



# Pushed for Being Better: On the Possibility and Desirability of Moral Nudging

Bart Engelen<sup>1</sup> · Thomas R. V. Nys<sup>2</sup>

Accepted: 19 December 2023  
© The Author(s) 2024

## 1 Introduction

Despite the initial focus on paternalistic nudges that promote the “health, wealth, and happiness” of nudgees (Thaler & Sunstein 2008), the literature now increasingly discusses the use of nudging techniques for other purposes, like promoting social, civic, or green behavior amongst nudgees. In this paper, we examine whether those can plausibly be labeled ‘moral nudges’. To do so, we ask whether the deliberate design of people’s choice environments can actually promote genuine *moral* thinking, feeling and acting. Our main aim is to analyze the alleged tension between nudging on the one hand and moral worth on the other and to argue that ‘moral nudging’ is not an oxymoron. In addition, we analyze which moral nudges are desirable. Given how nudges are increasingly popular tools in moral education and in attempts to make people behave in arguably moral ways (donating to good causes, performing civic duties, et cetera), it is important to assess whether and when nudges undermine, preserve or promote the moral worth of resulting actions.

In section 2, we conceptually clarify what moral nudging is and provide some tentative examples. In section 3, we go into the tension between nudging and what morality requires. Some argue that nudging might well have morally desirable outcomes but that nudged actions lack genuine moral worth. In section 4, we argue that moral nudging is not an oxymoron and that nudges can preserve and, in some cases, even promote moral worth. Our main claim is that nudges may not only help people to do the right thing but also to do it for the right reasons. When designed and implemented wisely, they can leave intact and, in some cases, even promote, people’s awareness of and attentiveness and responsiveness to the right moral reasons.

---

✉ Bart Engelen  
b.engelen@tilburguniversity.edu

Thomas R. V. Nys  
t.r.v.nys@uva.nl

<sup>1</sup> Tilburg Center for Moral Philosophy, Epistemology and Philosophy of Science (TiLPS), Tilburg University, Warandelaan 2, 5037 AB Tilburg, The Netherlands

<sup>2</sup> Department of Philosophy, University of Amsterdam, Oude Turfmarkt 143, 1012 GC Amsterdam, The Netherlands

In section 5, we analyze which kind of moral nudges are not only possible but also desirable. In section 6, we state our conclusions.

## 2 Moral nudging: what and how?

Before delving into the supposed tension between nudging and genuine moral thinking, feeling and acting, we should get a better understanding of nudging (section 2.1), discuss different kinds of nudges (sections 2.2 and 2.3) and provide a definition and possible examples of ‘moral nudging’ (sections 2.4 and 2.5).

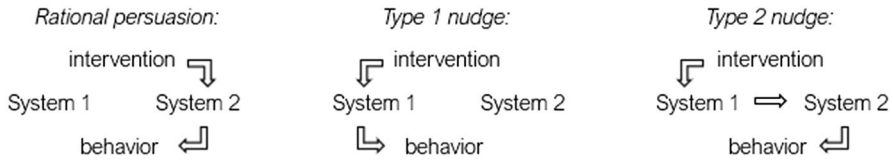
### 2.1 Nudging: what?

In line with the literature (see also: Congiu & Moscati 2022; Hansen 2016; Hansen & Jespersen 2013; Moseley 2020), we define nudging as deliberately (re)designing people’s so-called ‘choice architectures’ – the ways in which their choice options are framed and choice environments are designed – with the aim of influencing their behavior by tapping into their more automatic (i.e. less reflective) and ‘less-than-fully-rational’ psychological processes. Nudging is different from coercing or forcing (by leaving the options on the table and not making them prohibitively costly), from incentivizing (by not altering the material or financial costs or benefits attached to those options) and from merely persuading or informing (by targeting and relying people’s less reflective and less-than-fully-rational processes instead of their reflective and rational processes and capacities). This definition has two important aspects.

First, nudging is deliberate and intentional (see also: Hausman & Welch 2010). While choice environments may be inevitable and influential, regardless whether they are designed deliberately or not, nudging – as the deliberate (re)design of those environments – is not. However, once we are aware of the ways in which choice environments affect behavior, we can no longer evade the question whether and how to nudge. After all, refraining from nudging and leaving any choice environment ‘as is’ is also a decision with foreseeable consequences for people’s behavior.

Second, the distinguishing aspect of nudging is *how* it influences people, i.e. by targeting and relying on their less-than-fully-rational psychological mechanisms, such as (quasi-)automatic perceptual processes, emotional responses, cognitive heuristics and other deeply-rooted psychological tendencies like loss aversion or conformism. Nudges are based on (often recent) ‘behavioral insights’ from psychologists and behavioral economists about the ‘shallow’ decision-making processes that Daniel Kahneman (2011) labels “System 1” processes and that have been shown to influence human cognition, volition and action.<sup>1</sup> While all nudges target and rely on

<sup>1</sup> One does not need to endorse Kahneman’s ‘dual-processing’ theory or take a stance in recent debates on the nature and labeling of these processes or on how they relate to more reflective and controlled processes (Banerjee and John, 2021; Levy 2019) to agree that (1) there are less conscious, less controlled and less reflective processes in the human mind that partly influence what and how people think, feel and act, and (2) empirical insights about these processes can be employed to exert such influence, and (3) nudging is one way of doing that.



