The Hazards of Influence

(May 2017 draft)

Matthew X. Etchemendy

University of Chicago Law School

May 3, 2017

University of Chicago Law School
Room 403
1111 East 60th Street
Chicago, IL 60637

metchemendy@uchicago.edu
(650) 400-6114

Abstract. This paper considers the ethics of *non-rational influence*, defined approximately as psychological influence that bypasses or subverts the rational capacities. The question addressed is whether a substantial case can be made against exerting non-rational influence on others' beliefs, which is to say whether we can identify one or more broadly applicable, relatively strong *pro tanto* reasons to eschew such influence. Many popular arguments against non-rational influence are, I argue, non-starters: we cannot make a substantial case against non-rational influence on grounds that it represents a form of mind control, that it undermines autonomy, that it is disrespectful, or that it diminishes (or at least fails to enhance) the rational capacities of those we seek to influence. Instead, the most significant problem with non-rational influence is that it presents, at least in a wide range of cases, a greater risk of inducing *error*—that is, a greater risk of moving a person's overall set of beliefs further out of alignment with an accurate and complete account of the world. Concern over this risk-of-error problem, however, justifies only a fairly weak and radically context-sensitive case against non-rational influence; it is unclear at best whether this suffices to fully vindicate the widespread pre-theoretic distaste for non-rational influence. In short, non-rational influence may have a worse reputation than it deserves.

The Hazards of Influence[1]

We have a stake in the contents of each other's beliefs, and we have many ways of influencing and altering them. In this paper, I will be interested in one particular way of influencing each other's beliefs, which I will call *non-rational influence*. I will have more to say about what I mean by "non-rational influence," but roughly I have in mind psychological influence that—if I may employ an unoriginal, somewhat imprecise formula—bypasses or subverts the rational capacities. My subject will be the *ethics* rather than the mechanics of non-rational influence, a subject that has, as I see it, received less attention than it deserves.[2] Is there anything generally ethically suspect about exerting non-rational influence on others' beliefs, even if, as practically everyone seems to allow, it is sometimes the correct thing to do, all things considered?

The question is of considerable importance, because political action (and here I mean "political" in a very broad sense, to include the "politics" of private organizations) largely consists of efforts to exert psychological influence on others. Influence by verbal means, in particular, represents a major part, perhaps the bulk, of the work that lawyers, politicians, and scholars do. Often this involves *rational* influence in one form or another. We reason and argue with each other, trying to get our interlocutors to see things a certain way—typically, though not always, to see things our way, the way we think is right. Sometimes, too, we exert rational

[1] I am grateful to Mark Berger, Taylor Coles, Chiara Cordelli, Raff Donelson, Jerry Dworkin, Barbara Fried, Nadeem Hussain, Josh Kissel, Brian Leiter, Martha Nussbaum, Emma Saunders-Hastings, Steve White, and Jim Wilson for their helpful input and comments during various stages of this paper's development.

[2] For similar assessments regarding the relatively thin literature on the related, though in my view somewhat distinct, subject of *manipulation*, see Coons and Weber [2014], p. 1; Sunstein [2015], p. 215.

influence without arguing or reasoning at all. (Consider, for example, criminal penalties imposed to dissuade people from certain behaviors by altering the balance of reasons for and against such behaviors.) But efforts to alter others' ways of thinking often involve rather more than rational influence.[3] Charming and cajoling, for example, can be part of a lawyer's or a politician's repertoire, and we are generally all familiar with governments' use of emotionally provocative propaganda and advertisers' efforts to harness non-rational psychological processes to help sell products.

The question I want to ask is whether there are reasonably strong, broadly applicable *pro tanto* reasons to refrain from using non-rational influence to alter others' beliefs—reasons that are substantial enough, and that apply to a broad enough range of cases, to underlie something like a "rule of thumb" against the practice. Because this is a mouthful, I will sometimes just phrase the inquiry thus: can we make a substantial case against non-rational influence?

I am unsure whether we can. Many popular arguments against non-rational influence are, I think, non-starters: I believe, for reasons explained below, that we cannot make a substantial case against non-rational influence on grounds that it represents a form of mind control, that it undermines autonomy, that it is disrespectful, or that it diminishes (or at least fails to enhance) the rational capacities of those we seek to influence. In my view, the biggest problem with non-rational influence is that it presents, at least in a wide range of cases, a greater risk of inducing error, where by "inducing error" I mean moving a person's overall set of beliefs further out of alignment with an accurate and complete account of the world. I call this the risk-of-error problem. But concern over this problem justifies only a fairly weak and radically context-sensitive case against non-rational influence; it is unclear at best whether this suffices to fully

[3] *Cf*. Gaylin and Jennings [2003], p. 43; Frankfurt [2002], p. 28.

vindicate the widespread pre-theoretic sense that intentionally engaging in non-rational influence is somehow wrong or at least distasteful. In short, I conclude that non-rational influence may get a worse rap than it deserves.

The discussion proceeds as follows. Part I deals with definitional and methodological preliminaries. Part II considers and rejects various *prima facie* appealing bases on which to build a substantial case against non-rational influence. Part III is dedicated to laying out the best case I can perceive against non-rational influence, namely that non-rational influence presents heightened risks of inducing error. Finally, Part IV addresses the limits of the case laid out in Part III.

I. Preliminaries: Delimiting the Inquiry

Before getting into the thick of things, it is important to lay out the scope of the inquiry. First I will say what I mean by "non-rational influence," and then I will briefly explain my methodology and goals.

A. Non-Rational Influence Defined

"Non-rational influence" is not a term in everyday use. Even in the philosophical literature, it appears only rarely,[4] and I am not aware of anyone who has elsewhere set out to provide a definition (stipulative or otherwise) of this precise phrase. There is, however, a growing literature on *manipulation*, which may help provide some context and situate the present discussion in a broader scholarly context. It's standardly understood that manipulation is a form

[4] But see Blumenthal-Barby [2012], p. 355; Buss [2005] (*passim*); Noggle [1996], p. 49; Coons and Weber [2014], p. 10; Barnhill [2015], pp. 311–312.

of influence over people's psychology and, ultimately, actions,[5] and also that manipulation is ethically problematic, even if it is sometimes right to manipulate.[6] Beyond that, though, there is wide disagreement as to what manipulation involves. What's more, at least some instances of *rational* influence—like certain heavy-handed but not quite coercive offers or threats—intuitively seem to qualify as manipulative.[7] So manipulation and non-rational influence almost certainly aren't the same thing. Even so, some scholars working on manipulation have isolated and analyzed categories of influence that approximately match what I have in mind by non-rational influence. Moti Gorin, for example, describes what he calls "a dominant view of interpersonal manipulation" according to which it "occurs only if an influencer intentionally bypasses or subverts the rational capacities of the person he seeks to influence."[8] Like Gorin,[9] I

---

[5] See, e.g., Coons and Weber [2014], p. 1; Sunstein [2015], p. 216; Baron [2003], p. 48.

[6] See Coons and Weber [2014], pp. 1–4; Baron [2014], p. 108.

[7] Marcia Baron, for example, plausibly suggests that "[o]ne example [of manipulation] would be an offer to raise a student's grade if he has sex with you." Baron [2003], p. 41. This could of course involve an element of non-rational influence, but it need not; intuitively such an offer could involve solely *rational* influence of an ugly sort.

[8] Gorin [2014], p. 51; see also Buss [2005], p. 208 n.19 (suggesting that "the distinguishing mark of manipulative 'processes' is that they influence preferences (and beliefs) nonrationally" and mentioning in particular "the sort of influence that bypasses…rationality"); Noggle [1996], p. 49 (describing the view that "non-rational influences [are]…inherently manipulative" as "tempting" but "ultimately misguided"); Blumenthal-Barby [2012], p. 349 (describing a category of "nonargumentative influence" subdivided into two types, "reason-bypassing" and "reason-countering"). Cass Sunstein has carved out a thematically similar—though, it seems to me, importantly different—category of influence by proposing that "an effort to influence people's choices counts as manipulative *to the extent that it does not sufficiently engage or appeal to their capacity for reflection and deliberation*." Sunstein [2015], p. 216.

[9] Gorin [2014], p. 51.

think this is a flawed understanding of *manipulation*, but it is an ethically interesting category of influence in its own right, and it is basically what I have in mind by "non-rational influence." (In strictness, I would drop the qualification that the influencer must *intentionally* bypass or subvert the rational capacities of the target: I do not want to define away the possibility that a person might unintentionally exert non-rational influence.)

