

The Moral Permissibility of Nudges

Winner of the Gerritt and Edith Schipper Undergraduate Award for Outstanding Undergraduate Paper at the 63rd Annual Meeting of the Florida Philosophical Association

Valerie Joly Chock, *University of North Florida*

Advances in cognitive and behavioral science reveal that the way options are presented, what is referred to as ‘choice architecture’, strongly influences our decisions. We tend to react to a particular option differently depending on how it is presented to us. These discoveries inspired *nudging*, the idea that people’s decisions and behaviors should be influenced in predictable, non-coercive ways by making small changes to the choice architecture. Central to the debate on nudging is the question of whether it is morally permissible to intentionally nudge other people. Libertarian paternalists maintain that this can be the case.

In this paper, I differentiate between type-1 nudges and type-2 nudges according to the thinking processes involved in each. With this distinction in hand, I present the libertarian paternalistic criteria for the moral permissibility of intentional nudges. Having done this, I motivate an objection to type-1 nudges. According to this objection, type-1 nudges do not appear to be relevantly different than standard cases of manipulation, and manipulation is morally problematic. While I show that this objection fails, I argue that its evaluation raises a different challenge for Libertarian Paternalism. The libertarian paternalistic criteria fails because it ignores the moral distinction that exists between different kinds of nudges. That is, the distinction between what I call ‘counteractive’ and ‘non-counteractive’ nudges. I end by suggesting a revision of the criteria that avoids the problem.

NUDGING

Richard Thaler and Cass Sunstein define ‘nudge’ as “any aspect of the choice architecture that alters people’s behavior in a predictable way without forbidding any options or significantly changing their economic incentives.”¹² ‘Choice architect’ refers to a person responsible for presenting options.³ Their

decisions about how options are presented affect the decisions others make. This is supported by research that suggests that we commonly make irrational decisions due to systematic errors in how we reason.⁴ These errors result from heuristics, which are rules-of-thumb people use to make decisions. Heuristics, although sometimes accurate and useful, often lead to “severe and systematic errors,”⁵ called ‘cognitive biases’ that result when people choose what the heuristics suggest even when these are poor options.⁶ Thus, even when we make “good” decisions (outcome-wise), those decisions are still irrational (in some sense) when they are the result of biased decision-making.

Kahneman (2011) describes two thinking systems, System 1 and System 2, which Thaler and Sunstein refer to as ‘Automatic System’ and ‘Reflective System’ respectively. System 2 thinking is deliberate, controlled, effortful, and conscious. System 1 thinking is rapid, uncontrolled, effortless, and subconscious.⁷ Based on this distinction, it is possible to differentiate between two types of nudges: type-1 and type-2 nudges. The latter engages people reflectively—they engage System 2—whereas the former does not—they engage only System 1. Hansen and Jespersen (2013) offer the plate-size nudge as a type-1 nudge example.⁸ It consists of reducing the plates’ size in cafeterias, which typically reduces consumers’ food intake, and typically leads to a calorie intake reduction as well. This usually happens without engaging the subject’s reflective thinking because of biases that result in *mindless eating*.⁹ The typical behavior is to first fill the plate with food and then eat the food typically without reflecting about it.¹⁰ A type-2 nudge example with the same goal—reducing calorie intake—is placing nutritional labels besides the food along with placards encouraging low-calorie food consumption. Here, the nudge works by engaging diners’ reflective thinking when they read the information and reflect on their food choices (which typically results in diners consciously choosing lower-calorie foods or reducing the food amount they put on their plate). These examples illustrate the difference between type-1 and type-2 nudges. That is, type-1 nudges engage people subconsciously—they “work in the dark”¹¹—whereas type-2 nudges engage people reflectively.

LIBERTARIAN PATERNALISM

Libertarian Paternalism concerns how options should be presented. It is motivated, in part, by the unavoidability of choice architecture. When options are presented, they must be presented in some way. Further, any way of presenting the options inevitably influences people’s decisions. This, coupled with the fact that people often make irrational decisions, according to liber-

tarian paternalists, can make intentional nudges morally permissible. Since nudging will inevitably occur, choice architects should nudge well.

However, an intentional intervention is not morally permissible merely in virtue of being a nudge. Libertarian paternalists propose a criteria for the moral permissibility of intentional interventions (hereafter ‘LPC’), which gives a set of sufficient conditions for the moral permissibility of an intentional intervention. According to LPC, any intervention that is (i) a nudge, (ii) transparent, and (iii) aimed toward the welfare of those being nudged is morally permissible.¹² Here (i) requires that the intervention must preserve freedom of choice—it cannot forbid or attach incentives or consequences to options¹³; (ii) requires that the influence must be easy to resist—people must be able to choose a different available option¹⁴; and (iii) requires that the intervention must be justifiably intended to make people better off—it must have people’s best interest in mind. We can thus summarize LPC as the claim that intentional interventions are morally permissible if they preserve freedom of choice by means of transparent and welfare-aimed nudges.¹⁵

To see LPC applied to a case, consider the following example:

Cafeteria

Jess is a cafeteria manager who decides the way in which food is displayed. After reading some psychology studies, she learns that the food order at cafeteria lines substantially determines what ends on diners’ plates—the majority of people tend to select the food that is placed first in line.¹⁶ Aware of this, Jess now knows that whatever she decides regarding food placement will influence her diners’ choices. So, she asks herself: “How *should* I arrange the food?”

Jess could arrange the food: (a) So that diners are better off;¹⁷ (b) at random; (c) to maximize her profits; (d) by banning all unhealthy foods.

Option (d) limits the available options, so it fails to meet (i) of LPC. Thus, in this case, option (d) is not a permissible intervention under LPC. The remaining options (a–c), all meet condition (ii) of LPC because they do not make it significantly more difficult for diners to choose other available food. However, not all (a–c) satisfy (iii). Option (b) fails to satisfy (iii) because to arrange the food at random is to consciously ignore the diners’ best interests. Option (c) fails to satisfy (iii) because it either reflects the architect’s selfish purposes or fails to even consider the diners’ welfare. Therefore, from Jess’ set of options, (a) is the only option justified by LPC—and, thus, (a) is a

morally permissible option.

Although nudges have been used to positively influence behavior in various settings¹⁸, they have been heavily criticized. A common objection targets their moral permissibility by appeal to manipulation.¹⁹ This objection claims that although nudging can be morally permissible, there is a particular subset of nudges that is impermissible in virtue of being manipulative.

THE MANIPULATION OBJECTION

The charge from manipulation is that type-1 nudges are manipulative in a way that type-2 nudges are not, and that this makes a moral difference.²⁰ The objection claims that a nudge is manipulative if it influences people without sufficiently engaging their reflective and deliberative capacities.^{21,22} The charge from manipulation can be presented as follows:

The Manipulation Objection (TMO)

1. All interventions that engage people *only* subconsciously are manipulative.
2. Type-1 nudges engage people only subconsciously.
3. Type-1 nudges are manipulative. (1–2)
4. All manipulative nudges are morally impermissible.
5. Therefore, type-1 nudges are morally impermissible. (3–4)

Note that TMO makes no claims about the moral permissibility of type-2 nudges. This is motivated by the desire to keep some nudges while excluding others. The desire to leave type-2 nudges unscathed is itself motivated by the assumption that type-2 nudges are not manipulative (or at least not manipulative in the relevant sense), and thus not morally problematic, because they engage people consciously.

Subliminal messages²³ engage people only subconsciously.²⁴ Without getting into empirical details, suppose that subliminal messages are capable of influencing people's choices. Suppose also that, as part of a public health campaign, subliminal messages encouraging calorie intake reduction are introduced in TV programs. As a result, many people subconsciously reduce their calorie intake. However, they are unaware that they are doing so, largely because they are unaware of the subliminal message itself. Even though the outcome may be desirable, the intuition is that such a use of subliminal messages is manipulative. This intuition gives reasons to accept premise 1.

Premise 2 is true by definition and premise 3 follows from 1 and 2.

Premise 4 can be motivated by appeal to autonomy. Hausman and Welch (2010) define ‘autonomy’ as “the control an individual has over his or her own evaluations and choices.”²⁵ They argue that when nudges do not involve “rational persuasion,” the subject’s autonomy is diminished.²⁶ Here the degree of reflection that takes place in an individual’s decision-making is proportional to the amount of autonomy they exercise. Since manipulative nudges engage individuals only on subconscious levels, the degree to which they exercise their autonomy, if at all, is very low. The idea is that the choice architect potentially has more control over the individuals’ choices than the individuals themselves. It is such a diminishing of autonomy what makes type-1 nudges morally impermissible according to TMO.²⁷

LPC makes no distinction regarding which System (1 or 2) is engaged in nudging.²⁸ Thus, under LPC type-1 nudges and type-2 nudges are equally permissible. If TMO succeeds and all type-1 nudges are morally impermissible, then the LPC’s sufficiency claim is false because LPC fails to account for the moral difference between type-1 nudges and type-2 nudges.

IMPLICATIONS

TMO makes a moral distinction between type-1 nudges and type-2 nudges. This distinction is motivated by the desire to leave type-2 nudges unscathed, which is problematic. Such project rests on the assumption that type-2 nudges exist independently from type-1 nudges, but this is simply not the case. While type-1 nudges and type-2 nudges are exclusive—a nudge is either of the type-1 or type-2 variety—they are importantly related. It is impossible to cleanly separate them because while System 1 can—and often times does—operate on its own, System 2 always depends on System 1. For instance, in the nutritional labels type-2 nudge example, reading and reflection only come into the picture after the subject notices and reacts to the placards. However, without the initial reaction (System 1), there would be no reflection (System 2). There is simply no way to skip straight to engaging the subject’s System 2 thinking.

Given that the distinction between these types of nudges is not as clear as it is assumed by proponents of TMO, if the objection is successful, it not only undermines the moral permissibility of type-1 nudges but the moral permissibility of type-2 nudges as well. In other words, if type-1 nudges are morally impermissible, and type-2 nudges cannot exist without type-1 nudges, then there would be no permissible nudges at all. By arguing that type-1 nudges are impermissible, proponents of TMO inadvertently make the case

that all nudges are morally impermissible. This is a radical result. If TMO succeeds, many everyday interactions would be morally impermissible.

To appreciate the radical nature of TMO's conclusion, think about the prevalence of nudges in our everyday lives. People dress well for job interviews which nudges employers to take them seriously, students raise their hands which nudges professors to call on them, people make recommendations to friends which nudges them to make particular choices, etc.

All these 'everyday nudges' would be impermissible if we accept TMO's conclusion. To accept TMO's conclusion is to accept that we do something morally wrong much of the time. This is unreasonable. TMO proves too much. Therefore, its conclusion should be rejected.²⁹

This way of rejecting the objection relies on the absurdity that results from its conclusion. This rejoinder claims that something goes wrong in the argument for TMO, but it fails to identify what exactly that is. Thus, it is not a satisfying solution. Something else should be added to complement this rejoinder. Something that can help us point out what goes wrong.

EVALUATION

Rejecting that all type-1 nudges are impermissible is not the same as accepting that all type-1 nudges are permissible. There is an important difference between type-1 nudges like the plate-size nudge and interventions like subliminal messages that creates a moral distinction between them. In what follows, I address Thaler and Sunstein's attempt to make such a distinction.

1. Monitoring

Thaler and Sunstein condemn subliminal messages because it is impossible to monitor them.^{30,31} They justify this by appeal to LPC's transparency condition. With this understanding of transparency, they implicitly introduce unreasonable expectations about people's capacity to monitor nudges. On their account, for a nudge to be transparent, it must be practically monitorable; and for it to be practically monitorable, the nudge must be practically detectable. This brings about a problem. In the plate-size example, the nudge is not practically monitorable.³² Thus, if subliminal messages are impermissible because they are not practically monitorable, then nudges like the plate-size nudge and many 'everyday nudges' are impermissible too. Here, we can see that the transparency condition is extremely restrictive. This is sufficient

reason to not accept how practically monitorable a nudge is as a good divider between permissible and impermissible type-1 nudges. Thaler and Sunstein's solution is a possible one but it overly restricts the set of morally permissible nudges. Thus, their rejoinder to TMO is not satisfying.³³

