Free & Equal | A Journal of Ethics and Public Affairs

Renzo, Massimo 2025 Manipulation and Practical Agency. *Free & Equal: A Journal of Ethics and Public Affairs*, 1(1): 108–138. DOI: https://doi.org/10.16995/fe.17722

Open Library of Humanities

Manipulation and Practical Agency

Massimo Renzo, King's College London, UK,
massimo.renzo@kcl.ac.uk

Philosophers typically argue that manipulation is wrong because it impairs our practical reasoning. Recently, Sophie Gibert has challenged this view, proposing instead a reductive account of the wrong of manipulation. Gibert's account is reductive in that it dispenses with the idea that there is a distinctive non-moral feature of manipulation upon which its wrongness supervenes. Rather, the wrong of manipulation is nothing over and above the wrong of infringing other rights. In this paper, I raise a number of objections against Gibert's reductive account, arguing that it is both overinclusive and underinclusive. I then propose a novel formulation of the traditional view. I argue that, in addition to avoiding the sort of problems raised by Gibert, my reformulation of the view enables us to give a more nuanced treatment of central cases of manipulation, including some that are relevant to law and policymaking, such as the morality of nudging. My reformulation comes at a price. The view I offer rests on a fairly demanding picture of the kind of practical reasoning required to exercise our practical agency well. But since this picture is independently plausible, this is a price we should be happy to pay.

Manipulation and Practical Agency

Massimo Renzo

I. INTRODUCTION

What makes manipulation wrong? By far, the most popular answer philosophers have given to this question is that manipulation is wrong because it impairs or circumvents our practical reasoning.[1] Call this the "Reasoning View." In a recent paper, Sophie Gibert rejects this widely shared view, offering in its stead an alternative, reductive, account of the wrong of manipulation.[2] Gibert's account is reductive in that it dispenses with the idea that there is a distinctive non-moral feature of manipulation upon which its wrongness supervenes (such as the fact that it impairs or circumvents our practical reasoning). Rather, as she puts it, "the wrong of wrongful manipulation is nothing over and above the wrong of infringing other rights."[3] If she is correct, this is a significant result, one that should lead us to rethink the way manipulation is normally conceived. Ultimately, I remain unpersuaded, but I believe there are important lessons to be learned from engaging with Gibert's account.

In what follows, I start by raising a number of objections against Gibert's "Reductive View," arguing that the view is both overinclusive and underinclusive. Moreover, I argue that the view fails to account for cases of wrongful manipulation in which no one is wronged. I then move on to defending the Reasoning View. While Gibert's discussion is largely devoted to the objection that the Reasoning View is overinclusive, I believe the worries she raises on this front can be addressed relatively easily, by acknowledging the role played by conventions in shaping many social interactions. It's the objection that the Reasoning View is underinclusive that, to my mind, raises a more serious challenge. While this challenge should not lead us to abandon the view, it does call for a significant revision of the way in which it is normally formulated. This amendment comes at a price, as we shall see. The Reasoning View will need to rest on a fairly demanding picture of the kind of practical reasoning required to exercise our

[1] The claim, of course, is that manipulation is only *pro tanto* wrong. It's understood that, in some cases, it might be permissible all-things-considered.

[2] Sophie Gibert, "The Wrong of Wrongful Manipulation," *Philosophy & Public Affairs* 51, no. 4 (2023): 333–72, https://doi.org/10.1111/papa.12247.

[3] Gibert, "Wrongful Manipulation," 334.

practical agency (i.e. our capacity to respond to our reasons for action) well.[4] But since this picture seems to me independently plausible, this is a price we should be happy to pay.

Two caveats before we begin. First, while my aim is to defend a version of the Reasoning View, I should stress that I don't believe the Reasoning View can provide an exhaustive account of manipulation. This is because there are different ways in which we can interfere with other people's efforts to exercise their practical agency when we manipulate them.[5] It's only to be expected that different forms of wrongness attach to each of them. I will argue that the Reasoning View, correctly understood, has the resources to provide a convincing account of one of these types of manipulation. It would be a mistake, however, to think that it can explain the wrongness of all of them.

Second, while my aim is to defend a version of the Reasoning View, I will not provide a full defence of the view here. Doing so would require not only revising the picture of practical agency on which the view normally rests but also paying attention to a feature of manipulation that has so far been neglected in the literature, namely, the distinctive way in which the interference with our efforts to exercise our practical agency is carried out when we are manipulated. In this paper, I only tackle the first of these tasks. The second must wait for another day.[6]

II. GIBERT'S ACCOUNT

This is how Gibert formulates the Reasoning View:

> REASONING VIEW: Wrongful manipulation is wrong because it alters the target's practical reasoning in a particular way—a way that is specifiable in non-moral terms.[7]

The distinguishing feature of this view is that it proceeds by identifying a particular manner in which manipulation affects our practical reasoning and then uses this idea to explain why manipulation is wrong. So, for example, it is sometimes argued that

---

[4] I prefer the expression "practical agency" to "rational agency" because the latter is sometimes associated with a rationalistic notion of practical reasoning – one that requires conscious deliberation and reflection on our reasons for action, as well as detachment from our emotions. For a powerful criticism of this picture of practical reasoning, see Nomy Arpaly *Unprincipled Virtue: An Inquiry Into Moral Agency* (Oxford: Oxford University Press, 2002), https://doi.org/10.1093/0195152042.001.0001.

[5] In Massimo Renzo, "Why Manipulation Is Wrong," *Political Philosophy*, 2, no. 1 (2025), https://doi.org/10.16995/pp.17395, I distinguish four types of manipulative interference.

[6] Renzo, "Why Manipulation Is Wrong," pp. 191-6.

[7] Gibert, "Wrongful Manipulation," 336.

we are wrongfully manipulated when someone influences our deliberation by offering us reasons that bypass our practical reasoning (Gibert calls this the "Non-rational Influence View") or when someone interferes with our practical reasoning in a way that makes our reasoning worse (the "Quality of Reasoning View"). On these views, the type of treatment upon which the wrong of manipulation supervenes is explained in non-moral terms. (The claims that a certain influence bypasses our practical reasoning, or that it makes our reasoning worse, are not themselves moral claims.) The wrong of manipulation is thus basic, Gibert argues, in the sense that it is not to be explained by appealing to further moral claims. The wrong is explained in non-moral terms.

Why should we reject the Reasoning View? Gibert's central objection is that the view is overinclusive. The Non-rational Influence View, for example, seems bound to count as instances of wrongful manipulation behaviours that are clearly morally innocent. Wearing perfume on a date, dressing nicely for an interview, or baking cookies for an open house all constitute forms of non-rational influence. And yet, these clearly are not instances of wrongful manipulation.

The Quality of Reasoning View faces similar problems, given that we lack a general obligation not to worsen other people's practical reasoning. For example, I do not wrong you if during a financial negotiation I use tactics like flattering you or pretending that I'm not interested in your offer. Nor do I wrong you if I bluff in a poker game, or if I trash talk to disrupt your performance during a basketball game. Finally, an attorney does no wrong in choosing a charismatic witness over a less effective one, or in asking questions that will paint them in a favourable light, even if those questions are not directly pertinent to the case. These behaviours are not wrong despite the fact that they all involve making someone's reasoning worse in some respects.

On the other hand, and this is the second objection raised by Gibert to the Reasoning View, we can think of cases in which manipulating someone will *improve* their reasoning. For example, displaying disturbing pictures of lung disease on cigarette packages might help people to give the right weight to the risks associated with smoking.[8] Or consider the following case:

> "India and Sebastian are having a serious discussion about who will make career sacrifices next year to care for their kids. Sebastian has recently observed that India tends to undervalue her own preferences and needs, but that she is slightly less inclined to do so when she is angry at someone. Before their conversation, Sebastian ensures that India overhears him on the phone with his mother, divulging some

[8] Gibert, 341.

> of their marital issues that they have agreed to keep private. As a result, India gets angry and gives more appropriate weight to her preferences in their subsequent discussion."[9]

Here Sebastian seems to have wrongfully manipulated India, Gibert argues. Far from having worsened her reasoning, however, he has improved it. If so, the Reasoning View is underinclusive, as well as overinclusive.

Where does this leave us? Let's start with the objection of over-inclusiveness. Gibert is not saying, of course, that in business transactions, competitions or in court anything goes. Flattering you during a business negotiation might not wrong you, but lying about the value of my assets or forging documents does. Bluffing at poker is not wrong, but hiding cards in my sleeve is. And while painting a witness in a positive light is not a wrongful way to attempt to influence jurors, intimidating them is. So, the question is: How can we explain the difference between the cases in which worsening someone's reasoning is (*pro tanto*) wrong and those in which it is not? Since both types of cases feature an attempt to negatively impact someone's practical deliberation (either by circumventing it or by impairing it), this idea cannot be used to discriminate between them.[10] What does then?

