

# Chapter 11

## Neural and Environmental Modulation of Motivation: What's the Moral Difference?\*

Thomas Douglas

[This is a draft prepared for *Treatment for Crime*]

### 1. Introduction

A prison superintendent decides to have an aggression-lowering drug sprayed into the prison's corridors as a way of combating prisoner violence. An employer requires his employees to wear scalp electrodes that administer a motivation-enhancing electric current as they work. A well-meaning barista laces a dieting customer's drinks with a hunger-suppressing drug.

Most of us would find these interventions to be morally objectionable, in the sense of being *pro tanto* morally wrong—wrong, absent defeaters. Indeed, I surmise that most would find objectionable all instances of what I will call *conative neuromodulation*: the nonconsensual modulation of a person's motivational states and dispositions through neurophysical or neurochemical means. I take it that these means consist in the

---

\* I would like to thank audiences in Oxford and Roskilde and two anonymous reviewers for their helpful comments on earlier versions of this paper, Areti Theofilopoulou for her editing assistance, and the Wellcome Trust [grant number 100705/Z/12/Z] and Uehiro Foundation on Ethics and Education for their funding.

application of physical energy (e.g. electric current) or chemical agents, and notably not psychological techniques, to alter the brain states on which the targeted motivational states and dispositions supervene.

Conative modulation is not always thought wrong, all things considered. One application of conative neuromodulation that commands a measure of acceptance involves the nonconsensual imposition of motivation-influencing drugs on patients with severe psychiatric conditions who pose a serious risk to themselves or others.<sup>1</sup> Another involves the mandatory administration of motivation-promoting agents to sleep-deprived soldiers.<sup>2</sup> However, even these instances of conative neuromodulation are normally thought to be somewhat objectionable—that is, somewhat *pro tanto* wrong—even if all-things-considered justified in some cases.<sup>3</sup>

So conative neuromodulation is widely regarded as morally objectionable. However, what exactly explains its objectionability remains unclear. Some instances of conative neuromodulation may be objectionable in virtue of their covertness. (Example: suppose that the barista who laces his customer's drinks with a hunger-suppressing drug does so unbeknownst to the customer.) Similarly, some instances of conative neuromodulation might be objectionable because they risk the physical or mental health of the recipient.

---

<sup>1</sup> A number of jurisdictions allow for such interventions under mental health legislation, and the implementation of these provisions is not generally regarded as morally outrageous.

<sup>2</sup> I thank an anonymous reviewer for suggesting this example. In some cases, the mandatory administration of motivation-promoting agents to soldiers may be consensual because the soldier has, through signing up to military service, consented to their imposition. However, in other cases, their imposition is truly nonconsensual. This is most obviously so when the soldier was conscripted into military service.

<sup>3</sup> Indeed, there is controversy over whether such uses of conative neuromodulation are indeed justified and some influential ethical codes hold that they are not. See, for example, Council of Europe, *Convention for the Protection of Human Rights and Dignity of the Human Being with regard to the Application of Biology and Medicine* (Oviedo: Council of Europe, 1997).

(Example: suppose the aggression-lowering drug administered by the prison superintendent tends also to cause depression.) But conative neuromodulation need not be covert—its use could be announced in advance, and in many cases, its use may be directly observable: perhaps the prisoners can see the aggression-lowering drug sprayed into their prison’s corridors, perhaps the hunger-suppressing drug has a distinctive taste. Similarly, conative neuromodulation need not impose any significant health risks; it may consist in the administration of drugs with no significant adverse side-effects. Yet even where nonconsensual conative neuromodulation is announced in advance, can be detected, and poses no significant health risks, it seems, intuitively, to be morally objectionable.

Perhaps a more promising explanation for the objectionability of conative neuromodulation would appeal to its physical invasiveness. In order to exert an effect on an individual’s motivation, physical energy or a chemical agent needs to enter that individual’s body and produce at least some neural changes. Perhaps, then, these interventions violate some right to, or frustrate some important interest in, freedom from bodily interference.

This appeal to bodily interference might go some of the way towards explaining the objectionability of conative neuromodulation. But I do not think it goes all the way. For one thing, the kind of bodily interference involved in some forms of conative neuromodulation seems rather unimportant in comparison to the kind of *mental* interference that it also involves; the fact that conative neuromodulation alters aspects of the motivational psychology of those subjected to it seems more normatively significant than that it also involves, for example, forcing them to inhale a small amount

of some chemical agent or subjecting their brains to a small electric current. Thus, we might expect that at least part of its objectionability derives not from its physical invasiveness, but from its mental invasiveness—from the kind of mental interference that it involves.

Further support for this suggestion can be derived from contrasting conative neuromodulation with other interventions that are equally physically invasive, but lack the mental invasiveness of conative neuromodulation. Consider initially a more fully elucidated version of the first of the three cases with which I began:

The superintendent of a prison is concerned about a recent spate of attacks by some prisoners on others. She reads that another prison is tackling such violence by having a visible but scentless spray released into the air in the prison's corridors. By modulating neural activity in a particular region of the brain, the spray tends to reduce extreme aggressive impulses. The intervention is thought to have no other significant mental effects and to pose no health risks, and it complies, let us suppose, with all relevant laws and regulations.<sup>4</sup> The superintendent announces that the same spray will be administered in the corridors of her prison. This is done, and rates of prisoner violence fall as expected.