What, then, are the rational capacities? Here I will depart somewhat from Gorin's account.[10] By "rational capacities," I mean—again, approximately—the psychological capacities that endow one with an ability to adjust one's beliefs and other attitudes appropriately in light of the available reasons.[11] I have deliberately tried to keep this account of the rational capacities as minimalistic as realistically possible. Note, however, that this account takes for granted a *normative* account of rationality. (Non-normative alternatives are possible: the rational capacities could be conceived simply as capacities for reasoning, where reasoning is understood as "thinking through the possibilities, drawing inferences, and the like."[12] For a variety of reasons, I think the inquiry would be of less interest if we adopted an account of the rational capacities along these lines.[13]) Note also that on this picture, almost all humans have rational

[10] On which see Gorin [2014], p. 52.

[11] When I refer to reasons, I mean considerations that weigh in favor of some belief or other attitude. In other words, I am talking about normative, not explanatory, reasons for various mental states. For a discussion of the distinction, see Finlay and Schroeder [2015].

[12] Gibbard [1990], p. 49.

[13] Notably, if we were to keep the account truly non-normative, we would be unable to distinguish between *good* and *bad* inferences. Consider two people, both of whom see a jogger go by their respective houses late at night. The first is a psychologically healthy person with an active imagination; the second suffers from paranoid schizophrenia. Both may wonder, "Why did that jogger go past my house at this odd hour?" And both may

capacities, though some have more well-developed rational capacities than others: in other words, people vary in how well and how reliably they are able to recognize and weigh reasons, and to adjust their beliefs and other attitudes accordingly.

It would be tempting to spend a lot of time trying to specify in more detail what must occur (or fail to occur) if the rational capacities are to be bypassed or subverted.[14] But I do not think this is necessary: we can proceed with a fairly intuitive, if imprecise, idea of what non-rational influence, conceived as influence that bypasses or subverts the rational capacities, amounts to. Consider two television advertisements for a brand of cola, both of which are intended to make people adopt beliefs that will in turn make them more likely to buy the advertised product. Advertisement #1 explains that the advertised brand is usually preferred over competing brands in blind taste tests. If the ad succeeds in making some consumers want to drink the cola, it is presumably because those consumers think that if *most* people prefer the taste

---

consciously evaluate a number of possibilities, including (1) that the jogger is an innocent person who ventured out at night to avoid the heat of the day, and (2) that the jogger is a CIA spy. The psychologically healthy person considers the evidence and draws the inference that the jogger is probably not up to anything nefarious. The schizophrenic person draws the opposite inference. The two have come to opposite conclusions, but both carefully considered a similar range of possibilities, engaged in the same degree of conscious deliberation, and so on. Both, in short, equally engaged in *reasoning* (in a naturalistic, non-normative sense) before reaching a conclusion about the jogger's motivations. The difference is that the schizophrenic's reasoning processes were badly flawed, and it seems natural to say that the schizophrenic's thinking was less *rational* than that of the psychologically healthy person, and that the schizophrenic's *rational capacities* are seriously limited in comparison to those of the psychologically healthy person. By my lights, then, giving a person a drug that induces severe paranoia and thereby causes him or her to draw wildly flawed inferences, would be a straightforward case of non-rational influence.

[14] See Gorin [2014], pp. 52–57 (discussing the nature of rational capacities, as well as several conceptions of what is involved in subverting them).

of this cola over competing brands, the odds are good they too will prefer it. Given *prima facie* credible evidence that most people *do* prefer the taste of this cola, consumers infer that they will prefer it as well. This certainly seems like a straightforward case of rational influence. Advertisement #2, by contrast, shows a series of vignettes in which attractive and athletic people are seen drinking the product.[15] Most of us, I assume, will intuitively judge that Advertisement #2 is an effort at non-rational influence, or at least influence involving a significant non-rational component. It may make people believe that the product is superior to competing brands, but if it does, it is hard to avoid concluding that the audience's rational capacities must in some way have been bypassed or subverted.[16]

---

[15] *Cf.* Barnhill [2015], pp. 308–309 (describing a classic Coca-Cola advertisement something along these lines).

[16] The evolutionary psychologist Geoffrey Miller suggests a potential way in which ads of this kind could, in fact, lead people by rational means to believe that the product is worth buying: "The ad viewer himself need not believe that the brand has any logical or statistical link to the aspirational trait he wants to display. He must simply believe that other ad viewers from his social circle will perceive such a link. If I want to look tough, I don't need to believe that the Hummer H1 really looks tough; I need only believe that more gullible onlookers will think it looks tough, and credit me with toughness for owning it." Miller [2009], pp. 98–99. The basic idea, as I understand it, is this. Often when we buy products, we do so as part of a broader strategy of cultivating some image—in the Hummer example, toughness. Now, it would be irrational to think that tough people drive Hummers just because an ad shows tough people driving Hummers. But it is *not* silly to think that a massive ad campaign showing tough people driving Hummers will make *other people* think that tough people drive Hummers. And if I just want to seem tough to other people, this may give me a good reason to buy a Hummer. (Note, however, that the rational process at work here apparently relies on the ability of the ad to non-rationally influence at least some viewers.) This illustrates the complexities involved in actually determining whether and to what extent non-rational influence is at work in a particular situation, but I do not think it calls into question the existence of a meaningful distinction between rational and non-rational influence. I am grateful to Raff Donelson for useful discussion of these themes.

Of course, there are many other examples one could give, including many cases in which it would be quite hard to determine whether rational or non-rational influence is involved, or perhaps more properly speaking, the degree to which the influence involved is rational or non-rational. Moreover, it bears emphasizing that the basic units of analysis here, the things we are actually categorizing as rational or non-rational, are *instances of influence*: events in which a given input or stimulus produces a particular change in belief in a particular person at a particular time. In strictness, therefore, we cannot say that Advertisement #2, or a striking propaganda poster, or a certain piece of music, counts as rational or non-rational *per se*. In themselves, these are just objects or phenomena in the world. Exposure to them may affect different people's beliefs in different ways, or produce a variety of different changes to a single person's beliefs.

It will very often be the case that an assertion, text, image, or rhetorical move exerts a complicated mix of rational and non-rational influence. But this does not prevent us from coherently venturing claims about the kind of influence they are *intended* or *expected* to produce, or generalizing about the kind of influence they will *tend* to produce. Many of us would be inclined to point at a data-rich paper on crime patterns and say "rational influence!"; and to point at the Willie Horton ad and say "non-rational influence!"[17] This is not wrong, exactly; it gestures at a real and important difference between various strategies for altering people's beliefs about criminal justice policy. But we should not forget that such reactions will at best capture imperfect generalizations. Data-rich papers will sometimes, and to some extent, exert non-rational influence; lurid TV ads will sometimes, and to some extent, exert rational influence.

[17] I am grateful to Barbara Fried for suggesting the example of the (in)famous Willie Horton ad.

Notwithstanding the caveats and complications, however, my hope is that these rough definitions and simple examples provide sufficient clarity about the phenomenon under consideration for us to proceed with some fruitful ethical analysis.

B. Methodological Considerations

Now, on to some brief methodological points. The question I want to investigate, informally stated, is whether a substantial case can be made against non-rational influence. As noted above, when I speak of "making a substantial case against non-rational influence," I mean identifying one or more broadly applicable, relatively strong *pro tanto* reasons to eschew non-rational influence. By "broadly applicable" reasons, I mean reasons that weigh against as many instances and kinds of non-rational influence as possible. By "relatively strong" reasons, I mean reasons with significant force: just as one does not make a substantial case against going outdoors by pointing to the increased risk of suffering a fatal meteorite impact, one does not make a substantial case against non-rational influence by showing that some aspect of the world would be trivially better if we were to eschew it. That said, *relatively* strong reasons—strong enough to justify some kind of *prima facie* presumption or default rule against non-rational influence—will suffice. Knockdown or dispositive reasons are not required. Nor do I want to limit the inquiry to a search for narrowly "moral" reasons in particular, to the extent moral reasons are distinct from practical reasons in general.

I do, however, want to place one important initial restriction on the range of eligible reasons to eschew non-rational influence: we should exclude any that are essentially dependent or parasitic upon putatively widespread tendencies for people to dislike, or to be offended or upset by, non-rational influence. This exclusion is not motivated by the mistaken idea that such

tendencies are normatively irrelevant. They are, in fact, quite normatively relevant. Assuming most people strongly dislike being subjected to non-rational influence, it is fairly easy to make a certain kind of pragmatic case against exerting such influence: all else equal, it seems preferable to avoid doing things to people that they don't appreciate, or that offend them. (Policymakers and politicians, in particular, have ample reason to avoid courses of action that, if discovered and fully understood, would create widespread anger.) But while I am happy to admit that a case of this kind can be made against non-rational influence, this is not saying much. After all, in certain social contexts, a case of this kind could be made against interracial marriage, or women going without a male chaperone, or lenders charging interest on loans. The basic problem is that there can always be strong reasons to avoid doing things because of the *false* beliefs or *inapt* feelings of others. Such beliefs and feelings can be quite normatively relevant simply because they exist, whether or not they are misguided.[18] But I think it is clear enough why we should not be quite satisfied resting a case against non-rational influence on such grounds, and I will accordingly rule out any story about why we should eschew non-rational influence that relies on them.[19]

[18] *Cf.* Alexander [1992], pp. 173–176, 193 (addressing the ethical issues surrounding the hiring of employees whose job performance may be adversely affected by the negative reactions, including irrational or even immoral reactions, of customers and others with whom they will interact on the job).

