

Open Research Repository

Nudging for liberals

Item Type	Journal article
Authors	Moles, Andrés
DOI	https://doi.org/10.5840/soctheorpract201541435
Download date	2024-10-12 00:25:58
Link to Item	http://hdl.handle.net/20.500.14018/13485

Nudging for Liberals

Andrés Moles

Abstract: In this article I argue that anti-perfectionist liberals can accept nudging in certain areas: in particular, they can accept nudges aimed at helping people to discharge their nonenforceable duties, and to secure personal autonomy. I claim that nudging is not disrespectful since it does not involve a comparative negative judgment on people's ability to pursue their plans, and that the judgments that motivate nudging are compatible with treating citizens as free and equal. I also claim that despite being sometimes manipulative, nudging is easy to resist and so it can be employed to pursue legitimate goals.

Keywords: nudging; manipulation; anti-perfectionism; respect; liberalism; duties

Research in the behavioral sciences in recent decades strongly suggests that there are means other than rational persuasion by which governments can improve citizens' welfare without restricting any of their choices and without making their choices more costly. We now know that a variety of biases are pervasive features of human cognition. We also know that environmental settings have a deep impact on our choices and that seemingly trivial changes in the way information is conveyed, choices are arranged, or default rules are set can change the choices people make.¹ These biases make it possible that "choice architecture" can be arranged in a way that "nudges" people into making certain choices.

Richard Thaler and Cass Sunstein (T&S) have recently argued that nudging is compatible with libertarianism.² Their claim has been subject to serious criticism.³ In this article, however, I explore a different question: Is nudging ever morally permissible from a liberal point of view? This question is pressing, I think, because critics have suggested that nudging is disrespectful and manipulative.⁴ In this article, I endorse an

¹Richard H. Thaler and Cass R. Sunstein, *Nudge: Improving Decisions about Health, Wealth, and Happiness*, revised and expanded ed. (London: Penguin, 2009); Dan Ariely, *Predictably Irrational: The Hidden Forces That Shape Our Decisions* (New York: HarperCollins, 2009).

²Thaler and Sunstein, *Nudge*.

³Gregory Mitchell, "Libertarian Paternalism is an Oxymoron," *Northwestern University Law Review* 99 (2005): 1245-77.

⁴Jonathan Quong, *Liberalism without Perfection* (Oxford: Oxford University Press, 2011); Till Grüne-Yanoff, "Old Wine in New Casks: Libertarian Paternalism Still Vio-

anti-perfectionist version of liberalism inspired by John Rawls.⁵ Rawls observes that liberal societies are deeply divided by a variety of reasonable, yet incompatible, conceptions of the good. Given this fact, he argues that for political institutions to be properly justified, we should eschew appealing to controversial views of what makes a life valuable. The upshot of this argument is a commitment to anti-perfectionism, namely, the view that there are legitimacy-based reasons to restrain from appealing to comprehensive contested truths. Political institutions, Rawls insists, should be justified by appealing only to the interests of free and equal citizens committed to sharing the benefits and burdens of social cooperation. Citizens, according to Rawls, have two fundamental moral powers (with correlative interests in their development and exercise): having a sense of justice and the ability to rationally form, revise, and pursue a permissible conception of the good. If we accept this version of liberalism, manipulation is objectionable because it threatens people's ability to fulfill their second moral power, that is, to form and pursue their ends independently. Similarly, it is disrespectful because it fails to treat people as capable of living their lives according to their own lights.

The main claim of this article is that insofar as nudging satisfies the justificatory constraints that any other policy needs to satisfy, nudging is permissible. This means that in itself nudging is not morally problematic. I will proceed as follows. In the first section I will present the argument that motivates nudging. In section 2, I will briefly discuss the concept of "nudge," and will provide a characterization. In my account, a nudge is an intentional feature of choice architecture that would be ineffective on a person if she was perfectly rational. Nudges are also easily resistible. Sections 3 and 4, respectively, address the challenges that nudging is disrespectful and that it is manipulative. Section 5 illustrates the way in which a variety of nudges can be justified within an anti-perfectionist framework. Section 6 concludes.

1. The Argument for Nudging

The argument in favor of nudging is based on three claims. First, it denies the Millian epistemic argument, which holds that "individuals know their

lates Liberal Principles," *Social Choice and Welfare* 38 (2012): 635-45; Daniel Hausman and Brynn Welch, "Debate: To Nudge or Not to Nudge," *Journal of Political Philosophy* 18 (2010): 123-36; T.M. Wilkinson, "Nudging and Manipulation," *Political Studies* 61 (2013): 341-55.

⁵John Rawls, *Political Liberalism* (New York: Columbia University Press, 1996). Clearly not all liberals endorse this view. The arguments I defend assume, nevertheless, that this view is correct.

tastes and situations [including their preferences] better than officials do, they are in the best position to identify their own ends and the best means of obtaining them.”⁶ Mill’s “best judge of one’s own interests” principle is an empirical claim that we have reasons to doubt. Familiar cases such as weakness of the will, hyperbolic discounting, or the inability rationally to gauge probabilities cause people to make bad choices.⁷

Second, in many cases, individuals’ preferences are endogenous to social contexts. We tend to assume that people have stable preferences, and that (seemingly) arbitrary contextual changes would not affect those preferences. The assumption, however, is false. Consider these examples: people’s preferences are sensitive to the way in which information is framed. Even experienced medical doctors are more likely to choose a treatment when its effects are described as 200 lives saved out of 600, as opposed to 400 deaths out of 600.⁸ In addition, people’s preferences can also be altered when an irrelevant alternative is added. People might prefer going on an all-expenses-paid holiday to Rome over Paris, but when a third option such as “Paris without coffee” is added, people might switch to Paris over Rome.⁹ A third example: people seem to have a status quo bias. Whenever there is a default option, people tend to stick with it. These effects can be due to people taking the status quo as a suggestion (when they are not sure, they stick with it), or due to inertia and procrastination. For instance, enrollment in retirement savings plans dramatically increases when people are automatically enrolled (with an opt-out option), in contrast to when they need to opt in.¹⁰

In T&S’s view, it is often impossible to identify preferences prior to contexts. This claim has to be read with some caution. They do not argue that all preferences are endogenous, but rather that in certain contexts and domains, some of them are. Clearly, some of our preferences are strong enough to resist the effects of frames or default rules. Part of T&S’s argument includes criteria that identify the cases in which preferences are endogenous and subject to irrational biases. They suggest that people’s choices are suboptimal when people are inexperienced about

⁶Cass R. Sunstein, *Why Nudge? The Politics of Libertarian Paternalism* (New Haven: Yale University Press, 2014), p. 7.

⁷For hyperbolic discounting, see Jon Elster, *Explaining Social Behaviour* (Cambridge: Cambridge University Press, 2007); for probabilistic irrationality, see Thomas Gilovich, Dale Griffin, and Daniel Kahneman, *Heuristics and Biases* (Cambridge: Cambridge University Press, 2002). It is unclear whether the choices are only suboptimal or absolutely bad. I take it that T&S consider both.

⁸Daniel Kahneman, “Maps of Bounded Rationality,” *American Economic Review* 93 (2003): 1449-75.

⁹Ariely, *Predictably Irrational*, p. 10.

¹⁰T&S, *Nudge*, p. 43.

making certain choices, when feedback is confusing, and when the effects of choices (as opposed to their rewards) come slowly.¹¹ For instance, most of us have little experience (if any) of choosing retirement plans or mortgages. Choosing the best mobile phone plan is confusing, since it is difficult to anticipate exactly how much we will use the different services provided. Finally, “investment” or “sinful” goods, such as dieting or exercising, are hard to achieve, since the costs are paid now and the rewards come much later.

The third claim supporting nudging is that not all paternalism needs to be coercive. In T&S’s account, coercive paternalism involves increasing someone’s welfare by restricting some of her choices or significantly increasing their costs. Instead, they defend a form of paternalism that is libertarian in the sense of being “liberty-preserving.” In their words, libertarian paternalism “is a relatively weak, soft, and nonintrusive type of paternalism because choices are not blocked, fenced off, or significantly burdened.” Libertarian paternalism works by arranging choice environments in ways that make it more likely that people will make choices that make them better off “as judged by themselves.”¹²

There are two different ways in which people’s choices can be improved. First, a person can make a better choice by improving the process by which she reaches her decision. An environment can make relevant reasons more salient or can help obscure some distracting factors. Moreover, reducing the impact of cognitive biases can also help to improve the process of deliberation. Giving (good) advice is a paradigmatic case of improving decision-making in this sense. Implementing blind review procedures is also an example of enhancing decision-making by reducing biases. The second way of improving someone’s choice is to make sure that her decision aligns with her preferences. This case pays no attention to the process of decision-making: it looks only at whether people’s choices correspond to their preferences. Nudging is controversial because it employs the second strategy. It enrolls (or exploits) cognitive biases in order to improve choice, without enhancing the process of decision-making.¹³

2. A Characterization of Nudge

An apparent difficulty in the literature is that there is no clear definition

¹¹Ibid., pp. 10, 80-84.

¹²Ibid., p. 5.

¹³Notice, though, that sometimes more deliberation might be detrimental for preference alignment. In sports, one can perform better by not thinking too much. See Gerd Gigerenzer, *Gut Feelings* (New York: Viking, 2007).

of “nudge.” Although T&S define nudge “mainly by example,”¹⁴ they write that “[a] nudge ... is any aspect of the choice architecture that alters people’s behavior in a predictable way [without being coercive]. To count as a mere nudge, the intervention must be easy and cheap to avoid.”¹⁵ Later on, they provide a different definition. They introduce a counterfactual test to the effect that “a nudge is any factor that alters the behavior of Humans, even though it would be ignored by Econs.”¹⁶ “Econs” are, in T&S’s jargon, the “textbook picture of human beings offered by economists”: unbounded rational beings with unlimited cognitive resources, full information, and no time constraints to make decisions, who also come furnished with complete and consistent preferences.¹⁷

I am not too optimistic about the usefulness of working out a satisfactory definition of nudge. The political morality of nudges depends on substantive arguments: finding out what duties public officials owe to citizens and what their prerogatives are. These underlying principles are not available through conceptual analysis. Whether nudging is permissible or not depends on normative arguments, not on semantics. Nevertheless, it is important to be clear about how concepts are used. I will stipulate three features that nudges have. Although I believe these features capture well the central cases of nudging, there might be influences that my stipulation fails to capture and that some people want to classify as nudges. As noted, I do not believe anything morally significant hinges on this.

My account of “nudge” has three fundamental features. First, nudges operate through choice architecture. When agents design social environments, they can arrange aspects in a way that predictably influences choice. The prediction is an ex-ante judgment about the likelihood that people would choose one option over another. Second, besides being features of choice architecture, nudges are the deliberate intention of an agent. There are two reasons in support of this condition. First, for the purposes of political morality, the justificatory burden of intentional actions is higher than their nonintentional effects (which does not mean that the latter are unimportant). The second reason is that the objections of manipulation and disrespect, which will be addressed in the following sections, make more sense if nudging is intentional rather than unintentional.¹⁸ Third, nudges are easy to resist. Yashar Saghai argues that an influence is easy to resist if a person can “effortlessly oppose the pres-

¹⁴Hausman and Welch, “Debate,” p. 124.

¹⁵T&S, *Nudge*, p. 6.

¹⁶*Ibid.*, p. 8.

¹⁷*Ibid.*, p. 6.

¹⁸Notice that T&S disagree: since they define nudge in terms of its effects, it matters little whether the influence is intentional or not.

sure to get her to ϕ if she does not want to ϕ .”¹⁹ It is in virtue of being easily resistible that nudging is supposed to preserve free choice, in the sense that it goes deeper than merely preserving the choice set.

T&S explain that nudging is effective and easily resistible at the same time because it depends on a cognitive system that is effortless, automatic, and unconscious.²⁰ Saghai nuances this explanation and suggests that typical nudging operates through shallow cognitive processes. He suggests that these processes have the following three properties: “(1) they are fast; (2) the ‘cognitive miser’ is inclined to rely on them because they consume few resources; (3) they yield responses that are not the result of full-blown deliberation.”²¹ These features capture the counterfactual test mentioned above. The idea is that nudges are effective on people, but would not be on Econs. The reason is that Econs have no shallow processes, or if they do, these processes are always under their deliberative control. Notice, though, that reliance on shallow cognitive processes is not part of my definition of nudging. Although most (or perhaps all) nudges rely on them, it is possible that some do not.²²

Saghai explains that a person can easily resist influences that operate through these processes if she can become aware of the influence, if she can inhibit the response triggered by the influence, and when she is not in a situation that undermines these abilities.²³ For instance, features of choice architecture can grab our attention especially when they influence us against our settled preferences. They can also remove obstacles that would impede the satisfaction of our plans. Some influences increase the difficulty of taking certain actions. For instance, they might impede choice by creating distractions or overloading our cognitive systems. People perform worse at solving puzzles when they are tempted by

¹⁹Yashar Saghai, “Salvaging the Concept of Nudge,” *Journal of Medical Ethics* 39 (2013): 487-93, p. 489.

²⁰T&S, *Nudge*, pp. 19-22.

²¹Saghai, “Salvaging the Concept of Nudge,” p. 489. Notice that “shallow process” is a broader category than “automatic process” because some shallow processes might be partially conscious, even though they are not fully conscious. T&S appeal to a dual system theory, and argue that nudges rely on System 1 (*Nudge*, pp. 21-24; see also Sunstein, *Why Nudge?* pp. 26-34, and Daniel Kahneman, *Thinking, Fast and Slow* (London: Penguin, 2011)). This is unnecessary: we can characterize nudge without a commitment to such a controversial theory. For a critique, see Hugo Mercier and Dan Sperber, “Intuitive and Reflective Inferences,” in Jonathan Evans and Keith Frankish (eds.), *Two Minds: Dual Processes and Beyond* (Oxford: Oxford University Press, 2009), pp. 149-70; Arie Kruglanski and Gerd Gigerenzer, “Intuitive and Deliberate Judgments Are Based on Common Principles,” *Psychological Review* 118 (2011): 97-109.

²²This might be possible if, for instance, developments in cognitive science show that the idea of shallow processes is inadequate.

²³Saghai, “Salvaging the Concept of Nudge,” pp. 489-90.

chocolate rather than boring radishes.²⁴

The use of shallow processing also explains why Sunstein insists that nudging does not rely on incentives or fines.²⁵ Perfectly rational agents are sensitive to changes in prices and will respond to incentives according to their own utility function. Sunstein's argument is that since nudges should preserve free choice, they should not alter their costs. This includes both incentives and taxes. Allegedly, this makes nudging and libertarianism compatible. There are exceptions, though, in which small changes in prices can be best understood as nudges. For instance, people might react irrationally to a very small change in the cost of an option. One explanation of this phenomenon might be that people tend to value choices that are free much more than choices that are marginally costly.²⁶ People's perceived utility is larger when the choice's cost goes from one to zero than when the choice goes from two to one. Incentives might also affect people indirectly by inducing endowment effects.²⁷

The fact that an incentive can be, but need not be, a nudge introduces some level of epistemic uncertainty in determining whether a given influence counts as a nudge or not. Consider these two cases: Bob and Alfred go to a restaurant where there are miniature football goals placed in the men's urinals with a ball suspended from the cross bar to try to encourage accurate urination.²⁸ Is the arrangement a nudge? Imagine that Ben urinates absentmindedly, and because of the football goal his accuracy is increased. In order to provide a full explanation about the way he urinated, we need to cite influences that did not figure in his intentions or deliberations: they relied on shallow processes. Consider now the more difficult case of Albert. He takes part in the activity since it is great fun. He may not have urinated as straight were it not for the goal and ball but it does not follow that it was the choice architecture (the manner in which choices are presented) that influenced him; rather, it was the substantive content of those choices.

The fact that the goal in the urinal nudges Bob but not Albert means

²⁴Roy F. Baumeister, Ellen Bratslavsky, Mark Muraven, and Dianne M. Tice, "Ego Depletion: Is the Active Self a Limited Resource?" *Journal of Personality and Social Psychology* 74 (1998): 1252-65.

²⁵Sunstein, *Why Nudge?* pp. 58, 82-86.

²⁶Cristina Shampanier, Nina Mazar, and Dan Ariely, "Zero as a Special Price," *Marketing Science* 26 (2007): 742-57.

²⁷Endowment effect is the phenomenon by which "people tend to find it much harder to part with something once they have it in hand than to forego it when they have never had it in hand." Luc Bovens, "The Ethics of *Nudge*," in Till Grüne-Yanoff and Sven Ove Hansson (eds.), *Preference Change: Approaches from Philosophy, Economics and Psychology* (Dordrecht: Springer, 2009), pp. 207-20, at p. 208. See also Kahneman, *Thinking, Fast and Slow*, pp. 289-99.

²⁸I owe this example to Matthew Clayton.

that we can never be sure whether a feature of the choice architecture nudges a particular person. This is not a problem though, since choice architecture implies only an ex-ante probability of agents choosing one course of action over another. Thus, nudging relies on a statistical judgment. On the assumption that most of us are like Bob rather than like Albert, then designing urinals in this way counts as nudging.

There are two leading objections to nudging. One argues that it is disrespectful, the other that it is manipulative. I address these objections in turn.

3. Equality and Disrespect

Liberals might be concerned that nudging people fails to respect their agency. Recently, Jonathan Quong has argued that all forms of paternalism are wrong because they fail to treat people as if they were free and equal.²⁹ The objection seems especially plausible in the case of State paternalism. One can argue that the duty of treating people as free and equals accrues mostly (or exclusively) to government officials rather than to citizens in their daily life. Quong argues that paternalism is wrong since it involves a negative judgment about people's abilities to lead their own lives.³⁰ He defines paternalism as any act where (1) "An agent A attempts to improve the welfare ... of agent B with regard to a particular decision or situation that B faces," and (2) "A's act is motivated by a *negative judgment* about B's ability ... to make the right decision or manage the particular situation in a way that will effectively advance B's welfare ..."³¹

Liberals sympathetic to Rawls think that the government ought to treat every citizen as a free and equal member in a fair system of social cooperation.³² Citizens are free in the sense that their identity is neither attached to, nor dependent on, any substantive commitment they might have,³³ they are "self-authenticating sources of valid claims,"³⁴ and they can take "responsibility for their ends."³⁵ Citizens are also assumed to have a fundamental interest in developing their two moral powers: the

²⁹Quong, *Liberalism without Perfection*.

³⁰Although Quong claims that paternalism is pro tanto wrong, he doesn't defend this claim. He argues only for paternalism's prima facie wrongness. *Liberalism without Perfection*, p. 100 n. 66. I take this to mean that the burden of proof falls on those who want to justify paternalism. This weaker claim is compatible with the view that paternalism, when properly justified, is in no way wrong.

³¹Ibid., p. 80 (emphasis in original).

³²Rawls, *Political Liberalism*.

³³Ibid., p. 30.

³⁴Ibid., p. 32.

³⁵Ibid., p. 36. This means that in virtue of their goals, it is fair that they bear some burdens and benefits.

capacity for an effective sense of justice, and the ability to “form, revise, and pursue a conception of the good.”³⁶ Citizens are equal in the sense that they are assumed to have the capacities for justice, and for a conception of the good (and the abilities to be a fully cooperating member of society) to a minimum degree. Equal moral status is given to all those who are above the threshold. In Rawls’s words, basic equality is a “range property.”³⁷ Quong believes that paternalism is objectionable because it is inconsistent with treating people as free and equal. Paternalism “involves one person or group denying that another person or group has the necessary capacity, in a given context, to exercise the second of the two moral powers.”³⁸ Nudging looks like a clear violation of moral status because its main motivation is that (in some situations) people make bad choices. Notice, though, that Quong does not need to deny the empirical claims in support of nudging. He can accept that in fact people are neither very effective, nor equally able to act according to their preferences and convictions. He only needs to assert that regardless of people’s actual abilities and differences between them, we have respect-based reasons to ignore them. We ought to treat people *as if* they had the same abilities and competencies.

Quong’s objection to paternalism can be read in two different ways. First, paternalism can be a comparative wrong: it involves one agent treating another as having less ability to exercise her second moral power than someone else. Imagine that Ana takes Karl’s mobile phone when they go for drinks at night because she thinks that after a few glasses, he will make some calls that he will later regret. But she keeps her phone with her because she thinks that she will not engage in “drink and dial.” Ana treats Karl paternalistically because she expresses a comparative negative judgment about Karl’s ability to resist making regrettable calls when drunk.

Second, Quong’s objection can also be read as involving no comparison: an agent might assert that an individual just lacks the required abilities to exercise her second moral power without claiming that others possess those abilities.³⁹ Imagine now that Ana leaves her own and Karl’s phone at home, because she thinks that both she and Karl will drink and dial. She does not assert superiority over Karl (she might even think that she is more likely to dial than him), but still shows non-comparative disrespect for Karl’s ability for self-control.

What can advocates of nudging reply to the disrespect objection?⁴⁰ It

³⁶Ibid., p. 19.

³⁷John Rawls, *A Theory of Justice* (Oxford: Oxford University Press, 1999), p. 444.

³⁸Quong, *Liberalism without Perfection*, p. 101.

³⁹Ibid.

⁴⁰To be fair to Quong, his arguments address paternalism in general and do not focus

seems that the comparative reading of the objection has more force than the second one. This is because the first reading is nonegalitarian. Regarding the comparative reading, notice, first, that the nudger need not assert that she is better in any respect than the paternalized. Her alleged superiority is fully explained by the (cognitive) situation in which she finds herself. She can justify her knowing better “because [she] recognize[s] that [the paternalized is] currently in a context or state where people (including [her]) happen to make frequent mistakes.”⁴¹ Everyone might be in the same position with respect to their inability to perfectly revise and pursue their goals. We can think of nudges as, in principle, appropriate for everyone. This reply avoids the comparative reading of the objection.

The noncomparative reading of Quong’s objection can be addressed by noting that the second moral power does not specify exactly the extent to which citizens are able to form, revise, and pursue a conception of the good. It says only that they have such a capacity. Since moral status (which is a function of having the two moral powers) is a range property, there are weighty reasons not to pay attention to individual differences in the ability to exercise the second moral power.⁴² But it also leaves open the question about how efficient we should assume citizens to be in exercising this power. If we assume that citizens are perfectly rational and efficient, any departure from that assumption would be disrespectful. But why should we make this assumption? Instead, we could assume that citizens are less than perfectly rational and less than perfectly efficient at pursuing their goals (while staying above the minimal threshold that triggers the range property). The judgment that citizens are less than fully rational would not be disrespectful, because it is consistent with the requirement of treating citizens as free and equals. It does not deny that they have this moral power; it only makes a less demanding assumption about what it means for citizens to be free.

Weak forms of paternalism such as nudging can also be defended by noting that the second moral power involves a set of different abilities.⁴³ It involves the capacity to set up ends and goals for oneself, to respond to and create reasons and values, and the ability to deliberate about and act

on nudge paternalism. Even if the latter can resist Quong’s objections, his arguments might succeed against harder and stronger forms of paternalism.

⁴¹Christian Coons and Michael Weber, “Introduction,” in Christian Coons and Michael Weber (eds.) *Paternalism: Theory and Practice* (Cambridge: Cambridge University Press, 2013), p. 13.

⁴²Ian Carter, “Respect and the Basis of Equality,” *Ethics* 121 (2011): 538-71.

⁴³By “weak paternalism,” I mean paternalism that targets only the means that people employ to achieve their ends. Gerald Dworkin, “Paternalism,” *The Stanford Encyclopedia of Philosophy* (Summer 2014 ed.), ed. Edward N. Zalta; <http://plato.stanford.edu/archives/sum2014/entries/paternalism/>.

upon the means for realizing those values and goals. Advocates of nudging do not claim that citizens lack this set of abilities; rather, they claim only that in particular contexts people are not very good at acting upon the means to realize their ends. Nudging targets primarily instrumental choices and preferences, rather than settled or noninstrumental preferences, values, and goals. For this reason, it does not involve a negative judgment on a particular agent; if anything, it only involves a negative judgment on a species-wide feature.⁴⁴ Insofar as the judgment is located at the level of human cognitive architecture, I do not see why it would affect self-respect.⁴⁵

4. The Manipulation Objection

The second objection to nudging is that it is manipulative. Critics object that even though not coercive, nudging is problematic because it aims to “deliberately circumvent people’s rational and deliberating faculties, and instead [it] seek[s] to influence their choice through knowledge of the biases to which they are susceptible.”⁴⁶ Indeed it is odd to care about coercion but not about manipulation. Liberals object to manipulation and coercion very much on the same grounds: they both interfere with people’s independence. Coercion makes us act in ways that do not fully reflect our will, but rather someone else’s. Manipulation, on the other hand, “perverts the way the person reaches decisions, forms preferences or adopts goals.”⁴⁷

In this section, I will distinguish between two forms that the manipulation objection might take. First, following Robert Noggle, manipulation consists of an agent moving someone’s “cognitive levers” away from their ideal settings.⁴⁸ I argue that if manipulation is understood in this way, many nudges are not manipulative. Second, I distinguish a related kind of cases that involve an agent taking advantage of someone’s cognitive shortcomings. I am unsure whether they involve manipulation and will call them instead “cognitive opportunism.” Although many nudges are cases of cognitive opportunism, not all are. Thus, I suggest that some forms of nudging are neither opportunistic nor manipulative. Then I move on and argue that one of the wrong-making features of both manipulation

⁴⁴This does not mean, of course, that everyone is susceptible to the same biases at the same level. These cognitive biases are only statistically significant.

⁴⁵Notice that this is an empirical claim that needs to be tested.

⁴⁶Grüne-Yanoff, “Old Wine in New Casks,” p. 236; see also Hausman and Welch, “Debate,” pp. 128-29.

⁴⁷Joseph Raz, *The Morality of Freedom* (Oxford: Blackwell, 1986), pp. 377-78.

⁴⁸Robert Noggle, “Manipulative Actions: A Conceptual and Moral Analysis,” *American Philosophical Quarterly* 33 (1996): 43-55.

and opportunism is that the victim of manipulation or opportunism is subjected to someone else's. The last claim I make is that insofar as nudges are easily resistible and are used to pursue legitimate goals, their use is permissible.

There is a vast literature of manipulation and there are deep disagreements about what it is, and what makes it wrong.⁴⁹ Coercion changes the structure of the choices we face. Manipulation, by contrast, changes the internal levers that guide deliberation. This thought is at the core of Noggle's analysis of manipulation. According to him, a manipulative action is one that aims to lead the victim away from ideal levels of deliberation.⁵⁰

This view has four important features. First, it depends on someone's being made to fall short of the ideals that govern deliberation.⁵¹ There are some ideal standards about the kind of mental states we should have. When a person makes it her aim that someone else deviate from these standards, she engages in manipulation. Second, it accounts for the fact that manipulation does not need to be hidden from the agent. Although in many cases the most efficient way to adjust someone's cognitive levers involves doing it behind her back, it is possible to do it openly. For instance, imagine that when one's child does not get an extra helping of cake, she, exaggerating her crying, expresses how happy some more cake will make her. Her father gives in because he struggles to resist her crying. At the same time, he is perfectly aware that he is being manipulated. The child manipulates her father in this example because his decision-making is impaired: it falls short of what it should be.

Third, there is a variety of ideal mental states that a victim can fall short of.⁵² One can manipulate another person by meddling with her beliefs by introducing either false or irrelevant ones. Similarly, one can manipulate desires. This can be done either by giving someone a desire to do something she has no reason to want, or by changing the intensity of a rational desire. Imagine that Rose has a desire to visit her ill mother. Toni can manipulate her by either reducing the strength of her desire or by increasing it. In both cases she is manipulated because the force of her desire is not ideal. A third element that can be adjusted through manipulation is emotions. One can be induced to emotionally react in ways that do not fit the

⁴⁹For an excellent collection that illustrates the difficulty of developing a theory of manipulation, see Christian Coons and Michael Weber (eds.), *Manipulation: Theory and Practice* (Cambridge: Cambridge University Press, 2014).

⁵⁰Noggle, "Manipulative Actions," p. 47.

⁵¹Notice that Noggle specifies the "ideal standards" as the standards the manipulator thinks should guide the manipulated deliberations (ibid., p. 48).

⁵²Notice that the influence should directly target those mental states. Making someone's deliberation worse by breaking her skull is not an act of manipulation. See Anne Barnhill, "What is Manipulation?" in Coons and Weber (eds.), *Manipulation*, chap. 2, p. 65.

situation. A person can be made to feel guilty in circumstances that do not merit guilt. In the same way as manipulating desires, the intensity of a fitting emotion can be distorted. Four, manipulation is not a binary notion. Acts can be more or less manipulative depending on how much the manipulated party departs from the ideal levels. This account of manipulation explains why there is a presumption against manipulation. By diverting an agent from ideal forms of deliberation, the manipulator undermines the agency of the manipulated party.⁵³ Notice that manipulation focuses on ideals of deliberation. This makes manipulation consistent with making people choose what it would be rational for them to choose.

An objection might be that the first feature is too strong. Why should manipulation consist in diverting from ideal forms of deliberation rather than from the actual levels? If an agent does not make us worse off than we are, why should this count as manipulation? This objection is plausible because nudging typically does not introduce any new biasing element; it merely exploits already existing biases. Perhaps these cases are not instances of manipulation. It is possible that they illustrate a different kind of wrong. Cognitive opportunism is objectionable because, in general, not worsening someone's condition is insufficient for the moral permissibility of an action. Consider a variation of Victor Tadros's claw hammer case.⁵⁴ Imagine that a child is drowning in a lake, and that only David can save the child. He could grab him out from the water at the expense of getting his favorite shirt wet. David, however, decides to save the child by jamming a claw hammer into the child's eye, grabbing him out from the water. David's action is clearly impermissible. This is true notwithstanding the fact that the child is not worse off (and actually better off). Morality demands much more than merely not worsening others' situations. For a case closer to the debate about nudging, consider a situation in which a salesman uses a person's extreme tiredness as an opportunity to sell her a product. Even though the salesman did not cause her tiredness, and did not make her cognitive situation worse, he takes advantage of her not-ideal cognitive situation and gets her to buy this product.

Let's turn now to the question of whether nudging is manipulative or opportunistic. Some nudges can be manipulative. Some framing might affect deliberation adversely. We might discourage people from eating potatoes by indicating that they contain L-ascorbic acid (the biochemical name for vitamin C).⁵⁵ This frame manipulates because it creates the

⁵³For the purposes of this article, it makes little difference whether manipulation is only *prima facie* or *pro tanto* objectionable.

⁵⁴Victor Tadros, *The Ends of Harm* (Oxford: Oxford University Press, 2011), p. 161.

⁵⁵Adrien Barton, "How Tobacco Health Warnings Can Foster Autonomy," *Public Health Ethics* 6 (2013): 207-19, p. 211.

false impression that there are some dangerous chemicals in potatoes. Other nudges can also manipulate by anchoring people to irrelevant information that will affect their further deliberation.

Some nudges are rather opportunistic. They do not lower the standards of deliberation, but exploit our cognitive shortcomings. Cases in which certain foods are placed at eye level do not make deliberation worse, they just exploit the fact that people tend to reach for objects that are at eye level. Default rules increase the likelihood of people choosing a particular option because the status quo is sticky. But, they do not create the stickiness.

Importantly, however, some nudges are neither. Think about nudges that make relevant reasons salient. This form of nudging does not lower our cognitive levers—quite the opposite: it aims to bring our attention to features that are worth noting. Its aim is to improve deliberation by pointing out reasons to be considered. The thought is that people are sufficiently responsive to reasons and it is possible to get them to attend to those reasons. For instance, some frames make important information more salient. A case that comes to mind is the list of ingredients in packaged food, where some ingredients (such as the amount of sugar or fat) are written in bold letters. This nudge increases the likelihood that people will refrain from consuming certain products (assuming that they want to reduce their calorie intake). It does so by making some ingredients more salient.

A more difficult case is one in which emotional reactions are enrolled to generate the correct reactions: using pictures of starving children to remind us of the prevalence of global poverty and improve our emotional reactions, for example. They are supposed to encourage people to respond to reasons and make appropriate judgments, not to divert their attention. They are not manipulative because they do not make us fall short of the ideal response; rather, they calibrate them in a way such that people can be sensitive to information. It is true that they do not do it by rational persuasion, but there is no reason to think that nonrational persuasion ought to be manipulative.⁵⁶ For the same reason, they are not opportunistic; they alleviate a disadvantage we have, rather than taking advantage of it.

Manipulation and opportunism need not be impermissible. The argument developed here only suggests that there is a presumption against them, rather than a prohibition. One way to justify manipulation consists in showing that the person being manipulated can still exert her will. In other words, a wrong-making feature of manipulation is that it subjects one person to the will of another. If this feature can be neutralized, then manipulation might be easier to justify. If the characterization of nudging defended in section 2 is correct, then it seems that this objection can be

⁵⁶Noggle, “Manipulative Actions,” p. 49.

rebutted. A condition for an influence to be a nudge, I argued, is that it is easily resistible. This means that a person can (virtually) effortlessly resist the pressure to perform the action favored by the nudge. Easily resistible nudges ensure that a person's actions reflect her will. Another way to justify manipulation and opportunism resides in showing that the victim has a duty to act in certain ways. Imagine that a child is drowning in a shallow pond and the only way to save the child is for Joe to jump in. However, he is reluctant to jump because he fears that the water is too cold. It is permissible, I would argue, to lie to Joe, telling him that the water is warm. I now turn to a liberal account of some of the duties that nudging can be helpful in fulfilling.

5. Anti-Perfectionism

In this section, I move from how nudging does whatever it does to what it is supposed to do. As I explained in previous sections, most critiques target the means that paternalism uses to benefit people. In this section, I explore whether nudging is consistent with anti-perfectionism. I argue that nudging can be legitimately employed to increase compliance with principles of justice and to enable people to act on their settled conceptions. The next four subsections (5.a-5.b.iii) do not address paternalism. They rely only on the view that it is permissible to get people to discharge the duties they owe to others, rather than in efforts to improve their situation.⁵⁷ It is not until the last subsection (5.b.iv) that I return to the problem of paternalism.

Perfectionism is the view that it is permissible for the government “to promote or discourage particular activities, ideas, or ways of life on grounds relating to their inherent or intrinsic value.”⁵⁸ I would argue, however, that there are strong reasons to abandon perfectionism.⁵⁹ The main motivation for its rejection is that the state is bound to treat citizens as free and equal. An important part of being treated as free and equal consists in being able to endorse the constraints that apply to us. For many liberals, it follows that coercive actions need to be justified publicly.⁶⁰ Furthermore, Rawls combines the duty of public justification with the observation that liberal societies are characterized by competitive, irreducible (reasonable) pluralism.⁶¹ Citizens disagree sharply about

⁵⁷I remain agnostic on whether justice is a requirement of living well. Cf. Ronald Dworkin, *Justice for Hedgehogs* (Cambridge, Mass.: Harvard University Press, 2011).

⁵⁸Quong, *Liberalism without Perfection*, p. 27.

⁵⁹I cannot explore this issue at length here. I will only sketch some of those reasons.

⁶⁰Rawls, *Political Liberalism*.

⁶¹*Ibid.*, p. xviii. The fact of reasonable pluralism is the natural outcome of the free

“comprehensive” doctrines: views about “what is of value in human life, and ideals of personal character,” as well as other metaphysical or religious conceptions.⁶² From the fact of pluralism, Rawls derives the principle of liberal legitimacy according to which “the exercise of political power is fully proper only when it is exercised in accordance with a constitution, the essentials of which all citizens as free and equal may reasonably be expected to endorse in the light of principles and ideals acceptable to their common human reason.”⁶³ The idea is that public justification can only be given to a “free-standing conception” of justice: one that is fully justified by the liberal core of ideas of society as a fair system of cooperation between free and equal citizens. The state’s role is limited to distributing resources in a manner that allows each person to pursue her (permissible) conception of the good. Appeals to comprehensive doctrines (or comprehensive values) fall short of public justification because they cannot be reasonably expected to be endorsed by all citizens as free and equal.⁶⁴ The core of anti-paternalism is then an ideal of political autonomy, according to which citizens must be able to endorse the actions and constraints imposed by those in authority.

Anti-perfectionists complain that improving people’s well-being might not be a legitimate goal for the state. At this point, anti-perfectionism need not distinguish between nudging, other forms of paternalism, and rational persuasion. Imagine that government promotes family oriented life on the (true) grounds that family life is intrinsically worthwhile. Anti-perfectionists would object that this justification of family life is a comprehensive value that many reasonable citizens reject.⁶⁵ Thus, the state has no legitimate reason to promote this value. So, the question is: When is nudging permissible within an anti-perfectionist framework? In the remainder of this section, I will explore five main spheres in which nudging might be acceptable.

5.a. Enforceable duties

The government has a duty to ensure a fair distribution of resources that allows each person to pursue her conception of the good. There is a variety

exercise of human reason.

⁶²Ibid., p. 13.

⁶³Ibid., p. 137.

⁶⁴I assume that anti-perfectionism applies to all coercive decisions of the state, and not only to “constitutional essentials and matters of basic justice.” For the latter position, see *ibid.*, pp. 227-30; for a defense of the former, see Quong, *Liberalism without Perfection*, pp. 273-89.

⁶⁵Of course, demographic changes can provide nonperfectionist reasons for encouraging family life and procreation.

of strategies that the government can use to increase compliance, ranging from persuasion and nudging to threats of fines and imprisonment.

There are two kinds of reasons that make nudging preferable to coercion. On the one hand, nudges may be effective at increasing compliance. For instance, the U.K. government has produced a guide that aims to reduce tax evasion. The guide provides a variety of nudges that are effective at reducing fraud. It suggests that tax declaration forms should be easy to fill in, that the sense of honesty can be increased by forcing people to sign the declaration before they enter their details, and that people tend to comply with rules when they are aware that most people also comply with them.⁶⁶ These measures might be more effective in fighting against fraud than increased penalties (although they cannot replace them).

The benefits of nudging might go beyond efficacy. They also have a moral component: in general people welcome the opportunity to act in ways that enhance their sense of agency. Coercion tends to diminish this sense by restricting the options available, even if restricting those options is what justice requires. Another desideratum in favor of nudging is that choice plays a role in moral education. Discharging one's moral duties freely has some pedagogical benefits. Indeed, it is one of the ways in which people exercise their sense of justice. This idea can be better appreciated by pointing out that many of our duties are, at least partly, "action reasons."⁶⁷ We have an interest in living under conditions of justice. This claim is underdetermined, since it does not specify who should bring about such a state of affairs. However, it seems plausible that each of us has a reason to act in ways that bring about justice, which goes beyond the outcome of our actions. The fundamental interest in exercising and acting upon a sense of justice goes beyond the mere interest we have in inhabiting a just society. Nudging seems to score better than coercive threats in securing that people exercise their moral powers.

5.b. Nonenforceable duties

In contrast to the previous example, it may sometimes be undesirable for the government to attempt to enforce certain duties or political values. This might be due to different reasons. Some of them are related to feasibility problems. At other times, attempting to enforce a duty might violate the duty (or the value) it aims to enforce. Another reason for non-enforceability may be that enforcing one duty involves the violation of a

⁶⁶Cabinet Office Behavioural Insights Team (2012), "Applying Behavioural Insights to Reduce Fraud, Error and Debt"; https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/60539/BIT_FraudErrorDebt_accessible.pdf (accessed 17 June 2015).

⁶⁷Raz, *The Morality of Freedom*, pp. 145-46.

different duty (or the value that a duty is supposed to serve), which overrides the first duty.⁶⁸ Nudging allows governments to circumvent the problems associated with nonenforceability by making it more likely that citizens freely comply with those duties. The following examples illustrate this possibility.

5.b.i. Public goods: First, some duties are conditional upon other people performing certain actions. The production of some public goods is an example. We are under a duty to perform an action once others have contributed sufficiently to the production of a good x , but we have no duty to perform the action when others have not made their contribution. Because everyone is equally situated, it is difficult to get people to start producing the good. Nudging can help to increase assurance that the public good will be secured. A simple example might illuminate this thought. Imagine there is an “honesty box” close to a coffee machine. People are asked to make a small contribution towards the provision of coffee every time they make one. Ex hypothesi, a person is required to contribute only if others do, since she is not required to subsidize coffee for others.⁶⁹ The provision of coffee requires the assurance that free-riding is not widespread. There is a way to nudge people into being more cooperative: a banner picturing eyes close to the box seems to significantly increase cooperation.⁷⁰ Another example, the preservation of green areas and the environment above a certain level, might also be a conditional duty that can be nudged. Framing the efficiency of cars’ fuel consumption in “gallons per mile” instead of “miles per gallon” makes the cost of fuel more salient, nudging people into choosing greener cars.⁷¹ Stickers on cars with different colors (say ranging from green to red) might also nudge people into choosing more efficient cars.

5.b.ii. Injunctions: Nudging may play an important role in increasing compliance with a second class of nonenforceable duties. In this analysis, I follow Shelly Kagan.⁷² Rights are complex bundles of claims and liberties. In typical cases, persons have full rights to perform an action X , which involves both a liberty to X , and an injunction against others’ in-

⁶⁸Cf. Cécile Fabre, “Distributive Justice and Freedom: Cohen on Money and Labour,” *Utilitas* 22 (2010): 393-412, p. 400.

⁶⁹If no one cooperates, there will be no coffee for anyone, but morality does not demand the provision of coffee.

⁷⁰Melissa Bateson, Daniel Nettle, and Gilbert Roberts, “Cues of Being Watched Enhance Cooperation in a Real-World Setting,” *Biology Letters* 2 (2006): 412-14.

⁷¹T&S, *Nudge*, p. 204. Some of T&S’s examples are not proper nudges, because they involve taxes and penalties on fuel consumption or cap-and-trade schemes (see p. 196).

⁷²Shelly Kagan, *The Limits of Morality* (Oxford: Clarendon Press, 1989), pp. 219-30.

terfering with their X-ing.⁷³ That I have a right to my book means that I have a liberty to use my book, and a claim that others don't prevent me from using my book. Not all rights are like this, however. In some instances, a person might enjoy only an injunction against interference without having the liberty to perform an action. The duty to do X (or to refrain from X-ing) is unenforceable in virtue of the injunction. Nudging can increase the degree to which people fulfill their duties without violating the injunction.