**Figure 1** Types of interventions

such processes, nudged actions are not necessarily completely automatic or unconscious (see also: Congiu & Moscati 2022). To understand this, let us have a closer look at the means nudges employ.

## 2.2 Nudging: how?

A first, influential categorization of nudges distinguishes between ‘Type 1’ or ‘mindless’ nudges and ‘Type 2’ or ‘mindful’ nudges, which refers to Kahneman’s notions of ‘System 1’ (shorthand for less reflective processes) and ‘System 2’ (shorthand for more reflective processes). Let us quote Pelle Hansen and Andreas Jespersen (2013: 14) who first made this distinction in an influential paper.

Both types of nudges aim at influencing automatic modes of thinking. But while type 2 nudges are aimed at influencing the attention and premises of - and hence the behaviour anchored in - reflective thinking (i.e. choices), via influencing the automatic system, type 1 nudges are aimed at influencing the behaviour maintained by automatic thinking, or consequences thereof without involving reflective thinking.

Importantly, Type 2 nudges still invoke less-than-fully-rational (System 1) processes and thus differ from rational persuasion and straightforward provision of information, reasons or arguments. When you inform and try to rationally persuade someone, you target and rely on their rational capacities and on more reflective processes: you hope they understand, digest and consider relevant considerations and update their beliefs and desires accordingly. In contrast, nudges always work *via* System 1 processes. While Type 1 nudges rely on these to directly influence behavior, Type 2 nudges work more indirectly and target System 1 processes with the aim of drawing people’s attention to specific information, making information easier to digest and/or stimulating reflection. Figure 1 below visualizes the threefold distinction between rational persuasion, Type 1 and Type 2 nudges and shows how all nudges target less reflective (System 1) processes in their attempt to influence people’s behavior.

Footnote 1 (continued)

Different theoretical strands in the literature, for example, disagree whether these processes and resulting beliefs, preferences and choices should be called ‘irrational’ (Ariely 2010), ‘a-rational’ (Engelen 2019), ‘rational’ (Levy 2019), ‘boundedly rational’ (Gigerenzer 2001) or ‘ecologically rational’ (Gigerenzer 2001; Schmidt 2019). Without taking a stance in these debates, we use the rather broad label ‘less-than-fully-rational’, with ‘rational’ referring to conventional, decision-theoretic notions of ‘rationality’.

Take a government (or a charity) that aims to increase donations of organs (or of money).<sup>2</sup> It can use rational persuasion, for example by setting up informational campaigns that detail relevant considerations and provide arguments why donations matter. Alternatively, it can implement a Type 1 nudge, like changing the default (from an opt-in to an opt-out or from one-off to monthly payments). Finally, it can use Type 2 nudges, like adding emotionally charged pictures of individual beneficiaries (increasing salience and triggering emotional responses) or adding colorful donation buttons online (targeting perceptual processes to focus people's attention).<sup>3</sup>

### 2.3 Nudging: to what end?

While the first categorization is based on how nudges work, i.e. which psychological mechanisms they target and rely on (Type 1 or Type 2), the second is based on the aims they can promote. Despite most of the literature focusing on paternalistic nudges that benefit nudgees (like health and safety nudges) – in part because of nudging's initial justification in “libertarian paternalist” terms (Thaler & Sunstein 2008) – growing attention is paid to nudges aimed at other ends. Nudges can be employed to benefit nudgers (like nudges maximizing company profits), to benefit others, or promote some other value (see also: Hagman et al 2015).

Think of the aforementioned pro-social nudges to increase donor registrations (Thaler & Sunstein 2008: 177-179) or to raise charitable donations (Capraro et al 2019; Rühle et al 2020), cooperative nudges to stimulate cooperation in prisoner's dilemmas (Capraro et al 2019), civic nudges to increase tax compliance (John et al 2009; Niker 2018) or green nudges to reduce energy or water consumption (Byerly et al 2018).

In each of these cases, the same nudging techniques (changing defaults, increasing an option's salience, reframing language, appealing to emotions, conformism or norms, et cetera) are used not to benefit nudgees but to promote other people's interests or the public good. Quite often, a nudge serves multiple aims. Painting lines on the road to improve road safety serves the interests of both targeted drivers (pro-self; paternalistic) as well as other people in traffic (pro-social).