[19] A similar analysis applies to issues of consent. Perhaps it is often true that people do not or, if they were asked in advance, would not consent to being non-rationally influenced. This may be morally relevant, and it is *certainly* relevant to practical questions about whether and when to engage in non-rational influence. But it would be very unsatisfying simply to say that we should eschew non-rational influence because people would not consent to it, for the simple reason that we should like to know whether their failure to consent is a product of ignorance or mistake. Consider another analogy. It is widely agreed, I think, that it's usually wrong to make people metabolize foods or drugs in the absence of actual or at least hypothetical consent. Now imagine social circumstances in which most people would emphatically refuse to eat any food prepared by a member of some demographic group, call it

One final methodological point is in order. I will be focusing on non-rational influence on *beliefs*, simply because this makes for a more focused discussion. Thus, whenever I refer to non-rational influence in the balance of this paper, I am talking about non-rational influence on beliefs. (Similarly, whenever I refer to rational influence, I'm talking about rational influence on beliefs.) Strictly speaking, this means that all of my conclusions will be limited to the ethics of non-rational influence on beliefs. That said, I don't see this as an especially substantial limitation. By "beliefs," I simply mean attitudes that can be expressed using assertoric statements. The discussion is not, for example, intended to be limited to non-rational influence on non-normative or natural-factual beliefs. Similarly, the focus on beliefs does *not* limit the discussion to cases in which the influencer has no designs on also changing others' conative attitudes or their actions/behavior. After all, typically we influence each other's beliefs in the hope (or at least the expectation) that a change in their beliefs will produce downstream changes in desires, plans, hopes, fears, and other conative attitudes—and ultimately changes in behavior. For that matter, we sometimes try to alter people's beliefs by first making changes to their conative attitudes (e.g., making them more fearful so they are more likely to believe they should invest in insurance or home security).

Demographic X, perhaps because of an unshakeable belief that food is tainted by contact with such people. To be sure, in such circumstances it might make every bit of practical, even ethical, sense for restaurants not to hire food preparers of Demographic X. But if we wanted to ask quite generally, "What is wrong with publicly offering food prepared by members of Demographic X?," it feels awfully unsatisfactory to answer that few people would consent to eat food prepared by members of demographic X. This would be true even if we accepted that people have a basically absolute right to control what they eat. *Cf.* Buss [2005], pp. 224–225 (making very similar observations about consent-based objections to manipulation).

My guess is that the basic ethical analysis set out in this paper could be extended, with certain relatively modest alterations, to non-rational influence *simpliciter*, not just non-rational influence on beliefs. But to focus the discussion, I won't try to defend that broader claim here.

II. A Series of False Starts

Having dealt with the preliminaries, let us begin our consideration of some candidate reasons to eschew non-rational influence—i.e., some potential theories about what is wrong with exerting such influence. In this section, I'll be considering four candidate theories that are not, as I see it, satisfactory. These four theories are based, respectively, on notions of control, autonomy, respect, and the value of cultivating others' rational capacities. It is important to consider them, I think, for two interconnected reasons. First, the relevant literature suggests these are likely to be the most common *prima facie* objections to non-rational influence. Second, since my own account of what is problematic about exerting non-rational influence, as laid out in Part III, makes for only a weak and highly context-dependent case against the practice, I feel obliged to explain, at least briefly, why the most popular alternatives—which might be thought to motivate stronger objections to non-rational influence—are ultimately unsatisfactory.

A. Non-Rational Influence as Mind Control

Perhaps the most commonly articulated objection to non-rational influence, at least in general public discourse, is that it involves a worrisome exercise of control over others' minds: metaphors about the impropriety of "pulling people's strings" abound in this context. What exactly is meant by this is not always apparent; indeed, those who make such objections may not *themselves* always know quite what they have in mind. There are, after all, different ways in

which one thing, or person, may be said to control another. This much, however, is clear: whatever else it may be, "control is a causal phenomenon."[20] Accordingly, there is a certain austere sense in which non-rational influence plainly constitutes mind control. It represents an exercise of deliberate causal influence over others' beliefs. Just as obviously, however, the same is true of rational influence.[21] The only sensible conclusion seems to be that we possess a fair amount of control over each other's minds—at least in a certain minimal, causal sense—and that there is nothing broadly or generally objectionable about exercising such control. Indeed, the opportunity to be subject to such control is one of the great benefits of being a member of a social and language-using species.

One could try to argue that non-rational influence presents a problem not because it is a form of mind control, but because if we resort to it, we will end up exerting *too much* control over others' minds. There are, however, at least two problems with this line of thought. First, even if non-rational methods were generally more causally efficacious than rational ones, we would need an argument for why the former fall on the wrong side of some (possibly imprecise) boundary between acceptable and unacceptable degrees of interpersonal control. Second and more importantly, there is no apparent reason to believe that non-rational methods are generally

---

[20] Mele [1995], p. 10.

[21] I take for granted a naturalistic picture according to which human beings are, like macroscopic physical objects in general, subject to deterministic causal laws. (More precisely, I take for granted something akin to what John Martin Fischer and Mark Ravizza call "almost causal determinism," according to which "macroscopic events are not, strictly speaking, causally determined, but they are very *close* to being determined," Fischer and Ravizza [1998], p. 15 n.24.) Of course, even if one rejects such determinism or almost-determinism, it is hard to dispute that both non-rational and rational influence are forms of causal control. (One needn't be a determinist to agree with this, though all determinists will.)

more causally efficacious than rational ones. Experimental results and anecdotal narratives illustrating the power of non-rational influence can be spellbinding, but we cannot lose sight of the fact that *rational* methods of influence are also extremely powerful. Indeed, they are so powerful and so basic to our lives that it is easy to take their power for granted: the power of rational influence has a tendency, so to speak, to fade into the background of life. This extraordinary power presumably explains why corporations attract workers *mainly* by offering to pay salaries and provide concrete benefits, instead of by cooking up clever advertisements (along the lines of Advertisement #2 above) for unpaid positions.

Indeed, when others attempt to influence us rationally, we often experience such efforts as quite irresistible. No matter how much a mother may *want* to believe that her son did not commit a crime, she may find the conclusion that he did commit the crime utterly irresistible if she is shown rock-solid video evidence. She may wish she could forget what she saw, but unless and until she does, she experiences her newfound belief as inescapable. My point is not that rational methods *are* generally more powerful that non-rational ones. I am not sure this is true; indeed, I am not even sure how to give concrete content to it. My point is only that I cannot see any convincing way to mount an objection to non-rational influence based on the thought that non-rational methods represent a more powerful form of "mind control" than rational methods.

B. Non-Rational Influence as a Threat to Autonomy

A thematically related concern is that non-rational influence undermines or compromises autonomy. I suspect this is the most common objection to non-rational influence, at least among

scholars: it certainly enjoys a prominent place in the relevant literature.[22] Unfortunately, autonomy is a tricky subject because it has become so theory-laden and because there are so many different accounts of what autonomy involves.[23] This flowering of different theories is due in part to the fact that the word "autonomy" has only a thin core of shared meaning. It means "self-governance," as the etymology of the word plainly suggests.[24] This is not, I think, an especially confusing notion in itself, but it can seem overwhelmingly complex and difficult when one considers all the different *ways* in which an entity can be (or fail to be) self-governing, some of which have clear and substantial normative significance, and some of which do not.

Where the autonomy of persons is at issue, the most intuitive, everyday sense of "autonomy" probably concerns a person's social position. Asking whether a person is autonomous in this sense is approximately equivalent to asking whether he or she, rather than someone else, runs his or her life. Can a person do what he desires? More importantly, perhaps, could a person still do what she wanted if she wanted to do something different than she actually

---

[22] See Buss [2005], pp. 208–209 ("What, exactly, is wrong with…manipulating someone…? One popular answer—perhaps the most popular—is that treating someone in this way is incompatible with respecting her autonomy."); Blumenthal-Barby [2012], p. 352 ("Influence by nonargumentative means…has engendered a good deal of moral suspicion, and I think that it is fair to say that much of this comes from concerns about autonomy."); Sunstein [2015], p. 226 ("The most obvious objection to manipulation is that it can insult both autonomy and dignity.").