Compare this instance of conative neuromodulation, henceforth *Calming Spray*, with the following variant:

---

<sup>4</sup> I make this assumption in order to set aside the potential objection that the intervention is morally problematic in virtue of being unlawful.

The superintendent of a different prison is concerned about an outbreak of a viral respiratory disease amongst the prisoners. There is no effective treatment for the infection, but it is possible to halt its spread by spraying a visible but scentless spray into the air. The spray enters the bloodstream of those who inhale it and stimulates their immune response in a way that reduces their infectiousness to others, though it does not ameliorate their own condition. It has no mental effects and is thought to pose no health risks, and its use complies with all relevant laws and regulations. Moreover, the spray will be administered for a similar length of time to *Calming Spray*, would have bodily effects with similar longevity, and can be expected to prevent harm that is comparable in its moral weight to that prevented by *Calming Spray*. The superintendent announces that the spray will be administered in the corridors of her prison. This is done, and, as expected, the outbreak of the disease resolves.

Call the superintendent's intervention in this second case *Antiviral Spray*.<sup>5</sup> Many would, I take it, find *Antiviral Spray* somewhat morally objectionable. But I suspect most would find it less objectionable than *Calming Spray*. Yet, *Antiviral Spray* involves physical invasion of a similar kind and degree as *Calming Spray*. This suggests that at least some of the objectionability of *Calming Spray* derives not from its physical invasiveness, but from some feature of *Calming Spray* that *Antiviral Spray* lacks. One

---

<sup>5</sup> The spraying of aircraft passengers with pesticide on arrival in certain countries may constitute a real-life analogue of this practice, although it might be argued that this constitutes a lesser infringement on bodily integrity than *Antiviral Spray* since, although the pesticide spray is inhaled, its inhalation is arguably unintended and is not a means to the intended effect. I thank an anonymous reviewer for suggesting this example.

such feature—the most obvious such feature—is that *Calming Spray* interferes with the mental life of the prisoners in a way that *Antiviral Spray* does not.

It seems plausible, then, that conative neuromodulation is morally objectionable in part because of the kind of mental interference that it involves. This accords with a recent proposal made by the legal scholars Jan Christoph Bublitz and Reinhard Merkel regarding the legal regulation of conative neuromodulation and some other forms of mental influence.<sup>6</sup> Bublitz and Merkel note that these forms of influence would normally only be regarded as unlawful insofar as they violate legal rights over the body. This, they think, is puzzling, given that the most important effects of these interventions are mental, rather than physical.<sup>7</sup> They claim that the legal impermissibility of these neurointerventions ought to be grounded not only in their physical invasiveness, but also, at least in part, in the kind of mental interference that they involve. I am suggesting that a parallel point holds with regard to morality: the putative *moral* objectionability of conative neuromodulation must be grounded not only in its physical invasiveness, but at least in part, in the kind of mental interference that it involves.

It turns out, however, that it is not straightforward to account for the moral objectionability of conative neuromodulation in this way. Or so I will argue. The difficulty is that the kind of mental interference involved in some forms of conative neuromodulation seems no more morally problematic than that involved in certain environmental interventions that are plausibly morally innocuous.

---

<sup>6</sup> Jan Christoph Bublitz and Reinhard Merkel, ‘Crimes Against Minds: On Mental Manipulations, Harms and a Human Right to Mental Self-Determination’, *Criminal Law and Philosophy* 8 (2014), pp. 51–77.

<sup>7</sup> Bublitz and Merkel, ‘Crimes Against Minds’, p. 2: “Why are we so convinced that, while no one is allowed to do anything to another’s body without consent, they are to virtually mess up another’s mind? Why is scarring souls so different from scarring the surface of the skin?”

My argument will focus on comparison between one hypothetical instance of conative neuromodulation, namely the abovementioned *Calming Spray* intervention, and one particular environmental intervention that might be regarded as morally innocuous. I wish to argue that *if* one indeed finds this environmental intervention morally innocuous, then one should find *Calming Spray* similarly innocuous, at least with respect to the form of mental interference that it involves.<sup>8</sup>

The putatively innocuous environmental intervention that I have in mind is a variant of *Calming Spray*:

The superintendent of a prison is concerned about a recent spate of attacks by some prisoners on others. She reads that another prison was able to reduce prisoner violence by painting the corridors of the prison in a soothing shade of green; seeing the green walls tends to attenuate extreme aggressive impulses. The intervention is thought to have no other significant mental effects and to pose no health risks, and it complies with all relevant laws and regulations. Moreover, the walls will be painted green for the same length of time as the spray would be administered, would have mental effects with similar longevity, and can be expected to prevent harm that is comparable in its moral weight to that prevented by *Calming Spray*. The superintendent announces that the same technique will be used in her prison. The walls are painted in the soothing shade of green, and rates of prisoner violence fall as expected.

---

<sup>8</sup> I will not henceforth comment on questions of legal permissibility, though on certain views about the relationship between law and morality, what I will say will have a clear bearing on the law.

I will call the intervention described in this scenario *Calming Walls*.<sup>9</sup>

*Calming Walls* seems to me to be morally innocuous. I do not expect that all will share this moral assessment, and I will not defend it here. (Indeed, my subsequent argument might be thought to indirectly undermine it.) My argument will be directed only to those who *do* find *Calming Walls* to be morally innocuous. I will argue that, given this, it is difficult to offer a satisfying mental-interference based explanation for the moral objectionability of *Calming Spray*.