The question we will be considering here is whether and when such nudges can rightly be labelled ‘moral’ nudges. Can nudges promote what morality requires?<sup>4</sup>

<sup>2</sup> Other examples of the distinction between Type 1 and Type 2 nudges can be found in Hansen (2016) and Hansen & Jespersen (2013).

<sup>3</sup> This also illustrates how most interventions invoke multiple processes and thus influence people in multiple ways. A single intervention can work both as a nudge *and* as a piece of information or argument (see also: Congiu & Moscati 2022: 193-195; Levy 2019). Smartly designed labels also carry information, as do shocking pictures on cigarette packages. As such, it can be hard to assess how exactly an intervention influences an individual in any given instance. That said, we argue in section 4 that the ‘nudge-aspects’ of interventions are not at odds with people becoming aware of and attentive or responsive to moral reasons but can actually be key in promoting such attitudes.

<sup>4</sup> If morality requires helping others, the question is whether and when ‘prosocial nudges’ count as genuinely moral nudges. However, if one believes that there are also duties to oneself, one can also ask whether some of the ‘pro-self’ nudges plausibly count as moral nudges.

## 2.4 Moral nudging: what?

Relatively little attention has been paid to ‘moral nudging’, with the exception of work by Valerio Capraro and co-authors (2019), Lily Eva Frank (2020) and Fay Niker (2018). We define moral nudging as the *deliberate (re)designing of people’s choice environments with the aim of facilitating, encouraging and, if successful, promoting what morality requires from them (in terms of moral thinking, feeling and acting)*. Of course, the important question here is what exactly ‘morality requires from people’ and thus what ‘moral thinking, feeling and acting’ entails. In sections 3 and 4, we discuss how different ethical frameworks provide different answers to these questions and thus whether and how moral thinking, feeling and action can be nudged.

Just like ‘prosocial nudging’ promotes prosocial behavior, ‘moral nudging’ promotes moral behavior, i.e. the kinds of actions and decisions that have genuine moral worth. The question whether this is possible or not (i.e. an oxymoron) should be distinguished from the question whether it is morally permissible or even desirable. Perhaps ‘moral nudging’ is possible but impermissible; perhaps it is impossible (an oxymoron) yet morally justified (it could for example generate desirable outcomes without promoting truly moral behavior). We come back to this in sections 4 (where we settle whether moral nudging is possible) and 5 (where we analyze whether people have a duty to engage in it).

## 2.5 Moral nudging: tentative examples

Here are some tentative examples of moral nudges, borrowing some from the existing literature and adding some of our own.

Capraro and co-authors (2019) were first in explicitly labeling interventions as ‘moral nudges’. In their study, they measure the impact of so-called ‘mind-set nudges’ on prosocial and cooperative behavior. Right before participants have to make a decision, they are simply asked: “what do you personally think is the morally right thing to do in this situation?” or “what do you think your society considers to be the morally right thing to do in this situation?”. These nudges use a priming technique to trigger or activate participants’ moral considerations. The results show that asking such questions increases both prosocial behavior in subsequent Dictator Games and cooperative behavior in subsequent Prisoner Dilemma Games. In addition, participants donated 39 to 47% more to real-life charities after being primed in this way. A study that corroborated these findings also found that such moral nudges remained effective when implemented transparently (Gråd, 2021).

Some of the previously mentioned prosocial, civic and green nudges also tentatively qualify as moral nudges, even though they are not described as such. When morality requires helping others, serving the public good or engaging in pro-environmental behavior, the nudges below arguably promote exactly that.

- Policy changes from an opt-in to an opt-out system for organ donation, assuming that this boosts actual donor numbers and thus helps save innocent lives.
- Nudges by charities (which effectively promote worthwhile causes) to stimulate people to donate (more than they otherwise would) by using defaults, salience, anchors, decoys, haggling and framing (Ruehle et al 2020).
- Cleverly designed roads with patterns, lines and visual illusions to slow down drivers and reduce the harm they might cause to others.
- Cleverly designed and placed hand sanitizers to prevent infections in care facilities.
- Messages in hotel bathrooms that suggest that a vast majority of customers reuses its towels.
- Shocking pictures of intubated babies on cigarette packs to saliently remind smokers that their habit might inflict harm on (born and unborn) children.
- Salient ‘fair trade’ or ‘cruelty free’ labels to reduce harms done to farmers or animals (Sheehan & Lee 2014).
- Gamification to reduce people’s ecological footprints, like taking ‘musical stairs’ or using a shower device that offers salient and immediate feedback on the amount of hot water consumed by depicting a polar bear on a shrinking slab of ice (this device has been shown to save 9.3 liters of water and 20% of energy per shower; Staake et al 2016).
- Redesigning hiring practices to combat implicit bias and increase diversity (O’Meara et al, 2020).