Contributing to poverty alleviation campaigns seems a plausible example. Besides the enforceable duty to pay taxes, we also have a duty to fight poverty. One way to discharge this duty is by donating to agencies such as Oxfam. It is permissible, I believe, to nudge people into donating even if it would be wrong to force them to do so. This is because we have no liberty-right to keep all of our post-tax income given the level of poverty we witness in the world. Crucially, this is compatible with having a claim-right that others do not take our money and donate it. There are a variety of nudges that governments can employ to reduce poverty. Agencies can use emotional pictures of destitute children to raise money. They can exploit the status quo bias by suggesting a higher donation, while leaving open the possibility of donating less than the suggested amount. They might use schemes such as "give more tomorrow," in which one commits to an increment to the amount of money one gives to charity every year.⁷⁴

Similarly, one can increase the consumption of fair trade goods by making them the default option in cafeterias. Another area in which we might have a duty to donate is that of blood and organ donation. Many people think that the compulsory confiscation of blood and cadaveric organs is wrong. People might enjoy an injunction against confiscation for a variety of reasons that range from religious beliefs to intense fear of syringes. Again, this is consistent with a moral obligation to donate. So-called presumed consent schemes (with an opt-out clause) will help to increase the availability of organs for transplant.⁷⁵

5.b.iii. Nonpublic duties: The last examples of unenforceable duties I will discuss involve those in which regulation is unfeasible because of lack of transparency (or publicity). The main example I have in mind concerns attempts to neutralize implicit biases. Research in the last twenty years suggests that sexist and racist biases are widespread in our societies.⁷⁶ A

⁷³I use "claim" and "injunction" interchangeably.

⁷⁴T&S, *Nudge*, pp. 227-28.

⁷⁵*Ibid.*, pp. 184-92.

⁷⁶Mahzarin Banaji and Anthony Greenwald, *Blind Spot* (New York: Delacorte Press, 2013).

person has an implicit bias when she “has an automatic cognitive or affective association between [some stigmatized group] and some negative property ... or stereotypic trait ..., which are accessible and can be operative in influencing judgment and behavior without [her] conscious awareness.”⁷⁷ An important feature of implicit biases is that they are also harbored by individuals who reject the content of the stereotype. For instance, a common phenomenon is that of aversive racism: individuals who explicitly endorse egalitarian beliefs, but who display negative attitudes and feelings towards blacks.⁷⁸ The unconscious nature of implicit biases does not cancel the duty that individuals have not to behave in a racist manner, although it can mitigate the burdens that people bear or the degree to which they are to blame when they display these biases.⁷⁹ At any rate, given that we assume that citizens have a fundamental interest in exercising a sense of justice, they are also motivated to reduce their (implicitly) racist reactions.

The question, then, is what measures can be taken to neutralize biases. One option is to leave it up to individuals to neutralize them. This strategy is likely to fail. In order to correct for biases, several steps need to be taken.⁸⁰ First, many of us are not even aware that we are biased. People tend to underestimate how susceptible they are to harboring these influences.⁸¹ Second, when a person is aware of the bias, she needs to be motivated to neutralize the bias. Third, even if one is sufficiently motivated, one needs to know when the biased response might occur, and its direction and magnitude. Fourth, neutralizing biases requires that one is able to exert sufficient control over one’s responses. Some studies suggest that trying to exert direct control over suppressing implicit associations might create a “rebound effect” that increases biased responses.⁸²

⁷⁷Jules Holroyd, “Responsibility for Implicit Biases,” *Journal of Social Philosophy* 43 (2012): 274-306, p. 275.

⁷⁸Adam R. Pearson, John F. Dovidio, and Samuel L. Gaertner, “The Nature of Contemporary Prejudice: Insights from Aversive Racism,” *Social and Personality Psychology Compass* 3 (2009): 314-38. This does not mean that biases are equally distributed in society. People who intrinsically value racial equality tend to display less bias than those who are either uncommitted to racial equality, or those whose commitment to equality is motivated by social norms. Holroyd, “Responsibility for Implicit Biases,” pp. 280-82. It is possible that some people do not have biases, but unfortunately most of us do.

⁷⁹I remain agnostic about whether we also have a duty not to have biases even if those biases do not affect our attitudes and behavior.

⁸⁰Timothy Wilson and Nancy Brekke, “Mental Contamination and Mental Correction: Unwanted Influences on Judgments and Evaluations,” *Psychological Bulletin* 116 (1994): 117-42.

⁸¹John Bargh and Tanya Chartrand, “The Unbearable Automaticity of Being,” *American Psychologist* 54 (1999): 462-79.

⁸²Holroyd, “Responsibility for Implicit Biases,” p. 283.

Because direct control strategies are likely to fail, individuals might take indirect measures to reduce their biased responses. Nudging can help at each of these stages. One possibility is to arrange social environments in a way that offers counter-stereotypes. For instance, the government can promote associations between positive traits and stereotyped groups: in some places, there is a widespread implicit association between violence and blacks.⁸³ The government can redress this association by offering positive associations with black people. In the long run, this might help to dissociate the negative link between blacks and violence.

Similarly, there is some evidence that people can partly reduce their biased reactions when goals that include equal treatment are activated.⁸⁴ It is difficult for individuals to keep egalitarian goals activated while doing other tasks, because their cognitive resources are limited. However, environmental cues can work as a nudge by activating these goals. These strategies can help people to comply with their duty not to be racially and sexually biased.

5.b.iv. Autonomy: The last example is the most controversial of the list. It brings back questions about paternalism that I had abandoned in the preceding paragraphs. I intend to defend weak nudge paternalism. This is, I believe, compatible with the account of anti-perfectionism sketched above. At first sight, it might seem that paternalistic policies such as seatbelt regulations or anti-tobacco campaigns are incompatible with anti-perfectionism because they tend to appeal to the intrinsic values of the goals they promote, in this case the value of leading a healthy life.⁸⁵ In this section, I will explore a different reply.

In *A Theory of Justice*, Rawls argues that in the original position it is rational for the parties “to protect themselves against their own irrational inclinations by consenting to a scheme of penalties that may give them a sufficient motive to avoid foolish actions and by accepting certain impositions designed to undo the unfortunate consequences of their imprudent behavior.”⁸⁶ There are two ways to understand when the parties might adopt paternalistic policies. First, the parties have an interest in ensuring that their choices satisfy their settled preferences, goals, and values. They will have a motivation to avoid silly choices that go against this interest. Seatbelt regulations are a good example. Most of us do want to reduce

⁸³Patricia Devine, “Stereotypes and Prejudice: Their Automatic and Controlled Components,” *Journal of Personality and Social Psychology* 56 (1989): 5-18.

⁸⁴Thomas Webb, Paschal Sheeran, and John Pepper, “Gaining Control over Responses to Implicit Attitude Tests: Implementation Intentions Engender Fast Responses on Attitude-Incongruent Trials,” *British Journal of Social Psychology* 51 (2010): 13-32.

⁸⁵I will set aside issues involving passive smoking.

⁸⁶Rawls, *A Theory of Justice*, p. 219.

the risks of dying in a car accident. Failure to fasten a seatbelt increases the risk of death (or severe injury), and (for most of us) the value of fastening the seatbelt is purely instrumental. Thus, it seems reasonable for the parties in the original position to accept seatbelt regulations. The second way to understand Rawls's remarks appeals to a condition of autonomy inspired by Harry Frankfurt.⁸⁷ On this view, a condition for being an autonomous person is that there is sufficient identification between an agent's first- and second-order desires. Phenomena such as weakness of the will and addiction are classic examples: I have a desire to have another helping of ice-cream, and I also have a (second-order) desire to keep my desires for ice-cream under control. Paternalist measures can be autonomy-friendly by making it more likely that the hierarchy of desires obtains.⁸⁸

There is, however, a problem with this reply. Many paternalist policies cannot be tailored individually, and they might be objected to by agents who have different preferences and convictions.⁸⁹ Use the case of smoking. First, it is certainly possible for a reasonable person to endorse a conception of the good that involves smoking, regardless of its costs. Imagine Robin, a hedonist who thinks that the pleasure of smoking outweighs its costs. When later in life he gets cancer, he does not regret his smoking. Rawlsian-style paternalism cannot be justified to Robin since his autonomy is not compromised.⁹⁰ Second, there might be a smoker who desires not to smoke, but whose conception of the good gives constitutive value to the ability to face and resist temptation even if some of her choices are not maximally conducive to her ends.⁹¹ Betty wishes she did not smoke, but values the opportunity to struggle against the craving for a cigarette. Even if most of the time her resistance is futile, her conception of the good is reasonable and needs to be accommodated by anti-perfectionism.

These two cases present a serious challenge to Rawls's argument. But notice that they only address some forms of coercive paternalism. Nudg-

⁸⁷Harry G. Frankfurt, "Freedom of the Will and the Concept of a Person," *The Journal of Philosophy* 68 (1971): 5-20.

⁸⁸For a developed account of autonomy, see Andrés Moles, "Autonomy, Free Speech and Automatic Behaviour," *Res Publica* 13 (2007): 53-75; see also John Christman, "Procedural Autonomy and Liberal Legitimacy," in James Stacey Taylor (ed.), *Personal Autonomy* (Cambridge: Cambridge University Press, 2005), pp. 277-98.

⁸⁹In some cases, it might be possible to arrange individualized nudges. The account defended here is compatible with approving public funds to allow individuals to set up nudges that diminish their inability to act on their conception of the good. See also Sunstein, *Why Nudge?* pp. 99-100.

⁹⁰Even if he was physically addicted to tobacco, his autonomy would be unaffected because he does not regret his addiction. The conception of autonomy I assume here does not require an ability to do otherwise.

⁹¹William Glod, "Political Liberalism, Basic Liberties, and Legal Paternalism," *The Southern Journal of Philosophy* 48 (2010): 177-96.

ing might escape this challenge, because the costs of resisting the nudge imposed on Betty and Robin are low.⁹² Parties in the original position would want to insure themselves against irrational choices, but would also want to make sure that it is easy to avoid the insurance scheme if it does not accommodate their conception of the good.

A possible objection to this argument is that it smuggles a conception of the good that is incompatible with anti-perfectionism. One possibility is that the example succeeds, but only because smoking is addictive, and it is in our interests to be free from addiction. This line is inadequate, since Robin can be both addicted and autonomous. The argument is based on an account of anti-perfectionism that aims to accommodate (rather than disregard) people's reasonable conceptions of the good.⁹³ The goal is to find an arrangement that makes it easier for people to realize their permissible conceptions of the good, independently of their intrinsic merits.

At this stage, it is useful to differentiate between two different accounts of neutrality that anti-perfectionists can endorse. Some of them might think that the state ought to be neutral about the content of different conceptions of the good and also about the relationship between persons and their conception of the good. Defenders of this view might think that the promotion of personal autonomy lies beyond the reach of legitimate governments. A more restrictive account of neutrality requires that the state refrain from promoting some conception of the good, but affirms that there are weighty reasons to promote a particular relationship between persons and their conception of the good. For them, it is acceptable that the government promotes personal autonomy. This is still compatible with anti-perfectionism, since it rejects the perfectionist principle that "there is no fundamental principled inhibition on governments acting for any valid moral reason."⁹⁴ The position I describe here holds that the value of political autonomy is a side-constraint on the actions of public authorities, but insists that it is within their legitimate reach to promote personal autonomy.⁹⁵

Imagine a case analogous to the nudge against smoking. Imagine that a certain group of Catholics believe that they ought to refrain from premarital sex. However, given the weakness of the flesh, few of them succeed. Imagine that the government could devise a nudge that was effective with Catholics, but easily resistible for everyone else. I do not see

⁹²Some forms of coercive paternalism are also easy and cheap to avoid, such as small taxes. The main issue here is easy resistibility, rather than avoiding coercion.

⁹³Matthew Clayton, "Liberal Equality and Ethics," *Ethics* 113 (2002): 8-22, pp. 11-12.

⁹⁴Joseph Raz, "Facing Up: A Reply," *Southern California Law Review* 62 (1989): 1153-236, p. 1230.

⁹⁵See also Matthew Clayton, *Justice and Legitimacy in Upbringing* (Oxford: Oxford University Press, 2006).

why it would be wrong for the government to help them to act in line with their deepest convictions. Notice also that nudging is compatible with neutrality about autonomy: presumably those who do not value autonomy at all might be indifferent about being nudged.

One of the implications of this argument is that nudging has a narrower scope than the libertarian paternalism advocated by T&S. In my view, nudging is appropriate only to secure that people discharge their duties, or to secure that the interest they have in exercising their fundamental moral powers is fulfilled.

6. Concluding Remarks

While most critics of nudge paternalism argue that it is problematic for the manner in which it operates, I have criticized it mostly for its ends. I have also defended a version of nudging. I argued that nudges are not disrespectful, nor need be objectionably manipulative or opportunistic. They can be used, I suggested, when they promote legitimate goals. These goals include, first, making coercive institutions more effective and bringing additional benefits. Second, nudging can help to foster some conceptions of autonomy by making it more likely that people act on their long term, settled preferences, or by helping them to achieve identification between their lower- and higher-order desires. Although these conceptions of the good are controversial, the negligible costs of escaping nudging manage to accommodate some difficult cases. Third, nudging can play a more central role in securing that people comply with the unenforceable duties they are under. Although my account has a more limited scope than T&S's, it fits better with anti-perfectionist liberalism.⁹⁶

*Department of Philosophy and Department of Political Science,
Central European University
molesa@ceu.edu*

⁹⁶I am grateful to János Kis, Zoltán Miklosi, Tom Parr, Merten Reglitz, Simon Rippon, and Adam Slavny for illuminating comments and questions. Thanks to Eszter Timár for stylistic suggestions. Earlier versions of this paper were presented at MANCEPT, Central European University Political Science Departmental Seminar, CEU Summer University course on applied philosophy, the National Autonomous University of Mexico (UNAM), and the University of Rijeka. I am especially grateful for written comments from Matthew Clayton, Christophe Heintz, Danny Scoccia, Kalle Grill, and two anonymous referees for this journal.