The argument proceeds as follows. First, in sections 2 and 3, I consider two putatively mental interference-based objections that have been raised against conative neuromodulation, asking whether these succeed in fully accounting for the moral objectionability of *Calming Spray* while also accommodating the moral innocuousness of *Calming Walls*. I argue that they do not. Second, I consider whether the descriptive differences between *Calming Spray* and *Calming Walls* can explain the supposed moral

---

<sup>9</sup> Note that some might classify *Calming Walls* as a ‘nudge’. However, there are at least two grounds on which this classification might be disputed. First, on some definition, a nudge must operate via a change in ‘choice architecture’ (see, for example, Richard H. Thaler and Cass R. Sunstein, *Nudge: Improving Decisions About Health, Wealth and Happiness* (London: Penguin, 2009), at p. 6). The concept of ‘choice architecture’ has not been elucidated, but it could plausibly be understood in a way that excludes *Calming Walls*. For instance, on one reading, the term ‘choice architecture’ is to be read as literally referring to the physical topography of the agent’s environment, so that a change in choice architecture involves altering the locations of objects in the agent’s environment. On this view, *Calming Walls* arguably does not qualify as a nudge. Second, it is sometimes held to be definitional of nudges that they are ‘easy and cheap to avoid’ (Thaler and Sunstein, *Nudge*, p. 6). Again, this may not be the case for *Calming Walls*. On the other hand, some have understood nudges such that the category could plausibly accommodate *Calming Walls*. For example, T. M. Wilkinson takes the distinctive feature of nudging to be its reliance on “modern social science’s insights into human thinking and behaviour”: “nudges include any methods, apart from carrots or sticks, that rely on those insights” (see T. M. Wilkinson, ‘Nudging and Manipulation’, *Political Studies* 61 (2013), pp. 341–55, at p. 343). On this view, there seems no reason to exclude *Calming Walls* from the category of nudges.



difference between them. I argue that the descriptive difference that is most plausibly morally significant—that *Calming Walls* operates via perceptual channels, whereas *Calming Spray* does not—cannot, in the end, explain the putative moral difference. Section 5 concludes by presenting a trilemma that falls out of the preceding argument.

Since my argument is limited to comparing *one* instance of conative neuromodulation with *one* instance of environmental motivation modulation, my argument does not licence any general conclusions regarding these two classes of intervention. However, I do think it generalises some distance. For instance, it seems to me that the intuitions I have cited and the arguments I will invoke would apply equally to variants of *Calming Spray* and *Calming Walls* that were set not in a prison but in a university or a company, say.<sup>10</sup> Precisely how far the argument generalises is an issue I leave for further discussion.

## **2. Proper spheres of influence**

One objection that has been made to some forms of conative neuromodulation holds that they transgress a limit on the proper sphere of influence of the modulator. Conative neuromodulation acts directly on the brain states of those subject to it, and, at least in typical cases, these brain states fall outside the domain over which the modulator enjoys a (defeasible) permission to exert her influence.<sup>11</sup> By contrast, environmental

---

<sup>10</sup> Nothing that I say will be specific to the criminal justice context of these interventions. I chose that context only (i) to emphasise the robustness of intuitions against conative neuromodulation; even in criminal justice contexts, where we accept very significant intrusions, conative neuromodulation can seem morally objectionable, and (ii) to make clear that the interventions are nonconsensual.

<sup>11</sup> Reinhard Merkel, personal communication. Bublitz and Merkel ('Crimes Against Minds', pp. 71-2) advance a similar point though it is made in terms of 'legally protected interests' rather than proper spheres of influence: "Trivially, persons may paint their houses or design their stores in whichever way they want although this deliberately changes the minds of visitors

motivation modulation typically operates via altering features that lie within the influencer's proper sphere of influence. Consider, for instance, Thaler and Sunstein's now classic 'cafeteria nudge' example in which cafeteria staff arrange their various meal offerings such that healthier foods lie at eye level, while less healthy foods are placed higher or lower. The layout of foods in a cafeteria falls within the domain over which cafeteria staff enjoy a (defeasible) permission to exert their influence, so this intervention seems to involve no overstepping, by the staff, of limits on their proper sphere of influence.

This appeal to proper spheres of influence might seem promising as a means of driving a moral wedge between *Calming Spray* and *Calming Walls*, for intuitively the colour of the prison's walls falls within the superintendent's proper sphere of influence, whereas the chemistry of the prisoners' brains does not. However, on closer inspection this initial promise quickly dissolves. The fate of the appeal depends, I believe, on how, exactly, we seek to explain *why* the interference involved in *Calming Spray* transgresses a limit on the superintendent's proper sphere of influence.

There are two importantly different possibilities. On the first explanation, *Calming Spray* involves such a transgression because, or because of the way in which, it acts on the prisoners' bodily states; the bodily (including brain) states of the prisoners lie outside the domain over which the superintendent may permissibly exert the relevant

---

and perceivers in various respects. In these and many other cases, indirect interventions are expressions of legally protected interests which have to be balanced against the right to remain free from interventions.