While these examples fit the definition of nudges – as deliberate changes in choice environments that influence behavior by targeting less-than fully-rational (System 1) processes –, we analyze in what follows whether they count as genuine *moral nudges*. We first discuss the reasons why moral nudging arguably is an oxymoron (section 3) before moving on to our central claim that it is not (section 4). Our main argument is that, while nudges in some cases can be at odds with what morality requires, they are often compatible with, and can even promote, genuine moral thinking, feeling and acting. After having established that moral nudges are possible and real, we analyze whether and when such nudges are also desirable (section 6).

### 3 Why ‘moral nudging’ is arguably an oxymoron

The main argument why moral nudging is considered an oxymoron lies in the supposed tension between nudging techniques and what morality requires. In this section, we formulate this as clearly and charitably as possible.

The worry is that nudges preclude specific kinds of motivations and attitudes that are arguably required for *genuine* moral thinking, feeling and acting. The quasi-automaticity, thoughtlessness and effortlessness (induced by the less-than-fully-rational heuristics and biases that nudges rely on) are considered incompatible with the reflectiveness, attentiveness, autonomy, authenticity and/or agency that genuine moral decision-making requires. Even putting aside worries about the potentially detrimental impact of nudges on (moral) autonomy (Furedi 2011; Hausman &

Welch 2010) and agency (Bovens 2009), the worry is that nudges may make you ‘do the right thing’ but not *for the right reasons*. Morally worthy actions, the argument goes, are done for the right reasons and nudging is considered incompatible with the kind of genuine moral decision-making required for resulting actions to have moral worth.

We first flesh out this specific understanding of what morality requires in sections 3.1 and 3.2. Next, in section 3.3, we present the case that moral nudging is arguably an oxymoron because nudging precludes people from doing the right thing for the right reasons.

### 3.1 What morality requires: doing the right thing

Of course, there are different answers to the fundamental ethical question what morality requires. Whether moral nudging is an oxymoron (or not) depends on one’s preferred ethical framework.

In consequentialist or utilitarian approaches, morality requires doing the right thing, i.e. generating the best possible outcomes. What matters, ethically speaking, is whether actions (or rules, institutions or policies) have desirable outcomes. In these approaches, moral nudging is clearly *not* an oxymoron. If nudges help promote desirable outcomes, as the examples listed above presumably do, they count as moral nudges. According to John Stuart Mill’s famous remark in *Utilitarianism*, ethics is about doing the right thing, not about what motivates such actions.

It is the business of ethics to tell us what are our duties, or by what test we may know them; but no system of ethics requires that the sole motive of all we do shall be a feeling of duty; on the contrary, ninety-nine hundredths of all our actions are done from other motives, and rightly so done, if the rule of duty does not condemn them. (...) Utilitarian moralists have gone beyond almost all others in affirming that the motive has nothing to do with the morality of the action. (Mill 1863/1906: 26)<sup>5</sup>

Take effective altruists, who argue that you have a duty to do “the most good you can do” (Singer 2015). The amount of good you do is what matters, not the nature or quality of your underlying motives. Effective altruists do not care whether you are motivated by a reflective, authentic and/or autonomous sense of duty or nudged by some smart website design or emotionally exploiting charity campaign. If it is our

<sup>5</sup> We can distinguish between an action’s motive (its *ground* or *reason*: why does one engage in it?) and intention (its *aim* or *goal*: what does one try to realize or accomplish in the world?). As the quote makes clear, Mill believes that motives do not matter in evaluating the moral value of actions. According to him, it doesn’t matter whether you save a child from drowning out of a self-interested desire for money or reputation, as long as the goal of your action (its intention) is to save the child. However, Mill does believe that motives matter when evaluating people’s *characters*. People who tend to act out of beneficence (as a motive or reason for acting) are sympathetic to the plight of others and disposed to act accordingly. Such virtuous characters are key in promoting the “greatest good for the greatest number.” For an in-depth analysis of Mill’s own understanding of the differences between motives and intentions, see Ridge (2002).

moral duty to maximize good outcomes, consequentialists argue, we should design choice environments in ways that help us fulfil this duty, full stop.<sup>6</sup>

Note that one does not have to be a full-blooded consequentialist to appreciate this point. Take traffic, where “doing the right thing” arguably requires avoiding harms to others. Whether driving safely is motivated by carefully considering safety and harm or arises mindlessly because drivers are kept in line by clever road designs or smart lane assist technologies, is irrelevant to assessing its moral worth. In domains like this, where morality doesn’t require motivational purity, moral nudging is *not* an oxymoron.

### 3.2 What morality requires: doing the right thing for the right reasons

The supposed tension between nudging and morality only arises on a different understanding of morality. According to non-consequentialists, it is not only actions and their outcomes that matter but also the underlying intentions and attitudes that motivate these. Only actions motivated by the right kind of motives and performed *for* the right reasons, they claim, have genuine moral worth. To flesh this out, we discuss how Immanuel Kant and Aristotle understand moral worth and turn to Julia Markovits’ more general account of doing the right thing *for the right reasons*. In section 3.3 then, we discuss how nudges arguably undermine this.

