

DO MANIPULATORS ALWAYS THREATEN RATIONALITY?

Moti Gorin

I. INTRODUCTION

If there is a dominant view of interpersonal manipulation in the philosophical literature, it is the view that interpersonal manipulation occurs only if an influencer intentionally bypasses or subverts the rational capacities of the person he seeks to influence.¹ I believe this claim about the nature of manipulative influence draws plausibility from two main sources. First, there is a broad range of cases in which it is true that the rational capacities of the manipulated person are bypassed or subverted. Several such cases will be discussed below. Second, because manipulateness is viewed as a negative character trait, the concept “manipulation” is typically understood in a highly moralized manner. Consequently, it may be assumed that forms of interpersonal influence that are generally taken to be morally benign or even exemplary—for example, rational persuasion—cannot be used manipulatively. The thought is something like this: if manipulation is impermissible, *pro tanto* or otherwise, while rational persuasion is permissible, then rational persuasion cannot involve manipulation and manipulation cannot involve rational persuasion. Since rational persuasion, which is morally benign or even exemplary, always involves, or just is, engagement with the ratio-

nal capacities of the agent being influenced, bypassing or subverting these capacities is morally problematic. Therefore, manipulation always involves, or just is, the bypassing or subversion of an agent’s rational capacities, and this is what renders it morally wrong.

I do not mean to suggest that the line of reasoning sketched above is the only one available to someone wanting to defend an account according to which manipulation necessarily involves the bypassing or subversion of the manipulated agent’s rational capacities. However, I do think that if one were to defend what I will call the Bypass or Subvert View (BSV) of manipulation, one probably would want to appeal to cases and to emphasize the differences between typical cases of manipulation and typical cases of rational persuasion. In any case, I do not wish to defend BSV, for I think it is false. In what follows I will argue for this claim. After providing several interpretations of what it is to bypass or subvert a person’s rational capacities, I show that each interpretation is consistent with the presence of manipulation in the history of the process that led to the behavior in question. In arguing against BSV, I draw a rather surprising conclusion, which is that one agent may be manipulating another even when the only form of influence she uses is the provision of good reasons or sound arguments.

II. MANIPULATION AND THE RATIONAL CAPACITIES

Before moving on to criticize BSV, it is necessary to get clear about its central claim. I understand the claim as a disjunction:

BSV: interpersonal manipulation is a process of influence that necessarily either bypasses or subverts the rational capacities of the person whose behavior is being influenced.

I will begin by addressing the first disjunct—that is, the claim that manipulation *bypasses* a manipulee's rational capacities—and then move on to the second—that is, the claim that manipulation *subverts* a manipulee's rational capacities. In order to assess BSV, it will be helpful first to characterize the rational capacities in some way. There are many challenging philosophical questions about rationality and its realization in agents, for example, questions about what sorts of psychological states contribute to making up an agent's rational self. I cannot here address these questions or provide anything like an exhaustive list of the rational capacities. For current purposes I will rely on what should be a relatively uncontroversial characterization of the rational capacities.

Rational Capacities: those capacities that enable agents to assess and revise their beliefs in accordance with the basic canons of logic; to evaluate their epistemic and practical options against criteria generated by their beliefs, values, and preference sets; to make adjustments to these beliefs, values, and preference sets in light of new information; and to act in accordance with their judgments about what they have most reason to do.²

To say that one person bypassed the rational capacities of another may be taken to mean that, in influencing someone, the influencer made use of a means of influence that did not engage the influenced person's rational capacities at all. The sort of example that motivates this view usually involves a manipulator who has direct access to the

causal mechanisms underlying the behavior of the manipulee. Take, for example, Harry Frankfurt's famous would-be manipulator, Black, who, through the use of high-tech gadgetry, has the power to control the neurophysiological goings-on in the brain of Jones such that Black can immediately determine how Jones chooses to and indeed does act.³ Or consider Alfred Mele's case of Beth, an academic working in a department overseen by a dean who wishes Beth were more industrious. The dean hires a team of very capable psychologists who learn what makes Beth tick and then, via sophisticated brainwashing techniques, directly instill in Beth the mental states that make her a highly motivated and productive scholar while eradicating whatever values or preferences were earlier holding her back from diligently pursuing her work.⁴

Cases like Frankfurt's and Mele's describe manipulators who make use of means of influence that entirely bypass the rational capacities of the agents whose behavior they wish to control. The rational capacities of the influenced person play no role in the processes that determine her behavior. However, there are many cases of manipulation in which the rational capacities of the manipulee do play a mediating role in the process of influence. Here is one such case:

Election: Jones is campaigning to become President of the United States. He knows he needs substantial support among religious conservatives if he is to have any chance of winning the election. In order to appear more attractive to members of this demographic, Jones regularly invokes Scripture while advocating in favor of his political platform at public appearances. Jones is very skeptical about the existence of God, the truth of Scripture, and all other claims the belief in which constitute (or partly constitute) the religious orientation of the voters to whom he is trying to appeal.

If Jones's use of religious rhetoric plays a substantial role in the explanation of why some

religious conservatives vote for him, then it is plausible that he has manipulated these voters and that their voting behavior is a product of his manipulation. Nevertheless, Jones has not bypassed the rational capacities of these voters. In fact, the success of Jones's strategy crucially depends on these capacities. After all, Jones intends his audience to perceive his biblical references as providing them with a reason to vote for him. He rightly assumes that given the preferences, beliefs, and values of his audience, their belief that he is in the relevant sense "like them" will motivate them to vote for him. If he did not believe this, if he believed instead that religious conservatives were incapable of recognizing that he was providing them with an apparent reason to vote for him, he would not have appealed to their capacity to make the inferences he intended they make. Thus, *Election* shows that manipulation need not involve the bypassing of the rational capacities.

I believe the natural response to this case is simply to concede that the requirement that manipulation entirely bypass the rational capacities is too strict, and to then focus on whatever features of the case seem troubling. For example, despite Jones's engagement with voters' rational capacities, it remains true that he intended them to behave in ways they would be unlikely to behave if they were better informed about the features of the options about which they were deliberating. The thought is that although he did engage the voters' rational capacities, Jones did so in a way that subverted these capacities. Thus, though manipulation need not involve the wholesale circumvention of the rational capacities, it is still open that it involves their subversion.

Iia. Subversion as Active Interference

According to one possible conception of what it is to subvert the rational capacities, subversion is best understood as *active interference* with those capacities.

Active Interference Subversion: to interfere directly with a person in such a way as to generate psychological states the presence of which is incompatible with the proper functioning of the person's rational capacities.

Often, in trying to shape others' behavior, manipulators elicit psychological states that are incompatible with the manipulees' ability accurately to represent and assess their situations or to behave in a manner that is consistent with their assessments. The following two cases serve as paradigmatic examples of this:

Theater: I have grudgingly agreed to attend the opening of a play with you. Halfway to the theater, I "engag[e] your sublimated compulsive tendency to check the stove" and you turn back toward home. As a result, we miss the play, as I intended.⁵

Legislation: Some elected officials wish to pass legislation because doing so will allow them to tighten their grip on power while enriching their political patrons. They know this particular piece of legislation will be more likely to gain popular support if it is viewed by a fearful and anxious public as a security-enhancing measure. The officials or their representatives make fear-inducing statements through a compliant media before pushing publicly for their bill, which then passes with little public opposition.

In *Theater* the manipulator induces psychological states that impede the manipulee from acting in light of her considered judgments about what she has most reason to do. The manipulee initially decided to go to the play, and we may suppose this decision was the product of rational deliberation. After beginning to implement the plan that would allow her to carry out her intention, she finds herself strongly drawn toward another course of action that is inconsistent with her earlier plan, and the new course of action is not one on which she had rationally settled. Her rational capacities—in this case her capacity to act consistently with her judgment about what she has most reason to do—have been

subverted by the manipulator's stimulation of her compulsion.

In *Legislation* the psychological states induced by the officials interfere with the ability of citizens to evaluate the rationale for the bill and to assess the full ramifications of its passing. Fearful citizens are likely to assign disproportionate weight to the value of policies they perceive as promoting their safety, and thus by identifying the proposed legislation with the promise of security while simultaneously scaring citizens, the officials pervert the deliberations of the citizens whose support (or, more accurately, absence of opposition) they seek.

Commonplace examples like *Theater* and *Legislation* reinforce the view that manipulation is a matter of actively interfering with the rational capacities of the manipulee. There are countless examples of manipulation like this where manipulators "push the buttons" of manipulees, giving rise to psychological states whose effect is to overwhelm the manipulee's ability to assess and revise her beliefs in accordance with the basic canons of logic; to evaluate her epistemic and practical options against criteria generated by her beliefs, values, and preference sets; to make adjustments to these beliefs, values, and preference sets in light of new information; or to act in light of her considered judgments about what she has most reason to do.

However, though manipulators often actively interfere with the rational capacities of the people they are trying to influence, they do not always do so. Sometimes a manipulator will take a more hands-off approach and merely exploit an inherent flaw in the rational capacities of the manipulee. Drawing on recent research in behavioral economics, Cass Sunstein and Richard Thaler describe how an influencer's knowledge of others' cognitive biases can help the influencer shape the behavior of those she seeks to influence.⁶ To take just one example, research shows that the decisions of medical patients regarding

potential treatments can be strongly influenced by the way the information about the outcomes of the treatments is framed.⁷ The following imaginary (though realistic) example illustrates how this works.

Futile Treatment: Dr. Rasmussin's patient, Ms. Jackson, is very ill. Ms. Jackson is ninety years old and, in the judgment of Dr. Rasmussin, has a life expectancy of no more than six months. One of her non-life-threatening ailments is curable, but the treatment is very expensive and it requires the devotion of scarce medical resources. Dr. Rasmussin believes that providing this treatment to Ms. Jackson would be futile as she very likely will not live long enough to enjoy its benefits. Moreover, if Ms. Jackson gets the treatment, then some other younger or healthier patient who would enjoy its benefits will not receive it. In the judgment of Dr. Rasmussin, Ms. Jackson should not receive the treatment. However, Dr. Rasmussin knows that Ms. Jackson believes that when it comes to improving her health and extending her life, nothing should be regarded as futile. She is adamant that Dr. Rasmussin should provide the treatment. Dr. Rasmussin has no intention of providing the treatment but does not want to unnecessarily alienate or hurt his patient by expressing his unvarnished opinion about the futility of treating her. Instead, in discussing the matter with Ms. Jackson, the doctor makes use of a particular cognitive bias, sometimes referred to as "the framing effect." Rather than truthfully telling Ms. Jackson that 90 percent of the patients who receive the treatment survive, he truthfully tells her that 10 percent do not survive. Upon learning this, Ms. Jackson judges that the treatment is too risky and decides to "refuse" the treatment.

Assuming that in this case the framing effect played a decisive role in shaping Ms. Jackson's decision—that is, assuming that she would have continued to demand the treatment had Dr. Rasmussin framed the information in terms of survival rates rather than in terms of mortality rates—it is plausible that Ms. Jackson's decision to refuse the treatment and her remaining positively

disposed toward her doctor are products of Dr. Rasmussin's manipulating her.⁸ That is to say, intuitively Dr. Rasmussin manipulated Ms. Jackson into "refusing" the treatment and into agreeing to the course of action that Dr. Rasmussin favored.⁹ Dr. Rasmussin manipulated Ms. Jackson and yet Dr. Rasmussin did not directly interfere with Ms. Jackson's rational capacities, at least insofar as he did not stimulate psychological states whose presence is incompatible with or threatening to her ability effectively to deliberate about her options and to act in light of her considered judgments. Thus, if manipulation subverts the manipulee's rational capacities, it must do so in a way that does not require the direct interference with those capacities.

Iib. A Narrow Teleological Interpretation of "Subversion"

Cases like *Futile Treatment* suggest that if manipulation is to be understood as the subversion of the manipulated person's rational capacities, we need a conception of "subversion" that does not entail a manipulator's direct interference with a manipulee's rational capacities. Such a conception would cover cases of active interference but would be more inclusive in order to capture other cases where the manipulated person's rational capacities are impeded in some way, but where the presence of the impediment is not something for which the manipulator is responsible.

An account of subversion that focuses on the function of the rational capacities rather than on the etiology of the mechanism that undermines them will capture cases of direct agential interference like those described by Frankfurt and Mele as well as those like *Futile Treatment*, in which the manipulator merely exploits an already existing cognitive defect. On this view, to influence someone in a way that subverts her rational capacities is *Narrow Purpose Subversion*.

Narrow Purpose Subversion: to cause a behavior-underlying change in the person via a process

that impedes the person's rational capacities from fulfilling their function.

This construal of what it is to subvert the rational capacities explains the judgment that Dr. Rasmussin has indeed influenced Ms. Jackson via a process that subverted her rational capacities. By providing Ms. Jackson with information framed in terms of mortality rates rather than in terms of survival rates, Dr. Rasmussin succeeded in getting Ms. Jackson to make a decision she would not otherwise have made, given her set of beliefs, values, and preferences. Dr. Rasmussin decided to exploit the framing effect because he knew that Ms. Jackson's background attitudes would make it rational for her to insist on the treatment. Considering the significant weight Ms. Jackson places on the value of medical interventions, to decide against such an intervention merely on the basis of how information is presented to her rather than on the substance of that information is plausibly to have behaved irrationally, and thus by targeting one of her cognitive biases Dr. Rasmussin impeded Ms. Jackson's rational capacities from fulfilling their function. Given the set of her beliefs, values, and preferences, in medical contexts Ms. Jackson aims to maximize her chances of improving her health and extending her life. Dr. Rasmussin's intervention undermined her ability to achieve this aim. And because the function of the rational capacities narrowly understood is to help an agent achieve her ends, which ends are products of her set of attitudes, Dr. Rasmussin has subverted Ms. Jackson's rational capacities.

Though the *Narrow Purpose Subversion* view is an improvement on the *Active Interference* view insofar as the former is able to account for a wider range of cases that intuitively qualify as cases of manipulation, it too succumbs to counter-examples. Some cases of manipulation do not undermine the ability of the rational capacities to fulfill their function but actually enhance this ability or even

supplant it. For example, a doctor may exploit the framing effect in order to get a severely depressed patient to make a decision that is consistent with the patient's considered judgments, but which is difficult for her to make while in the midst of a bout of depression.¹⁰ In such a case, the depression undermines the ability of the rational capacities to fulfill their function while the doctor's focused use of the framing effect *enhances* this ability. Here the framing effect functions as a kind of proxy for the rational capacities.

The *Narrow Purpose Subversion* version of BSV also fails to capture some cases of manipulation in which the rational capacities of the manipulee are in no way inhibited from fulfilling their function. Consider the following example:

Lucrative Suicide: After a long period of philosophical reflection, Jacques becomes convinced that in the absence of God, life has no meaning. He also firmly believes that if life has no meaning, he has no reason to continue living, for a life without meaning would be for Jacques little more than a stretch of suffering and boredom. But Jacques believes in God and he believes that God's existence lends meaning to life. Thus, he is motivated to continue living his life. James stands to inherit a nice sum of money upon the death of his cousin Jacques. James sets out to convince Jacques that his theism is unfounded with the intention that Jacques's acceptance of this claim will lead to his suicide. James finds the most powerful anti-theistic arguments available and presents them to Jacques who, after a period of reflection, sees the arguments to the end—the very end.

James does not stimulate psychological states that are incompatible with Jacques's ability carefully to reflect upon his attitudes. Nor does James exploit some inherent cognitive bias of Jacques's or otherwise hinder Jacques's rational capacities from fulfilling their function. Given Jacques's considered beliefs, values, and preferences, his action is rendered rational. James does nothing to

undermine Jacques's ability to deliberate calmly and clearly about his options or to act in light of his considered judgment about what he has most reason to do.

IIc. A Wide Teleological Interpretation of "Subversion"

Perhaps what the case of Jacques and his conniving cousin shows is not that manipulation need not impede the rational capacities from fulfilling their function, but rather that the function of the rational capacities should be understood in some other way. Thus far I have assumed that the purpose of the rational capacities is to help agents achieve their ends, given their current attitudes, values, and preferences. This conception of the rational capacities opens the door to cases in which a manipulator appeals to propositional attitudes with problematic content—for example false beliefs—in order to get the agent who holds these attitudes to behave in ways that are internally consistent with the agent's other attitudes and preferences but which are, from a more objective standpoint, unreasonable. Given Jacques's beliefs, desires, values, and so on, his acquisition of the belief that there is no God may have made it rational for him to kill himself. Nevertheless, we may want to say that his suicide was unreasonable. Perhaps Jacques should not have believed that in the absence of God life lacks meaning, or that suicide is the appropriate response to a meaningless existence. Perhaps he should not have allowed abstract metaphysical arguments to move him to take such drastic action even if a warrant for such action was the upshot of his rational deliberation.

When James convinces Jacques that there is no God, he provides Jacques with a motivating reason to take his own life—that is, a reason that plays a role in explaining Jacques's subsequent behavior.¹¹ What he arguably does not provide, however, is a reason that justifies Jacques's suicide, a reason an appeal to which renders Jacques's action not only

consistent with the attitudes he does have, but consistent with the attitudes he *ought* to have. The thought here is that the function of the rational capacities is best understood at least in part in terms of their linking up with whatever reasons there are, irrespective of whether or not these reasons currently play any role in the agent's deliberation or action. On this view, to influence an agent in a way that subverts her rational capacities is *Wide Purpose Interference*.

Wide Purpose Interference: to cause a behavior-underlying change in the agent via a process that impedes the agent's rational capacities from fulfilling their function, where the function of the rational capacities is to guide an agent toward behavior that is supported by whatever reasons there are, irrespective of whether or not these reasons currently play any role in the agent's belief and preference sets.

In her essay on manipulation in politics, Claudia Mills articulates a view of manipulation that moves in the direction just sketched. According to Mills, manipulation

in some way purports to be offering good reasons when in fact it does not. A manipulator tries to change another's beliefs and desires by offering her bad reasons, disguised as good, or faulty arguments, disguised as sound—where the manipulator himself knows these to be bad reasons and faulty arguments. A manipulator judges reasons and arguments not by their quality but by their efficacy. A manipulator is interested in reasons not as logical justifiers but as causal levers. For the manipulator, reasons are tools, and a bad reason can work as well as, or better than, a good one.¹²

According to this account, James has manipulated Jacques because he has knowingly disguised a bad reason or faulty argument to commit suicide as a good reason or sound argument to do so. But has James done this? It seems not. Rather than presenting God's non-existence as a good reason for suicide, he exploited Jacques's belief that it was such a reason. Thus, Mills's proposal needs to be

amended to say that a manipulator either knowingly offers bad reasons or arguments as good ones or exploits the manipulee's already mistaking the former for the latter. A person who deliberates on the basis of false beliefs or who makes fallacious inferences will often arrive at mistaken conclusions about what she ought to believe or to do. Thus, an influencer who provides defective arguments or reasons or who exploits the presence of false beliefs or the tendency to reason in a defective manner can fairly be said to subvert the rational capacities of the person she influences. A teleological interpretation of rational capacity subversion that takes a broader view of the purpose of the rational capacities can make sense of the intuition that James has subverted Jacques's rational capacities, and thus it provides a more compelling account of the relation between the rational capacities and manipulation.

But this account of manipulation will not work, either. To see why, notice that there is a tension between, on the one hand, Mills's observation that manipulators judge reasons and arguments by their causal efficacy and not their justificatory quality and, on the other hand, her central claim that manipulation is a matter of passing bad reasons or arguments off as good ones. She rightly points out that as a causal lever, "a bad reason can work as well as, or better than, a good one," but she does not note that the converse of this is true as well. That is, *as a causal lever, a good reason can work as well as, or better than, a bad reason*. If a manipulator is indifferent to the justificatory quality of reasons, caring only about their causal efficacy, then it seems that she will use good reasons—that is, reasons that really do justify—when these are more effective at bringing about the behavior at which she is aiming. When a manipulator makes use of good reasons or arguments, the justificatory quality of the causally effective reason or argument will be merely incidental for her. Yet it

does not follow from this that she knowingly disguises a bad reason as a good one.