2. Counteracting

A more promising response makes a distinction between the *ways* in which interventions engage our biases. Subliminal messages are morally problematic because they exploit our biases in a way that other type-1 nudges do not—they decrease our sensibility to reasons. Biases prevent our rationality status from being optimal. Nudges are intended to *counteract* already operant biases that prevent us from making rational choices. Subliminal messages do not *counteract* biases. Rather, they merely *activate* them. Thus, in a sense, while some nudges (“counteractive”) elevate our rationality status by counteracting our biases and not affecting our sensibility to reasons, other nudges (“non-counteractive”) lower it even more under its already suboptimal level by activating additional biases on top of the already operant ones and decreasing our sensibility to reasons.

Suppose Jess wants to influence diners to consume healthy desserts (e.g. fruit) over less healthy options (e.g. cupcakes/donuts) because doing so will improve their welfare. She knows diners are biased to fill their plates with what is presented first. And she now learns about subliminal messages. Thus, she considers two options to nudge her customers: (i) placing healthier desserts first in line or (ii) placing subliminal messages throughout the cafeteria. The former is a counteractive nudge. This nudge relies on changing the environment to prevent people's already operant bias (tendency to choose what is placed first in line) from leading them towards poor choices (unhealthy over healthy food). This nudge utilizes the same already operant bias to influence people to make better choices not by eliminating the bias but by counteracting it with a change in the environment (arranging the food so that healthy foods are first). Further, counteractive nudges do not decrease people's sensibility to reasons because, although people might not recognize how the choice architecture affects their choices about what they eat (or even recognize that this is doing so at all), they can at least recognize the choice architecture itself. They recognize that food is placed in some way and that they have options to choose from, even if that all happens at System 1 level without reflection. The idea here is that, if people started to think about and list all the possible reasons they could have for choosing the food they put on their plate, they

could at some point list the placement of the food as a reason. What matters here is not that they *actually* list the food placement as a reason for choosing the food they put on their plate but that there is a *potential* for them to do so.

Subliminal messages do not rely on utilizing already operant biases or changing the environment but rather on activating *additional* biases (tendency to be susceptible to subconscious auditory/visual message). These non-counteractive nudges decrease people's sensibility to reasons because people are unable to recognize the subliminal message within the choice architecture. Thus, the potential to list the subliminal message as a reason for choosing the food is non-existent. It is not only that they cannot *actually* list the subliminal message as a reason but that there is no *potential* for them to do so. Given this distinction, it is possible to offer an account of permissible nudges based on how nudges work. A nudge is *permissible* when it does not decrease our sensibility to reasons and *impermissible* when it does.³⁴ This distinction is compatible with the intuition that motivates TMO. That is, that the relationship between autonomy and reflection plays a role in the moral permissibility of nudges. The difference is that, for proponents of TMO, *actual* reflection about and determination of the reasons for our choices is necessary for exercising one's autonomy, and thus for a nudge to be permissible. In my view, the *potential* for such reflection and determination is sufficient for exercising one's autonomy, and thus for a nudge to be permissible.

Whether this rejoinder defeats premise 1 (All interventions that engage people only subconsciously are manipulative) or premise 4 (All manipulative nudges are morally impermissible) of TMO depends upon how we understand manipulation. Central to the manipulation literature, is the question of whether manipulation is inherently impermissible. This is a difficult question with no clear answer. But independently of what answer one adopts, my rejoinder does enough to prove that TMO is unsuccessful by showing that there is no moral problem with counteractive nudges even when they are type-1 nudges. My rejoinder can be a response to TMO for both those who take manipulation as inherently impermissible and those who do not. On the one hand, my rejoinder allows for counteractive nudges to be manipulative but permissible. The possible position here is that counteractive nudges are manipulative because they engage individuals subconsciously, but they are morally permissible because they do not decrease one's sensibility to reasons. Thus, they do not prevent one from exercising one's autonomy. If manipulation is not inherently impermissible, then premise 1 is true and my rejoinder defeats premise 4 by showing that not all manipulative nudges are morally impermissible. On the other hand, my rejoinder also allows for counteractive nudges to be permissible by not being manipulative. The possible

position here requires a shift in the focus of the definition of ‘manipulative nudge’ from merely engaging the individual subconsciously to not allowing the individual to exercise their autonomy.