In contrast to Mill, Immanuel Kant argues in his *Groundwork* that an action has “true moral worth” only if it expresses a genuinely good will and is done “not from inclination but from duty” (Kant 1785/1998, 4: 398). On a popular reading of Kant, we should do our duty ‘aus Pflicht’, *from* or *out of duty*, and not because of some other motive such as self-interest (Herman 1981; Markovits 2010). If we merely conform to the moral law, instead of acting out of respect for it, our actions lack moral worth (Johnson & Cureton, 2021).

[A]n action of this kind, however right and amiable it might be, has still no genuinely moral worth. It stands on the same footing as the other inclinations – for example, the inclination for honor, which if fortunate to hit on something beneficial and right and consequently honorable, *deserves praise and encouragement, but not esteem*; for its maxim lacks moral content, namely, the performance of such actions, not from inclination, but *from duty*. (Kant 1785/1998, 4: 398; italics ours)

<sup>6</sup> From the fact that consequentialists consider moral nudging to be a possibility and not an oxymoron, one should not infer that they always consider it morally desirable. If nudges undermine the motivation or agency of people doing the right thing or generate ‘reactance’ (nudges doing the exact opposite of what nudgers were hoping for; Entwistle 2021), they fail, also on a consequentialist reading. One can easily imagine someone taking longer showers because they want to see the polar bear drown, for example. Or take a consequentialist argument for ‘moral bio-enhancement’: the promotion of moral behavior using, for example, oxytocin and ‘selective serotonin reuptake inhibitors’ (Persson & Savulescu 2012). Some consequentialists object to this, arguing that it undermines moral agency because it robs people from the opportunity to have their moral judgements govern their decisions (Harris 2013), or to make decisions that truly express their will (Huang 2018).



The worry then is that (prosocial, civic or green) nudges target and rely on “non-moral motives” (Herman 1981) or default people into doing the right thing and that, as a result, their actions lack true moral worth on this Kantian understanding. Lily Frank (2020, 381) summarizes this objection well.

The Kantian-style worry is that actions performed as a result of interaction with moral technologies may not have moral worth at all because they are not performed out of good will or out of respect for the moral law (...) and are performed merely in accordance with duty rather than from duty.

Take the horror-inducing cigarette packs that scare pregnant women out of smoking or the shower device that makes us want to save the cute digital polar bear. In these cases, we might be nudged to do our duty, but we are not acting *out of* duty but out of disgust, or some misplaced desire to win a silly game. As we have seen, nudges crucially target, trigger and work *via* System 1 processes that do not count as the kind of moral motives (having a good will, respect for the moral law) that motivate morally worthy actions.

Like Kant, Aristotle stresses the ethical importance of the appropriate kinds of motives and attitudes.

[I]t is no easy task to be good. (...) anyone can get angry – that is easy – or give or spend money; but to do this to the right person, to the right extent, at the right time, with the right motive, and in the right way, that is not for everyone, nor is it easy; wherefore goodness is both rare and laudable and noble. (Aristotle, *Nichomachean Ethics* II.9, 1108b)

For Aristotle, a genuinely virtuous person (*phronimos*) not only does the right thing (sheer luck or habit can cause this as well) but also has a specific kind of character and practical knowledge (*phronesis*) that allows her to ‘know’ what to do. Aristotle understands virtuous action as produced “according to the right reason” (*kata ton orthon logon*) (1138b24), where “the motivating reason is felt “at the right times, with reference to the right objects, towards the right people, with the right motive, and in the right way” (Aristotle, *Nichomachean Ethics* II.4 1106b20-22)” (Niker 2018: 157). Virtuous action should be performed for its own sake (VI.5 1140b) and requires “a certain level of affective and perceptive engagement as part of the practical reasoning process” (Niker 2018: 156). Virtuous people discern the morally salient features of the situation they encounter, decide on a virtuous action themselves in full knowledge that it is the right thing to do (Khan 2005: 42).

Again, the objection against moral nudging is that it is antithetical to the specific kind of epistemic and motivational state that genuine virtuousness requires. If you need nudges to donate to money or organs, you are not a genuinely virtuous person, the objection goes. Although in some way ‘laudable,’ you did not act ‘in the right way,’ for its own sake, showcasing practical wisdom (Frank 2020: 379-380).

Julia Markovits (2010: 203) summarizes these non-consequentialist approaches to morality as follows.

Morally worthy actions are ones that reflect well on the moral character of the person who performs them (...). When we do the right thing because it hap-

pens to suit us, or happens to be in our interest, our action has no moral worth. This is intuitive. Morally worthy actions must be performed for the right (motivating) reasons.