For Gibert, it's the fact that in some cases you have independent rights that are violated when I interfere with you in certain ways, whereas in other cases you don't. For example, you have a right not to be lied to, but not a right that I don't flatter you. This explains why my attempt to manipulate you by lying about the value of my assets constitutes a form of wrongful manipulation, whereas my attempt to do so by flattering you doesn't. Similarly, when we play poker, you have a right that I don't use hidden cards, but not a right that I don't bluff. And when we are in court, you have a right that I don't intimidate jurors, but not a right that I don't choose charismatic witnesses and paint them in a favourable light. So, it's the fact that in some cases your rights are being violated that explains why you are wrongfully manipulated.

We thus get to Gibert's preferred view:

> REDUCTIVE VIEW: Manipulation is wrong if, only if, and because it influences the target's practical reasoning—though not by providing additional reasons—in a way that infringes one or more of their other rights—specifically, their rights against interference.[11]

---

[9] Gibert, 359.
[10] Gibert, 353.
[11] Gibert, 334.

The view is reductive, Gibert clarifies, because "it implies that the wrong of wrongful manipulation is nothing over and above the wrong of infringing other rights. On the Reductive View, all wrongfully manipulative acts are wrong in virtue of some other wrong they involve—namely, a rights-infringement." In this sense, "the wrong of wrongful manipulation is non-basic: it cannot be explained in non-moral terms."[12]

III. REJECTING THE REDUCTIVE VIEW

The ambition of the Reductive View is to provide an account of the wrong of manipulation without appealing to the idea that manipulation involves interfering with someone's practical reasoning in the way suggested by the Reasoning View. A good way to test the view is thus to consider cases in which the conditions of the Reductive View are fulfilled but this kind of interference is absent. Consider:

> ***Gossip***: You know the identity of my biological father. I don't. Suppose it is generally agreed in our society that this sort of information ought not to be shared with people, unless they explicitly request it, given how distressing it can be to have it. Still, ever unable to let a good bit of gossip go to waste, you decide to tell me. As a consequence of that, I decide to contact my biological father.

Here, by revealing the information, you affect my practical reasoning in a way that infringes one of my rights against interference. The Reductive View would thus have to say that this is an instance of wrongful manipulation. But this seems implausible. While I am certainly wronged by your conduct, intuitively your behaviour does not seem to be an instance of wrongful *manipulation*. Why not? What's missing? It's tempting to reply that, although you have affected my practical reasoning in a way that infringes my rights against interference, you haven't affected it in the distinctive way required for the behaviour to count as manipulative. Manipulative behaviour has a certain shape. It involves an effort to influence someone's practical reasoning so as to induce a distinctive kind of response – say, by leading them to act against what they have reasons to do or some such. The reason why *Gossip* does not seem an instance of manipulation is that this component is absent. Despicable as it is, your gossiping is not aimed at producing this kind of response. You just can't keep your mouth shut.

Notice that my point here is not that manipulation needs to be intentional. If it was, Gibert could reply that since the intentionality condition is contested by some writers,[13] any counterexamples built around it are going to have limited force. My point

[12] Gibert, 334–35.

[13] Kate Manne, "Non-Machiavellian Manipulation and the Opacity of Motive," in *Manipulation: Theory and Practice*, eds. Christian Coons and Michael Weber (Oxford: Oxford University Press, 2014), 221–245,

in the text, however, is accepted even by philosophers like Kate Manne, for whom manipulation is not necessarily intentional, in the sense that it can be the product of motives that are opaque to the agent. In describing her central case of a woman who sends expensive presents to her relatives to make them feel guilty about not visiting her, Manne writes that "[h]er gifts are effectively *designed*... to make them all feel guilty for not paying her more attention."[14] This sort of design is what's missing in *Gossip*. Whether you think that the *mens rea* of manipulation is intention, recklessness, or something else,[15] the *actus reus* must involve something like the design Manne talks about. The aim of the conduct, whether conscious or unconscious, must be to induce in the target a certain response. Take that away, and you get cases like *Gossip*, which, intuitively, do not seem to be cases of manipulation.

If I'm right about this, the Reductive View turns out to be overinclusive, which is especially interesting since this is Gibert's main criticism of the Reasoning View. The Reductive View, however, seems also underinclusive. To see this, consider the example of business negotiations. As Gibert notices, a number of manipulative strategies are deemed not wrongful during a business negotiation, including flattering the counterpart, overwhelming them with useless information and surrounding them by good looking negotiators they might be especially keen to please.[16] Other strategies, such as revealing embarrassing private information about them or intimidating them through a violent fit of rage, are clearly wrongful. As we have seen, what explains this difference, according to the Reductive View, is the fact that whereas we have an independent right to privacy and an independent right not to be intimidated, we don't have a right not to be flattered or not to be surrounded by good looking people. But suppose now that instead of arranging to have conventionally good-looking people in the meeting, your counterpart arranges for someone who bears a striking resemblance to your recently deceased mother to be there. That seems to me an instance of wrongful manipulation. But which further right of yours is being violated here? We don't have an independent right to not be exposed to people who resemble those who are dear to us. If this is an instance of wrongful manipulation, it's tempting to think that it's because it involves playing with your emotions in a way that is meant to negatively

https://doi.org/10.1093/acprof:oso/9780199338207.003.0011; Moti Gorin, "Do Manipulators Always Threaten Rationality?," *American Philosophical Quarterly* 51, no. 1 (2014): 51–61.

[14] Manne, "Non-Machiavellian Manipulation," 225, italics in the original.

[15] On this question, see Marcia Baron, "The Mens Rea and Moral Status of Manipulation," in *Manipulation: Theory And Practice*, eds. Christian Coons and Michael Weber (Oxford: Oxford University Press, 2014), 98–120, https://doi.org/10.1093/acprof:oso/9780199338207.003.0005.

[16] The last one is a recurring example in Gibert's discussion, though not one she herself uses in the negotiation scenario (Gibert, "Wrongful Manipulation," 352).

interfere with your ability to effectively exercise your practical reasoning during the negotiation process. Of course, we might say that you have a right against *that*. But this will not do, since this right is precisely the right against having one's practical reasoning compromised that the Reasoning View appeals to. The Reductive View cannot help itself to it. The view must appeal to some further right in the background, but it's unclear what that right might be.

Gibert might reply that the right in question is the right to not be gratuitously upset. But the worry with this reply is similar to the one we've encountered in discussing *Gossip*. The violation of this right might explain why you are wronged by being exposed to someone who has a striking resemblance to your deceased mother. But unless this interference is designed to induce a specific response, such as making a mistake in the negotiation, the wrong you suffer when the violation of this right affects your practical reasoning is not the wrong of manipulation.

To see this, consider a version of the case in which it's just a coincidence that someone at the meeting resembles your mother. Suppose their presence upsets you, and that the (nonculpable) violation of your right to not be gratuitously upset affects your practical reasoning, say, by inducing you to accept a lowball offer. You still would not be suffering the wrong of manipulation. (Just as you wouldn't if what throws you off during the negotiation is the fact that your mother's favourite song happened to be playing in the waiting room before the meeting.) Gibert's view cannot vindicate this idea, however, since her conditions for wrongful manipulation are met in this version of the case. Thus, once we appeal to a right not to be gratuitously upset to answer the objection of underinclusiveness, Gibert's view becomes, once again, overinclusive.

A final worry about the Reductive View is that it's unable to vindicate cases in which we act wrongfully by manipulating someone, despite the fact that no one is wronged by our conduct. And yet such cases seem to exist. Consider:

> Jealous Guy: Yoko and Guy are in an exclusive relationship. Because she has once made Guy jealous by flirting with someone else, Yoko consents to Guy making her jealous, if he wants. Guy flirts with someone at a party, exclusively to make Yoko jealous. Flustered, Yoko goes on to embarrass herself at the party, making a scene and getting angry at the host for no good reason.

Making Yoko jealous seems an instance of wrongful manipulation (assuming, as I do here, that doing so serves no good purpose), yet Yoko is not wronged by it.[17] Yoko

[17] Some might be tempted to deny that this is an instance of manipulation since Yoko fully understands what Guy is doing and why. But manipulation is not necessarily covert. As Gibert notices, some athletes use trash talk to get each other off their game, a tactic that seems manipulative even if its purpose is well

is not wronged because she has consented to this treatment, and thus she has waived whatever rights normally ground her claim against these treatments. However, the fact that Yoko is not wronged does not mean that Guy does not act wrongfully in manipulating her. If there are no good reasons to make Yoko jealous, doing so seems wrong.