[23] As Gerald Dworkin wryly noted over 25 years ago, "[a]bout the only features held constant from one author to another are that autonomy is a feature of persons and that it is a desirable quality to have," adding that "[i]t is very unlikely that there is a core meaning which underlies all these various uses of the term." Dworkin [1988], p. 6; see also Buss [2005], p. 196; Taylor [2009], p. 1; *cf.* Mele [1995], p. 237.

[24] See, e.g., Oshana [2005], p. 183; Gaylin and Jennings [2003], p. 30 (discussing the etymology of the word).

does, or would somebody stop her?[25] Although non-rational influence *can* compromise autonomy in this straightforward (and straightforwardly important) sense—as when a person is charmed into joining an oppressive cult—it can also *enhance* autonomy of this kind, as when it gives people the courage they need to extract themselves from slave-like social conditions.[26] There is no apparent systematic relationship between non-rational influence and autonomy of this sort.

What about autonomy conceived as the ability to control the content of one's own mind? We can call this "autonomy as psychological self-determination." Does non-rational influence undermine autonomy as psychological self-determination, and if so, does it matter? The answer may depend on what is meant by self-determination, since this term too is amenable to different interpretations. If self-determination means an absence of control by others, then we are back to the worries discussed and rejected in Part II.A. More plausibly, however, the *sine qua non* of psychological self-determination is not the *absence* of external causal control (including causal control by others), but rather the *presence* of some kind of "internal" control or determination.

But what kind of ethically significant "internal" control is non-rational influence liable to compromise? Often when we speak of self-determination, what we have in mind is *choice*. But if psychological self-determination requires choice, then failures of psychological self-determination are both ubiquitous and ethically unremarkable. When I look at my watch and see

[25] The latter formulation is important to capturing the fact—and it certainly appears to be a fact—that a slave would suffer from a massive and systematic lack of autonomy even if he or she were rather satisfied with life. See Oshana [2006], pp. 53–57; Christman [2005], p. 282.

[26] *Cf.* Fleming [1989], p. 83 (arguing that "deprogramming" cult members may sometimes be morally justified on grounds that it could restore a subject's autonomy); Sunstein [2015], pp. 230–231 (suggesting that manipulation could potentially enhance autonomy).

the hands indicating that the time is 3:50 P.M., I come to believe that it is 3:50 P.M., and choice simply does not enter into the process. It is true that I probably made an initial choice to look at the watch (though this clearly could have happened by accident as well), but the broader point is that the typical experience of belief- (and, for that matter, desire-) formation is not one in which choice features prominently. In this connection, note how strange—even vaguely creepy—it sounds when people talk about "choosing to believe" something or another. We generally *don't* think of beliefs as something we choose, and when someone says he or she "chooses" to believe something (say, that God exists, or that he or she has a future in show business), it's hard not to suspect that he or she actually harbors some nagging doubts about the proposition in question.[27]

An alternative gloss we could put on the notion of self-determination is this: a belief (or other attitude) is a product of self-determination, in the relevant sense, if it came about as a result of the subject's other attitudes. Consider again the example of looking at my watch. I see that the hands indicate that the time is 3:50 P.M. I then come to believe that it is 3:50 P.M. Although I did not *choose* to acquire this belief—phenomenologically speaking, the process was "automatic"—I certainly would not have come to that belief if I didn't have certain other prior beliefs. For example, I would not have come to the belief that it is 3:50 P.M. if I didn't already believe that my watch is reliable. Hence my belief that it is 3:50 P.M. was causally determined by particular, contingent features of my psychology, notwithstanding the absence of choice.

Perhaps this represents a kind of self-determination, but even if so, almost no interesting cases of non-rational influence threaten or undermine it. A cola ad featuring attractive young people drinking the beverage probably will not give a viewer any inclination to drink the cola if,

[27] The same is true of many conative attitudes. We usually don't *choose* to like our friends; our liking for them just *happens* as a result of the experiences we go through together.

as it happens, the viewer thinks the people in the ad look vain and unlikeable. Graphic warnings on cigarette packages won't make a person believe he or she should stop smoking if he or she lacks a strong desire for long-term self-preservation. In other words, non-rational influence is ordinarily dependent on the details of the target's psychology. The possible existence of freak cases where this is not so—radical brain surgery, perhaps—is not especially relevant if we are looking for *broadly applicable* reasons to eschew non-rational influence.

There is likely to be a temptation, at this point, to say that psychological self-determination requires that our mental states be causally dependent on some relevant *subset* or *aspect* of our psychological economy.[28] Indeed, one might even be inclined to posit that our mental states are not self-determined except when they are a product of rational reflection, or even rational reflection that has not been compromised by some form of subversive malfunctioning.[29] If this were a reasonable way of understanding psychological self-determination, then non-rational influence would, of course, systematically undermine autonomy as psychological self-determination. But it is highly questionable whether this *is*, in the end, a reasonable way of understanding psychological self-determination, for it relies on the tacit conflation of the self *tout court* with the rational part of the self. (Man is a rational animal, but much else in addition.) It is hard not to think, then, that this is not so much a conception of

[28] A story of this kind could, though strictly needn't, involve identifying the relevant subset or aspect of our psychological economy with our "true" or "real" selves. As Marina Oshana has observed, however, "attempts to elucidate the true or the real self have proven notoriously difficult. Even if it made sense to base autonomy upon some core of the individual, it would be difficult to delineate an aspect of the individual (such as rationality) that could lay a legitimate and noncontentious claim to be that core." Oshana [2006], pp. 51–52.

[29] Dworkin appears to have been attracted, at one point, to a view of this kind, but later distanced himself from it. See Dworkin [1976], p. 27.

psychological self-determination—and thus, in a sense, self-governance or autonomy—but rather something closer to a (contestable) conception of *ideal* self-government.[30]

But in any event, for our purposes, this conception of autonomy gets us nowhere. If we are working with a conception of autonomy according to which we are autonomous precisely when, and to the extent that, our attitudes are a product of rational thought processes, the notion of autonomy can do no independent explanatory work in an argument against non-rational influence. To ask why we should eschew non-rational influence is then, in substance, just to ask what is wrong with compromising a person's autonomy in this sense.

Nor will it help to say that self-governance requires that our attitudes be the product of psychological processes we *endorse*, or *would* endorse if we thought about it, or *should* endorse, and that we do not, would not, or should not endorse non-rational psychological processes. Again, this would involve either begging the question outright (if the claim is that we *should not* endorse non-rational psychological processes) or saying, in so many words, that what is wrong with non-rational influence is that it does not sit well with us. (Again, the fact that people disapprove of non-rational influence may be ethically significant, quite regardless of whether they're right to do so, but the question we really want to answer is whether their disapproval is justified.)[31]

---

[30] *Cf.* Noggle [2005], p. 96 (criticizing certain "substantive" approaches to autonomy that "risk[] conflating authenticity/autonomy with some other notion—such as moral agency, rationality, or some sort of mental health," on grounds that they "change[] the question . . . from one about the person's relation to one of her own psychological elements to the question of whether the element, in itself, exemplifies some other property that has nothing to do with the person to whom it belongs").

[31] See *supra* Part I.B.

Although I admit I have not considered every possible autonomy-based objection to non-rational influence,[32] I have tried to address some of the most promising ones. As far as I can tell, the kinds of autonomy non-rational influence does systematically undermine—like autonomy as an absence of external causal control—have no clear ethical significance; and non-rational influence does *not* systematically undermine the kinds of autonomy that *do* seem to have clear ethical significance—like autonomy as the ability to do what one wants without external social interference. In my view, therefore, autonomy is not the place to look if we wish to make a substantial case against non-rational influence.

C. Non-Rational Influence and Respect

Just as the notion of autonomy frequently appears in contemporary moral philosophy, including in the literature most relevant to the present inquiry, so too does respect.[33] Might deliberate non-rational influence be disrespectful in some way or another? Respect, like autonomy, comes in many flavors,[34] but so far as I can tell, the only way in which non-rational influence might be seen as disrespectful is that it fails to respect the *rational capacities* of others,

---

[32] For example, perhaps one could argue that non-rational influence is more likely than rational influence to produce the kind of structural psychological fragmentation addressed by many "hierarchical" or "coherentist" theories of autonomy. See Buss [2014] (providing an overview of such theories); Taylor [2005], pp. 4–10 (summarizing and critiquing well-known hierarchical accounts of autonomy). But I cannot see how such an objection could be made to work, because I cannot discern any evidence that non-rational influence is more likely than rational influence to make us adopt beliefs or other attitudes that clash with our higher-order attitudes or otherwise render our overall psychological economy less coherent.

[33] See, e.g., Sunstein [2015], pp. 217, 220, 226–227.