Following Markovits then, for an action to have moral worth, its justifying reason (what makes it the right thing to do) should also be its motivating reason (what makes someone do the right thing). When justifying and motivating reasons do *not* align, you can still do the right thing but not for the right reasons.

According to what I will call the Coincident Reasons Thesis, *my action is morally worthy if and only if my motivating reasons for acting coincide with the reasons morally justifying the action* – that is, if and only if I perform the action I morally ought to perform, for the (normative) reasons why it morally ought to be performed. (Markovits 2010: 205; see also Gorin 2018: 243-244).

### 3.3 Why nudges preclude acting for the right reasons

On this non-consequentialist account of morality, nudges are arguably incompatible with what morality requires as they undermine the moral worth of resulting actions. According to what we call the “Incompatibility Claim”, nudged actions lack moral worth as nudges preclude people from acting *for the right reasons*. It states that there is an inherent tension between the practice of nudging and the kind of motives and attitudes required for true moral worth. There are different ways of spelling out this claim.<sup>7</sup>

On a first reading, nudges – even when they promote doing the right thing – inhibit the kind of authenticity and autonomy required for moral agency. Acting for the right reasons means acting on *your own* reasons why something is right or wrong, reasons you understand and endorse, instead of outwardly complying to whatever someone else wants you to do. When nudged, the reasons, values and considerations you act on are those of the nudger who has deliberately designed your choice environment. Someone else is pulling your strings and pushing your buttons. A well-known objection to nudging claims that it “imposes the will of one agent on another” (Hausman & Welch 2010: 133) and involves a substitution of values (Rebonato 2012; White 2013). When you are defaulted into organ donation, it’s not really you who decides to register. Even if it is the right thing to do, you are not really doing it for the right reason (in the case of defaults, you are not even doing anything yourself).

On a second reading, the processes and factors that play a key causal role in how (both Type 1 and Type 2) nudges work are completely arbitrary and irrelevant, morally speaking. Tweaks to your choice environment and the quasi-automatic System 1 processes that this triggers do not count as good moral reasons to act. The disgust that cigarette pack pictures trigger and even the eagerness to win the polar bear

<sup>7</sup> While there have been quite extensive debates about the extent to which and the reasons why nudges are arguably at odds with authenticity and autonomy – which arguably relate to the issue of moral worth in complex ways – we largely put these issues aside and focus on moral worth specifically.

shower game, may lead people to do the right thing (motivating reasons) but they are not what *makes* taking shorter showers or stopping smoking the right thing (justifying reasons).

Type 1 nudges, which Fay Niker (2018: 158) calls “*automatic-behavioral nudges*”, look like they are incompatible with doing the right thing *for the right reasons* as they target and rely on System 1 processes to directly change our behavior without affecting “how we see the reasons for behaviour”. The automaticity of the psychological mechanisms that these nudges invoke means that people are nudged to do the right thing but “not in a manner that requires or develops the exercise of the aspects of practical reason that are characteristic of virtuous action and reaction” (Niker 2018: 158). Doing the right thing, yes, but not for the right reasons. When you are defaulted into donating an organ or money, your action lacks moral worth, the argument goes.

This worry, however, also applies to Type 2 nudges, which also work via System 1 and thus crucially invoke causal factors that are completely arbitrary or irrelevant, from a moral point of view. Cleverly designed websites, cigarette packages and shower devices influence behavior by triggering quasi-automatic perceptual and emotional processes that direct and focus people’s attention and thus count as Type 2 nudges. But the causal factors at play here (the design elements, the perceptual, cognitive and emotional processes) motivate resulting actions (motivating reason, what drives the action) but do not count as ‘right reasons’ (justifying reason, what makes the action the right thing to do).

What is more, nudges can even promote actions – perhaps even the right ones – for the (*morally*) *wrong* reasons. Think of how nudges can exploit sexist biases to induce men to hire more women. One can, for example, nudge employers by using software that makes pictures of female applicants ‘sexier’. While the outcome might be morally desirable, it comes about for all the wrong reasons.

On a third reading, nudges are similar to other interventions like financial incentives in that they can *crowd out* intrinsic motivation (see: Gråd et al 2021), impair self-determination and shift the locus of control “from inside to outside of the person affected” (Frey 2012: 92). The key claim here is a causal one: when exposed to nudges, people’s moral motivations might be pushed away and replaced by non-moral motivations. Thi Nguyen (2020) makes this argument for a specific nudging technique: gamification. When something is gamified (as the polar bear shower device does), people become distracted from their initial aims and values (to adopt a greener lifestyle). Their attention and efforts are redirected to another target (to score points or win a game). Moral motivations are crowded out by nonmoral ones. Gamifying nudges then “can amplify our motivation to act, but in order to do so, it needs to alter the goal” (Thi Nguyen 2020: 200). In other words, nudges gamifying moral behavior can make (more) people do the right thing but crowd out their attentiveness and responsiveness to the right reasons.

