

Interpersonal Manipulation

Michael Klenk

Delft University of Technology

m.b.o.t.klenk@tudelft.nl

Abstract: This article argues that manipulation is negligent influence. Manipulation is negligent in the sense that manipulators do not chose their method of influence because for its potential to reveal reasons to their victims. Thus, manipulation is a lack of care, or negligence, exclusively understood exclusively in terms of how one influences. That makes the proposed account superior to the most influential alternative, which analyses manipulation disjunctively as violation of several distinct types of norms. The implication is a paradigm shift in understanding manipulation in terms of what manipulators intend to do toward a focus on what they fail to do or intend.

Keywords: Manipulation; Rationality; Autonomy; Intention; Nudging.

What is manipulation? The question has attracted much recent attention.¹ Manipulation is an arguably problematic and pervasive social phenomenon.² It thrives in close relationships, perpetuated even by those who otherwise abstain from the nefarious conduct more frequently studied by ethicists, like coercion or torture.³ It also plays an increasingly consequential role publicly, where deepening psychological insights and increasing technological sophistication allow for unprecedented manipulative potential in (hyper-targeted) advertising, design, and politics.

¹ Gorin 2014b; Noggle 2018b; Cohen 2018; Krstić and Saville 2019; Coons and Weber 2014b.

² Wood 2014; Coons and Weber 2014a; Baron 2003.

³ Manne 2014; Noggle 1996.

The most influential account of manipulation suggests that manipulation is a kind of trickery, that is, *the attempt to influence someone so that they violate what the manipulator thinks are norms of belief, desire, or emotion* (see Noggle 1996). This view is promising because it accounts for many paradigmatic cases of manipulation (as will be shown below). It is also wide enough to avoid counterexamples that crippled alternative accounts that explained manipulation in terms of hidden influence, autonomy subversion, or the bypassing of rationality.⁴

However, the view that manipulation is trickery fails where it matters most. The trickery account fails to provide a unified account of the norms that manipulation arguably undermines. Several paradigmatic cases of manipulation – pressuring tactics, the 'hijacking' of someone's proper functioning, and paternalizing nudges – do not involve trickery. The trickery account can be broadened to include both norms applicable to the manipulatee (e.g. to adopt reasonable beliefs) and norms applicable to the manipulator (e.g. beliefs honest intentions in influencing others) to explain these cases. But then the account loses its simple elegance and, more importantly, it mischaracterises manipulation as *aiming at* violating those norms, when the violation of those norms is a mere side effect of manipulation.

Therefore, I introduce and defend a simpler and yet more adequate account of manipulation. Accordingly, manipulation is analysed in terms of the manipulator's grounds for choosing a particular means of influence. Manipulators

⁴ Noggle 1996, 2018a; Gorin 2014b; Barnhill 2016; Mills 2014; Sunstein 2016; Cholbi 2014; Baron 2014; Hanna 2015; Klenk 2021.

are *negligent* about whether their influence reveals reasons to the manipulatee. Violating a norm about the appropriate intention behind interpersonal influence is thus shown to be necessary and sufficient for manipulation. The resulting account also explains why manipulation can be unintentional, and it provides a superior basis for (empirically) measuring and normatively evaluating manipulation than the trickery account.

I will proceed as follows. Section I briefly recaps the trickery account of manipulation and Section II presents three important counterexamples to the view. Section III draws two important implications, which have gone unnoticed in the literature and which motivate the search for an improved account. Section IV presents and defends the novel negligence view of manipulation. Section V spells out implications for (morally) evaluating manipulation and assessing it empirically.

I Manipulation as trickery

The view of manipulation as trickery says that manipulative action is the attempt to get someone's beliefs, emotion, or desire fall short of the ideals that in the view of the influencer govern the target's beliefs, emotions, or desires (Noggle 1996).⁵

On the trickery view, and to stick with Noggle's helpful metaphor, manipulation is the adjustment of "psychological levers" away from what the manipulator thinks are ideal settings for the target. Noggle writes (1996, 44):

⁵ The trickery account is very influential and – in basic outline - accepted by several commentators in the current debate (e.g. Baron 2003, 2014; Coons and Weber 2014b). Deviators focus on small additions. For example, Barnhill (2014) accepts the trickery view with the slight modification that manipulation must hurt the manipulatee's self-interest. And while Mills (1995) defends a narrower version of the trickery view she expands her view to match Noggle's account in later works (Mills 2014).

There are certain norms or ideals that govern beliefs, desires, and emotions. Manipulative action is the attempt to get someone's beliefs, desires, or emotions to violate these norms, to fall short of these ideals.