A similar objection can be brought against Robert Noggle's account of manipulative action, on which account "manipulative action is the attempt to get someone's beliefs, desires, or emotions to violate [relevant] norms, to fall short of these [relevant] ideals."¹³ According to Noggle, manipulators aim to bring about behavior that falls short of the manipulator's epistemic, conative, and emotional ideals. Insofar as an attempt to get someone to behave in ways that fall short of one's ideals is equivalent to an attempt to get someone to behave in ways one believes to be unsupported by reasons, Noggle seems to be committed to something like the *Wide Purpose Interference* interpretation of BSV.¹⁴

But if manipulators sometimes traffic in what they take to be good reasons, then even the wide teleological interpretation of what it is to subvert the rational capacities fails. According to this interpretation, to subvert the rational capacities is to cause a behavior-underlying change in the agent via a process that impedes the agent's rational capacities from fulfilling their function, where the function of the rational capacities is enlarged to include the agent's satisfaction of the demands of the objectively reasonable. Here are two counter-examples to the claim that manipulation necessarily involves the subversion of the rational capacities, where the latter are understood according to the *Wide Purpose Interference* definition:

Trust Me: Suppose I intend to tell you a lie two months from now. The lie is going to be so egregious that I am not very confident that you will believe it when the time comes. In order to gain your trust, over the next two months I offer you sensible advice, I convince you about various matters by constructing sound arguments, I make many true and easily verifiable claims, I criticize others when they lie, and so on. As it turns out, I must permanently leave the country just before the two months are up and consequently never deliver the lie.

Global Warming: Candidate Green is running for President of the United States. In her view, the most significant problem we face today is global warming and her presidential run is motivated *exclusively* by her desire to implement policies that will significantly decrease the quantities of greenhouse gases that are being released into the atmosphere. Green knows she cannot win the election if she openly makes the reduction of greenhouse gases the only—or even the central—plank of her political platform. Green consults with polling experts to determine which issues most exercise voters and she decides in advance that she will adopt the policies that will help her win office, irrespective of whether these are policies that she personally supports. Green adopts the most popular positions—with the exception of her position on greenhouse gases—and after arguing persuasively in favor of these policies, goes on to win the election. Most of the citizens who voted for her oppose her environmental policies, which policies she begins aggressively to implement once elected.

In *Trust Me* it is plausible that when I give you a sound argument tomorrow or next week, I am manipulating you; when you make a good choice due to my having given you sensible advice, your choice is (at least partly) the product of manipulation; when you judge that I do not tell you things I believe to be false, your judgment has been manipulated. If a year from now you read my journal and discover my plot, it will be perfectly reasonable for you to judge that I was manipulating you during these two months, that your coming to trust me was a product of my manipulating you by giving you sage advice and good arguments. And yet while I manipulate you, I fully engage your rational capacities. The process of influence that leads to your trusting me did not bypass your rational capacities, it did not actively interfere with these capacities, and it did not exploit an inherent flaw in these capacities or otherwise hinder these capacities from guiding you toward reason-supported behavior. You had good reason to

trust me during those two months, as I made every effort to help you form true beliefs and to behave in accordance with the dictates of practical reason. Moreover, whatever reason you might otherwise have had to withhold your trust—namely, that I intended to lie to you—you do not have here, as the state of affairs that generates this reason would not obtain due to my having to leave the country.

Similarly, in *Global Warming*, Green engages fully with the rational capacities of voters and she neither actively interferes with the proper functioning of these capacities nor inhibits them in any way from fulfilling their function. We may go further to stipulate that the popular policies were the objectively correct policies, so that the rational capacities of the voters really did guide them toward behavior that is supported by objective reasons. Thus, citizens who vote for Green on the basis of her platform behave in a reason-supported manner. Nevertheless, it seems that Green acted manipulatively when she constructed a political platform strictly on the basis of its popularity and only because doing so would allow her to implement her favored unpopular emissions policies.

Global Warming and *Trust Me* are counter-examples to the *Wide Purpose Interference* account of rational capacity subversion. They show that an agent can do what she has good reason to do, do it in light of those reasons, and yet be manipulated into doing it. Because manipulators care only about the causal

efficacy of reasons and not about their justificatory qualities, they will appeal to good reasons when they judge that these will be effective in bringing about the outcome they seek.