This is in line with how consent operates in situations that do not involve manipulation. If you consent to me punching you in the face, you make it the case that I do not wrong you by doing so. But punching you would still be wrongful, insofar as it causes harm for no good reason. Now, if Guy's conduct is wrongful but Yoko is not wronged by it, how can we explain that? Once again, the Reasoning View has a natural answer to this question: Guy's conduct is wrongful because undermining Yoko's practical reasoning in certain ways is wrong, even if she lacks a right against it.

Perhaps Gibert might reply here that she is only interested in cases of manipulation that are wrongful *in that they wrong someone*.[18] If so, her view is simply not meant to cover cases like *Jealous Guy*. That's an acceptable answer, as far as it goes. But if it's true that the Reasoning View provides a unified account that deals both with cases like this and cases in which the target of manipulation is wronged, reasons of parsimony and elegance would support it over the Reductive View. Of course, this is assuming that the Reasoning View can be defended from Gibert's criticisms. Time now to move to that question.

IV. WHY THE REASONING VIEW IS NOT OVERINCLUSIVE

Gibert's main objection to the Reasoning View is that it's overinclusive. To make this point, she considers a number of situations in which someone's practical reasoning is negatively affected in the way described by the Reasoning View without this constituting an instance of wrongful manipulation. Examples include wearing perfume on a date, dressing nicely for an interview, choosing attractive salespersons to man a food stand or flattering a competitor in a financial negotiation. Some of these cases are best characterized as instances of non-wrongful manipulation; others don't seem instances of manipulation to begin with. Regardless, the wrong of manipulation is absent from all of them.

known to its victim. Similarly, if you and I are competing for the same job, I might succeed in making you nervous before your interview even if you are fully aware of what I'm doing. Indeed, the effectiveness of these strategies is typically enhanced when the victim is aware of being manipulated.

[18] Gibert does write that she "assume[s] that when manipulation is *pro tanto* wrong, this is because it *pro tanto* wrongs the person being influenced" ("Wrongful Manipulation," 341), though that assumption is never motivated.

I will argue that some versions of the Reasoning View –more precisely, of the Quality of Reasoning View– have a natural way to deal with this objection.[19] Before we consider this line of reply, however, it's worth pausing to spell out exactly where the Reductive View and the Reasoning View come apart. Gibert argues that within the Reasoning View, the wrong of manipulation is basic in the sense that "the features of wrongful manipulation that make it wrong can be defined non-morally, without reference to any other wrong that's involved."[20] Now, I take it that defenders of the Reasoning View can be assumed to share the standard view of wronging, according to which A wrongs B by φ-ing when A violates a directed duty not to φ that A has toward B, i.e. a duty that correlates to a claim B has that A does not φ. If so, the very notion of wrongful manipulation that Gibert has in mind (i.e. manipulation that wrongs someone) necessarily involves appealing to the idea that someone's right is being violated.[21] In this respect, it's potentially misleading to characterize the Reasoning View as saying that "the features of wrongful manipulation that make it wrong can be defined non-morally." For an instance of manipulation to count as wrongful, according to the Reasoning View, it's not enough that someone's practical reasoning is negatively impacted in the way described by the view. It must be the case that the target has a claim against their practical reasoning being impacted in this way and that this claim is being violated. What sets the Reductive View apart from the Reasoning View is not that the former involves appealing to the idea that someone's rights are violated and the latter does not, but rather that the former grounds the wrong of wrongful manipulation in the violation of *some further rights to non-interference*, such as the right that we have not to be lied to about the value of someone's assets in a business transaction or the right that jurors in our trial are not intimidated.[22]

With this in mind, defenders of the Reasoning View can address the cases discussed by Gibert by pointing out that they involve special practices or conventions in virtue of which the right against having our practical reasoning impacted in certain ways is suspended. This is why we are not wronged by conduct that has that effect on us. Basketball players know –and if they don't, should know– that a certain amount of trash talk is tolerated on the court and they are expected to plan accordingly in preparing their game. Similarly, negotiators in a business transaction should know that certain types of mind games are tolerated and that this is something they ought

[19] In talking about the Reasoning View, I will always have in mind the Quality of Reasoning View.

[20] Gibert, "Wrongful Manipulation," 333.

[21] I'm intentionally leaving aside here the possibility of manipulation that does not wrong anyone mentioned above. Since this does not seem to be the notion of manipulation that Gibert is interested in, it would not be fair to rely on it to criticize her view.

[22] As Gibert herself makes clear on pp. 334–35.

to consider in conducting their negotiation. So, defenders of the Reasoning View can deal with Gibert's cases by noticing that, although we normally do have a right against having our practical reasoning interfered with in a way that makes our reasoning worse, we lack that right in circumstances that are regulated by particular conventions and social practices. The fact that we lack that right is what explains why we are not wronged in these cases.

This is, of course, a common feature of social life. For example, I have a right that you don't lie to me, but if you are insincere when you declare that you are unwilling to consider offers lower than $500,000 for an apartment I'm keen to buy, I don't think you wrong me. It's understood that, in this context, the right to be told the truth does not apply in light of an ongoing practice that regulates this sort of exchange. Given the practice in question, an honest answer to the question of whether you would consider an offer of $480,000 is not to be expected, and I'm not wronged when I'm not offered one. It would be a mistake, though, to conclude from this observation that we lack a general right not to be lied to. What we should conclude is rather that certain practices and conventions carve out exceptions in this general right and its correlative duty. The same is true for the right against having our practical reasoning negatively impacted in the cases discussed by Gibert. It would be a mistake to conclude from those cases that we lack a general right of this sort. What we should conclude is rather that certain practices and conventions carve out exceptions in this right and its correlative duty.

Gibert is aware of this sort of move, as she does argue against a version of it. Her focus, however, is on views that attempt to explain the absence of the relevant right by appealing to consent or forfeiture. Gibert's arguments against these views are to my mind persuasive, but I don't think they can be successfully extended to cover views that appeal to conventions.

Consider, for example, Gibert's discussion of consent.[23] Gibert correctly points out that in order to consent to something one needs to *a)* understand that by acting in certain ways one is consenting to that thing, and *b)* have a sufficiently fine-grained control over what one is consenting to. Neither condition, however, is satisfied in the cases she offers as counterexamples to the Reasoning View. For example, I'm not wronged by someone who's trying to entice me to buy a house by baking cookies before the viewing, even if I happen to be unaware of this strategy, and thus I cannot consent to it. And my capacity to control what I'm consenting to in situations like this is very limited. For example, I cannot announce that I do not consent to being influenced by the delicious smell, thereby making it the case that I'm wronged by the

[23] Gibert, "Wrongful Manipulation," 341–43.

estate agent. In contrast, in cases of genuine consent, I normally do have this kind of control. For example, I can neutralize the normative implications of raising my hand while someone is asking for volunteers if I clarify that I'm only swatting a fly.

And these are not the only problems for consent-based accounts, Gibert argues. Suppose a plausible story can be offered to explain how, by behaving in certain ways, we thereby consent to having our practical reasoning interfered with in the way described by the Reasoning View. The story must also be able to explain why the behaviour in question waives our right with respect to specific treatments (and not others) and against specific agents (and not others). For example, say that we agree that by going out in public I consent to being non-rationally influenced by the good looks of street vendors to buy their product. Still, I do not consent to being non-rationally influenced to buy their products in other ways – say, by being bullied into it. And if I bump into a potential suitor who is keen to exploit some insecurities of mine to lure me into agreeing to a romantic getaway, well, I certainly do not consent to *that* merely by leaving the house! But if it's by choosing to go out that I waive my right against others interfering with my practical reasoning, once I'm out, the right is no longer present, Gibert argues. And if the right is no longer present, we cannot appeal to it to explain why I'm wronged by these behaviours. The right is no longer there to stop bullying salespersons and dodgy suitors. Consent-based accounts are thus insufficiently fine-grained to do the job they are meant to do.

Forfeiture-based accounts are discussed more concisely, but they raise similar worries, according to Gibert. If we try to explain why my right against having my practical reasoning negatively interfered with (in the ways the Reasoning View indicates) is forfeited in virtue of the fact that I've engaged in acts of wrongdoing, we will also need to explain why the right in question is forfeited with respect to specific treatments (and not others) and against specific agents (and not others).[24]

I believe that, once suitably understood, the appeal to convention can be used to answer all these objections. To begin with, the way conventions operate is largely independent of how specific individuals understand them to operate. If the convention is that members of my college are permitted to read magazines that have been left in the common room, I am not wronged if another member reads the copy I've left there. I'm not wronged even if I happen to be unaware of the convention. If someone had a responsibility to alert me to it and they didn't, I might have a complaint against that. But, unlike in the case of consent, the way in which my rights are altered when a certain convention is in place does not depend on the fact that I have an accurate understanding of how the convention operates.

[24] Gibert, 356–57.

Nor do conventions normally depend on us being able to control how the normative landscape changes when they operate. Conventions independently determine which rights are suspended and which are not, and against whom. If the convention is that only members of my college are permitted to read the magazines left in the common room, we don't need to come up with an explanation of why I'm wronged by non-members when they read my copy, or of why I'm wronged by members that use the magazine to wrap their sandwich, rather than to read it. Conventions and practices carve out exceptions in the relevant right by identifying *ab initio* which agents and behaviours are affected by its suspension and which aren't.

The same can be said for the cases discussed by Gibert. There's an ongoing practice among basketball players that allows trash talking their opponents, one among salespersons that allows them to use their charm and good looks to entice customers to buy their products, and so on. If this is the way these practices operate, we don't need a further story to explain why other forms of interference, such as bullying someone into making a purchase or luring them to go on a date, are wrong. Here too, the relevant conventions and practices carve out exceptions in the relevant right by identifying *ab initio* which agents and behaviours are affected by its suspension and which aren't.

Gibert briefly acknowledges that appealing to convention might "go some distance toward explaining" the fact that the relevant right is absent in the cases we are interested in.[25] However, she dismisses the sort of strategy outlined above because she thinks that if we are to argue that conventions suspend our right only with respect to certain agents and certain behaviours, we need an explanation of why the convention has precisely that shape. As she puts it,

> "[t]he objector needs an independently plausible theory which predicts that the boundaries of legitimate influence will be precisely what the Reductive View predicts they will be. This is, at best, a tall order. It does not seem as though we can scrutinize the act, as performed by the individual, and thereby reveal why it shapes the normative landscape in precisely this way, or even describe the way in which it shapes the landscape without reference to moral phenomena. On the Reductive View, there is no such mystery, for there is no global right against reasoning-subverting influence to be given up. The normative landscape is as it is because individuals are only wronged by forms of reasoning-altering influence that infringe their other rights."[26]

[25] Gibert, 343.
[26] Gibert, 358.

Now, if the objections discussed in the previous section are sound, the Reductive View does not correctly identify all and only the cases of wrongful manipulation, and thus we should not want our account of convention to track the verdicts of the Reductive View. But even if I'm wrong about this, I'm not quite sure how to understand Gibert's point in the quote. Conventions and practices are normally regarded as binding regardless of whether we can provide an accurate explanation of why they have the shape that they do. To be sure, valuable conventions will be supported by adequate justifications. But these justifications are not expected to univocally determine a specific set of norms, such that the rights of participants are exactly what they are and could not have been otherwise. Trash talk is accepted in basketball and not in tennis. Practices such as painting jurors in a certain light or dressing formally at job interviews are accepted in the contexts discussed by Gibert and not others. I take it that we can agree that these facts are the product of binding conventions, even if we lack an independent theory capable of predicting why these conventions have the shape they have, and even if we concede that they could have easily had different ones. (A possible world in which trash talk is accepted in tennis and not in basketball certainly doesn't seem too remote from ours.)

A central component of the justification of convention is after all that they serve a valuable coordinative function, thereby enabling participants to successfully pursue their goals.[27] The specific shape that conventions take generally plays a less significant role in their justification. This is why very different practices about how transparent lawyers are required to be in court, or how pushy salespersons are allowed to be in a street market, can work perfectly well in different jurisdictions and social contexts.

If I am right about this, we can agree that the right against having our practical reasoning interfered with is suspended in certain circumstances in light of given practices and conventions, even if we lack an independent account which predicts why the right is suspended exactly in the way it is. This is enough to rebut Gibert's objection to the argument outlined in this section.

[27] David Hume, *A Treatise Of Human Nature*, 2nd ed. (Oxford: Oxford University Press, 1978), https://doi.org/10.1093/oseo/instance.00046221; David Lewis, *Convention* (Cambridge, Mass.: Harvard University Press, 1969). This is compatible, of course, with the claim that conventions serve further roles; see Andrei Marmor, *Social Conventions: From Language to Law* (Princeton: Princeton University Press, 2009), https://doi.org/10.1515/9781400831654; David Owens, *Bound by Convention: Obligation and Social Rules* (Oxford: Oxford University Press, 2022), https://doi.org/10.1093/oso/9780192896124.001.0001.

V. WHY THE REASONING VIEW IS NOT UNDERINCLUSIVE

The main focus of Gibert's discussion is the objection that the Reasoning View is overinclusive. But I think it's the objection of underinclusiveness that raises a more serious worry for the view. Consider then India and Sebastian's case. By sharing his marital secrets with his mother, Sebastian angers India. However, he does so to induce India to give the right weight to her own interests in discussing childcare arrangements, something she would have otherwise failed to do. This is thus a case in which someone's reasoning is improved, rather than worsened, by manipulation, Gibert argues. And if so, far from being the key to understanding why manipulation is wrong, the idea that manipulation worsens its victims' reasoning is not even a necessary condition of manipulation.

But has Sebastian's intervention really improved India's reasoning? In one respect it has, obviously. Since India will now give the right weight to her own needs, she will avoid the mistakes she is prone to make when she doesn't. In another respect, however, it hasn't. Being angry about someone's indiscretion is *not* a good reason to value one's own needs more. Inducing someone to value their needs more as a response to feelings of anger is thus a way of inducing a form of bad reasoning.[28]

Notice that the problem here isn't that feelings are not a good reason to revise the assessment of our own needs. Feeling tired *is* a good reason to give more weight in our deliberation to our need to rest; being worried about cardiovascular diseases *is* a good reason to give more weight in our deliberation to our need to exercise more. The problem in India's case is that the kind of rational support that feelings can provide to certain types of deliberation is missing in the reaction Sebastian intends to produce in India. While motivating you to exercise more by making you worry about cardiovascular diseases is a way to induce an appropriate response to your reasons for action, motivating India to prioritize her own interests in allocating childcare responsibilities by making her angry about an unrelated issue is not. [29] This is a form

[28] I take this to be the most natural interpretation of Gibert's case, though I should stress that she does not explicitly say that it's the anger that causes India to value her needs more. Gibert only writes that "Sebastian has recently observed that India tends to undervalue her own preferences and needs, but that she is slightly less inclined to do so when she is angry at someone," and that "[a]s a result [of Sebastian's stratagem], India gets angry and gives more appropriate weight to her preferences in their subsequent discussion." This is worth noticing because we could think of other interpretations of the case. I come back to some of them below (see footnote 29 and footnote 41).

[29] One might be tempted to reply that in breaching India's trust, Sebastian gives her evidence that he doesn't sufficiently respect their agreements, and feelings of anger *about that* might rationally support India's choice to give more weight to her own needs in the subsequent discussion about childcare arrangement. Notice, however, that this is not the case Gibert has in mind. In her version of the case, Sebastian mentions the indiscretions to his mother simply because he has observed that India is less

of bad reasoning in the same way in which arriving at the correct solution to an equation through an invalid mathematical step is a form of bad reasoning. Even if the answer is correct, the method used to reach it undermines the rationality of the cognitive process.

The Reasoning View thus has the resources to accommodate cases like Sebastian and India's, once correctly understood. The view should be interpreted as concerned not only with the question of whether the target of manipulation ends up acting as they have reason to, but also with the question of whether in acting as they have reason to, they are motivated by sound reasons. This is because there are two ways in which one can undermine our practical reasoning when we try to respond to our reasons for action: they can do so by inducing us to act *against what we have reason to* do or, as in Sebastian and India's case, by inducing us to act as we have reason to *for the wrong reasons*.[30]

Gibert's neglect of this distinction is hardly surprising, I should stress, since defenders of the Reasoning View are themselves generally oblivious to it.[31] In formulating their view, they seem to have in mind precisely the version of the view Gibert criticizes. For example, Jason Hanna argues that whether an attempt to influence our behaviour counts as manipulative depends on whether it leads us to assign the appropriate weight to our reasons in our deliberation (or, as he sometimes puts it, whether the influence in question brings us "closer to the ideal deliberative standards").[32] Thus, for Hanna, whether using brightly coloured cigarette warning labels is manipulative depends on whether doing so will encourage us to better appreciate the reasons we have not to smoke. If it does, it's not manipulative.[33]

In a similar vein, Robert Noggle writes that a tactic is manipulative when it aims to produce a "faulty mental state." Noggle does not explain how he understands the notion of faulty mental state, but the examples discussed by him make clear that, on this point, he thinks like Hanna. For instance, he writes that:

---

inclined to undervalue her needs "when she is angry at someone." It's the mere feelings of anger that are meant to produce this effect. Sebastian might as well have produced those feelings by reminding India of an upsetting episode from the day before.

30 Renzo, "Why Manipulation Is Wrong," 185-91.

31 A notable exception is Moti Gorin, "Paternalistic Manipulation," in *The Routledge Handbook of Paternalism*, eds. Kalle Grill and Jason Hanna (New York: Routledge 2018), 241–45, https://doi.org/10.4324/9781315657080-20.

32 Jason Hanna, "Libertarian Paternalism, Manipulation, and the Shaping of Preferences," *Social Theory and Practice* 41, no. 4 (2015): 627–28, 636, https://doi.org/10.5840/soctheorpract201541434.

33 Hanna, "Libertarian Paternalism," 628.

> "[A]n emotional appeal only seems manipulative if the emotion is inappropriate. Thus, it does not seem manipulative to get someone to fear something whose very real dangers that person fails to appreciate."[34]

And again:

> "[Inducing someone to pay] selective attention only seems manipulative if the influencer induces the target to pay too much attention to the wrong things (or too little attention to the important things). If, instead, the influencer seeks to get the target to pay the right amount of attention to something important, then the influence does not seem manipulative."[35]

But it's clearly possible to induce someone to fear something whose dangers they fail to appreciate in a way that counts as manipulative. Knowing that you tend to underestimate how easy it is to contract a certain virus, I might lie to you and say that a new variant, much more contagious than the previous one, has just appeared. Even if I do so to induce you to correct your assessment of the risks associated with the virus, this seems a clear case of manipulation.[36] Similarly, it's possible to manipulate someone into paying the right amount of attention to something important. For example, knowing that you tend to ignore the speed limit, I might mislead you to think that speed cameras have been recently installed in your area. Again, even if I do so to ensure that you will pay the right amount of attention to your speed, I seem to have clearly manipulated you.

This is why a plausible account of manipulation needs to pay attention not only to whether certain influences lead us to act as we have reason to, but also to whether they do so by enabling us to be moved by the right reasons. That a certain influence has rendered us more responsive to reasons is not enough to conclude that it is not manipulative. For that to be the case, the influence must enable us to correctly grasp which reasons should motivate us in the situation at hand. Only if the Reasoning View

---

[34] Robert Noggle, "Pressure, Trickery, and A Unified Account of Manipulation," *American Philosophical Quarterly* 57, no. 3 (2020): 243, https://doi.org/10.2307/48574436.

[35] Noggle, "Pressure, Trickery," 243.

[36] Interestingly, in an earlier paper on nudging, Noggle seems to be aware of this (though he doesn't explicitly address the question of motivation). Discussing the case of a physician who helps a patient quit smoking by triggering in them an irrational disgust for tobacco, Noggle argues that this constitutes an instance of manipulation, because the disgust produced by the physician is based on an irrational aversion (Robert Noggle, "Manipulation, Salience, and Nudges," *Bioethics* 32, no. 3 (2017): 169, https://doi.org/10.1111/bioe.12421). This is exactly the right thing to say here. But if so, Noggle should conclude that getting someone to fear something whose danger they fail to appreciate *can be* manipulative.

is formulated in this way, can it handle cases like India and Sebastian's and the ones just introduced.

In Gibert's counterexample, Sebastian's intervention induces in India a flawed response to her reasons for action because, as we have seen, being angry about someone's indiscretions is not a good reason to value one's own needs more. In my counterexamples, my intervention induces in you a flawed response to your reasons for action, because the existence of a new variant is not a good reason for you to revise your assessment of how risky the virus is (since there is no new variant), nor is the presence of speed cameras in your area a reason to control your speed (since there are no cameras in your area).[37]

Importantly, reformulating the Reasoning View along the lines I suggest enables it not only to deal with these counterexamples, but also to provide a more nuanced treatment of central cases of manipulation, including some that are relevant to law and policymaking. Consider, for instance, cigarette packages that are designed to induce smokers to quit.[38] Is this an instance of manipulation? The answer to this question will depend on what the design involves exactly. If, as in Hanna's version of the case, it's the use of brightly coloured warning labels that is meant to influence smokers to quit, then the design is manipulative. There might be an evolutionary explanation of why bright colours are likely to induce this sort of reaction, but the mere fact that the label is, say, bright red, rather than grey or brown, does not provide one with a reason to quit smoking. The case (discussed by Gibert) of cigarette packages displaying disturbing pictures of lung disease is different, since the revulsion smokers might feel at the sight of those pictures *is* a good reason to quit smoking. Thus, insofar as the aim of the

[37] One might reply that *the belief* that there is a new variant is a reason for you to revise your assessment of how risky the virus is. But even if we accept this reply (worries about it are discussed in Maria Alvarez, *Kinds of Reasons: An Essay in the Philosophy of Action* (Oxford: Oxford University Press, 2010), 124–64, https://doi.org/10.1093/acprof:oso/9780199550005.001.0001), surely our interest in responding correctly to our reasons for action is not limited to avoiding the charge of irrationality narrowly conceived as involving inconsistencies between our apparent reasons for action. In exercising our practical agency, our concern is also to reason so as to adequately respond to the substantive reasons we have. For this distinction, see Thomas Scanlon, *What We Owe to Each Other* (Cambridge, Mass.; London: Belknap Press of Harvard University Press, 1998), 30–32, https://doi.org/10.2307/j.ctv134vmrn; John Broome, *Normativity, Rationality, and Reasoning: Selected Essays* (Oxford: Oxford University Press, 2021), https://doi.org/10.1093/oso/9780198824848.001.0001. Clearly, when defenders of the Reasoning View argue that manipulation undermines our efforts to adequately respond to our reasons for action, they have in mind not only the charge of irrationality narrowly conceived.

[38] This is a classic instance of nudging. I will outline how the view defended in this paper helps us address the question of the morality of nudging in section VI.

pictures is to vividly remind us of how repugnant the effects of lung disease are, their use does not seem manipulative.[39]

Now a complication. I have argued that there is an important sense in which Sebastian undermines India's effort to adequately respond to what she has reason to do. This is because, although his intervention leads her to give the right weight to her own interests, it does so in a way that lacks rational support. But consider now:

> Happy Place: India and Sebastian are about to discuss how to allocate childcare responsibilities for next year. Sebastian knows that India tends to stress out about this sort of decision, and that when she does, she gets in her head and doesn't do very well at prudential reasoning. To avoid this, Sebastian ensures that India is as calm as possible before the discussion. He puts on some relaxing music, avoids distressing topics over dinner and even recalls happy memories from a recent vacation. This relaxes India and enables her to give appropriate weight to her interest in the subsequent discussion.

It's natural to think that India is not being manipulated by Sebastian in this version of the case. But what's the difference between this case and the original one?[40] If a principled answer to this question cannot be offered, then it looks as if the right response to Gibert's argument would not be to argue (as I've done) that the Reasoning View can handle India's case, but rather to deny that India's case is an instance of manipulation to begin with.

I believe that there is, indeed, an important difference between these two cases. In the original version of the case, Sebastian's intervention aims at using India's anger as a motivational force that will lead her to assign more weight to her interests when she deliberates. In *Happy Place*, by contrast, Sebastian's intervention aims at removing psychological impairments that get in the way of India's efforts to assess her interests.

The first strategy is manipulative because being angry about something unrelated to the weighing of her own interests is not a good reason for India to value those interests more. Thus, although Sebastian is ultimately trying to promote India's well-being, he does so in a way that involves producing a faulty response to her normative

---

[39] This is not to say that the use of the pictures is not problematic. Only that if it is, this is not because it constitutes an instance of wrongful manipulation. For a defence of the view that rationally persuading others to do something for their own good can be objectionably paternalistic, see George Tsai, "Rational Persuasion as Paternalism," *Philosophy and Public Affairs* 42, no. 1 (2014): 78–112, https://doi.org/10.1111/papa.12026.

[40] I'm grateful to an anonymous reviewer for pressing this objection.

situation. India is motivated to act by something that is not a good reason for the conduct in question.