By contrast, directly intervening into another's brain by, e.g. secretly spraying Oxytocin or putting a magnetic coil over another's head is not normally an expression of a legally protected interest."

kind of influence. This explanation has the virtue of rendering the implied objection to *Calming Spray* easy to motivate. It is easy to see why the bodily states of the prisoners might lie outside the proper sphere of influence of the superintendent. They may do so, for example, because the prisoners hold property rights over their bodies. However, on this explanation, the objection is irrelevant to our question: we were seeking an explanation for why *Calming Spray* is more problematic than *Calming Walls* in virtue of the kind of *mental* interference that it involves, but the present objection appeals only to *bodily* interference.

On the second explanation, *Calming Spray* transgresses a limit on the superintendent's proper sphere of influence at least in part in virtue of the kind of *mental* interference that it involves. It cannot be that *Calming Spray* lies beyond the proper sphere of influence of the superintendent merely because it intentionally affects the mental states of the prisoners, for there are many kinds mental effects that a prison superintendent is defeasibly permitted to produce. A prison superintendent may, for example, permissibly cause a prisoner to believe that he will be released tomorrow, if that is indeed the case. Nor can the problem be merely that it has the particular mental effect that it has, for that effect—a reduction in aggression—is one that the superintendent would plausibly be permitted to bring about in other ways, for example, through an educational intervention or, I would think, *Calming Walls*. Perhaps, however, there is some feature of the *manner in which* the superintendent brings about these mental effects that explains why *Calming Spray* transgresses a limit on the superintendent's proper sphere of influence.

The difficulty here is that it is certainly not *obvious* why the kind of mental interference involved in *Calming Spray* takes it beyond the superintendent's proper sphere of

influence. The proponent of the present objection owes us an explanation for why it should be thought to do so. However, it seems to me that none of the most promising candidate explanations appeal to the particular role or situation of the prison superintendent. Rather, they all purport to show that that the relevant kind of mental interference is *pro tanto* wrong *regardless* of who performs that interference. But if that is so, then there is no need to appeal to the proper sphere of influence the prison superintendent. The objector to *Calming Spray* can instead appeal directly to the general wrong-making features of the mental interference that it involves.

Thus, it seems that the appeal to proper spheres of influence either collapses into an appeal to bodily interference, or, if it invokes mental interference, is redundant, because the explanation for why the mental interference involved in *Calming Spray* oversteps the superintendent's proper sphere of influence will also provide a direct explanation for its objectionability.<sup>12</sup> Our question then remains: *is* there any direct explanation for the objectionability of the kind of mental interference involved in *Calming Spray*?

### **3. Arational influence**

Perhaps the most prominent concern that has been raised about conative neuromodulation in the literature to date holds that such modulation is objectionable in part because it involves intentionally influencing or overriding the recipient's choices in ways that do not engage the rational capacities of the recipient. Thus, for example, John Harris objects to certain forms of conative neuromodulation (those intended to produce moral improvement) on the grounds that these interventions operate

---

<sup>12</sup> Of course, it may be that the appeal to proper spheres of influence is also redundant on the first understanding, where the prison superintendent's proper sphere of influence is taken to be limited so as to exclude bodily invasion. It may be that the prohibition on bodily invasion is more general and not at all connected to the role or situation of the prison superintendent.

directly on the mainsprings of action, on emotions or other dispositions, and in particular they operate by cutting out or bypassing what they perceive as a dangerously paralyzing or dilatory process that might somehow get between an impulse and the moral action it impels. This sometimes, but not necessarily, dilatory process is thought or reflection.<sup>13</sup>

Harris does not specify what, exactly, he intends to include within the categories of ‘thought’ and ‘reflection’, but it seems plausible and charitable to interpret him as including all mental processes that are rational, in the sense of being reasons-responsive.<sup>14</sup> That, in any case, is how I will interpret him.

It is important to note that, if an objection like Harris’ is to succeed when directed against *Calming Spray*, it cannot be understood as maintaining that this intervention leaves no space for rational engagement on the part of the prisoners. There may even be a sense in which this intervention in fact *relies* on their rationality. *Calming Spray* may succeed in preventing violent behaviour because, in attenuating aggressive impulses, it leaves the prisoners who would otherwise have been prone to such impulses with more scope to govern their conduct through rational processes—processes that typically do not result in a decision to act violently.

---

<sup>13</sup> John Harris, ‘What It’s Like to Be Good’, *Cambridge Quarterly of Healthcare Ethics* 21 (2012), pp. 293–305, at p. 294. See also Robert Sparrow, ‘Better Living through Chemistry?’, esp. at pp. 26–7. This concern is sometimes presented, as by Harris, as being a concern about interventions that *bypass* rationality.

<sup>14</sup> Alternatively, he may intend to refer only to deliberative processes—processes involving conscious thought. However, interpreting him thus seems less charitable, for it is difficult to see why an intervention that engages the recipient’s reasons-responsive but non-deliberative processes should be thought more problematic than one which engages reasons-responsive deliberative processes.

Rather, the point must be that the initial effects of the intervention on the motivational states of the recipient are not mediated by rational processes.<sup>15</sup> The intervention changes motivational states, and these then exert an influence on rational processes, rather than the other way around. This sets *Calming Spray* apart from certain other approaches that the superintendent might have taken to diminishing prisoner violence. Suppose that, rather than administering the *Calming Spray*, the superintendent had chosen instead to institute classes in which prisoners would read literature chosen to inspire empathy for the victims of violence. And suppose these classes had succeeded in reducing violent outbursts in the prison. In this case, the superintendent would, as in *Calming Spray*, have exerted an influence on the motivational states of the prisoners. However, in this case, the primary motivational change would have been the upshot of *rational* engagement with literature, a feature that plausibly makes this strategy for aggression-reduction less problematic than *Calming Spray*.

