What's Wrong with Manipulation in Education?

Ron Aboodi

The University of Toronto

1. INTRODUCTION

Education is an attempt to influence, and so is manipulation. However, manipulation in education can be morally problematic. I argue that one of the important potential problems is the danger of hindering *deliberative projects*, a notion that I will explain in the following section. This danger can render some manipulative educational practices morally wrong all-things-considered, or at least *pro tanto* wrongful in the sense that there's a reason to avoid them—a reason that may be outweighed by countervailing reasons. (Often, we cannot determine whether an instance of manipulation is right or wrong all-things-considered, without weighing the relevant reasons against each other.)

There have been various philosophical attempts to capture the essence of manipulation. Here are some rough formulations of a few proposed definitions: (1) Jason Hanna argues that (the central type of) manipulation involves the intentional use of nonrational means of influence that affect the target's deliberation for the worse.[1] (2) Robert Noggle argues that manipulating must involve getting someone to make a mistake.[2] (3) John Tillson defines manipulating as attempting to persuade with the attitude of choosing the best means for the sought persuasion, irrespective of any other consideration.[3]

My argument doesn't depend on any particular account of what counts as *manipulation*. I will conform to the ordinary usage of this term. As varied as the ordinary usage is, some general claims can safely be made. First, coercing someone or straightforwardly convincing her doesn't amount to manipulating her. Second, while it seems that morally innocuous manipulations are conceptually possible, often manipulation in education is morally *worrisome*, in the sense that we *suspect* (*prima facie*) that it is at least *pro tanto* wrongful. Third, not all manipulations are deceptive.[4] For example, a teacher

could deploy manipulative methods to inculcate true beliefs that the teacher believes in. Even in such a case, the deployment of manipulative methods seems suspicious. Fourth, non-deceptive manipulations can be worrisome even when they are legitimately *paternalistic*, in the sense that they aim at the good of the recipients of education.[5] Note that I will use the term *student* to refer to any recipient of education, and the term *educator* in a broad sense that includes not only teachers but also parents and relevant policy makers.

I'll focus on Nondeceptive Paternalistic Manipulation in Education—hereafter "NPME"—in situations where paternalism is justified, mainly because identifying the problems with such NPME is harder than ill-willed or deceptive manipulation. Here are a few examples of worrisome NPME:

> I. A teacher deters students from questioning certain values (whose adoption would be good for the students) by intentionally instilling shame or fear (without coercing the students).

> II. A mother influences her son to reject a particular view, which happens to be false, merely by an ad hominem attack on one of its supporters. (Assume that she should know better and that her ad hominem attack doesn't provide any evidence for the falsity of the view.)

> III. A father repeatedly plays on his daughter's guilt in order to make her work harder in school.

And yet, some NPMEs (or practices that *seem* like NPMEs) seem innocuous:

> IV. A teacher controls the release of materials in order to get students to appreciate the appeal of a popular yet wrongheaded argument before exposing them to its shortcomings.

> V. An instructor uses body language, tone of voice, and images in a Power-Point presentation that appeal to non-deliberative mechanisms in order to influence the students to pay more attention, maintain their focus, or to better remember the content.

How do we draw the line between such innocuous practices and *pro tanto* wrongful NPMEs? Moral evaluation is too complex for hard-and-fast rules, and different NPMEs could be problematic for different reasons.[6] Nevertheless, I will argue that there's one factor that can explain (to a large extent) the differences among the intuitive degrees of seeming *pro tanto* wrongness in examples *I-V*: the danger of hindering deliberative projects.

I explain what I mean by *deliberative projects* in the following section. Section 3 explains how NPME might hinder the students' future deliberative projects. Section 4 argues for the significance of this danger in education, by comparing it to the most ambitious, well-defended, and prominent philosophical diagnoses of the wrongness of manipulation. Section 5 touches on remaining ethical questions in regard to influencing the *objective quality* of the students' deliberative projects.