The second strategy is not manipulative because here Sebastian is not trying to get India to treat psychological features that lack normative significance as if they were reasons-giving. Rather, he is neutralizing the effects of psychological mechanisms that prevent India from accurately assessing features of the situation that do have genuine reason-giving force. That this process happens through the employment of arational means, such as the use of relaxing music, does not compromise India's response to her reasons for action because, at the end of the day, her process of deliberation only includes reasons that have genuine normative force. Ultimately, India is moved by a sound understanding of what she has reason to do. Sebastian's intervention only eliminates the noise produced by her anxiety about making important decisions.

Now, I don't pretend that drawing the distinction between *inducing someone to act for the wrong reasons* and *making the right reasons salient by removing impairments on our capacity to deliberate* (through arational means, when necessary) will be always straightforward. But the distinction is important and should have a central role in any sound version of the Reasoning View. It's this distinction, I submit, that we ought to use to interpret cases that are hard to adjudicate on intuitive grounds.[41]

VI. ACTING FOR THE RIGHT REASONS

I have argued for a version of the Reasoning View that rules out as manipulative not only influences that aim to get us to fail to act as we have reason to, but also influences that aim to get us to do what we have reason to do for the wrong reasons. Only if understood in this way, can the Reasoning View avoid counterexamples like India's case or the cases I have raised in the previous section. Let me now suggest a second amendment to the way the view is often formulated.

[41] But couldn't we offer an interpretation of Gibert's original case along the lines of *Happy Place*? Perhaps India has a peculiar psychology whereby she needs to be in a state of anger in order to be able to correctly assess her reasons for action. If so, couldn't we understand Sebastian's intervention in the original case as also aimed at neutralizing the effects of psychological mechanisms that prevent India from accurately assessing features of the situation that have genuine reason-giving force? This interpretation strikes me as far-fetched, but notice that even if correct, it would not undermine my rejection of Gibert's argument that the Reasoning View is underinclusive. For if we interpret her original case along the lines of *Happy Places*, we should think that it is *not* a case of manipulation (for the same reason why we think that *Happy Places* is not a case of manipulation). But if so, the case could not be used to draw conclusions about the morality of manipulation.

Gibert presents the Reasoning View as stating that we are manipulated when someone interferes with our practical reasoning "in a way that makes our reasoning worse,"[42] and similar formulations are adopted by defenders of the view.[43] But in order to assess whether someone's reasoning has been made worse, we need a baseline. Worse relative to what? The most obvious answer is: "Worse relative to how the victim would have reasoned in the absence of manipulation." But many of the cases considered so far, such as the India's case, or the counterexamples I've raised in the previous section, show that this answer is problematic. Valuing one's needs more as a response to feelings of anger might be a bad way for India to respond to her reasons for action. But is it a worse way to respond to those reasons than continuing to undervalue her needs (which is what would've happened without Sebastian's intervention)? Hard to believe. Similarly, reducing one's speed because of a mistaken belief in the presence of speed cameras might be a bad way for you to respond to the reasons you have to drive safely. But is it a worse response to those reasons than driving recklessly? Again, that seems implausible.

The Reasoning View, however, is not committed to the idea that manipulating someone is a matter of making their reasoning worse. The view is best understood as stating that manipulation undermines (or perhaps negatively interferes with) its target's efforts to exercise their practical agency well so as to adequately respond to their reasons for action. And it's clearly possible to do that and yet improve how the target responds to their reasons for action overall, as the cases discussed in the previous section illustrate.

This is in line with how we treat other forms of interference with people's efforts to exercise their practical agency. For example, intoxicating someone is a matter of compromising, through the use of certain substances, their efforts to exercise their capacity to control their behaviour, not a matter of making their capacity to control their behaviour worse. If I spike your drink, I intoxicate you even if, had I not done that, you would have gotten much drunker by yourself. Similarly, brainwashing someone is a matter of forcibly inducing them to adopt unfounded beliefs and attitudes, not a matter of making their holding of unfounded beliefs and attitudes worse. If I forcibly induce you to believe that I'm the son of God and that the world will end on Tuesday, I've brainwashed you even if, had I not done that, you would've gone on to adopt beliefs that are much more absurd.

---

[42] Remember that in this paper we are focusing exclusively on the version of the Reasoning View that Gibert calls "Quality of Reasoning View."

[43] See, for example, the quote from Hanna, "Libertarian Paternalism," 631.

With the introduction of this second amendment, we now have a much more plausible formulation of the Reasoning View – one that, as we have seen in the previous section, has a great deal of explanatory power. It is not entirely surprising, however, that defenders of the view have not adopted it. This is because of how demanding it is. While it's clear that we should care about doing what we have reason to do, it's less clear why we should care that in doing what we have reason to do, we are motivated by certain reasons rather than others. Of course, in some cases, we should care about it because our motivation will say something about our character or the value of our conduct. If I take care of my child only to avoid social sanctions, rather than because I love him, that says something bad about me. It also compromises the value of my conduct, in that what I do fails to express proper concern for my child, despite benefiting him.[44] But leave these cases aside and take a situation in which our motivation does not reflect poorly on us or our conduct. Is there a sense in which we fall short in how we respond to our reasons for action, if, as in India's case, we're motivated by a feature of the situation that fails to provide adequate support for our conduct?

Before considering this question further, let me slightly amend Gibert's example. I do so because, in its current formulation, we might be tempted to think that India's behaviour does, after all, say something bad about her, since failing to give sufficient weight to her interests could be plausibly taken to reflect a lack of self-respect. Best to pick a case in which this feature does not interfere with our discussion. So, let's imagine that India's problem is not that she tends to undervalue her own interests, but rather that she tends to underestimate the burdens of childcare. Given how burdensome an unequal distribution of childcare duties would be, India has reasons to suggest an equal one, and the anger produced by Sebastian's phone call leads her to do just that. Indeed, to make the case even clearer, let's leave aside for now the fact that Sebastian is trying to influence her behaviour. So, imagine that the effect produced by the phone call is accidental.

Here's our case:

> Phone Call 1: Sebastian and India are about to have a conversation about the distribution of childcare responsibilities. India, who tends to underestimate how burdensome these responsibilities are, is likely to suggest an unequal distribution

[44] Perhaps, in some cases, being motivated in the wrong way might even affect the permissibility of my conduct. See Jeff McMahan, "Intention, Permissibility, Terrorism, and War," *Philosophical Perspectives* 23, no. 1 (2009): 345–72, https://doi.org/10.1111/j.1520-8583.2009.00175.x; Victor Tadros, *The Ends of Harm: The Moral Foundations of Criminal Law* (Oxford: Oxford University Press, 2013), ch. 7, https://doi.org/10.1093/acprof:oso/9780199554423.001.0001; Ralph Wedgwood, "Defending Double Effect," *Ratio* 24, no. 4 (2011): 384–401, https://doi.org/10.1111/j.1467-9329.2011.00508.x.

> in Sebastian's favour. Before their conversation, believing to be alone in the house, Sebastian is on the phone with his mother and divulges some of the marital issues he and India had agreed to keep private. India, who, unexpectedly, is home, overhears him and gets angry. When they talk later, she suggests an equal distribution of responsibilities. She does so because she is upset and (correctly) believes that this suggestion will end the conversation quickly.

Let's stipulate that India has indeed reasons to suggest an equal distribution of childcare responsibilities. Is there a sense in which she fails to adequately respond to this reason in virtue of the fact that, in suggesting it, she's motivated by the desire to keep the conversation short, rather than by the fact that she appreciates that this is the distribution she has reason to insist on?

Philosophers who have considered this question tend to answer negatively. The most prominent example is Joseph Raz, who distinguishes between "conforming" with the reasons for a given action, i.e. performing the action in the circumstance in which the reason calls for doing so, and "complying" with those reasons, i.e. performing the action in the circumstance in which the reason calls for doing so *because one realises that it does*. Here's how Raz poses the question:

> Do [reasons for action] aim at action, so that if the action occurs all is as well as it should be? Or do they aim at one's reasoning as well, so that they demand, as it were, to figure in one's reasoning and/or in one's motivation? The view that reasons for action are always reasons for compliance fits well with the idea that practical reasons are guides to action. If one is not guided by them, then one is failing to behave as one should. If reasons for action are understood as reasons for conformity, then one may still talk of reasons for action as guides for behaviour, but only in the sense that, other things being equal, it is legitimate, i.e. alright, for them to figure in one's reasoning or motivation. They are guides in the sense that the Michelin guide to Paris is a guide. I may use it, but I do not have to. I do not even have to be aware of its existence. There is absolutely nothing wrong in using another guide, if it is as good. The important thing is that I get to see the things which are worth seeing in Paris. Similarly, on this understanding of reasons the important thing is that the act for which the reason is a reason gets done (unless the reason is defeated). It does not necessarily matter if it is done for this or some other (good) reason.[45]

[45] Joseph Raz, *Practical Reason and Norms* (Oxford: Oxford University Press, 1999), 180, https://doi.org/10.1093/acprof:oso/9780198268345.001.0001.

Raz takes the latter view. He writes that "reasons for action are, barring special circumstances, merely reasons to conform," and that "there is no loss, no defect, blemish, or any other shortcoming, in conformity with reason achieved not through compliance with it, but for other reasons."[46] If he's right about this, we should believe that as long India conforms with the reason she has to suggest an equal distribution of childcare responsibilities, her response to this reason is not compromised by the fact that she's not motivated by it. If reasons for action only call for conformity, and there is no loss, no defect, blemish, or any other shortcoming, in conforming with a reason achieved not through compliance with it, but for other reasons, then India does not fall short in responding to the reason in question. By suggesting an equal distribution of childcare responsibilities, she conforms with it, regardless of what motivates her.

This is also true in a version of the case in which Sebastian engineers this outcome. Consider:

> Phone Call 2: Like *Phone Call 1*, but in this case, Sebastian is aware that India is home and ensures that she overhears what he says. He does so because he knows that India tends to underestimate how burdensome an unequal distribution of childcare responsibilities would be for her. He correctly predicts that, if angry, she will want to keep the conversation short, and will thus suggest an equal distribution.

Once again, if reasons for action only call for conformity, there is "no defect, blemish, or any other shortcoming" in how India responds to the reason she has to suggest an equal distribution of responsibilities. Her response to this reason is not compromised by the fact that, in acting as the reason calls for, India is not motivated by it. (In this respect, there is no difference between *Phone Call 1* and *Phone Call* 2.) But if so, the Reasoning View cannot be formulated in the way required to handle the counterexamples introduced in the previous section. For that would require taking the view that adequately responding to our reasons for action involves not only acting as they call for, but also doing so because we are motivated by them in the right way, i.e. because we appreciate which features of the situation provide adequate support for our conduct. India's anger about Sebastian's indiscretion is not such a feature. Although

[46] Raz, *Practical Reason*, 182. See also, Joseph Raz, *Engaging Reason: On the Theory of Value and Action* (Oxford: Oxford University Press, 2000), 92, https://doi.org/10.1093/0199248001.001.0001; John Gardner, "Justifications and Reasons," in *Offences and Defences: Selected Essays in the Philosophy of Criminal Law* (Oxford: Oxford University Press, 2007), 91–120, https://doi.org/10.1093/acprof:oso/9780199239351.003.0005.

it explains why India wants to keep her conversation with Sebatian short, in the sense that it makes her behaviour intelligible, her anger does not provide rational support for the choice of an equal distribution of childcare responsibilities.

If we want a formulation of the Reasoning View that can vindicate the claim that Sebastian manipulates India in *Phone Call* 2 (and the other cases discussed above), we thus need to reject Raz's view that merely conforming with our reasons for action is enough to adequately respond to them. We need to explain why there is a sense in which India falls short in responding to the reasons she has to suggest an equal distribution of childcare responsibilities when, instead of being motivated by the recognition of this reason, she's motivated by her anger.

VII. PRACTICAL AGENCY

We have seen that to defend the Reasoning View from the charge of being underinclusive, we need to adopt a formulation of the view that enables it to handle cases like *Phone Call* 2, as well as the other counterexamples discussed in the previous section. And we have seen that doing that requires formulating the view so that it rules out as manipulative not only influences that aim to induce us to fail to act as we have reason to, but also influences that aim to induce us to do what we have reason to do for the wrong reasons. Finally, we have seen that this, in turn, requires subscribing to a picture of practical agency according to which adequately responding to our reasons for action is not simply a matter of conforming with them, i.e. of acting as they require, but also a matter of being motivated by the recognition of what they call for when we so act. How plausible is this picture? In this final section, I will sketch an account of how we might go about defending it.

Necessarily, here I will have to paint with a broad brush. Reasons of space prevent me from developing the account or considering possible objections.[47] But since the argument for the Reasoning View outlined so far rests on a picture of practical agency that, on the face of it, seems quite demanding, completing the argument requires giving a sense of how one might go about justifying such a picture. So, our questions are: Why should we care that, in responding to the reasons for action that apply to us, we are motivated by certain reasons rather than others, when failure to do so neither reflects poorly on our character nor does it compromise the moral value of our conduct?

[47] I do so in Massimo Renzo, "Motivation and Alienation," in *Engaging Raz: Themes in Normative Philosophy*, eds. Andrei Marmor, Kimberley Brownlee, and David Enoch (Oxford: Oxford University Press, 2025), 172–192, https://doi.org/10.1093/9780198925378.003.0010, on which this section draws.

In which way do we fall short in exercising our capacity to respond to our reasons for action when we are not motivated in this way?

To answer these questions, start with the idea that reasons are facts that call for a certain response from us through our practical reasoning, i.e. our capacity to form beliefs, desires, emotions, intentions to act, etc.[48] Exercising our practical agency well is a matter of adequately grasping and responding to these facts. What happens, then, when we fail to be motivated by the reasons that apply to us in the right way? In some cases, the problem is that, even though we recognize that we ought to be motivated by certain reasons, we can't help acting for different ones. Essentially, these are cases of weakness of the will: We φ for reason T, against our best judgment that we ought to φ for reason R. (I realize that I ought to go for a run because I would enjoy it and it would be good for me, yet I go to spite my neighbour, who is home with a broken leg.) Now, strictly speaking, weakness of the will normally refers to cases in which we act against our best judgment, i.e. cases in which we φ rather than υ-ing, despite believing we have most reason to υ. But if acting against our best judgment about what we have reason to do is a defective exercise of our practical agency, plausibly so is acting against our best judgment about which reasons should move us to act. Thus, we can use our favourite account of weakness of the will to explain why, in these cases, we fall short, *qua* rational agent.

In other cases, however, our failure to be motivated by the right reasons has nothing to do with weakness of the will. These are cases where we fail to recognize that we ought to be moved by those reasons to begin with. We φ for reason T because we don't realize that we ought to φ for reason R. Here, the problem lies in our defective appreciation of our reasons for action.

If exercising our practical agency is a matter of grasping and adequately responding to normative facts, then when we are not moved to act by the salient reasons for action that apply to us (and this is not explained by weakness of the will),[49] we are failing to recognize some of these facts for what they are. This might be because we fail to recognize the very existence of such facts (we fail to appreciate that we have a reason R to φ), or because, despite acknowledging their existence, we fail to appreciate their normative force (we know that we have a reason R to φ, but we fail to appreciate that R is what should move us to φ).[50] Either way, our response to these facts is compromised,

[48] Classic formulations of this picture include Thomas Scanlon, *What We Owe to Each Other*; Raz, *Engaging Reason*; Joseph Raz, *From Normativity to Responsibility* (Oxford: Oxford University Press, 2011), https://doi.org/10.1093/acprof:oso/9780199693818.001.0001.

[49] In the rest of the discussion, I will take this qualification for granted.

[50] For simplicity, in this formulation, R refers to a single reason. But of course, we often have more than one salient reason to act. R could also refer to a set of reasons.

in that even if we happen to act in accordance with what they require, our conduct is not informed by an adequate grasp of how we are to be moved by them.

This is what happens in *Phone Call 2*. India conforms with the reason she has to suggest an equal distribution of childcare responsibilities. But insofar as she's moved to do so by her anger, rather than by the fact that she appreciates that this is the distribution she should favour, the way in which she responds to this reason is in one sense flawed. Although she ends up acting as the reason calls for, she fails to appreciate how she is supposed to be moved by it. Now, this is clearly not a moral failure. Reactive attitudes such as blame or resentment, for example, seem inappropriate. So, what kind of failure is it? It's a failure to understand her normative situation, and thus, in an important sense, a failure to understand herself.

Grasping and being adequately moved by the salient reasons that apply to us is a way of knowing ourselves and our place in the world. Ignoring those reasons, and the role they should play in moving us to act, is failing to understand our own position within the network of values, requirements, and permissions that structure our lives in the normative space we inhabit. Ultimately, we might say, this is a failure of self-knowledge.

The question of self-knowledge has been traditionally conceived as the problem of securing epistemic access to our mental states, where these are understood in non-normative terms. At least since Descartes, the way in which philosophers have framed the question has been, roughly: "How can we have epistemic access to the feelings, desires, beliefs and intentions we happen to possess?" However, this way of framing the question neglects the normative dimension of self-knowledge, and this surely is a problem if we're interested in a notion of self-knowledge that is well suited for rational agents, i.e. agents whose distinctive feature lies in the capacity to respond to normative considerations.

An earlier tradition, going back to Plato and Socrates, recognizes this important point. This is why, within this tradition, the injunction to know oneself is tied to a certain ideal of what it is to be a good practical agent, one that is dedicated to the pursuit of a good life in light of a sound understanding of one's own nature.[51] And if we conceive self-knowledge along these lines, it becomes easy to see why self-knowledge requires having epistemic access not only to the feelings, desires, beliefs and intentions *we happen to possess*, but also –crucially– to the *reasons we in fact have to feel, desire, believe and intend*. When we are confused or mistaken about these reasons, our self-knowledge is compromised.[52]

[51] Brie Gertler, *Self-Knowledge* (London: Routledge, 2010), 27–28, https://doi.org/10.4324/9780203835678.
[52] Here I'm grateful to Josiah Ober for helpful discussion.

Once we attend to this idea, it's easy to see how it is not only correct, but obviously so. Part of what it is to grasp that we have a reason to φ is ultimately to grasp something about ourselves: That φ-ing is called for, in the circumstances at hand, as something *we* ought to do, desire, feel or believe (absent countervailing considerations). In this sense, improving our understanding of what we have reason to do (or intend to do), desire, feel or believe is a way of improving our self-understanding. In grasping how we are to be moved by those reasons, we understand something about ourselves and about how we are connected to the normative world around us.

This is why we fall short in exercising our capacity to respond to our reasons for action when we merely conform with them. Even when it doesn't reflect poorly on our character, our failure to be motivated by our reasons in the right way constitutes a defective exercise of our practical agency, insofar as those reasons make certain demands on us and we fail to recognize those demands for what they are. We fail to do so even if, as it happens, our behaviour is not at odds with them. In merely conforming with our reasons, the way in which we exercise our practical agency manifests a lack of self-knowledge,[53] in that it is not informed by an adequate understanding of how our own reasons bear on how we should act. In this sense, we are alienated from our own reasons for action.[54]

An advantage of this picture is that it enables us to offer a compelling explanation of the wrong of manipulation in cases where being manipulated is beneficial to us, including cases of nudging.[55] Philosophers notoriously struggle to explain why such forms of manipulation are morally problematic. A common answer is that, even if they lead us to act as we should, nudges wrong us because they undermine our autonomy.[56] But it's unclear how plausible this answer is. After all, in deciding to act as the nudge directs us to, we are still responding to our own understanding of what we have reason to do. And while nudges affect our perception of the options available to us, this can't

---

[53] Provided, once again, that our behaviour is not an instance of weakness of the will.

[54] I further discuss this notion of alienation in Renzo, "Motivation and Alienation", especially 177-81.

[55] Cass R. Sunstein, *The Ethics of Influence: Government in the Age of Behavioral Science* (Cambridge: Cambridge University Press, 2016), https://doi.org/10.1017/CBO9781316493021.

[56] Luc Bovens, "The Ethics of Nudge," in *Preference Change: Approaches from Philosophy, Economics and Psychology*, eds. Till Grüne-Yanoff and Sven Ove Hansson (Berlin: Springer, 2009), 207–20, https://doi.org/10.1007/978-90-481-2593-7_10; Daniel M. Hausman and Brynn Welch, "Debate: To Nudge or not to Nudge," *Journal of Political Philosophy* 18 (2010): 123–136, https://doi.org/10.1111/j.1467-9760.2009.00351.x; T. M. Wilkinson, "Nudging and Manipulation," *Political Studies* 61 (2013): 341–355, https://doi.org/10.1111/j.1467-9248.2012.00974.x; Jeremy Waldron, "It's All for Your Own Good," *New York Review of Books*, 9 October 2014; David Enoch, "How Nudging Upsets Autonomy," *Journal of Philosophy*, 121 (2024): 657–685, https://doi.org/10.5840/jphil20241211241.

be enough to undermine our autonomy, since it's always the case that our perception of the options available is affected by forces outside our control, whether natural or man-made.[57]

If the view outlined in this paper is plausible, however, we can offer an alternative account of why nudges wrong us. Nudges wrong us because, when we are nudged, we are intentionally led to merely conform with what we have reason to do, and this constitutes a flawed response to our reasons for action.[58] To take a famous example, if what moves me to choose the salad, rather than the burger, is the fact that it's been placed in a certain location in the cafeteria, I fail to be moved by the salient reasons I have to choose the salad –say, the fact it's the healthy option, or the fact that eating meat is wrong–, even though I conform with those reasons. The problem with nudges then is that, even if our autonomy is not affected by them, they induce in us a sort of alienation from our reasons for action. If I have health-related reasons to order the salad rather than the burger, but instead I do so only because of how the two dishes have been displayed in the cafeteria, I'm wronged (*pro tanto*) because I'm intentionally led to act in a way that is not informed by a sound understanding of what I have reasons to do.[59] My action is shaped by a failure to grasp how I should be moved by my own reasons for action and thus, ultimately, by a failure to understand myself.

VIII. CONCLUSION

Gibert raises an important challenge to the Reasoning View, the traditional approach according to which manipulation is wrong because it impairs our practical reasoning. In its place, Gibert encourages us to adopt her Reductive View. Here, I have argued that the Reductive View should be rejected as both underinclusive and overinclusive. Moreover, I've argued that the view fails to vindicate cases of wrongful manipulation that do not wrong anyone.

The main focus of this article, however, has been providing a defence of the Reasoning View. I have argued that, once we pay attention to the role that conventions play in regulating our interactions, the view can be defended against Gibert's objection of being overinclusive. It's the objection of being underinclusive that, to my mind,

---

[57] Here I'm adapting an objection raised by Sarah Buss to the broader idea that manipulation undermines autonomy. Buss "Valuing Autonomy and Respecting Persons: Manipulation, Seduction, and the Basis of Moral Constraints," *Ethics* 115, no. 2 (2005): 195–235, https://doi.org/10.1086/426304.

[58] A similar line of thought is outlined in Gorin, "Paternalistic Manipulation," 245. I'm not persuaded by the arguments Gorin offers in support of it, but I lack the space to address them here.

[59] The same alienation would be produced if the location of the salad was accidental, of course. But I'm only wronged when my alienation is produced by someone acting with the relevant *mens rea*.

raises a more serious worry for the Reasoning View. Gibert is right that existing formulations of the Reasoning View are unsuccessful on this front. Far from providing an exhaustive account of why manipulation is wrong, these views cannot even vindicate the idea that undermining the victims' reasoning is a necessary condition of manipulation.

To avoid this problem, we need to adopt a more robust formulation of the Reasoning View, one that rules out as manipulative not only influences that aim to get us to fail to act as we have reason to, but also influences that aim to get us to do what we have reason to do for the wrong reasons. This, in turn, requires subscribing to a picture of practical agency according to which adequately responding to our reasons for action is not simply a matter of acting as they require, but also a matter of being motivated by the recognition of what they call for when we so act.

I've sketched a possible way in which we might go about defending this robust picture of rationality and explained why it enables us to offer a novel account of the wrong of manipulation. Of course, more needs to be said to develop this picture, but I hope I've done enough to give a sense of its appeal. Regardless of how persuasive you'll find it, however, the main lesson of this paper is that something like this picture has to be offered, if we want to defend a plausible version of the Reasoning View.[60] To avoid counterexamples like those discussed in this paper, the Reasoning View must rest on a picture of practical agency that is more demanding than its defenders have acknowledged so far.

[60] Examples of alternative views that could play this role are outlined in Susanne Mantel, *Determined by Reasons: A Competence Account of Acting for a Normative Reason* (New York: Routledge, 2018), https://doi.org/10.4324/9781351186353, and Mark Schroeder, *Reasons First* (Oxford: Oxford University Press, 2021), especially Ch. 10, https://doi.org/10.1093/oso/9780198868224.001.0001.

Acknowledgments

Thanks to Sophie Gibert and David Owens for helpful comments on the first half of the paper, which originally appeared as a contribution to the blog PEA Soup. I'm also very grateful to two anonymous reviewers for extensive, insightful feedback. The research for this essay was financially supported by a Research Fellowship of the Leverhulme Trust (RF-2020-351\8).

Competing interests

The author has no competing interests to declare.