What norms matter? The relevant norms or ideals are the ones that the manipulator envisions for the manipulatee. This retains a parallel with deception (where it matters what the deceiver takes to be the truth, from which he deviates), and it avoids the potential problem of committing to and identify objective norms that govern belief, desires, or emotions.⁶ Thus, the trickery view classifies manipulation as an "intentionally characterised" action (Noggle 1996) and specifies it quite broadly in terms of attempting one's victim to violate *some* belief, desire, or emotion-related norm. In effect, the breadth of the different norms for emotions, beliefs, and desire that we recognise gives the account tremendous breadth and explanatory power. Thus, the trickery account avoids the mistake of trying to shoehorn manipulation into the mould of necessary violation of some allegedly more basic capacity or process.⁷

The trickery view requires that the manipulator aims, attempts, or intends to have the manipulatee violate these norms, though it is not clear how to characterise these notions precisely in the context of manipulation. There are two main questions. First, is the aim, attempt, or intention to make the manipulatee *adopt a faulty* mental state? Or simply to make the manipulatee *adopt a particular mental state*? Moreover, can one follow an aim without being, in the moment, consciously aware of it and thus 'accidentally' manipulate? Must a manipulator

⁶ Though, of course, that depends a lot on one's metaethical predilections. Hanna 2015 thinks that objective norms are the relevant ones for manipulation, though Noggle 1996 already suggests powerful arguments against that view.

⁷ Compare the discussion of accounts that define manipulation in terms of hidden influence, deception, autonomy-undermining, or rationality subversion and the relevant counterexamples in Gorin 2014b and Klenk 2021.

have the norm violation she aims for explicitly in mind, and thus be able in the first place to understand what norms and ideals govern her victim? The trickery account does not give answers to these questions, though they seem relevant in light of recent arguments that suggest that unintentional or at least unconscious manipulation is possible (Baron 2014; Manne 2014). I shall return to these points below.

2 Trickery is not necessary for manipulation

First, it is helpful to discuss some salient counterexamples to the trickery account.

2.1 Pressuring

Pressuring tactics like charming, peer pressure, or guilt tripping appear manipulative but cannot easily be explained by the trickery account.⁸ At least, it is not obvious which norms or ideal a manipulator who uses pressuring tactics aims to violate. Two examples illustrate the point. Consider:

Emotional blackmail. Mark is a horrible cook, as Carola began to understand after being married to him for years. Like in everything else in their relationship, he does not take rejection lightly. At lunch, he tells her "Eat the food I cooked for you or I will hurt myself."

Peer pressure. Taylor seeks a cheap place to stay in an expensive city and ends up in a fraternity with a heavy drinking culture. He abhors drinking, but feels that he must participate to keep his place in the frat house.

⁸ As will be clear later, I think that there is indeed a more basic phenomenon, or common factor, that explains why these seemingly different cases of manipulation are cut from the same cloth.

Both cases seem like paradigmatic cases of manipulation. In both cases, however, it is false that the manipulator aims to make the manipulatee's beliefs, emotions or desires fall short of some ideal. The emotional blackmailer does not aim at changing the manipulatees beliefs

Unless proponents of the trickery account can reveal relevant norms that the manipulator intends to undermine, both cases are counterexamples to the view that trickery is necessary for manipulation.

One can try to make sense of pressuring cases by a disjunctive account, arguing that there are two distinct forms of manipulation (Baron 2003): manipulation as trickery, and manipulation as pressuring. Obviously, however, one then abandons the theoretical elegance of the trickery account that explained manipulation in terms of one specific intention: to make someone undermine norms about belief, emotion, or desire.

Moreover, the idea of 'pressuring' would itself requires an analysis that is not available yet. Precisely *what* is going on in a case like peer pressure, for example, remains opaque. Thus, without a further elucidation of the norms the manipulator intends to undermine, it seems that pressuring cases cannot be accounted for by the trickery account.

2.2 Hijacking

In a case of what I will call *hijacking*, the manipulator uses the 'proper functioning' of a victim's cognitive or affective to reach manipulative ends. Hijacking seems manipulative. For example, Gorin (2014a) has documented such cases at great length to combat the view that manipulation necessarily subverts or bypasses rational capacities. As Gorin argues, manipulation sometimes proceeds precisely

by hijacking those rational capacities. The point applies more widely to the trickery account when the proper functioning of a manipulatees mental faculties are exploited by the manipulator. Two examples illustrate the point:

Hijacking affect A politician is convinced by Achen and Bartels's (2016) claims about the irrationality of voters. He makes sexist comments to ignite the opposition. As intended, the opposition is angry and voices their concerns, much to the dismay of the politician's base and hitherto undecided voters. This only helps the politicians cause.

Hijacking reason Another politician is not convinced by the irrationality of voters, and finds that voters are very much concerned with the saving environment. The politician proceeds to give good arguments for the protection of the environment, and she is voted into office. The politician herself, however, does not care about the environment themselves at all.

Both cases seem to be cases of manipulation. However, it is false that the manipulators aim to make the manipulatees fall short of the ideals that govern their emotions or beliefs, respectively. Arguably, it is reasonable to react toward sexist comments with anger. Similarly, it is reasonable to accept good arguments for a true conclusion, if anything is.