[34] See Dillon [2015].

or (equivalently, I think) to respect others as *rational beings*. If one resorts to non-rational influence, the thought might go, one seems to be taking the view that rational influence is not sufficient to get the job done, and this involves a kind of insult to a person's rational capacities: if a person's rational capacities are in good working order, wouldn't one just try to influence him or her rationally?

There is something to this thought. Often we wish to persuade someone, and we face a choice of methods. A thought along the lines of, "He's a clear thinker; he'll listen to reason" often precedes, and motivates, an attempt to use rational methods, or at least mainly rational methods.[35] And this seems like a respectful thought: it is good to be a clear thinker and to have a sharp capacity to recognize and weigh reasons properly. If, however, one hopes to influence a person but concludes that she is muddle-headed and not reliably rational, then rational methods will seem less attractive. In short, it seems fair to say that one's evaluation of a person's rational capacities will have at least some bearing on, or connection with, one's choice to use more or less rational methods of influence.

Of course, the connection is rather loose, and there are almost certainly cases in which those who resort to non-rational methods do not have a low opinion of their targets' rational capacities at all. (I offer a clear example, albeit for somewhat different purposes, in Part III.)[36] Even setting aside such difficulties, however, there is a more fundamental problem: it is not clear what any of this has to do with respect.

---

[35] Often, but not always. As George Tsai has emphasized, even efforts to rationally persuade by means of sincere argumentation can be motivated by a negative, possibly disrespectful assessment of the target's "capacity to gather or weigh evidence." Tsai [2014], p. 79.

[36] See *infra* text accompanying note 46.

Quite plausibly, it is disrespectful to treat people as less rationally capable than they actually are. But what if one's decision to use non-rational methods is based on an accurate assessment of the target's rational capacities? Is it disrespectful to tailor one's interactions with people to their actual, relevant characteristics, even if those characteristics fall short of some ideal? This does not seem to be true in general: if, for example, I try to assist a person who seems too weak to lift some object, he or she may insist I have underestimated his or her strength, and even be offended. But if the person really *isn't* up to the task, it is unclear why my attempt to help should be seen as disrespectful. The rules of etiquette or tact may sometimes require treating people as more rationally capable than they in fact are; failing to do so may be akin to giving an excessively candid opinion about the appearance of a friend's infant child. But this seems like an awfully weak basis for a case against non-rational influence; among other things, it would suggest that furtive or secretive non-rational influence is unproblematic, since what people do not know cannot bruise their egos.[37] If there is a significant problem with non-rational influence, it cannot simply be that attempts to exert non-rational influence come off as disrespectful to people with an inflated estimation of their own rational capacities.

D. Non-Rational Influence and the Cultivation of Rational Capacities

Another potential objection to non-rational influence involves the value of developing and maintaining the rational capacities. Some interactions give us practice in reasoning, and

[37] To be clear, my point is not that it is impossible to disrespect people if they don't find out about it: that is clearly false. My point is that tact- or etiquette-based justifications for treating people as if they are more capable than they really are do not seem to apply unless actual offense is given. By analogy, it may sometimes be wrong (specifically, tactless or rude) to give a candid opinion about the appearance of a friend's baby, but it isn't wrong to *think* that the baby is not, in fact, especially charming, if that happens to be the truth.

some do not. For example, engaging in rational political debate gives us an opportunity to hone our reasoning skills; being exposed to non-rational political propaganda plausibly does not. There are some complexities this simple picture omits, but let us suppose that when we exert rational influence on some person, we are at least more likely, in the main run, to improve that person's rational capacities than if we had exerted non-rational influence instead. And let us suppose—quite plausibly—that this is a good thing,[38] and justifies a preference in favor of rational influence. If these premises are correct, it seems we have a broadly applicable reason to favor rational over non-rational influence.[39]

But it seems unlikely that this line of thought can serve as the basis for a substantial case against non-rational influence. The problem with resting our case on non-rational influence's (putatively) poor likelihood of cultivating the rational capacities is that this is a *weak* reason to disfavor non-rational influence. We often face a choice of different means to some end, some of which will, as a sort of "side benefit," cultivate a relevant capacity or skill, and others of which will not. For example, lifting heavy objects makes us stronger; relying on pulleys or machines does not. Riding a bicycle improves our physical endurance; taking a car or subway to our destination does not. Reading improves our literacy; watching films does not (though it may cultivate some different skill, like pictorial or filmic "literacy"). These are fine reasons to favor manual lifting, cycling, and reading over the alternatives, but they are pretty weak, as the examples are intended to show. Lots of us *do* prefer manual work, cycling, and reading when all else is equal (and even when all else is not quite equal), at least partly for the very reason that

[38] On the value of the rational capacities, see Part III below.

[39] Similar lines of argumentation have been used to critique Sunstein and Thaler's libertarian-paternalist program. See Rebonato [2012], pp. 217–220.

these methods improve useful capacities. But most of us (correctly) think it is perfectly fine to do a lot of work using mechanical assistance, to ride the subway five days a week, to go to the movies with friends, and so on, even though each of these choices involves giving up an opportunity to improve some skill or capacity. (And even though the relevant capacity may in fact atrophy if it isn't practiced with at least some regularity.)

The point is not just that it is crazy to avoid these methods *at all costs*; the point is stronger than that: for the most part we comfortably accept these choices as normal, unremarkable parts of life. Of course, it is always possible to argue that we should give tremendous weight to the cultivation of rational capacities in our interactions with others, but I find it hard to believe that a very convincing case can be made for such a strong proposition. There simply are too many other important priorities. A slightly polemical way of putting the point might be this: the world is not one gigantic classroom, and it would be unwise to adopt general-purpose maxims of action as if it were otherwise.

III. Risk of Error

So far, I've canvassed various potential objections to non-rational influence and found them insufficient to ground a substantial case against the practice. But I think there is a stronger story to be told about why it often makes sense to eschew non-rational influence. As I see it, the major problem with non-rational influence is that it involves heightened risks of inducing error, as compared to rational influence. By "inducing error" I mean moving a person's overall set of beliefs further away from an accurate and complete account of the world.[40] On this picture, gaining a new true belief or ceasing to hold a false belief moves one *closer* to an accurate and

[40] Further away, that is, than they already are.

complete account of the world; by contrast, gaining a new false belief or ceasing to hold a true belief moves one *further* from an accurate and complete account of the world. Hence, an instance of influence—rational or non-rational—induces error when it makes the target acquire a false belief or cease to hold a true belief. The hypothesis is that non-rational influence is more likely to do so than rational influence. I will call this the risk-of-error problem.

Let us start with the basics. It is generally good for us to have true beliefs and generally bad for us to have false beliefs. Yes, there are critiques of the purported value of true belief, and we can readily allow exceptions to the general rule, but few would argue that true belief is not *in the main run* preferable to false belief. In short, it is generally good for our beliefs to be closer rather than further from an accurate and complete account of the world. Note that this provides a pretty straightforward ethical case against lying and deception: lies and deception are intended to produce false beliefs, and therefore are intended to produce an outcome of a kind that is generally bad for the affected person. This is not the only objection one can make against lying and deception, and it only provides *pro tanto* (not decisive) reasons to eschew lying and deception. Still, it makes a fair bit of sense to think that, at least as a general rule of thumb, it is bad to make people believe false things, and that strong reasons are needed to justify deliberately doing so. The very same considerations justify a strong, *pro tanto* policy against using non-rational influence to instill false beliefs.[41]

[41] I would venture to guess that an essentially identical line of thought can justify a strong, *pro tanto* policy against some instances of non-rational influence that are *not* intended to instill false beliefs: namely, those intended to instill non-cognitive attitudes that, though not false, are in some sense the wrong ones to have. As discussed in Part I.B, however, I have chosen to limit my discussion in this paper to non-rational influence on beliefs.

But non-rational influence does not always produce false beliefs, nor is it always intended to do so.[42] A defense lawyer might use a carefully cultivated, charming, and authoritative tone of voice at trial, and this could subvert the jurors' rational capacities—at least to some degree—by causing them to give insufficient weight to the less charismatic prosecutor's arguments. For all that, however, the defendant might well deserve to be acquitted. Similarly, gruesome graphical warnings on cigarette packages may owe most of their effect (to the extent they have an effect) to non-rational influence. But again, no false information is necessarily communicated and no false beliefs need be produced; the goal is to make people adopt the true beliefs that smoking frequently leads to horrific health problems and that they should not smoke. In other words, individuals and institutions can pursue non-rational influence without endorsing the idea of "noble lies" or otherwise misleading people for the sake of some further, greater good. For this reason, we cannot straightforwardly object to non-rational influence on the grounds that it produces or is intended to produce false beliefs, any more than we can object to assertoric discourse in general because some assertions are lies.[43]

---

[42] This is similar, though not identical, to the observation that non-rational influence does not always amount to what Robert Noggle calls "manipulative action," which Noggle understands to be "the attempt to get someone's beliefs, desires, or emotions to violate" or "fall short of" norms or ideals governing beliefs, desires, and emotions—in other words, to lead people's attitudes "astray." Noggle [1996], p. 44. (This is a point that Noggle himself has emphasized. *Id.*, p. 49.)