Thus, counteractive nudges are not manipulative because they allow the individual to exercise their autonomy. If manipulation is inherently impermissible, then premise 4 is true and my rejoinder defeats premise 1 by showing that counteractive nudges are not manipulative.

Disproving TMO proves that not *all* type-1 nudges are morally impermissible. However, this is not the same as proving that all type-1 nudges *are* morally permissible. According to my response, *some* (non-counteractive) type-1 nudges are impermissible. Thus, even though my rejoinder is sufficient to reject TMO, it does not eliminate the problem for LPC

because, if correct, it has the consequence that meeting LPC is insufficient for an intentional intervention to be morally permissible. One way of solving the problem is to add the condition that the intervention must be (iv) counteractive. This revision takes care of the challenge from manipulation while maintaining that most libertarian paternalistic interventions are morally permissible given that this revision minimally restricts the set of nudges that are morally permissible under the existing criteria.

CONCLUSION

Independently of whether this revision of LPC is successful, my main conclusion holds—i.e. TMO fails as an objection against the moral permissibility of all type-1 nudges. Perhaps my rejoinder to TMO is not more convincing/satisfactory than the other ones considered in this paper. Nonetheless, it seems to get at something more reasonable and practically applicable than the rest. The rejoinder to TMO from considering the implications fails to explain what goes wrong with the objection. Thaler and Sunstein’s rejoinder places an unreasonable expectation over nudges and significantly restricts the set of permissible nudges. In contrast, my rejoinder explains what goes wrong with the argument for TMO while also suggesting a possible solution that does not place an unreasonable expectation over nudges and just minimally restricts the set of permissible nudges. All while complementing the intuition that TMO’s implication—i.e. that all nudges, including ‘everyday nudges’, are impermissible—is unreasonable.³⁵

BIBLIOGRAPHY

- Ariely, D. (2010). *Predictably Irrational: The hidden forces that shape our decisions*. New York: Harper Perennial.
- Barnhill, A. (2014). What is Manipulation? In *Manipulation: Theory and Practice*, eds. Coons, C. & Weber, M., 98–120. Oxford University Press.
- Bovens, L. (2008). The Ethics of Nudge. In *Preference Change: Approaches from Philosophy, Economics, and Psychology*, chapter 10, eds. T. Grüne-Yanoff and S.O. Hansson, 207–219. Springer, New York.
- Faden, R. & Beauchamp, T. (1986). *A History and Theory of Informed Consent*. Oxford University Press.
- Hagman, W., Andersson, D., Västfjäll, D., & Tinghög, G. (2015). Public Views on Policies Involving Nudges. *Review of Philosophy and Psychology*, 6(3), 439–453.
- Hansen, P.(2016). The Definition of Nudge and Libertarian Paternalism: Does the Hand Fit the Glove? *European Journal of Risk Regulation*, 7(1), 155–174.
- Hansen, P. & Jespersen, A. (2013). Nudge and the Manipulation of Choice: A Framework for the Responsible Use of the Nudge Approach to Behaviour Change in Public Policy. *European Journal of Risk Regulation*, 4(1), 3–28.
- Hausman, D.M. & Welch, B. (2010). Debate: To nudge or not to nudge. *The Journal of Political Philosophy*, 18(1): 123–136.
- Kahneman, D. (2011). *Thinking Fast and Slow*. New York: Farrar, Straus and Giroux.
- Moore, GE. (1939). “Proof of an External World.” *Proceedings of the British Academy* 25: 273–300. Reprinted in Baldwin (1993), pp. 147–70.
- Nagatsu, M. (2015). “Social Nudges: Their Mechanisms and Justification.” *Review of Philosophy and Psychology*. 6(3): 481–494’
- Raz, J. (1988). *The Morality of Freedom*. New York: Oxford University Press.
- Shafir, E. (2013). *The behavioral foundations of public policy*. Princeton: Princeton University Press.
- Sunstein, C.R. (2014). Nudging: A Very Short Guide. *Journal of Consumer Policy*. Springer, 37(4), 583–588.
- Sunstein, C.R. (2015). Do People Like Nudges? *Administrative Law Review*, 68(2), 177–232.
- Sunstein, C.R. (2015b). The Ethics of Nudging. *Yale Journal on Regulation*, 32, 414–450.

- Sunstein, C.R. (2016). People Prefer System 2 Nudges (kind of). *Duke Law Journal*, 66, 121–168.
- Thaler, R.H. & Sunstein, C.R. (2009). *Nudge: Improving decisions about health, wealth, and happiness*. Penguin Books.
- Tversky, A. & Kahneman, D. (1974). *Judgement Under Uncertainty: Heuristics and Biases*. *Science* 185: 1124–1131.
- Wansink, B. & Hanks, A.S. (2013). *Slim by Design: Serving Healthy Foods First in Buffet Lines Improves Overall Meal Selection*. *PLoS ONE* 8(10).
- Wansink, B., Just, D., & Payne, C. (2009). *Mindless Eating and Healthy Heuristics for the Irrational*. *American Economic Review: Papers & Proceedings*, 99(2): 165–169.
- Wansink, B. (2004). *Environmental Factors that Increase the Food Intake and Consumption Volume of Unknowing Consumers*. *Annual Review of Nutrition*, 24: 455– 70.
- Wilkinson, T.M. (2013). *Nudging and Manipulation*. *Political Studies*, 61(2), 341–355.

ENDNOTES

- 1 Thaler and Sunstein (2009,6). Note that Thaler and Sunstein give the original definition of ‘nudge’ in terms of economic incentives. However, they accept that there are other kinds of incentives as well. Thus, a charitable interpretation must be inclusive of these other kinds. Note also that the definition given here treats ‘nudge’ as a noun. However, Thaler and Sunstein—as well as most authors who engage with the subject—also use the term as a verb to refer to the action of influencing people by means of nudges.
- 2 Nudges steer people in certain directions while maintain their freedom of choice. Things like suggestions, warnings, defaults, and recommendations are nudges. Things like fines, mandates, threats, bans, and direct instruction are not. A nudge is analogous to how a GPS works because it suggests the best option while still allowing people to go in a different direction if they so choose.
- 3 Salespeople, doctors, waiters, website designers, policy makers, and professors are all examples of choice architects whonudge.
- 4 This research was pioneered by the ‘heuristics and biases’ work of psychologists Daniel Kahneman and Amos Tversky, which is rooted in dual-process theories. See Tversky and Kahneman(1974).
- 5 Tversky and Kahneman (1974,1124).
- 6 Some examples of biases are - Anchoring: tendency to prefer initial suggestions; Default effect: tendency to favor the default (preselected) option; Availability: tendency to weight recent/ easily recalled information more heavily than that which is not readily recalled.
For more examples, empirical evidence, and in-depth discussion on heuristics and cognitive biases, see Tversky an Kahneman (1974), Kahneman (2011), Ariely (2010), and Thaler and Sunstein (2009).
- 7 Thaler and Sunstein write that “one way to think about this is that the Automatic System is your gut reaction and the Reflective Systemis your conscious thought.” (2009,21).
- 8 Hansen and Jespersen (2013, 15). For empirical evidence about-the plate-size default see Wansink(2004).
- 9 The term ‘mindless eating’ was coined by Brian Wansink. It refers to consuming food without paying close attention to what and

- how much is being eaten. See Wansink et al. (2009).
- 10 Hansen and Jespersen claim that this is a type-1 nudge because there is usually no conscious decision or choice made in this sequence of behavior in regard to how much to eat” (2013, 15).
- 11 Bovens (2008, 3).
- 12 This follows from Thaler and Sunstein(2009).
- 13 This means that it should alter the choice context (how options are presented), not the choice content (what options are presented).
- 14 Thaler and Sunstein write that the nudge must be “easy and cheap to avoid.” (2009,19).
- 15 This theory is ‘libertarian’ because it does not limit freedom of choice, and ‘paternalistic’ because it focuses on what is best for people.
- 16 For empirical evidence see Wansink and Hanks(2013).
- 17 In this context, what would make the customers better off can be defined in terms of health.
- 18 See Shafir (2013) for how nudges have been applied to public policy. See Thaler and Sunstein (2009) for specific examples on the increase of retirement savings (105–19), organ donations (177–84), healthy food consumption (262–3), and recycling (267–8). As well as on the reduction of environmental pollution (185–98), energy consumption (258–61), speeding (261–2), and urine spillage in public restrooms (268), among other things.
- 19 See Bovens (2008), Nagatsu (2015), Hansen and Jespersen (2013), Hausman and Welch(2010).
- 20 People in general tend to share this perception. For empirical evidence, see Sunstein (2015 and 2016) and Hangman et al(2016).
- 21 This is an overarching understanding of ‘manipulation’ drawing from various sources within the literature. Sunstein (2015b, 443–444) considers accounts by Barnhill (2014), Wilkinson (2013), Faden & Beauchamp (1986), and Raz(1988).
- 22 Nagatsu claims that “nudges are ethically problematic to the extent that they change individual behavior in such a way that is not responsive to the agent’s reasoning process” (2015,487).
- 23 The word ‘subliminal’ means ‘below threshold’. More specifically, below the threshold of consciousness. Here I am using ‘subliminal messages’ to refer to either auditory or visual messages presented below the average limits of human perception. In other words, to a signal or message that is typically unperceived consciously, yet perceived subconsciously.
- 24 Whether or not subliminal messages are nudges is contested.

Some critics argue that subliminal messages are not nudges. However, some others, including Thaler and Sunstein, find it difficult to differentiate the two because all it takes for an intervention to be a nudge is for it to preserve freedom of choice. That is, to not limit options or attach incentives or consequences to any of them. Since subliminal messages do not do any of this, it is reasonable to think that they are nudges. However, what is commonly resisted is that they are morally permissible. Thus, Thaler and Sunstein argue that even if nudges, subliminal messages are not instances of Libertarian Paternalism. My argument in this paper assumes that subliminal messages are nudges. More specifically, type-1 nudges given that they engage us only subconsciously.

- 25 Hausman and Welch (2010,128).
- 26 Hausman and Welch (2010,127).
- 27 In contrast, Hansen and Jespersen claim that type-2 nudges facilitate and increase freedom of choice and “empowerment” by means of reflective engagement (2013,24).
- 28 Recall that LPC holds that for an intentional intervention to be morally permissible, it has to be (i) a nudge, (ii) transparent, and (iii) aimed toward the welfare of those being nudged.
- 29 This evaluation of TMO follows the ‘Moorean Shift’, which is a way of objecting to an argument by appealing to commonsense. That is, by showing that rejecting its conclusion is more reasonable than accepting the conjunction of its premises. The move was originally proposed by GEMoore in response to skepticism. See Moore(1939).
- 30 Thaler and Sunstein claim that the “manipulation of this kind is objectionable precisely because it is invisible and thus impossible to monitor” (2009, 246).
- 31 Note that a person with the right technological equipment could detect messages that would be undetectable otherwise. Thaler and Sunstein’s claim, then, does not mean that subliminal messages are in fact “impossible” to monitor but perhaps only that they are practically undetectable.
- 32 This is the case even more so if the nudgee is unaware of the concept of nudging and the biases that influence their thinking.
- 33 Note, too, that there is another way in which Thaler and Sunstein’s distinction restricts the set of morally permissible nudges even more. It rules out a lot of type-2 nudges for the reasons I mention above having to do with type-2 nudges’ reliance on type-1 nudges. This is the case because a great

34 deal of type-1 nudges are not practically monitorable. Thus, if
non-practically monitorable type-1 nudges are impermissible,
then all the type-2 nudges that rely on them are impermissible
as well.

35 Nudges need not increase our sensibility to reasons in order to
be permissible. All that matters is that they do not decrease our
sensibility.

36 Thanks to Jon Matheson for providing helpful feedback on this
paper.