Therefore, both cases are counterexamples to the view that trickery is necessary for manipulation, unless, as before, proponents of the trickery account can produce norms that the manipulator attempts to undermine in these cases.

Note that proponents of the trickery account cannot make sense of manipulative hijacking by invoking norms about the outcome of an interaction. The thought might be this. Manipulative hijacking proceeds by some however-

defined proper process (e.g. it appeals to someone's rational capacities in a situation where that is required), but it may lead to outcomes that conflict with the ends of the manipulatee. So, the manipulator intentionally violated a norm about the consequences of the influence. Perhaps that is what Barnhill (2014) has in mind when she specifies that the condition of 'violating self-interest' must be part of the trickery account. As the Hijacking reason case shows, manipulation may sometimes lead to outcomes that are norm-compliant from the manipulatees perspective as well as from an objective standpoint (cf. Gorin 2014b, 2014a).

2.3 Paternalizing

Paternalising tactics are influences where the influencer intends the person being influenced to adopt a mental state or perform an action that is in the interest of the person being influenced. Paternalising tactics are of notorious fame from the nudging debate (Thaler and Sunstein 2009).

Not all nudges are manipulative. Informational nudges plausibly do not qualify (Noggle 2018a; Engelen and T. Nys 2020). But some nudges are manipulative, like as salience nudges (Noggle 2018a). Consider the following case:

Cafeteria A school cafeteria manager is faced with a decision about how to arrange the food. Given his knowledge that placing certain food items in certain places will boost or limit consumption of the respective items, he decides to promote healthy foods.

Arguably, we have here a case of manipulation. But what is the salient norm that the cafeteria manager aims to undermine? Per definition, paternalizing nudges are aimed at making their 'victims' behave in ways that are beneficial for

them (there is some debate, immaterial for present purposes, about how the 'victims's' self-interest is determined in nudging).⁹ At best, and I discuss that more below, the cafeteria manager accept the irrational belief-forming mechanism as a side effect.

Many forms of interpersonal influence fit the schema of the pressuring, hijacking, and paternalizing cases just discussed. These are common tactics, and arguably manipulative. However, unless proponents of the trickery account can show which norms the users of such tactics aim to undermine, trickery (understood in the technical sense, as the intentional undermining of a norm for belief, desire, or emotion) is not a necessary condition of manipulation.

3 Implications for the trickery account

So far, I presented counterexamples to the view that trickery, understood as the intentional violation of norms for belief, emotion, or behaviour, is a necessary condition for manipulation. In this section, I draw two important lessons from the counterexamples above.

The first lesson is that proponents of the trickery account cannot convincingly accommodate the counterexamples above by broadening their account by identifying and incorporating other types of norms that manipulation allegedly violates (to wit, in addition to norms that pertain to belief, emotion, or behaviour). The resulting theory would be, as I will show, ad-hoc and, therefore, unappealing.

I must note that my discussion of this point cannot be exhaustive: I cannot discuss *all possible* norms that might be violated in the cases above, thus leaving

⁹ See Engelen 2019; Engelen and T. Nys 2020; T. R. V. Nys and Engelen 2017 for helpful overviews over ethical issues to do with nudging that are tangential to the current debate.

room for proponents of the trickery account to argue that their account covers the examples above.

Nonetheless, the schematic discussion will already reveal that the trickery account would have to incorporate types of norms so different that it raises doubt whether the mere intention to violate *any* norm is characteristic or at least illuminating of manipulation. To elaborate, consider that in a given case of manipulation, proponents of the trickery account can try and find norms that the manipulator wanted the manipulatee to violate. Indeed, in all cases of manipulation, some norm is violated (as will become clear later, too). But proponents of the trickery account have to appeal to very different types of norms to explain the counterexamples above. For example, they would also have to invoke norms that apply to the manipulator (thus extending their account), in addition to the norms that apply to manipulatees (as their account suggests). I illustrate this through a discussion of the hijacking cases.

In the hijacking cases, the manipulator makes the manipulatee do something that is norm-abiding *given the manipulatee's information or situation*. In order to account for the manipulation in the case, the proponent of the trickery account now needs to point to a norm that applies to the manipulatee that the manipulator intends to violate. But what norm could that be? One norm that could be violated pertains to the motivation of the influencer, the manipulator, and it invokes information that is not available to the manipulatee. The voters in Hijacking reason arguably violate a process norm about the manipulator's motives (e.g. 'do not act on disingenuous influences') (cf. Bělohrad 2019). Had the voters in Hijacking reason known about the politicians true disgust for environmental

policies, they might not have voted for her. But this move not only broadens the scope of relevant norms that are relevant in an account of manipulation, it would also lead to vicious circularity. If Hijacking reason is a case of manipulation because the voters violated the norm not to act on disingenuous influences, then that comes close to vicious circularity: manipulation is the attempt to make someone violate a norm about not being manipulated. Therefore, the relevant norm cannot be a norm about the grounds on which one acts. Instead, the relevant norm must be a norm that the *manipulator* violates – not one that the manipulator intends the manipulatee to violate. Clearly, the manipulator in Hijacking reason manipulates a norm about abstaining from disingenuous influences. For proponents of the trickery account, however, that claim leads to a dilemma that drastically increases the scope of norms that manipulators may aim to violate. Either they accept that the norms that are violated are norms that apply to the manipulatee *given full information* – which is implausible – or they accept that the norms that are violated are norms that apply to the manipulator. That, however, would be a striking departure from the trickery account, which only refers to norms as they apply to the manipulatee. Therefore, a plausible way for proponents of the trickery account to deal with the counterexamples requires them to significantly broaden their account to include both norms that apply to the manipulatee (e.g. believe on good reasons) as well as norms that apply to the manipulator (e.g. have honest intentions).

The second lesson suggested by the counterexamples is that the trickery account reflects a deeper error in thinking about manipulation. The trickery account operates within the paradigmatic view of manipulation as characterised

by an intention of specific content. It erroneously focuses on what the manipulator *intends to do*. However, I suggest that manipulation, at its heart, is characterised by what the manipulator *fails to do*. Focusing on norms that the manipulator *intends to violate* may be on the wrong track for an account of manipulation. If instead we focus on norms that the manipulator effectively violates we retain a central and valuable insight of the trickery account – the relation between manipulation and violation of norms – and can avoid the multiplication of norms that the trickery account would require to make sense of the counterexamples above. Thus, in full generality, the association of manipulation with norm violation is not illuminating and the concrete proposal made by Noggle, and adopted by others, is too narrow.

Instead, I suggest that the association of manipulation with norm violation is on the right track *and* that we can explain all cases of manipulation in terms of a single norm violation when we focus on norms that govern permissible intentions to influence. I illustrate this through a discussion of paternalizing cases.

In paternalizing cases, manipulatees arguably violate norms about properly grounded beliefs, desires, or emotions. Manipulatees violate norms that apply to the grounds on which the act, think, or feel. As Noggle (2018a, 169) puts it, paternalizing manipulations fail to draw attention to facts that are "true and relevant" for the manipulatee's decision. However, it is not plausible that manipulators in nudging cases *aim, attempt, or intend* their manipulatees to violate such process norms. The end of paternalistic nudgers is to get manipulatees to perform some action, or adopt a certain mental state, not that they form a *faulty* mental state, or perform an unreasonable action. Like other manipulators,

paternalists do not care much norms, but they care about making people act in a certain way. At the same time, to stick with the assumption, they act manipulatively, and a nudging cafeteria manager who *aimed for his students to violate norms about how to ground ones food choices* ultimately seems as manipulative (or not) as the one who *aimed his students to eat more salad*. The reference to norms that apply to the manipulatee is not required to explain these cases.

So, the trickery account must invoke both norms that apply to the manipulator and norms that apply to the manipulatee and only the former are required to explain manipulation in the pertinent cases that make trouble for the trickery account. The latter norms about manipulator's intentions are also able to explain other paradigmatic cases of manipulation. A parsimonious understanding of manipulation would can therefore do with only the latter norms and this suggests that we should understand manipulation exclusively in terms of a violation of norms about what influencers ought to intend when influencing others.

Before spelling out the novel conception of manipulation in more detail, it is worth noting a dialectical point that should give further support to the account of manipulation developed in this paper. If the above is correct, and if trickery is sufficient (albeit not necessary) for manipulation, we might have to conclude that there are two types of manipulation: one kind of manipulation characterised in terms of norms about the manipulatee that the manipulator intends to violate (the trickery-based account of manipulation), and another kind of manipulation characterised in terms of norms about proper influence (which is spelled out below). That conclusion can be avoided in two ways. My preferred way is to show

that there is a common factor behind both types of norm violations. This is what I will attempt to show in the next section. Alternatively, it could be shown that trickery is insufficient for manipulation. I cannot defend this latter point in detail in this paper, but it is worth noting that there are borderline cases that put pressure on the claim that trickery is sufficient for manipulation. Consider:

Internalist manipulator Kelly is a staunch internalist about epistemic justification and he wants Bart to vote Biden. Unfortunately, the only information that Bart's other epistemic commitments (e.g. about whom to trust) allow Bart to process supports voting Trump. As election day approaches, Kelly appeals to Bart's enthusiasm for donkeys and his firm belief that anyone associated with donkey's ought to be voted into office and so gets him to believe that Biden is the right candidate after all.

Bart, the potential manipulatee, has a reasonable belief and wants to do the reasonable right thing (vote Biden, let's suppose), and there is ample evidence for supporting Biden (let's suppose), even if it is not available to Bart. The trickery view would classify this case as a case of manipulation. From Kelly's internalist perspective, Bart violates a norm for belief when he believes that Biden is the right candidate. Nonetheless, I am unsure as to whether Kelly manipulated Bart. Irrespective of the sufficiency of trickery for manipulation, I will suggest that there is a deeper underlying phenomenon in the next section.

4 Manipulation as negligent influence

I have argued that the most promising and most developed account of manipulation, the view that manipulation is trickery, fails. I will now defend a

superior alternative. Manipulation can be understood satisfactorily and *exclusively* in terms of norms about the manipulatee's intentions about the means or method of influence. The account I develop thus switches the 'domain' of norms that a manipulator violates from norms as they apply to the manipulatee to norms as they apply to the manipulator. This account maintains a central insight of the trickery account – that manipulation concerns the violation of norms about desire, belief, or emotion – but it is more explanatorily powerful in identifying a single norm that all manipulation cases violate.

Common to all cases of manipulation is the negligence of the manipulator, where 'negligence' is used as a (to be defined) term of art. The manipulator is negligent in that he or she does not care whether or not her chosen means of influence help her victim to adequately assess her situation. There is such a norm of 'care' in interpersonal interaction. That is the norm that manipulators violate. In other words, the manipulator does not care whether her chosen means of influence reveals reasons to the manipulatee.

Of course, we have already seen that manipulators sometimes hijack their victims's capacities and thus they expressly appeal to reasons to make their victims behave, feel, or think as desired. Manipulators, however, are not ultimately motivated by the reason-revealingness of their method of influence but rather by the fact *that it is an effective method* to get them what they want. That is, I propose that manipulators violate the following conditional:

Negligence I: Had a method of influence not revealed reasons for doing x to subject S, then S's influencer would have chosen a different method.

The negligence conditional is motivated by considerations about the nature of persuasion, and manipulation is understood in relation to persuasion. Like all social influences, persuasion is aimed at change (in someone's behaviour, thought, desires, etc.). Its characteristic feature, I argue, is that it finally aims at *reasonable change*, that is, change grounded in an appreciation of reasons. A persuader aims at change in the interlocutor insofar as the interlocutor makes that change based on an understanding of his or her reasons for doing so.

Coercer, similarly, care a great deal about reasons. They care to identify good reasons for their victims to do as they please. A credible and efficient coercive threat will provide good reason for the victim to follow suit. A threat that does not give them reason to follow suit will not be effective in coercion. Of course, unlike persuaders, coercers do not care about reasonable change per se.

Compared to persuasion and coercion, manipulation appears furthest away from reasonable influence (to wit, an influence aimed at or dependent on there being reasons for the outcome the manipulator wants to effect). We can understand manipulation without talking about reason at all. A persuader has reasons in mind, and tries to reveal them to the interlocutor. A coercer, similarly, relies on there being good reasons for the effect he wants to create by coercion (Schelling 1997). In contrast, a manipulator never is concerned with reason as a final end. Sometimes, showing you that you indeed have reasons to do what the manipulator wants is the most effective way for the manipulator to achieve her aims. But she does not pursue the revealing of reasons as a final end. Manipulation indeed lies between these extremes, as earlier writers have suggested (Beauchamp 1984; Faden and Beauchamp 1986).

The account of manipulation is not complete yet. Using any method that necessarily reveals reasons to the manipulatee will ensure that one cannot manipulate. For example, imagine you are connected to Mr Spock and your only means of influencing him is through written proofs. Such proofs necessarily count as reason revealing for Mr Spock, and so you cannot manipulate Mr Spock at all. But that seems wrong. Hijacking cases have shown that you could use adequate processes for nefarious outcomes. To capture this, I propose the following second condition:

Negligence II: Had the influencer chosen a method of influence M, M would have revealed reasons to subject S.

Thus, manipulation is careless influence in the sense that manipulative action is characterised by the violation of Negligence I and Negligence II:

Manipulative action: M intends to manipulate S iff M violates Negligence I or Negligence II.

There are some important differences to the trickery account. First, the negligence account makes it straightforward to account for pressuring, paternalising, and hijacking tactics. In all three, manipulators do not choose their method of influence to induce change based on a proper understanding of the situation, but they choose whatever method works best to induce that change (they will still instrumentally care about reasons, though unlike a coercer). The carelessness gives us a unified account of manipulation. The norms that matter for understanding manipulation are norms about *how one ought to go about influencing others* or, more concretely, norms about legitimate means of influence.

What these norms legislate will depend on the attributes of the person one is trying to influence.

The account of manipulation as careless influence preserves a conceptual analogy between manipulation and bullshitting (in Frankfurt's technical sense). Bullshitting is characterised as a lack of regard for the truth of one's assertions (or implications thereof). It contrasts with truthful assertion and deception (Fallis 2010). Fallis suggests that deception, and bullshitting, must be understood as including a reference to a context such that deception and deception requires a context where truthful assertion is expected. I am not sure whether such contextualising is needed in the case of manipulation, at least when it comes to *identifying* manipulation. It seems that you manipulate when you are careless in your influence (i.e. you violate Negligence I and Negligence II), independently of whether you are in a context where influencing requires no reason revealing. Lovers can manipulate, and parents, as do bantering friends. Often, the contexts where such interactions will take place will not require reason revealing, and yet it seems reasonable to allow that these influence (insofar as they are careless) are manipulative. It rather seems that unlike deception, which carries a stronger moral undertone, bullshitting and manipulation are more alike because you can do either independently of the context. There does seem to be a contextualised element of manipulation nonetheless: but it concerns its *evaluation* rather than the identification question, as I discuss in section 5.

A further interesting feature of the account of manipulation as negligent influence is the fact that it allows for unintentional or at least unconscious manipulation. As Baron (2014) and Manne (2014) suggest, as well as several

others who recognise the manipulativeness of guilt trips (cf. Noggle 2018b), it seems that some actions can be manipulative even though they are perpetrated without the express aim to have someone violate some norm or ideal. The carelessness account makes sense of this. These are cases where people do not try hard enough to be persuasive, they just don't care: they sulk, charm, and pressure because they are, consciously or unconsciously, consumed by a need to get what they want. There is a popular image that manipulation is artful influence. That is true in that those who unintentionally manipulate are skilful or effective in getting others to do what they want, but they are not therefore purposeful in planning out their scheme.

The view of manipulation as careless influence retains much of appeal behind the trickery account and indeed it can be seen as a development of it. Before outlining how it helps us toward a better base of measuring and evaluating manipulation, I need to address a complication and an objection.

The complication concerns a crucial but unclear and potentially controversial term at the heart of the carelessness account. What is reason revealing? Anything that helps an agent act (or think, or feel) to be reason responsive. This means that the agent is helped to conduct herself in ways that are supported by reasons *and* that she does so because she recognises those reasons. Reason revealing is, as it were, an internalist notion: something that is reason revealing helps an agent conduct herself in a reason supported way by responding to those reasons. It is still unclear what 'helping' means, but that can be understood in terms of an increased likelihood for the agent to be responsive to those reasons. So, attempting to pick a reason revealing method is thus just the attempt to influence the other

in such a way – through words, a picture, an insinuation, a display of emotion – that she is helped to conduct herself in ways that are supported by reasons because she recognises those reasons (not because she just happens to act accordingly, or because she does so for the wrong kinds of reasons). One might now think that the carelessness account makes the basic mistake of committing to an overly rationalistic picture, where only rational arguments can count as non-manipulative (because reason-revealing) influences. But that is not so. Many different ways for being reason-responsive have been described in the literature, many of which are affective rather than cognitive in nature (cf. Fischer and Ravizza 1998).

The objection concerns an alternative lessons that one might have drawn from studying the manipulative tactics of pressuring, paternalizing, and hijacking. My objection to the trickery account was that we need to introduce rather different types of norms to account for all these cases, some pertaining to the manipulatee (to wit, norms about that the manipulatee ought to do, feel, or think) and some pertaining to the manipulator (to wit, norms about the proper motives for influencing someone else). I suggested that the latter suffice. But even if that is true, the norms I described are not the only ones. Might we not explain manipulation in terms of the failure to be motivated by reasonable change? That is, whenever you influence someone while being negligent about whether or not the desired change is reasonable (for that person), then you manipulate (cf. Gorin 2014b). The account is similar to mine – it spells out negligence in terms of the *outcomes* of an interaction, rather than the means. However, this condition should be resisted. For one, it is not needed in addition to the one I introduced above.

Someone who cares about the process but not about the outcome does not seem manipulative. Someone who does not care about the process but about the outcome *does* seem manipulative. Second, the condition alone would have implausible results because it would only allow mutualistic or altruistic influence to count as non-manipulative. It seems perfectly possible to influence someone in a non-manipulative way without caring about that person's ultimate ends. A marketer that puts up a poster outlining the (actual) benefits of a new medication need not care about the ends of all people that come across the poster to avoid being manipulative.

Thus, I take it that this sketch of an account of manipulation as negligent influence has at considerable attraction, and some clear advantages over the trickery account. It accounts for plausible cases of manipulation in a parsimonious way, including cases of pressure, and it explains how unintentional manipulation is possible.

5 Implications for evaluating and assessing manipulation

The final part of my defence of the negligence account concerns the beneficial ways in which the carelessness account as an answer to the *identification* question reflects on the *evaluation* question, i.e. the question of why, if at all, manipulation is morally wrong.

Manipulation on the negligence conception is a binary affair: you either manipulate or you do not. But that does not mean that the evaluation of manipulation is binary, too. Rather, it seems plausible to invoke the idea of context-dependency that plays a role in the identification of deception, but now in the context of the evaluation of manipulation. Manipulation is morally wrong

insofar as it violates norms that legislate persuasion over manipulation in a given context. Clearly, manipulation is not always morally problematic. Lovers, for example, may be manipulating each other legitimately and without moral residue (cf. Buss 2005). Also, it may be that parents are quite often manipulating their young children. They just want them to do something, without taking care to lay out their reasons. Interestingly, the manipulation as negligent influence account thus explains why there are norms to increasingly interact with children in 'reason revealing' ways as they get older. On my account, this suggests that manipulation is less and less morally acceptable as we interact with increasingly agential interlocutors.

The negligence account may also explain why manipulation is frequently seen as wrong because it undermines autonomy. Though there is, arguably, no conceptual connection between manipulation and autonomy loss (Klenk and Hancock 2019), the carelessness account is well situated to explain why manipulation frequently (but not necessarily) leads to autonomy loss. That is because manipulative influences are chosen for being effective (in creating some change) but not necessarily for being reason revealing. If the latter is only a secondary thought, there is less opportunity for the manipulatee to accurately understand her situation and to govern her own affairs (cf. Baron 2014).

Furthermore, the negligence account explains the intuition that manipulation undermines people's agency or fails to respect them as agents and, therefore, is morally problematic. Manipulators indeed treat manipulatees like objects or automatons who sometimes happen to 'run on' reasons. The negligence approach can account for that intuition. But unlike other proposals, who locate the

wrong of manipulation in the failure of the manipulator to respect or care for her interlocutor's ends, the negligence approach is less demanding. It does not require that all our social influences originate in an attempt to make others or help others do what they have reason to do. We can want others to do something for us just because it is good for us. But to avoid manipulation, we should be motivated to influence them so that they do that because they see their reasons for it. The respect for someone else's agency that may make manipulation morally problematic is located in the failure to cater to others's reason responsiveness.

A final consideration in favour of the negligence conception of interpersonal manipulation concerns its assessment. Several writers have speculated that manipulation may not have been a focus of much philosophical and especially legal attention because of its intentional nature and consequent problems with assessing it (cf. Sunstein 2015). The negligence conception of manipulation may improve the situation because it allows for a conditional analysis of manipulation. That, in turn, is amenable to causal modelling (Pearl, Glymour, and Jewell 2016). Roughly, a given interpersonal influence can be identified as manipulation if the (counterfactually modelled) presence of reason revealing means of influence goes together with the manipulator's choice of a means of influence that is not reason revealing. Such causal models for the assessment of manipulation may provide interesting insights and shed new lights on debates within the ethics and policy of technology, which have often focused on questions about (technology-mediated) manipulation without a methodology able to further substantiate such claims.

6 Conclusion

Interpersonal manipulation is an important and ubiquitous social phenomenon. It has only begun to receive sustained attention in the philosophical literature, driven partly by increasing concerns about technologically-mediated influences. In this paper, I have argued that what emerged as the most promising account of manipulation, the trickery view, is misguided. The trickery view suggests that a manipulator intends some norms that pertain to the beliefs, desires, or emotions of the manipulatee. However, several plausible cases of manipulation like peer pressure or paternalizing influences cannot be explained by the trickery account. I have shown that rather than trying to extend the trickery view by looking for further norms that manipulators may intend to violate in these cases a more parsimonious and explanatory powerful account explains manipulation in terms of negligent influence. Accordingly, manipulation is characterised by the *lack* of intention to reveal reasons to one's interlocutor, rather than the explicit intention to harm one's interlocutor (by making him or her violate some norms of belief, desire, or emotion). The negligence account of manipulation represents an important shift in our understanding of manipulation away from undue focus on what manipulators want toward what they fail to do.

The present paper has aspired to demonstrate the need for and the initial persuasiveness of the negligence account of manipulation. Important questions remain. Future work should be directed at clarifying the account and spelling out its implications for the assessment of manipulation. Regarding the former, there are important open questions about the notion of reason revealing that is at the heart of the negligence account of manipulation and how to understand and

demarcate different means of influence in interpersonal interaction. Regarding the latter, it would be interesting and fruitful to assess and demonstrate the strength of the negligence account in a concrete causal model of manipulation.

References

- Achen, Christopher H., and Larry M. Bartels. 2016. *Democracy for Realists: Why Elections Do Not Produce Responsive Government* / Christopher H. Achen, Larry M. Bartels. Princeton studies in political behavior. Princeton: Princeton University Press.
- Barnhill, Anne. 2014. "What Is Manipulation?" In Coons and Weber 2014.
- Barnhill, Anne. 2016. "I'd Like to Teach the World to Think: Commercial Advertising and Manipulation." *JMB* 1 (3-4): 307–28.
<https://doi.org/10.1561/107.00000020>.
- Baron, Marcia. 2003. "Manipulativeness." *Proceedings and Addresses of the American Philosophical Association* 77 (2): 37. <https://doi.org/10.2307/3219740>.
- Baron, Marcia. 2014. "The Mens Rea and Moral Status of Manipulation." In Coons and Weber 2014, 98–109.
- Beauchamp, Tom L. 1984. "Manipulative Advertising." *Business and Professional Ethics Journal* 3: 1–22.
- Bélohrad, Radim. 2019. "The Nature and Moral Status of Manipulation." *Acta Analytica*. <https://doi.org/10.1007/s12136-019-00407-y>.
- Buss, Sarah. 2005. "Valuing Autonomy and Respecting Persons: Manipulation, Seduction, and the Basis of Moral Constraints." *Ethics* 115 (2): 195–235.
<https://doi.org/10.1086/426304>.
- Cholbi, Michael. 2014. "The Implications of Ego Depletion for the Ethics and Politics of Manipulation." In Coons and Weber 2014, 201–20.

- Cohen, Shlomo. 2018. "Manipulation and Deception." *Australasian Journal of Philosophy* 96 (3): 483–97. <https://doi.org/10.1080/00048402.2017.1386692>.
- Coons, Christian, and Michael Weber. 2014a. "Manipulation: Introduction." In Coons and Weber 2014, 1–16.
- Coons, Christian, and Michael Weber, eds. 2014b. *Manipulation: Theory and Practice*. Oxford: Oxford University Press.
- Engelen, Bart. 2019. "Nudging and Rationality: What Is There to Worry?" *Rationality and Society* 31 (2): 204–32. <https://doi.org/10.1177/1043463119846743>.
- Engelen, Bart, and Thomas Nys. 2020. "Nudging and Autonomy: Analyzing and Alleviating the Worries." *Rev.Phil.Psych.* 11 (1): 137–56. <https://doi.org/10.1007/s13164-019-00450-z>.
- Faden, Ruth R., and Tom L. Beauchamp. 1986. *A History and Theory of Informed Consent*. New York: Oxford University Press. <http://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&db=nlabk&AN=150188>.
- Fallis, Don. 2010. "Lying and Deception." *Philosophers' Imprint* 10 (11): 1–22. Accessed February 11, 2020.
- Fischer, John Martin, and Mark Ravizza. 1998. *Responsibility and Control: A Theory of Moral Responsibility*. Cambridge: Cambridge University Press.
- Gorin, Moti. 2014a. "Do Manipulators Always Threaten Rationality?" *American Philosophical Quarterly* 51 (1): 51–61. Accessed June 04, 2019.

- Gorin, Moti. 2014b. "Towards a Theory of Interpersonal Manipulation." In Coons and Weber 2014, 73–97.
- Hanna, Jason. 2015. "Libertarian Paternalism, Manipulation, and the Shaping of Preferences." *Social Theory and Practice* 41 (4): 618–43. Accessed November 27, 2019.
- Klenk, Michael. 2021. "Manipulation (Online): Sometimes Hidden, Always Careless." *Review of Social Economy*.
<https://doi.org/10.1080/00346764.2021.1894350>.
- Klenk, Michael, and Jeff Hancock. 2019. "Autonomy and Online Manipulation." *Internet Policy Review*. <https://policyreview.info/articles/news/autonomy-and-online-manipulation/1431>. Accessed February 28, 2020.
- Krstić, Vladimir, and Chantelle Saville. 2019. "Deception (Under Uncertainty) As a Kind of Manipulation." *Australasian Journal of Philosophy* 97 (4): 830–35. <https://doi.org/10.1080/00048402.2019.1604777>.
- Manne, Kate. 2014. "Non-Machiavellian Manipulation and the Opacity of Motive." In Coons and Weber 2014.
- Mills, Claudia. 2014. "Manipulation as an Aesthetic Flaw." In Coons and Weber 2014.
- Noggle, Robert. 1996. "Manipulative Actions: A Conceptual and Moral Analysis." *American Philosophical Quarterly* 33 (1): 43–55.
- Noggle, Robert. 2018a. "Manipulation, Salience, and Nudges." *Bioethics* 32 (3): 164–70.

- Noggle, Robert. 2018b. "The Ethics of Manipulation." In *Stanford Encyclopedia of Philosophy: Summer 2018*, edited by Edward N. Zalta. Summer 2018.
- Nys, Thomas R. V., and Bart Engelen. 2017. "Judging Nudging: Answering the Manipulation Objection." *Political Studies* 65 (1): 199–214.
<https://doi.org/10.1177/0032321716629487>.
- Pearl, Judea, Madelyn Glymour, and Nicholas P. Jewell. 2016. *Causal Inference in Statistics: A Primer*. Hoboken, New Jersey: John Wiley & Sons.
- Schelling, Thomas C. 1997. *Strategy of Conflict*. Cambridge, MA: Harvard Univ. Press.
- Sunstein, Cass R. 2015. "Manipulation, Welfare, and Dignity: A Reply." *SSRN Journal*. <https://doi.org/10.2139/ssrn.2661534>.
- Sunstein, Cass R. 2016. "Fifty Shades of Manipulation." *Journal of Behavioural Marketing* 213. <https://doi.org/10.2139/ssrn.2565892>.
- Thaler, Richard H., and Cass R. Sunstein. 2009. *Nudge [Electronic Resource]: Improving Decisions About Health, Wealth, and Happiness / Richard H. Thaler, Cass R. Sunstein*. Revised and expanded edition. New York: Penguin Books.
- Wood, Allen W. 2014. "Coercion, Manipulation, Exploitation." In Coons and Weber 2014, 17–50.