Why are interventions whose primary motivational effects are mediated by rational processes—henceforth simply, *rational interventions*—less problematic than arational interventions? Perhaps because they treat the objects of the influence as rational agents, and thus as the moral equals of the influencing agent, in a way that arational interventions do not. Robert Sparrow has argued that education, the paradigmatic rational intervention,

acknowledges a fundamental moral equality between educator and educated. Implicit in this relationship is the requirement that the educator must, if called upon to do so, be able to justify the norms that have shaped

---

<sup>15</sup> See, for discussion, Thomas Douglas, ‘Enhancing Moral Conformity and Enhancing Moral Worth’, *Neuroethics* 7 (2014), pp. 75–91.

the educational project and its content with reasons that the person being educated should accept. In doing so, the interaction is also open to the possibility that the person being educated will respond with counterarguments of sufficient force to change the mind of the educator. This requirement to be responsive to the demands of reason persists even if it is, in fact, impossible for the person being educated to offer arguments in reply to the activities of their educators in this particular case.

By contrast

[b]iomedical interventions to reshape the agency of others . . . operate in an instrumental or technical mode. They involve a subject acting towards an object and as such are fundamentally structured by a profound inequality”.<sup>16</sup>

Similarly, Elizabeth Shaw suggests that, under certain conditions, an arational intervention

would amount to treating her [the recipient] as if she were a puppet, an automaton or a robot—as something less than human. In other words it would ‘objectify’ her.<sup>17</sup>

I wish to suggest, however, that insofar as we should be concerned about *Calming Spray* and other forms of conative neuromodulation in virtue of their arational nature, we should also be concerned about some environmental interventions, including *Calming*

---

<sup>16</sup> Robert Sparrow, ‘Better Living through Chemistry?’ at p. 26. Sparrow is here developing a theme from Jürgen Habermas’ *The Theory of Communicative Action, Vol 2. Lifeworld and System: A Critique Of Functionalist Reason* (Boston, MA: Beacon Press, 1987), and *The Future of Human Nature* (Cambridge: Polity Press, 2003).

<sup>17</sup> Elizabeth Shaw, ‘Direct Brain Interventions and Responsibility Enhancement’, *Criminal Law and Philosophy* 8 (2014), pp. 1–20, at p. 12.

*Walls*. *Calming Walls* also has primary motivational effects that are presumably not mediated by rational processes. It is not that the greenness of the walls gives the students reasons to be calm—reasons to which the students then rationally respond. Rather, the students simply see the walls and then, via some brute subconscious mechanism, become calmer. At least, that is one possible elucidation of the mechanism of action, and the one that I will henceforth assume.<sup>18</sup> Thus, we see that what has been regarded as the most objectionable feature of conative neuromodulation—namely, that its primary motivational effects are arational—is a feature that *Calming Spray* shares with *Calming Walls*.

#### **4. Nonperceptual influence**

Up to this point, my strategy has been to begin from moral concerns that have been ascribed to interventions like *Calming Spray*, and show that these apply as much to *Calming Walls* as to *Calming Spray*.

In doing so, I hope to have cast some doubt on the suggestion that there *is* a morally significant difference between *Calming Walls* and *Calming Spray*. However, I have certainly not refuted this suggestion, as it may be that there is a morally significant

---

<sup>18</sup> Note that the ‘brute’ rider contained in this assumption is intended to rule out the possibility that the calmness is produced through a subconscious but reasons-responsive aesthetic sensibility. On certain views in aesthetics, the fact that the walls are of a particular colour could give perceivers of those walls reason to respond in certain ways, for example, by experiencing calmness. For instance, on an aesthetic realist view, perceivers could have reason to experience calmness because the wall colour has the objective property of calmingness. An aesthetic sensibility that reliably responded to such reasons could be regarded as a rational process. I assume that either (i) the wall colour generates no reason for the prisoners to experience calmness (for instance, because the relevant views in aesthetics are false, or because the experiencing of calmness is not suitably within the control of the prisoners), or (ii) if it does generate such a reason, the prisoners’ calmness is not a response to that reason. I thank Christopher Bennett for pressing me to address this issue.



difference between the two interventions that has not been noticed in existing criticism of conative neuromodulation,<sup>19</sup> or that has been noticed by others but not by me.

In order to diminish the likelihood that I have missed some plausible explanation of the moral difference between *Calming Walls* and *Calming Spray*, in this section, I will approach the matter from a different tack. Rather than beginning with the putatively problematic features of interventions like *Calming Spray* and examining whether *Calming Walls* and *Calming Spray* differ with respect to those features, I will begin with the descriptive differences between these two interventions and consider whether these descriptive differences can explain the putative moral difference.

Of course, there are *many* descriptive differences between *Calming Walls* and *Calming Spray*. An exhaustive analysis would examine the moral significance of each of these. I cannot pursue such an exhaustive method here. Instead, I will consider only one descriptive difference between the two interventions: the difference that I believe holds most promise as a potential ground for explaining the supposed moral difference. That difference has to do with the role played by perceptual processes in mediating the primary motivational effects of *Calming Walls*.