[43] One might worry that even when non-rational influence does not produce false beliefs, it will at least produce beliefs that are not justified, and so will fail to produce knowledge. (I am grateful to Mark Berger and Steve White for pressing me on this point.) It would be difficult to address this concern fully without take positions on some very basic and controversial questions in epistemology, but a few points are worth mentioning. First, non-rational influence *can* produce beliefs that *are* justified, at least in an important sense. For example, a person exposed to sufficient evidence to justify the belief that smoking is unwise may, under the influence of wishful thinking, peer

But although non-rational influence does not always produce (and is not always intended to produce) false beliefs, it might still present heightened *risks* of doing so. By analogy, drunk driving does not always cause physical harm or property damage (in fact, it usually does not), and is almost never intended to do so. But there is still a very substantial case to be made against drunk driving on grounds that it is more likely than sober driving to cause such harms. The problem, simply put, is that drunk driving is risky, usually too risky to be justified.

Does non-rational influence present heightened, substantial risks of producing false beliefs, at least in a wide range of cases? Intuitively, it might seem clear that it does. Our rational capacities can be viewed as (perhaps among other things) a tool or system for avoiding false beliefs. Indeed, they are arguably our most important tool for doing so. One way of looking at the point is this: if we had to pick a long-term strategy for bringing our beliefs into closer alignment with the truth, it would be better to choose to have sharp rather than dull or unreliable rational capacities. Non-rational influence, by definition, involves bypassing or subverting these capacities. In this sense, being subjected to non-rational influence is like taking

---

pressure, commercial advertising, etc., fail to arrive at that belief. If being subject to non-rational influence in the form of gruesome warning labels causes this person to conclude that smoking is unwise, then this would appear to be a case in which non-rational influence has produced a new true belief that is, at least in one very important sense, justified. Second, even when non-rational influence *does* produce unjustified belief, it will often do so by replacing an unjustified *false* belief with an unjustified *true* belief. Although it is likely preferable, insofar as it is feasible, to replace unjustified false beliefs with justified true beliefs, it is hard to see why there would be anything wrong with replacing unjustified false beliefs with unjustified true beliefs. Third, even where non-rational influence causes a *justified* false belief to be replaced with an *unjustified* true belief, justification may in many cases be forthcoming quickly. (I discuss a case that fits this description shortly below, with the example of Adam and Ben.)

an injection of a substance that will either bypass or subvert the immune system, which can in an analogous manner be viewed as a tool for avoiding various harmful biological influences.

We can allow that in *some* cases we do and should want our rational capacities to be bypassed or subverted. The literature on manipulation is replete with examples of situations in which people might rationally desire to be subject to non-rational influence.[44] The most obvious cases, like people who pay hypnotists to help with an addiction or phobia, do not clearly involve non-rational influence *on beliefs*, but there are probably cases where it would be rational to want one's beliefs to be shaped by non-rational processes. Be that as it may, it is understandable that in the absence of unusual circumstances one would generally want to avoid being subject to non-rational influence, precisely in order to minimize the risks of being led into error.[45] By analogy,

---

[44] See, e.g., Noggle [1996], p. 49.

[45] One might argue that by conceding this point, I have given up the main reason I cited for disregarding objections to non-rational influence grounded on the fact that people generally do not, or would not, endorse or consent to such influence. (See *supra* note 19 and accompanying text.) After all, did I not claim that the problem with such theories is that they seem weak and uninteresting if people would be *mistaken* not to endorse or consent to non-rational influence? And if people *aren't* generally mistaken to feel this way, as I am apparently now allowing, doesn't that mean that (by my earlier logic), there is usually a good *pro tanto* reason not to exert non-rational influence on them?

The situation is not so simple. Because the question before us is whether there is a substantial case to be made against exerting non-rational influence, the perspective from which to evaluate the problem is that of the would-be influencer. And we are now considering cases in which the influencer is trying to spread a belief the influencer takes to be true, but which the target (obviously) does not presently take to be true. Under these circumstances, and from the influencer's perspective, there is a comparatively "subjective" sense in which it would be perfectly justified for the target to resist the influence, but also a comparatively "objective" sense in which it would not.

Consider an analogy. Imagine you are playing a casino game in which $1,000 has been placed randomly into one of two boxes: Box 1 or Box 2. One box will be opened, and if the prize is inside that box, you get to keep it. At

it is reasonable to want one's immune system to work well, and generally to resist having it be bypassed or subverted, precisely in order to minimize the risks of bodily sickness—though of course there are special cases in which we may quite rightly want our immune systems to be suppressed, such as when receiving an organ transplant.

This might seem sufficient to ground a pretty solid (if defeasible) rule against non-rational influence. But this is to look at the problem of risk from the perspective of would-be subjects or targets of influence: really what it justifies, if anything, is a general rule in favor of *avoiding being subject to* non-rational influence. The important perspective for our purposes, however, is that of would-be influencers: we want to see if we can justify a rule (at least of a defeasible, *prima facie* sort) against exerting non-rational influence, not just a rule against being

---

the start of play, the game is set up so that Box 1 will open: that is the "default" play. But if you would prefer to have Box 2 open instead, you can press a button to make it so. The only catch is that if you press the button, the prize will be reduced to $950. Of course, it would be pretty silly to press the button. Now suppose your friend Jane is watching over your shoulder. "Push the button!" she urges. As it happens, she actually *knows* the prize is in Box 2: she caught a glimpse of the money being put in Box 2, due to a negligent error on the casino's part. (Jane doesn't, however, want to explain this to you, for fear that a casino employee might overhear and reset the game.) You hesitate, but Jane says, "If you don't press the button, I'll do it *for you!*"

Would you be justified in not endorsing Jane reaching past you and and pressing the button? In a comparatively subjective sense, yes; as you see it, Jane would not be accomplishing anything except reducing your expected winnings. But in another, comparatively objective sense, you should not disapprove of Jane pressing the button: crucially, your disapproval only makes sense from a position of ignorance. So, should Jane refrain from intervening? It seems odd to say yes, at least if we bracket moral or legal concerns about taking advantage of the casino. You will be richer when the boxes open, and you will almost certainly thank Jane when she later explains the situation to you. That said, Jane might be well advised not to push the button if she harbored legitimate doubts about the prize's location; and this observation anticipates the core line of argument in the remainder of Part III.

subject to non-rational influence. It may seem picky to emphasize the distinction, but where risk assessments are at stake, it emphatically is not: risk is always evaluated against some background state of knowledge, and the correct risk assessment may vary quite starkly depending on the relevant perspective. And an influencer's epistemic situation may be quite different from that of the target(s) in relevant respects.

Consider the following case. Adam is walking along a lonely road, trying to hitch a ride. He walks past Ben, a driver stopped at a traffic light. As it happens, Adam is harmless and would even chip in for gas if given a ride. Unfortunately, there has been a recent series of robberies along this stretch of highway involving people pretending to be hitchhikers, and both Adam and Ben know this. It may well be, therefore, that there is simply no way for Adam to convince Ben to give him a ride using wholly rational means. Adam could, of course, tell Ben he is harmless, but his testimony is almost worthless from Ben's perspective: Ben has no particular reason to believe Adam, since claiming to be harmless is just what a robber would do, too. Suppose, however, that Adam makes a point of looking extra-piteous and downtrodden, and that Ben is (and *knows* that he is) a soft-hearted sucker who has an extremely hard time saying "no" to piteous-looking people even when he has good reason to do so.

It might well be sensible for Ben to avert his eyes from Adam to avoid the emotional pull of Adam's piteous appearance: Ben knows the sight of a piteous-looking person tends to subvert his rational capacities, to make him discount the good reasons he has for being cautious. Ben has an obvious reason to want to neutralize this source of non-rational influence: it makes him more likely (from his vantage point) to make a terrible decision. But the situation is different from Adam's perspective. Even if Adam knew his piteous appearance would subvert Ben's rational capacities, the fact remains that Adam, unlike Ben, *knows* he is harmless, and therefore—again,

unlike Ben—knows that if he exerts non-rational influence on Ben, it will not lead Ben astray. In fact, it will lead Ben to help a deserving person and even make some money, something Ben himself will see as a "win-win" in retrospect. In other words, there is a significant risk-of-error problem from Ben's perspective, but not from Adam's perspective. Consequently, there really is no strong reason for Adam not to put on a piteous face and exert non-rational influence, at least so far as I can tell. Adam is not engaged in nefarious "mind control"; he is not compromising Ben's autonomy; he certainly is not being disrespectful; and although he is not helping Ben *solve* his soft-heartedness problem, it is not as if he is making the problem appreciably worse.

Note, however, that this is a rather special situation. Adam has what is sometimes called "privileged access" to the information Ben lacks—namely, information about his own intentions.[46] Realistically there is no chance Adam could turn out to be wrong about his own harmlessness. But would-be influencers are not always so lucky: they cannot, that is, always be sure that the things they *think* are true really *are* true. For example, if Adam were not trying to convince Ben to give him a ride, but trying to get him to vote for a particular presidential candidate, it would probably be appropriate for Adam to be more cautious about using non-rational methods. Adam might believe that the progressive candidate is better than the conservative one, but ironclad certainty is rarely warranted when it comes to judgments of this kind. For all Adam knows, Ben might know more than Adam about what makes a good president, or might just be better at evaluating different candidates' character traits. It might, therefore, be very foolish indeed for Adam to bypass or subvert Ben's rational capacities in this

[46] See Gertler [2015] (discussing the notion of privileged access). Note that the example's usefulness does not depend on a strong claim to the effect that Adam's belief that he intends no harm is utterly infallible; all that is required here is that Adam is justified in believing to a *practical* certainty that he intends no harm.

context, assuming that Adam wants (as he apparently should) for Ben to have the correct beliefs about which candidate is best.

Marcia Baron points to problems of this kind when she suggests, in a discussion of manipulation, that "the presumption needs to be on the side of viewing others as rational beings, with whom one can engage in dialogue, and *from whom one might learn—possibly even learning that one is deeply mistaken*."[47] Of course, there are many circumstances in which dialogue is not a realistic possibility: advertisers and propagandists, for example, can choose between rational and non-rational methods, but the influence is going to be unidirectional either way. Still, Baron's point can be generalized. Even when we believe something and want to make others believe the same, we may not be so justifiably confident in our beliefs that we should be prepared to bypass or subvert others' rational capacities in order to spread the message, and in so doing bypass or subvert the "independent check" or defense against error their rational capacities provide. Even if we *intend* to cause people to adopt true beliefs, the use of non-rational methods plausibly makes it less likely that the would-be targets of influence will (correctly) realize that the beliefs *we* think are true are, in fact, false. This is basically irrelevant in the Adam/Ben scenario because for practical purposes Adam is justified in having absolute confidence that he will not rob Ben. But as would-be influencers, we cannot be justified in assuming that all or even most cases are like this, unless we rightly have ironclad confidence in our views on a broad range of subjects. This, I think, explains why a readiness to use non-rational influence under all circumstances and without a second thought does, and should, come across as arrogant or presumptuous.[48]

---

[47] Baron [2003], p. 50 (emphasis added).

[48] *Cf. id.*, p. 50 (arguing that "being manipulative is a vice because of its arrogance and presumption").

The theory, then, is this: the major problem with non-rational influence is that it generally presents a greater risk of inducing error than rational influence does. Insofar as this is the primary problem with non-rational influence, the objectionableness of any particular attempt at non-rational influence will largely depend on how risky it is under the circumstances—just as drunk driving is most objectionable when the roads are busy, or when the driver is very drunk rather than only slightly drunk.

IV. The Limits of the Theory

The theory described above posits that we have the following broadly applicable, reasonably strong *pro tanto* reason to eschew non-rational influence: compared to rational influence, it presents a greater risk of inducing error. I think this theory does have significant merit, but it also has considerable limits. The risk that non-rational influence will induce error will vary greatly from case to case, as will the gravity of any error that might be induced. (Needless to say, potentially countervailing reasons, such as the importance of spreading some putatively true belief, as well as the difficulty of achieving the same end through more rational means, will also vary tremendously from case to case.)

Of course, all of this is rather platitudinous, and it certainly does not mean there is no sense in adopting, so to speak, a general rule of thumb disfavoring the use of non-rational influence—any more than the wide variation in the risks of driving drunk under different circumstances means there is no sense in making it a rule not to drive drunk. But why think we should generally be reluctant to exert non-rational influence because of the risk of inducing error, rather than simply feel unconcerned about it except in special circumstances where the risk is especially acute? Perhaps the better analogy isn't drunk driving, but *plain* driving: after all,

sober driving itself presents heightened risks in comparison with various alternatives, but we generally accept these risks as an unproblematic part of daily life. The issue is remarkably hard to adjudicate, and I can see no way forward except to proffer another suggestion modeled on Baron's work on manipulation: that we strive to know "when it is appropriate to try to bring about a change in another's [beliefs] and [do] this for the right reasons, for the right ends, and only in instances where it is warranted (and worth the risks)"[49]—the major pertinent risk being, as I've argued, the risk of inducing error.

A large part of the difficulty in justifying any general rules here, even weak rules subject to various caveats and exceptions, is that under the risk-based account developed here, it is far from clear whether there is any sense in attempting to offer suggestions, rules, or maxims of conduct for an amorphous and unspecified "we," given the great differences in our (that is, we humans') epistemic situations and rational capacities. If not for these differences, most attempts to influence by non-rational means might plausibly be seen as reflecting a reckless disregard for the risks of spreading error, or arrogance about the superiority of our own epistemic situation or rational capabilities. But this is not a remotely accurate description of the real world: there *are* profound differences in the quality of our epistemic situations and rational capabilities, and these have correspondingly profound implications for whether and when the risk-of-error problem is severe enough to offer a really significant reason against exerting non-rational influence.

I admit that this arguably has inegalitarian implications: on the theory offered here, those who are *generally* justifiably confident in their ability to locate the truth have less reason to eschew non-rational influence than those who are not. Insofar as people do vary in this respect, therefore, this suggests that with regard to the appropriateness of exerting non-rational influence,

[49] *Id.*, p. 48.

what is wise for thee may not be wise for me (or vice versa). I happen to think people do vary considerably in this respect, and that this is one roadblock in the way of making broad claims about the ethics of non-rational influence. But even if there were no average differences in the degree to which different people could be justifiably confident in their beliefs,[50] it would still be difficult to offer any particularly broad claims about the acceptability of non-rational influence under the risk-based theory on offer here, if only because any *given* person's level of justifiable confidence will vary tremendously from belief to belief, and from context to context. This implies that even if non-rational influence does not, as a general matter, present greater risks of inducing error for some influencers than for others, the risk-of-error problem will still vary greatly in severity from specific case to specific case. Furthermore, some people will be justified in having great confidence about their beliefs on a given subject, whereas other people would not be justified in having great confidence about their beliefs on the very same subject. After all, expertise varies, and this low-level diversity would interfere with any effort to simplify the ethics of non-rational influence even if, counterfactually, high-level equality could be taken for granted.

It may seem uncharitable or hasty to end such a short and preliminary inquiry by suggesting that our pre-theoretic intuitions about the ethics of non-rational influence may be misguided. So I will instead end with a few conciliatory remarks.

First, it seems to me that our negative intuitions about the ethics of non-rational influence are driven by certain highly salient areas in which non-rational influence is common: namely, commercial advertising and political campaigns. And in these two contexts there is unusually good reason to doubt whether influencers are both (1) in a good position to judge to truth of the beliefs they are trying to spread, and (2) likely to be motivated to avoid inducing error in others.

[50] Average, that is, over the range of each person's various beliefs on different subjects.

The problem is that in these contexts, influencers have an immediate and tangible stake—whether in money or power—in spreading certain beliefs, largely regardless of whether those beliefs are true.

The goal of advertising is to sell products; helping people *actually* make the right buying decisions is often (though not, in fairness, always) quite incidental. Likewise, the goal of political campaigns is to win, and one suspects that most political candidates are not overly concerned about whether, for example, they are really the best for the job—and, in any case, that they are absurdly biased in their own thinking about such matters. If true beliefs are to prevail and error is to be avoided in these contexts, it is probably because of the sound functioning of *consumers'* and *voters'* rational capacities. When advertisers and politicians engage in non-rational influence, this error-avoidance mechanism is (by definition) bypassed or subverted. Insofar as our intuitions about the ethics of non-rational influence are driven by our views on advertising and electoral politics, therefore, it is understandable why we would instinctively consider it a bad practice. Of course, none of this is to say that the rare advertiser or politician who *is* justifiably confident in the truth of the beliefs he or she hopes to spread should eschew non-rational influence. The point is just that the average advertiser or politician is probably less likely to be justifiably confident in the truth of the beliefs he or she is trying to spread than, say, the average teacher, parent, or friend. In sum, it may be that our pre-theoretic tendency to be disturbed by non-rational influence in advertising and political campaigns is, in fact, justified, but that it unjustifiably bleeds over into our *overall* views on the ethics of non-rational influence.