2. DELIBERATIVE PROJECTS

What I mean by *a person's deliberative project* is, roughly, her interrelated ongoing efforts to manage her conduct by committing and adhering to deliberative beliefs, ideals, policies, plans, goals, and the like.[7] My view on the nature of personal ideals, policies, and plans is inspired by Michael Bratman's work.[8] I see them as commitments whose primary function is to constrain the person's practical deliberations. The state of adhering to a policy, for instance, amounts to treating certain choices as settled in deliberation, at least tentatively. Once a person commits to a policy, it acquires a "default" status of non-reconsideration in related deliberations. Nevertheless, most commitments are *tentative* in the sense of allowing reconsideration when there's some justification to do so. All of this applies also to ideals, which have a higher level of generality than policies, and to plans, which have a lower level of generality than policies.

Managing one's conduct (in the ordinary use of this locution) differs from merely habitual behavior, instinctive or uncontrolled reactions, and being moved non-reflectively by emotional mechanisms or non-deliberative inclinations. I will refer to these types of motivations as *non-deliberative*. Nev-

ertheless, non-deliberative motivations can be part of a person's deliberative project if the person endorses or relies upon them when forming her plans or other commitments (without viewing such reliance as problematic). For instance, a person may consider a particular habit as conducive to one of her goals and rely on the workings of this habit when forming policies and plans. In such a case, if we interfere with this person's habit, we'd be interfering with her deliberative project. To take another example, a person may consider anger as legitimate in a particular type of situation. When she reflects on this type of situation she envisions being moved by anger as supported by her ideals, beliefs, and goals. As a result, in the relevant type of situation, calming down this person's anger might interfere with her deliberative project. By contrast, when a person's anger interferes with fulfilling her deliberative commitments—given that she hasn't endorsed or relied upon her anger when forming them—calming her down would probably be *conducive* to her deliberative project.

This explanation of the notion of *deliberative project* will suffice for my central argument, despite leaving room for further specification. (For instance, what unifies and how should we individuate deliberative projects? Should we sometimes acknowledge sub-personal or multi-personal deliberative projects? These are good questions for future research.)

3. THE RELEVANT THREAT TO DELIBERATIVE PROJECTS

Why are deliberative projects normatively significant? Some might argue that an obligation to respect people's deliberative projects stems from a deontological obligation to respect their autonomy or rational agency. By contrast, some consequentialists might explain the importance of respecting deliberative projects on the basis of some impersonal consequentialist principle.[9] Arguably, managing one's conduct as one sees fit is an important aspect of leading a good life. In this article I wish to remain neutral concerning the grounds of the normative significance of deliberative projects. Whichever ethical theory best explains this significance, educators should care about the students' deliberative projects.

This caring can be divided into two attitudes. One is a paternalistic attitude that strives to shape deliberative projects in an objectively good way. (Perhaps this attitude fits better with a consequentialist approach.) I leave this topic for future research, making only some preliminary remarks in section 5. The remainder of sections 3 and 4 will focus on the second attitude: a favorable, respectful attitude towards the actual deliberative projects that the students pursue at present and in adulthood, at least so long as these projects will conform to some normative standards of morality, and possibly *reasonableness*. Many views on the aims of education would endorse facilitating the successful fulfillment of such legitimate deliberative projects. An even broader consensus could be reached concerning the following weaker claim: educators have a *pro tanto* reason to *avoid hindering* the relevant deliberative projects.

How can educators facilitate (or at least, avoid hindering) future deliberative projects when they don't know the content of the students' future deliberative commitments (ideals, goals, plans, and so on)? I will focus on one proposed policy that is particularly relevant to the discussed danger in manipulation. It is the policy of cultivating (or at least, not hindering) the development of multi-purpose virtues (or character qualities) that are generally conducive to fulfilling deliberative projects. These virtues include intellectual virtues such as *open-mindedness* and *critical thinking*, and executive virtues such as *perseverance*, *courage*, and *patience*. I mean to use these terms here in a sense that renders these virtues sufficiently versatile to be generally conducive (and minimally detrimental) to succeeding in the broadest possible variety of (legitimate or reasonable) deliberative projects.[10]

Due to the nature of NPMEs, however, they might have side-effects that hinder the development or enactment of such multi-purpose virtues. To see this, focus on how such manipulation typically works. How can you influence students without straightforwardly convincing them or coercing them? What's left, usually, is either to mobilize their non-deliberative motivations (such as emotional mechanisms), or to divert their deliberative efforts (by concealing information, deceiving, or other types of trickery). Both of

these techniques risk conflicting with the cultivation of the relevant virtues. Systematic manipulative mobilizations of emotions sometimes risk creating emotional traumas or smaller emotional hindrances to the development or enactment of these virtues (especially the executive ones). Diversions of deliberative efforts sometimes risk forming bad deliberative habits, contrary to critical thinking. When such manipulative techniques (of both types) are deployed on children by someone in a position of authority or power, or by someone they look up to, their effects are more likely to persist longer and intensify, including the negative side effects. If the students imitate the educator, the problem is multiplied.

The greater the expected risk of such side effects, the more *pro tanto* wrongful the manipulation. This may be illustrated using the examples from my introduction. The manipulative mobilization of emotions can be illustrated with example *I*: deterring students from questioning certain views by instilling fear or shame. Such a practice would normally run against the cultivation and development of open-mindedness, critical thinking, and courage. The negative emotions might become associated with attempts to question or challenge the views of authoritative figures and social or institutional norms.

The danger of mobilizing emotional mechanisms can also be illustrated by example *III*. The guilt that the father repeatedly induces (in order to make his daughter work harder) risks creating relevant emotional problems. For example, it might contribute to the formation of a psychological problem of chronic guilt and low self-esteem, which runs against the cultivation or enactment of various executive virtues (such as *perseverance*). I suspect that many of us would estimate that this risk (in example *III*) is lower than the one in example *I*. I suspect that, to the same people, example *III* seems intuitively less worrisome than example *I*. If so, the risk seems to explain the worry (or at least part of it).

Now consider manipulative diversions of deliberative efforts by fallacious reasoning. Such manipulation normally runs against the cultivation and development of critical thinking (which is generally conducive to deliberative projects). This can be illustrated with example *II*, where the mother

argues against a particular view, which happens to be false, merely by an *ad hominem* attack on one of its supporters (an attack that doesn't provide any evidence for the falsity of the view). The most significant worry I see here is that, by doing so, the mother is setting a bad example. Namely, it might lead her son to deploy *ad hominem* fallacies in his reasoning in other areas, and in his argumentation with others, contrary to critical thinking.

The degree of the *pro tanto* wrongness in examples *I-III* seems to correlate with the risk of the side effects that I described above. This correlation seems to apply also to the remaining examples, which intuitively seem non-problematic: normally, when a teacher uses graphics, jokes, and references to popular culture to keep the students interested (example *V*), the risk of hindering their deliberative project seems minimal or nonexistent. This applies also to the case of getting students to appreciate the appeal of a popular yet wrongheaded argument by delaying their exposure to its shortcomings (example *IV*). If example *IV* seems somewhat more suspicious, I think this is partly because there's a chance that the teacher's plan would fail, leaving the students with uncritical acceptance of the wrongheaded argument (contrary to critical thinking).

These comparisons suggest that the risk of hindering deliberative projects provides (a large part of) the explanation of the intuitive degree of suspected *pro tanto* wrongness in each of examples *I-V*. In any regard, I hope I have said enough to establish that this risk is an important factor to take into account when we evaluate manipulations in education.

4. COMPARISON TO OTHER DIAGNOSES OF THE WRONGNESS OF MANIPULATION

To better appreciate the significance of the problem that I identified and characterized, it would be helpful to compare it to the most ambitious, well-defended, and prominent philosophical accounts of the wrongness of manipulation outside of the philosophy of education.