We saw in the previous section that both *Calming Walls* and *Calming Spray* can be understood to achieve their primary motivational effects through arational means. However, there is nevertheless a clear difference in the way in which they bring about those effects. *Calming Walls* operates via perceptual mechanisms whereas *Calming*

---

<sup>19</sup> This would be unsurprising since existing discussions of conative neuromodulation have typically contrasted it with deliberative forms of behaviour change, not with arational forms of environmental motivation modulation.

*Spray* does not. In *Calming Walls*, the students *see* the walls painted green, and this then sets in train some brute subconscious process which attenuates strong impulses towards aggression. In *Calming Spray*, there is no such role for perception. (The motivational change is instead the upshot of a chemically or physically induced change in the neurochemical bases of aggression.) Let us refer to this descriptive difference as the perceptual-nonperceptual difference.<sup>20</sup>

This distinction between *Calming Spray* and *Calming Walls* opens the door to the possibility that there is also a moral difference between the two. It provides a descriptive base on which the moral difference might supervene. However, that is not to say that it *explains* the moral difference. A satisfying explanation for why *Calming Spray* is more morally problematic than *Calming Walls* would do more than merely point to some descriptive difference between the two cases. It would, at the very least, point to some descriptive difference whose moral significance is independently plausible (for example, because there is a plausible explanation for why the difference should be morally significant, or because its moral significance is able to account for a range of other intuitive judgements). In what follows, I consider two independent grounds for thinking that the perceptual-nonperceptual difference is morally significant.

---

<sup>20</sup> This distinction between perceptual and nonperceptual forms of influence may map on to Bublitz and Merkel's distinction between direct interventions—which 'bypass' (p. 70) psychological processes—and indirect interventions—which are 'mediated. . . by internal processes on the part of the addressee' (p. 69). However, though Bublitz and Merkel suggest that an intervention should be understood as direct if it reaches the brain 'by other routes than sensual perception' (p. 69), they do not commit themselves to the view that involvement of perception is sufficient for an intervention to qualify as indirect. They write that 'Tentatively, indirect (or external) interventions are those stimuli which are perceived sensually . . . and pass through the mind of the person, being processed by a host of psychological mechanisms' (p. 69, my italics). See Bublitz and Merkel, 'Crimes Against Minds', pp. 69-70.

An initial suggestion would be that the perceptual-nonperceptual difference is morally significant because it is relevant to the *transparency* of *Calming Walls* and *Calming Spray*: because it engages perception, *Calming Walls* is transparent in a way that *Calming Spray* is not, and this explains its moral innocuousness.

A similar thought has been invoked in defence of nudge techniques. Nudges, as understood by Thaler and Sunstein in their eponymous book on the topic, are held by them to be transparent in two important respects. First, they are announced in advance. Second, the use of the nudge can be monitored.<sup>21</sup> For example, those subjected to the abovementioned cafeteria nudge can monitor how they are being nudged by observing how the foods are arranged.

It might seem that, by contrast, *Calming Spray* will not be transparent in this second respect. Though it was, in my initial scenario, announced in advance, its use cannot be monitored because it does not operate via perceptual means.

This is incorrect, however. Though *Calming Spray* does not *operate via* perceptual means, it can still be perceived (I specified that the spray was visible). There is no difference between *Calming Spray* and *Calming Walls* in the perceptibility of the stimulus. Thus, in this respect, *Calming Spray* is as amenable to monitoring as is *Calming Walls*. Rather the difference lies in the role played by perception in bringing

---

<sup>21</sup> Several authors suggest that this second kind of transparency may set nudges apart from subliminal advertising. See Thaler and Sunstein, 'Nudge', pp. 244-7; L. Bovens, 'The Ethics of Nudge', in T. Grüne-Yanoff and S. O. Hansson (eds.), *Preference Change: Approaches from Philosophy, Economics and Psychology* (Berlin and New York: Springer, 2009), at pp. 216-17; Wilkinson, 'Nudging and Manipulation', p. 344.

about motivational change: in *Calming Spray* it is irrelevant, whereas in *Calming Walls* it's crucial.

Still, it might be argued that there is a further important respect in which *Calming Spray* is not monitorable. Though the existence of the spray in the air can be seen, the process via which that spray reduces aggression is entirely opaque to those influenced by it. This *process* cannot be monitored, though the presence of the chemical that initiated it can. By contrast, in *Calming Walls*, those subject to the intervention can observe at least the first stage of the process via which the green-painted walls diminish their aggression—namely, their formation of perceptual beliefs regarding the colour of the walls.

It seems doubtful that this difference is significant, however. Although those subject to *Calming Walls* can see the wall colour that initiates their reduction in aggression and can observe what we might think of as the first step in the process via which their aggression is reduced—namely their formation of perceptual beliefs about the wall colour—beyond that, the process via which their aggression is reduced is, I take it, entirely opaque; recall that I assumed above that the remainder of this process is subconscious. Given this, *Calming Walls* seems barely more monitorable than *Calming Spray*.

A second suggestion would be that the perceptual-nonperceptual difference is morally significant because it maps on to a difference in resistibility. Those subjected to *Calming Walls* can resist its effects more easily than can those subjected to *Calming*

*Spray* in virtue of the fact that the former operates via perceptual channels while the latter does not.

Again, however, this suggestion runs into difficulties. There are three points at which the prisoners who are the targets of *Calming Walls* might resist its effects. First, they might simply avoid the stimulus—that is, they might avoid looking at the green walls. Second, they might look at the walls but resist forming the perceptual belief that they are such-and-such shade of green. Or third, they might resist the influence of this perceptual belief on their aggression. Let us examine these possibilities in turn, in each case considering whether there is reason to think that *Calming Walls* is indeed resistible in the specified way, and that *Calming Spray* is not in some parallel respect resistible.

Regarding the first possibility, it seems likely that the prisoners could avoid looking at the walls only by taking extreme measures (keeping their eyes closed for long periods of time, for example). But it may be that exposure to *Calming Spray* could also be avoided by taking extreme measures (for example, wearing a gas mask all day or holding one's breath for long periods of time), so, though it may be in some sense possible to avoid exposure to *Calming Walls*, this may not set *Calming Walls* apart from *Calming Spray*—whether it does will depend on how we fill out the details of these interventions. Let us simply stipulate that the relevant stimulus can be avoided at similar costs in the two cases. This rules out an appeal to ease of stimulus avoidance to explain the moral difference between these cases, yet intuitively, the moral difference between the two interventions remains.

The second possibility was that prisoners might look at the green-painted walls, but resist forming the perceptual belief that they are green. The difficulty here is that there is little reason to believe that the formation of perceptual beliefs can be resisted in this way. There is lively debate on the question whether our beliefs are ever within our direct voluntary control.<sup>22</sup> But even those who subscribe to direct doxastic voluntarism—the view that at least some of our beliefs *are* within our direct voluntary control—typically do not cite perceptual beliefs as plausible candidates for such beliefs. It is widely accepted, and I think very plausible, that we do not have direct control over such beliefs.<sup>23</sup> Moreover, the only remotely plausible way of maintaining that perceptual beliefs *are* within our direct voluntary control involves effecting a divorce between perceptual beliefs and the receipt of the perceptual information on which they are based. *If* one regarded perceptual beliefs as judgments made in response to receipt of perceptual information, one could perhaps credibly maintain their voluntariness;<sup>24</sup> one could hold that though we cannot voluntarily control what perceptual information we

---

<sup>22</sup> See, for some seminal contributions to this debate, Bernard Williams, ‘Deciding to Believe’, in H. E. Kiefer and M. K. Munitz (eds.), *Language, Belief, and Metaphysics* (Albany: SUNY Press, 1970); E. M. Curley, ‘Descartes, Spinoza, and the Ethics of Belief’, in M. Mandelbaum and E. Freeman (eds.), *Spinoza: Essays in Interpretation* (LaSalle, IL: Open Court Publishing, 1975); William Alston, ‘The Deontological Conception of Epistemic Justification’, in *Epistemic Justification: Essays in the Theory of Knowledge* (Ithaca: Cornell University Press, 1989). Carl Ginet, ‘Deciding to Believe’, in M. Steup (ed.), *Knowledge, Truth, and Duty: Essays on Epistemic Justification, Responsibility, and Virtue* (New York: Oxford University Press, 2001); Nishi Shah, ‘Clearing Space for Doxastic Voluntarism’, *The Monist* 85 (2002), pp. 436–45; Sharon Ryan, ‘Doxastic Compatibilism and the Ethics of Belief’, *Philosophical Studies* 114 (2003), pp. 47–79.

<sup>23</sup> For arguments that perceptual beliefs, specifically, are not subject to direct voluntary control, see Louis P. Pojman, ‘Believing and Willing’, *Canadian Journal of Philosophy* 15 (1985), pp. 37–55, at p. 41; Brian Weatherson, ‘Deontology and Descartes’s Demon’, *Journal of Philosophy*, 105 (2008), pp. 540–69, at pp. 556–61. Of course, we may be able to voluntarily control perceptual beliefs indirectly, for example, by controlling whether we are exposed to the external stimuli that produce them. However, as noted above, we have a similar form of control in relation to *Calming Spray*.

<sup>24</sup> John Cottingham describes a Platonic model of belief which allows for such a divorce. See his ‘Descartes and the Voluntariness of Belief’, *Monist* 85 (2002), pp. 343–60, esp. at p. 345.

receive given exposure to a particular stimulus, we can voluntarily control how we cognitively respond to that information.

On this model, however, it is not clear that *Calming Walls* exerts its effect via influencing the students' perceptual beliefs. Perhaps the mere receipt of perceptual information regarding the colour of the walls is enough to set in train a subconscious process that results in the attenuation of aggressive impulses. If this is so, then whether perceptual beliefs are within our direct control will be irrelevant to the resistibility of *Calming Walls*. What matters is whether the perceptual information that the students receive due to their exposure to the green walls is within their voluntary control, and it seems clear that it is not.<sup>25</sup>

We may, of course, be able to control the perceptual information that we receive and the perceptual beliefs that we form *indirectly*, that is, by taking other actions which affect the likelihood that we will acquire those beliefs or receive that information. For example, we may be able to control whether we are exposed to the relevant external stimuli. However, it is plausible that, with respect to *Calming Walls*, the only form of indirect control of perceptual information and beliefs available to the prisoners involves taking extreme measures, such as keeping one's eyes closed for long periods to time, and, as stipulated above, similar forms of control are available with respect to *Calming Spray*.

Finally, the third possibility was that the prisoners can resist the influence of the perceptual belief (or information) that the wall is such-and-such shade of green on their

---

<sup>25</sup> Weatherson, 'Deontology and Descartes's Demon', pp. 557-8.

levels of aggression. However again, there seems no reason to suppose that this can be directly resisted. As noted above, I assume that, once the relevant perceptual belief has been formed (or the relevant perceptual information has been received) aggressive impulses are then attenuated through a brute subconscious process, and there is no reason to suppose that we can directly and voluntarily control those processes.

Now, again, it may be that the subconscious processes via which *Calming Walls* operates could be *indirectly* resisted.<sup>26</sup> That is, it may be that the prisoners could take steps to make themselves more resilient to the calming effect of the green-painted walls. Perhaps, for example, they could foster feelings of rage that tend to negate that effect. However, as with the point about stimulus avoidance, similar points may apply to *Calming Spray*, depending on how we spell out the mechanisms of that intervention. Let us stipulate that the effects of *Calming Spray* can be indirectly resisted in the same ways as can those of *Calming Walls*. Still, the intuitive moral difference between the two interventions remains, and we can no longer appeal to a difference in indirect resistibility to explain this difference.

Taking stock: it is plausible that the influence of *Calming Walls* can be avoided only by taking those rather extreme steps required to avoid the stimulus itself, or by indirectly undermining the calming effect of seeing the walls (for example by fostering counter-acting rage). But, with the stipulations introduced in this section, the influence of *Calming Spray* can also be avoided in these ways. With these stipulations, *Calming Spray* is no more resistible than *Calming Walls*. Yet the intuitive moral difference between the interventions remains. Thus, like the appeal to transparency, the appeal to

---

<sup>26</sup> I thank an anonymous reviewer for pressing me to consider this point.



resistibility seems unable to explain why the different role of perception in *Calming Spray* and *Calming Walls* should be thought morally significant.

## **5. Implications**

I have argued that the dominant objections that have been raised against conative neuromodulation fail to establish that *Calming Spray* involves an objectionable form of mental interference that *Calming Walls* does not. I have also considered whether the involvement of perception in the mechanism of *Calming Walls*, but not that of *Calming Spray*, can explain the putative moral difference between the two interventions. I rejected what I take to be the most plausible grounds for thinking that it can: that the perceptual mechanism of *Calming Walls* makes it more transparent than *Calming Spray*, and that it makes it more resistible.

None of this absolutely rules out the possibility that the supposed moral difference between *Calming Walls* and *Calming Spray* can be explained, and in a satisfying way, by a difference in the kind of mental interference that is involved. It is possible, for example, that I have missed some basis for supposing that the perceptual-nonperceptual difference is morally significant. It is also possible that there is some other descriptive mental difference between the two cases that is able to explain the putative moral difference between them. However, I believe the considerations that I have discussed offer the most promising prospects for providing a satisfying mental interference-based explanation for this putative moral difference. Thus, I believe that my argument, if it goes through, substantially increases the credibility of the view that there is no satisfying way of explaining why *Calming Spray* is more problematic, in terms of its mental effects, than *Calming Walls*.

Where would the correctness of this view leave us? It would leave us with three options.

First, we could accept that *Calming Spray* is, like *Calming Walls*, not seriously problematic in terms of its mental effects. On this view, the objectionability of *Calming Spray* must derive chiefly from its physical invasiveness. But this implies, counter-intuitively, that, *Calming Spray* is not (much) more problematic than *Antiviral Spray*, a similarly physically invasive intervention that lacks the mental effects of *Calming Spray*. It also leaves us with the puzzling result that, even though the mental effects of *Calming Spray* seem more morally significant than its bodily effects, it is the bodily effects which explain its greater objectionability than *Calming Walls*.

Second, we could accept that *Calming Walls* is, like *Calming Spray*, seriously problematic in terms of its mental effects. But this is, I think, also counterintuitive. Moreover, this concession will cast a moral shadow on a wide range of environmental interventions—including many proposed nudge techniques—that seem similar to *Calming Walls* with respect to their morally significant features, but that are widely thought to be morally innocuous.

Each of these first two options involves conceding that *Calming Spray* is not more objectionable, in virtue of the kind of mental interference that it involves, than *Calming Walls*. A third option would avoid this concession: one could hold that there is a moral difference between *Calming Walls* and *Calming Spray* with respect to their mental effects, though there is no satisfying way of explaining it. One might claim, for example, that the bare intuition that *Calming Spray* is more problematic than *Calming*

*Walls* is itself sufficient to establish that there is a moral difference between them, and that the intuition that *Calming Spray* is more problematic than *Antiviral Spray* is sufficient to establish that this difference must be attributable in part to a difference in the kind of mental interference that they involve.

Each of these options has some awkward implications and it is not clear to me which is least problematic. But what does seem clear is that each of these options warrants further consideration.

What also seems clear is that the moral parallels between conative neuromodulation and environmental forms of motivational influence suggest that it would be fruitful to bring the ethical discussions regarding these two kinds of intervention into closer proximity. To date, there has been little overlap in these discussions, perhaps because they have been separated by disciplinary boundaries—discussion of conative neuromodulation has taken place largely within the confines of neuroethics and criminal justice ethics, while discussions environmental interventions—chiefly focused on nudges—have been discussed primarily within political philosophy and public health ethics. But if my arguments above are correct, or at least roughly on track, many of the arguments offered in discussion of the one kind of intervention could have interesting application to the other.