Each of these readings is based on the oft-heard objection that nudges – which crucially target and rely on people’s System 1 processes – exploit their cognitive deficiencies and vulnerabilities and influence them ‘behind their backs.’ Instead

of repeating this criticism, we want to draw out its implications for moral worth.<sup>8</sup> After all, the quasi-automatic nature of nudged actions is arguably at odds with the autonomy (Furedi. 2011; Hausman & Welch 2010), agency (Bovens 2009) and the motives and attitudes needed for actions to have moral worth. While nudges rely on your “inattentiveness” (Glod 2015: 602), morality requires a certain kind of attentiveness, namely to the right reasons, i.e. reasons that are 1) authentically yours and 2) distinctly moral in nature.

While these worries apply to specific nudges and their immediate effects, another set of concerns focus on the longer run. One worry is that nudgees, when exposed to nudges over and over again, may come to rely on them, reducing their autonomy over time (Furedi 2011; Waldron 2014). Frank (2020: 377) summarizes this worry: “As individuals offload moral decision-making to technologies or the built environment, the worry might arise that some of the skills, cognitive or affective, necessary for (...) moral innovation will be lost.”

A different worry is that nudges prevent people from *showing* their moral worth and expressing their moral selves, as people capable of doing the right things for the right reasons. Imagine that you are concerned about long waiting lists for organ donation, which leads you to opt in and register for donation. As soon as a general opt-out policy is implemented, you are defaulted into registration along with everyone else. While this may boost donor numbers overall, it also reduces your opportunity to express your moral self. The ‘long run’ corollary here is that nudges can lead to a weakening – or even disappearing – of our moral selves over time. The disgust triggered by cigarette packs and the sexism invoked by dubious hiring software are nonmoral or even immoral considerations that can, in time, drown out whatever moral reasons people have for doing the right thing.

In sum, the Incompatibility Claim holds that nudged actions cannot have genuine moral worth as they are not performed for the right reasons. Nudges target, trigger and rely on the causal force of heuristics, biases or norms that in no way constitute genuinely moral motives. When nudged, your actions are motivated by quasi-automatic factors with no moral relevance, not by considerations that are authentically yours and moral in nature. As such, in all of the above examples, nudged actions lack genuine moral worth, because what motivates them is not what justifies them.<sup>9</sup>

<sup>8</sup> Chris Mills summarizes these claims in a way that neatly ties into the Kantian and Aristotelian terminology discussed before: “Heteronomous behaviour can be caused by *any reason for action that motivates an individual contrary to (e.g., by overriding or subverting) their authentic will*. Heteronomy specifically threatens the *independence of an individual’s will* by disregarding her decision-making competency, thus *bypassing part of what makes her decision her own* (...). Critics may suggest that choice architecture is necessarily heteronomous because it seeks to exploit heuristics and cognitive biases in our reasoning. Accordingly, choice architects pursue a programme of manipulation that undermines the independence of an autonomous agent’s will by subverting the flaws in her decision-making competency to bring about particular outcomes.” (Mills 2015: 497-498, emphasis ours)

<sup>9</sup> According to Luc Bovens (2009: 10), when I am nudged to do something, my action is not rational and autonomous: “what is driving my action does not constitute a reason for my action – i.e., it is not a feature of the action that I endorse as a feature that makes the action desirable.” When applied to morality, the objection is that what is driving nudged action is not a feature of the action that (I endorse as what) makes the action moral.

This is true for citizens being defaulted into organ donation, people lured into taking shorter showers, passers-by seduced into charitable giving and pregnant women scared into giving up smoking.

## 4 Why moral nudging is not an oxymoron

In this section, we present our case against the Incompatibility Claim and argue that nudges, when designed and implemented well, can leave intact (section 4.1) and even promote (section 4.2) the attentiveness and responsiveness to the right kind of reasons needed for moral worth.

Before developing our main argument, remember the moral ‘mindset nudges’ that Capraro et al (2019) employed. If subjects donate more after being asked to consider what they take to be the right thing, it is hard to argue that this priming technique undermines the moral character of their decisions. Even if primed, it is their own, distinctly moral considerations that motivate their actions. In our view, such a nudge not only leaves intact the moral worth of people’s decision to donate; it also promotes it, as it makes people think about their own moral commitments in ways they would likely not, absent the nudge. Such Type 2 nudges thus trigger attentiveness to the right reasons. If this holds true, then moral nudging is *not* an oxymoron, as there is at least one example of a genuine moral nudge. In what follows, we argue that more nudges work like this, while admitting that not all nudges always do, since some indeed preempt or corrupt moral worth.

### 4.1 Nudges can leave intact acting for the right reasons

Before tackling the Incompatibility Claim; let us first explain briefly why nudges do not necessarily undermine or erode (moral) autonomy or agency. (For a lengthier discussion, see Engelen & Nys 2020; Levy 2019; Mills 2015; Sunstein 2016.). First, their easy resistibility and mild influence attenuate most of these worries. Second, the less-than-fully-rational processes they invoke are compatible with autonomy and agency: if they were not, their often inevitable influence would hardly leave any room at all for making autonomous decisions (Engelen & Nys 2020).