I begin with Noggle's account, according to which the wrongness of manipulation stems from inducing a mental state that falls short of ideal-

ly-rational functioning, constituting a mistake of some sort.[11] As Noggle explains, his account implies that it's necessarily *pro tanto* wrongful to place the healthier food at eye level in order to influence the customers to choose it, as in Thaler and Sunstein's famous Cafeteria case.[12] The only reason for the wrongness, according to Noggle, is that the salience of location of the salad influences the customers without constituting a consideration that *really* favors the salad. (Noggle is assuming that both the salad and the cake are equally available despite the fact that the cake is on a lower shelf.) Making the decision in an ideally rational manner, according to Noggle, requires disregarding such irrelevant information.

To examine the significance of the problem that Noggle characterizes in the context of education, let's go back to some examples from my introduction. In example *V,* the use of images for the purpose of influencing the students to pay more attention would turn out to be no less problematic, on Noggle's account, than the Cafeteria case (assuming the images don't constitute a consideration that really favors paying more attention to the presentation). But intuitively, such use of images is often morally innocuous, or at least insignificantly problematic. The same applies to example *IV* (controlling the release of materials in order to get students to appreciate the appeal of a popular yet wrongheaded argument before exposing them to its shortcomings). This examination suggests that the problem that Noggle characterizes—by itself, and insofar as it doesn't involve any threat to the students' deliberative projects—is insignificant for the evaluation of manipulation in education.

Furthermore: let's return to example *II.* The only relevant problem that Noggle's account helps us identify with this deployment of *ad hominem* is that it induces a mental state that falls short of ideally rational functioning (in Noggle's sense). But the negative impact of the bad example that the mother is setting on the development of the son's critical thinking seems much more significant. This illustrates, once again, that the problem that Noggle identified is less significant than the danger I identified, at least in the context of education.

Hanna argues that the (deontological) wrongness of manipulation stems from what he calls "worsening the manipulee's deliberative position" (in comparison to the manipulee's position had she not been manipulated). This amounts to causing the target to (1) act on the basis of bad reasons, (2) fail to adequately reflect on reasons, (3) attend to irrelevant considerations, or (4) place more (or less) weight on certain considerations than those considerations properly merit.[13] Moti Gorin takes a similar position, arguing that the wrongness of manipulation stems from deliberately causing the manipulee to behave in a way that is detached from *the reasons that ought to govern* her behavior.[14] In what follows I'll use the locution "distancing from reasons" to denote either (a) *worsening the manipulee's deliberative position* in the sense that's relevant to Hanna's account or (b) *causing detachment from reasons* in the sense that's relevant to Gorin's account.

Let's examine how Hanna's and Gorin's accounts can help diagnose example *II*. Notice that—when successful—this manipulation won't lead the son to believe in a falsehood. Indeed, there's a sense in which it might result in a *distance* between the son's rejection of the false view that his mother attacks and *the real reasons* to reject it. However, once again, the danger of hindering the son's future deliberative project seems more significant (than this distance, by itself), as illustrated before by focusing on the potential effect on the son's critical thinking.

Furthermore, notice that the normative significance of distancing the manipulee from particular reasons would rise with the relevance of these reasons to the manipulee's deliberative project. If, for example, these are reasons for believing that Napoleon preferred red wine over white, distancing from them doesn't seem as worrisome as reasons that are important for the fulfillment of the manipulee's deliberative project. This suggests that at least part of the normative significance of *distancing from reasons* is derived from that of deliberative projects.

Eric Cave diagnoses the problem with what he calls "motive manipulation."[15] One necessary condition for this type of manipulation, as Cave defines it, is that it works by mobilizing some "non-concern motive"—that

is, a motive that either (a) doesn't constitute a pro- or con-attitude or (b) is inaccessible to consciousness.[16] Cave argues that the (pro tanto) wrongness of motive manipulation stems from a principle that he calls *Modest Autonomy*: "individuals should refrain from activities that threaten to undermine the capacity of others to manage their concerns."[17] Putting aside the plausibility of this principle and Cave's argument, the capacity to manage concerns is generally conducive to deliberative projects, so undermining this capacity is likely to hinder the manipulee's deliberative project. It doesn't seem, however, that Cave's account can be used to accurately capture the problem with cases like example *II* above (which is not exactly a case of *motive manipulation*(. It's doubtful that the mother (in this example) undermines the son's capacity to manage concerns. More generally, undermining this capacity is not the only way to hinder the fulfillment of the manipulee's deliberative project. These considerations raise the possibility that the capacity to manage concerns inherits its normative significance from that of deliberative projects.

In sum, each problem that the prominent diagnoses of the wrongness of manipulation characterize might lead to hindering deliberative projects. But insofar as it does *not* hinder them, the relevant problem seems less significant for the moral evaluation of manipulations in education than the danger that I identified. Furthermore, none of these diagnoses captures all the typical ways in which manipulation in education might hinder deliberative projects.

5. INFLUENCING THE QUALITY OF DELIBERATIVE PROJECTS

Despite my argument in the previous section, the ideals of *rationality* and *responsiveness to reasons* remain relevant as candidate objective standards for the *quality* of deliberative projects. Arguably, education should aim to improve this quality, for example by influencing the students to be more responsive to reasons when forming their deliberative commitments. If so, the impact on the *quality* of deliberative projects should also be taken into consideration when we evaluate manipulations in education. This consideration is distinct from the impact on the successful fulfillment of the deliberative commitments that the students will actually pursue. These two considerations

might push in opposite directions: perhaps some educational practices are likely to objectively improve one's deliberative project (living up to objective standards) but are also likely to hinder its successful fulfillment (living up to one's own standards) or to hinder it in the scenario that it evolves in certain directions.

Evaluating such cases—and, more broadly, the effect of manipulation on the objective quality of deliberative projects—requires additional research. Such research must enter the debate on the aims of education, which falls beyond the scope of this paper. Nevertheless, two relevant points can already be made. First, educators should be cautious with paternalistic educational measures that restrict the scope of potential directions in which the students' deliberative projects may evolve. Remember that there's controversy about whether the reasons to which one should be responsive stem from one's personal preferences, or rather from independent, objective normative truths. Among those who believe in objective normative truths, there's controversy about their content. These considerations—as well as our human fallibility—call for some amount of uncertainty concerning our relevant normative views. Additionally, the relevant controversies might raise political problems for imposing substantive normative views in education.

Second, if we accept that rationality and responsiveness to reasons are objective standards for any deliberative project, notice that they favor the development of critical thinking (and trivially so, some would argue) among other intellectual multi-purpose virtues of the kind that I discussed in section 2. Plausibly, hindering the development of such virtues threatens the *quality* of deliberative projects, not only their actual *fulfillment*.

It would be interesting to examine (in future research) also whether this point applies to other candidate objective standards, such as *autonomy*, *authenticity*, and other normative, spiritual, or religious ideals. (Plausibly, there are multiple objective standards, so that we can see them as dimensions along which deliberative projects could improve.)

6. CONCLUSION

The central claim that I have been defending in this paper is that the risk of hindering deliberative projects is an important factor to take into account when we evaluate educational practices that might seem problematically manipulative. The greater the expected hindring, the more *pro tanto* wrongful the relevant practice is.

As we've seen in the previous section, focusing on deliberative projects has the potential to advance the debate on the aims of education. Arguably, these aims should include facilitating the fulfillment of legitimate deliberative projects, as well as paternalistically improving them. One question to examine in future research is whether these goals are self-standing or rather derivative of other aims of education. Another set of questions for further research pertains to the category of multi-purpose virtues that I have only started to explore in this paper: which multi-purpose virtues should be cultivated in order to meet the aforementioned goals, how to conceptualize these virtues in light of these goals (perhaps reconceptualizing some familiar virtues such as critical thinking), and how to cultivate these virtues.[18]

---

1 Jason Hanna, "Libertarian Paternalism, Manipulation, and the Shaping of Preferences," *Social Theory and Practice* 41, no. 4 (2015): 618–643.

2 Robert Noggle, "Pressure, Trickery, and a Unified Account of Manipulation," *American Philosophical Quarterly* 57, no. 3 (2020): 241–252.

3 John Tillson, "Wrongful Influence in Educational Contexts," in *Oxford Encyclopedia of Philosophy of Education* (Oxford: Oxford University Press, forthcoming), ed. Kathryn Hytten. I might add the following definition of *manipulating,* from the only article (to date) whose central focus is *manipulation in education*: making someone do something that is different in kind from what she intends to do. [Vance Kasten, "Manipulation and Teaching," *Journal of the Philosophy of Education* 14, no. 1 (1980): 53–62.]

4 Shlomo Cohen, "Manipulation and Deception," *Australasian Journal of Philosophy* 96, no. 3 (2018): 483–497.

5 Paternalism is often legitimate toward children in parenting and elementary schools [see for example Andrew Franklin-Hall, "On Becoming an Adult: Autonomy and the Moral Relevance of Life's Stages," *The Philosophical Quarterly* 63, no. 251 (2013): 223–247].

6 For a broad survey of the potential moral problems in manipulation, see: J. S. Blumenthal-Barby, "A Framework for Assessing the Moral Status of Manipulation," in *Manipulation: Theory and Practice*, eds. Christian Coons and Michael Weber (Oxford: Oxford University Press, 2014). (See the rest of this collection for additional philosophical articles about manipulation.) For potential moral problems of manipulation in teaching see Vance Kasten, "Manipulation and Teaching."

7 The term "deliberative project" has been used in a different context in: David Enoch, *Taking Morality Seriously: A Defense of Robust Realism* (Oxford: Oxford University Press on Demand, 2011). The pursuit of what John Rawls calls "a life plan" counts as a deliberative project. [John Rawls, *A Theory of Justice* (Cambridge, MA: The Belknap Press, 1971).] But I'm not committing here to any of Enoch's or Rawls's conceptual requirements or assumptions. By the term "deliberative belief" I refer to beliefs that guide a person's deliberative project by treating certain questions as settled. The person who holds a deliberative belief considers herself as having an answer to the relevant question. This differs from merely ignoring the question or being unaware of the possibilities that give rise to this question.

8 See for example Michael E. Bratman, *Structures of Agency: Essays* (Oxford:

Oxford University Press, 2007).

9 Such an account could be similar to J. S. Mill's defense of liberty on utilitarian grounds in his *On Liberty*.

10 My justification of the importance of these virtues—as being generally conducive to deliberative projects—has implications for how these virtues should be conceptualized (at least in this context). For example, the relevant sense of *open-mindedness* allows forming and adhering to commitments despite the willingness and ability to reconsider them when the time is right.

11 Robert Noggle, "Manipulative Actions: A Conceptual and Moral Analysis," *American Philosophical Quarterly* 33, no. 1 (1996): 43–55; Noggle, "Pressure, Trickery."

12 Robert Noggle, "Manipulation, Salience, and Nudges," *Bioethics* 32, no. 3 (2018): 164–170; Richard Thaler and Cass Sunstein, *Nudge: The Gentle Power of Choice Architecture* (New Haven: Yale University Press, 2008).

13 Hanna, "Libertarian Paternalism," 630–631.

14 Moti Gorin, "Paternalistic Manipulation," in: *The Routledge Handbook of the Philosophy of Paternalism*, eds. Kalle Grill and Jason Hanna (New York: Routledge, 2018), 236–247.

15 Eric M. Cave, "What's Wrong with Motive Manipulation?" *Ethical Theory and Moral Practice* 10, no. 2 (2007): 129–144.

16 Cave, "What's Wrong," 131–132.

17 Cave, "What's Wrong," 138.

18 This work was conducted as part of a research project of Nikkonna (registered Amuta). I thank each person who helped. I received helpful feedback from Shlomo Cohen, Yuval Eylon, Zeev Goldschmidt, David Kashtan, three anonymous PES referees, and the participants in my session at the 2021 PES