III. CONCLUSION

The account of manipulation according to which manipulation essentially involves the bypassing or subversion of the manipulated agent's rational capacities is attractive. The Bypass or Subvert View (BSV) postulates a unifying property of manipulation that both organizes our intuitions about a wide range of cases and purports to explain how manipulation differs from the proffering of good reasons or arguments. However, I have argued that each of several interpretations of BSV is vulnerable to counter-examples, and therefore that the dominant view of interpersonal manipulation is false. Moreover, some of the counter-examples—*Trust Me* and *Global Warming* in particular—reveal an interesting and perhaps rather surprising truth about manipulation: that the provision of reasons and arguments—good reasons and sound arguments—can be used manipulatively.¹⁵ One implication of this result is that insofar as manipulation is thought to be morally problematic,¹⁶ providing others with good reasons and sound arguments can sometimes be morally problematic.

The University of Pennsylvania

NOTES

I wish to thank Baruch Brody, George Sher, Mark LeBar, and two anonymous referees at this journal for providing me with helpful comments on earlier drafts. The paper is stronger as a result of feedback I received from audiences at the 2012 Pacific Meeting of the American Philosophical Association, the 2012 Workshop in Applied Ethics and Policy at Bowling Green State University, and a 2012 Works in Progress meeting held by the Philosophy Department at Rice University.

1. This view has been advanced in one form or another by a number of philosophers. See Baron (2003), p. 50; Beauchamp and Childress (2008), pp. 133–134; Blumenthal-Barby and Burroughs (2012); Cave (2007), p. 138; Greenspan (2003), p. 164; Mills (1995), p. 100; Stern (1974), p. 74. Thomas Scanlon

argues that manipulation is morally objectionable because “[c]onsidering the matter from the point of view of potential victims of manipulation, there is a strong generic reason to want to be able to direct one’s efforts and resources toward aims one has chosen and not to have one’s planning co-opted.” Insofar as the rational capacities play a central role in helping an agent direct her energies toward aims she has chosen, manipulation will, on his view, subvert these capacities. Scanlon (1998), p. 298.

2. This conception of the rational capacities is similar to Eric Cave’s conception of the capacities that render an agent “modestly autonomous.” See Cave (2007), p. 138. Thus, my arguments regarding the relation between manipulation and the bypassing or subversion of the rational capacities apply to Cave’s account of motive manipulation, as he maintains that manipulation is wrong because it violates Modest Autonomy.

3. Frankfurt (1969), pp. 835–837. In the case described by Frankfurt, Black does not manipulate Jones because Jones does what Black wants him to do on his own. But if Black were to intervene in the causal processes leading up to Jones’s behavior, his doing so would constitute a case of manipulation amenable to BSV. I wish to thank an anonymous referee with this journal for pushing me to clarify this point.

4. Mele (1995), p. 145.

5. Cave (2007), p. 132.

6. Sunstein and Thaler describe some of these methods in their article “Libertarian Paternalism Is Not an Oxymoron” (2003), pp. 1159–2012; and also in their book *Nudge* (2008).

7. Sunstein and Thaler (2003), p. 1161. The paper Sunstein and Thaler cite to support their claim about the efficacy of framing medical outcomes in terms of survival versus in terms of mortality is Redelmeier et al. (1993).

8. I leave it open for now whether or not what Dr. Rasmussin did was morally permissible. At this stage in the argument, I am concerned with establishing that certain instances of influence are instances of manipulation, and not with establishing anything about manipulation’s ethical status.

9. Ms. Jackson did not really *refuse* the treatment because it was not genuinely open to her to accept the treatment, that is, it was not being offered to her. Dr. Rasmussin’s antecedent decision to refuse to provide the treatment rendered Ms. Jackson’s decision otiose, though, of course, she did not realize this.

10. It may be objected that in this case it is not clear that the doctor is manipulating her patient. I suspect this objection is grounded in the judgment that manipulation is morally wrong and therefore that this cannot be a case of manipulation because what the doctor does is not morally wrong. I share the intuition that the doctor does not act wrongly (all things considered), but nevertheless I think she does manipulate her patient. In other words, I do not hold that manipulation is always wrong or that it is even always *pro tanto* wrong. An adequate defense of this claim would require more space than I have here. In short, my view is that in this case manipulation is not even *pro tanto* wrong because (1) the manipulee is not sufficiently responsive to reasons, and (2) the manipulation is aimed at behavior that is supported by good reasons from the point of view of the manipulee. I thank an anonymous referee with this journal for raising the objection.

11. Parfit and Broome (1997), pp. 99–146.

12. Mills (1995), pp. 100–101.

13. Noggle (1996), p. 44. The relevant norms are epistemic, conative, or emotional, depending on whether the manipulated state is a belief, a desire, or an emotion.

14. Though I cannot adequately address them here, I believe Noggle’s account runs into problems because he overlooks cases of paternalistic manipulation—that is, manipulation that is aimed at bringing about behavior that the manipulator believes *is* supported by good reasons or, in Noggle’s favored

terminology, behavior that does not fall short of the manipulator's ideals. Sometimes the upshot of manipulative action is behavior the manipulator believes is reason-supported and consistent with the behavioral norms of the manipulator. Noggle's account cannot accommodate this fact.

15. I begin to sketch an alternative account of manipulation, one that can distinguish between the manipulative use of good reasons and sound arguments and their non-manipulative use in "Towards a Theory of Interpersonal Manipulation" (forthcoming).

16. The question of the moral status of manipulation is a difficult one. I regret that I do not have space to take it up here. See note 9 above for a very brief and incomplete statement of my view.

REFERENCES

- Baron, Marcia. 2003. "Manipulativeness," *Proceedings and Addresses of the American Philosophical Association*, vol. 77, no. 2, pp. 37–54.
- Beauchamp, Tom, and James F. Childress. 2008. *Principles of Biomedical Ethics* (6th edition) (New York: Oxford University Press).
- Blumenthal-Barby, Jennifer, and Hadley Burroughs. 2012. "Seeking Better Health Care Outcomes: The Ethics of Using the Nudge," *American Journal of Bioethics*, vol. 12, no. 2, pp. 1–10.
- Cave, Eric. 2007. "What's Wrong with Motive Manipulation?," *Ethical Theory and Moral Practice*, vol. 10, no. 2, pp. 129–144.
- Frankfurt, Harry. 1969. "Alternate Possibilities and Moral Responsibility," *Journal of Philosophy*, vol. 66, no. 23, pp. 829–839.
- Gorin, Moti. Forthcoming. "Towards a Theory of Interpersonal Manipulation," in *Manipulation: Theory and Practice*, ed. Michael Weber and Christian Coons (New York: Oxford University Press).
- Greenspan, Patricia. 2003. "The Problem with Manipulation," *American Philosophical Quarterly*, vol. 40, no. 2, pp. 155–164.
- Mele, Alfred. 1995. *Autonomous Agents* (New York: Oxford University Press).
- Mills, Claudia. 1995. "Politics and Manipulation," *Social Theory and Practice*, vol. 21, no. 1, pp. 97–112.
- Noggle, Robert. 1996. "Manipulative Actions: A Conceptual and Moral Analysis," *American Philosophical Quarterly*, vol. 33, no. 1, pp. 43–55.
- Parfit, Derek, and John Broome. 1997. "Reasons and Motivation," *Proceedings of the Aristotelian Society*, suppl. vol. 71, pp. 99–146.
- Redelmeier, Donald A., Paul Rozin, and Daniel Kahneman. 1993. "Understanding Patients' Decisions: Cognitive and Emotional Perspectives," *JAMA*, vol. 270, no. 1, pp. 72–76.
- Scanlon, Thomas. 1998. *What We Owe to Each Other* (Cambridge, MA: Harvard University Press).
- Stern, Lawrence. 1974. "Freedom, Blame, and Moral Community," *Journal of Philosophy*, vol. 71, no. 3, pp. 72–84.
- Sunstein, Cass R., and Richard H. Thaler. 2003. "Libertarian Paternalism Is Not an Oxymoron," *University of Chicago Law Review*, vol. 70, no. 4, pp. 1159–1202.
- . 2008. *Nudge* (New Haven, CT: Yale University Press).