Importantly, the Incompatibility Claim assumes that actions, in the absence of deliberate nudges, have moral worth, which vanishes when nudges affect behavior. Let us accept that assumption for now and analyze whether nudges can *leave intact* the moral worth of actions that are initially performed for the right reasons.

Suppose that, being aware of the environmental harms of long, hot showers, you have adopted a 5-minute shower habit. Now suppose your gym installs the polar bear shower device or a 6-minute time lock. It seems implausible to claim that these interventions somehow destroy or reduce the moral worth of your actions. You would not be trying to win the polar bear game, as winning isn’t a challenge to you at all (even if indeed, as Thi Nguyen says, you could be ‘distracted’ by it). Similarly, you would remain perfectly unaware of the 6-minute shower lock, because it would

never come into effect. Or take a careful driver who developed a habit of not speeding and then encounters cleverly painted lines on the road. In each of these cases, the nudge does not *override* the original moral motive, but rather works *alongside* it. The motivating moral consideration remains in place when a nudge is introduced. The less-than-fully-rational processes that nudges invoke – whether they be Type 1 (painted lines) or Type 2 nudges (polar bear game) – do not necessarily displace or preempt the motivational force of one’s initial moral reason for acting.

When moral motivations are effective absent nudges, they do not necessarily lose strength when nudges are implemented. In the only available empirical study on potential ‘crowding out’ effects, Erik Gråd and co-authors (2021) tested how much nudges affect charity donations.<sup>10</sup> Their “results show no indication that nudges crowd out prosocial behavior. Instead, donations increased in all (...) conditions where people were nudged compared with the (no-nudge) control condition.” (Grad et al, 2021: 11) While moral motives were present absent any nudge (i.e. some participants in the control condition did donate to charity), nudges increased both the amount of donors and the average donations. Participants who did not perceive nudges as manipulative increased their donations, while those who did think of them as manipulative *did not lower* their donations (Grad, 2021: 7).

When moral reasons motivate your actions in non-nudged environments, it is hard to imagine why these would disappear when nudges are implemented. If you are so committed to taking shorter showers or donating to charity that you act on these commitments without any nudges being in place, then those motivations and the moral quality of the resulting actions do not vanish whenever nudges are implemented.

The gym example also shows that nudges that trigger nonmoral motivations do not necessarily preclude expressing one’s best self. Even if the nudge is strong enough to ensure – on its own – that you do the right thing, that doesn’t make it impossible to express your moral commitments. You can, for example, refuse to look at the polar bear or stop well ahead of virtually drowning it. Or when you find out about your government’s opt-out policy for organ donation, our prototypical example of a Type 1 nudge, nothing prevents you from expressing support for that policy or actively registering an advance directive. In fact, a lot of governments encourage citizens to do exactly that so that citizens can express their moral commitments explicitly and publicly.

Interestingly, a correct understanding of what Kant meant by ‘acting from duty’ supports this line of reasoning. Suppose that you are committed to helping your friend move house, and that, when the day arrives, it turns out to be a very pleasant experience: the sun is shining, her friends and parents are nice, and the moving itself is quick and painless. Does the fact that you enjoy the day (which would

<sup>10</sup> They investigated three kinds of nudges: 1) ‘default nudges’ (having to opt out of donating, a clear Type 1 nudge), 2) social nudges (being told that around 80% of participants donate, which can work both as a Type 1 nudge, increasing the salience of a social norm, and a Type 2 nudge, inducing reflection on what is expected in situations like this) and 3) ‘moral nudges’ (like those by Capraro and co-authors, a clear Type 2 nudge).

provide a sufficient motive for you to help out) detract from the moral worth of your actions? It seems not, as you would also have helped in less pleasant circumstances as well. Likewise, in our examples, you do not behave morally *only because of* the nudges and whatever nonmoral considerations these trigger. Your moral motives can still move (and suffice to move) your actions. Even if a shower device makes saving water fun, it does not impair the moral worth of taking shorter showers, which can still be done for and because of all the right reasons, which can remain fully in place and continue to motivate people.

Indeed, as Marcia Baron (1995: chapter 5) has pointed out, acting from duty does not require respect for the law to be the only or even the so-called ‘primary’ motive. Nelson Potter (1997: 494) summarizes Baron’s interpretation of Kant as follows: “Each individual action need not be done from duty (as a primary motive) for it to be the case that one ‘obeys the law from duty.’ For one may adopt a general, overarching maxim from duty and be committed to acting accordingly.” You can help your friends move, while ‘acting from duty’ *whilst* taking pleasure in it, as long as you do not help just because you like it.

Also, