pollination and mycorrhizal biology of orchids. *Journal of Experimental Botany* 59: 1085–96.

1. The traditional analysis presupposes that the deceptive behaviour did not involve bypassing of the victim’s agency: she was not drugged, hypnotized; she was not subjected to brain surgery, thought-insertion; etc. (Mahon 2007). [↑](#footnote-ref-1)