Second, many of us have a strong aversion to the use of non-rational influence in pedagogy. This aversion, too, may well be justified. In contrast to advertising and politics, what sets pedagogical contexts apart from the main run of human interactions probably *isn't* that

teachers are especially unlikely to be justifiably confident in the truth of the beliefs they hope to spread. (We may suppose, as we would certainly hope, that this is not the case!) More plausibly, what is special about pedagogy is that a core part of a teacher's job is to hone students' rational capacities. In Part II.D, I argued that we could not justify a general case against non-rational influence on grounds that the use of non-rational influence represents a lost opportunity to cultivate the rational capacities. That, however, was a broad generalization about human interaction across a wide range of contexts. The analysis plausibly differs for teachers.

Why think pedagogy is special in this respect? To return to an analogy discussed in Part II.D, it would be hard to make a general case against using mechanical assistance to lift heavy objects on grounds that manual labor does more to improve physical strength: we cannot reasonably focus with morbid intensity on the cultivation of physical strength when there are so many other important priorities. That being said, there is a *very* good case to be made against *personal fitness trainers* permitting or assisting their clients' use of levers and pulleys to lift heavy weights in the gym: after all, the primary point of fitness training is to improve physical strength. There is, I think, a similar case to be made that non-rational influence in pedagogy has especially heavy downsides. That said, those of us in whose imaginations pedagogy looms large—saliently including, I suppose, a majority of those who read academic philosophy—should not allow our aversion to non-rational influence in pedagogy to bleed unjustifiably into our views about other kinds of human interaction. In brief, it seems right to be leery about non-rational influence in the classroom, but, to return to the theme on which I ended Part II.D, it would be a mistake to fall into the trap of treating the whole world as if it should be governed by norms appropriate to classrooms.

Third and finally, people tend to be overconfident in their beliefs,[51] and confidence often correlates poorly with accuracy.[52] For this reason, the proposal that influencers should not be particularly reluctant to engage in non-rational influence when they are justifiably confident in their beliefs may seem quite dangerous and potentially destructive. If this advice were taken to heart, people might underestimate the risk-of-error problem and thus resort too readily to non-rational influence, simply because they are often poor judges of how likely it is that their views are correct. The best remedy for this, it would seem, would be to spread a message of epistemic humility, and to urge people to recognize that they are probably bad judges of whether they are good judges. It would be fair to worry, however, that such a message would largely fall on deaf ears. So perhaps the ethics of non-rational influence is a subject on which it is best for the conventional wisdom to be what it is, even if the actual reality is more complex and murky.

[51] See Dunning [2005], pp. 7–9.

[52] See *id.*, pp. 8, 43–44, 48.

References

Alexander, Larry. 1992. "What Makes Wrongful Discrimination Wrong? Biases, Preferences, Stereotypes, and Proxies," *University of Pennsylvania Law Review* 141 (1): 149–219.

Barnhill, Anne. 2015. "I'd Like to Teach the World to Think: Commercial Advertising and Manipulation," *Journal of Marketing Behavior* 1 (3–4): 307–328.

Baron, Marcia. 2003. "Manipulativeness," *Proceedings and Addresses of the American Philosophical Association* 77 (2): 37–54.

—. 2014. "The *Mens Rea* and Moral Status of Manipulation," in *Manipulation: Theory and Practice*, eds. Christian Coons and Michael Weber, 98–120. Oxford: Oxford University Press.

Blumenthal-Barby, J. S. 2012. "Between Reason and Coercion: Ethically Permissible Influence in Health Care and Health Policy Contexts," *Kennedy Institute of Ethics Journal* 22 (4): 345–366.

Bratman, Michael E. 2005. "Planning Agency, Autonomous Agency," in *Personal Autonomy: New Essays on Personal Autonomy and Its Role in Contemporary Moral Philosophy*, ed. James Stacey Taylor, 33–57. Cambridge: Cambridge University Press.

Buss, Sarah. 2005. "Valuing Autonomy and Respecting Persons: Manipulation, Seduction, and the Basis of Moral Constraints," *Ethics* 115 (2): 195–235.

—. 2014. "Personal Autonomy," in *The Stanford Encyclopedia of Philosophy* (Winter 2014 Edition), ed. Edward N. Zalta, http://plato.stanford.edu/archives/win2014/entries/personal-autonomy/.

Christman, John. 2005. "Procedural Autonomy and Liberal Legitimacy," in *Personal Autonomy:*

New Essays on Personal Autonomy and Its Role in Contemporary Moral Philosophy, ed. James Stacey Taylor, 277–298. Cambridge: Cambridge University Press.

Coons, Christian, and Michael Weber. 2014. "Introduction: Investigation the Core Concept and Its Moral Status," in *Manipulation: Theory and Practice*, eds. Christian Coons and Michael Weber, 1–16. Oxford: Oxford University Press.

Dillon, Robin S. 2015. "Respect," in *The Stanford Encyclopedia of Philosophy* (Fall 2015 Edition), ed. Edward N. Zalta, http://plato.stanford.edu/archives/fall2015/entries/respect/.

Dunning, David. 2005. *Self-Insight: Roadblocks and Detours on the Path to Knowing Thyself*. New York: Psychology Press.

Dworkin, Gerald. 1976. "Autonomy and Behavior Control," *Hastings Center Report* 6 (1): 23–28.

—. 1998. *The Theory and Practice of Autonomy*. Cambridge: Cambridge University Press.

Finlay, Stephen, and Mark Schroeder. 2015. "Reasons for Action: Internal vs. External," in *The Stanford Encyclopedia of Philosophy* (Winter 2015 Edition), ed. Edward N. Zalta, http://plato.stanford.edu/archives/win2015/entries/reasons-internal-external/.

Fischer, John Martin, and Mark Ravizza. 1998. *Responsibility and Control: A Theory of Moral Responsibility*. Cambridge: Cambridge University Press.

Fleming, Patricia Ann. 1989. "The Moral Status of Deprogramming," *Journal of Applied Philosophy* 6 (1): 77–86.

Frankfurt, Harry. 2002. "Reply to John Martin Fischer," in *Contours of Agency: Essays on Themes from Harry Frankfurt*, eds. Sarah Buss and Lee Overton, 27–31. Cambridge: MIT Press.

Gaylin, William, and Bruce Jennings. 2003. *The Perversion of Autonomy: Coercion and Constraints in a Liberal Society* (Revised and Expanded Edition). Washington: Georgetown University Press.

Gertler, Brie. 2015. "Self-Knowledge," in *The Stanford Encyclopedia of Philosophy* (Summer 2015 Edition), ed. Edward N. Zalta, http://plato.stanford.edu/archives/sum2015/entries/self-knowledge/.

Gibbard, Allan. 1990. *Wise Choices, Apt Feelings*. Cambridge: Harvard University Press.

Gorin, Moti. 2014. "Do Manipulators Always Threaten Rationality?" *American Philosophical Quarterly* 51 (1): 51–61.

Lenman, James. 2016. "Reasons for Action: Justification vs. Explanation," in *The Stanford Encyclopedia of Philosophy* (Spring 2016 Edition), ed. Edward N. Zalta, http://plato.stanford.edu/archives/spr2016/entries/reasons-just-vs-expl/.

Mele, Alfred R. 1995. *Autonomous Agents: From Self-Control to Autonomy*. Oxford: Oxford University Press.

Miller, Geoffrey. 2009. *Spent: Sex, Evolution, and Consumer Behavior*. New York: Penguin.

Noggle, Robert. 1996. "Manipulative Actions: A Conceptual and Moral Analysis," *American Philosophical Quarterly* 33 (1): 43–55.

Oshana, Marina A. L. 2005. "Autonomy and Free Agency," in *Personal Autonomy: New Essays on Personal Autonomy and Its Role in Contemporary Moral Philosophy*, ed. James Stacey Taylor, 183–204. Cambridge: Cambridge University Press.

—. 2006. *Personal Autonomy in Society*. Aldershot: Ashgate.

Rebonato, Riccardo. 2012. *Taking Liberties: A Critical Examination of Libertarian Paternalism.* New York: Palgrave Macmillan.

Sunstein, Cass R. 2015. “Fifty Shades of Manipulation,” *Journal of Marketing Behavior* 1 (3–4): 213–244.

Taylor, James Stacey. 2005. “Introduction,” in *Personal Autonomy: New Essays on Personal Autonomy and Its Role in Contemporary Moral Philosophy*, ed. James Stacey Taylor, 1–29. Cambridge: Cambridge University Press.

—. 2009. *Practical Autonomy and Bioethics*. New York: Routledge.

Tsai, George. 2014. “Rational Persuasion as Paternalism,” *Philosophy and Public Affairs* 42 (1): 78–112.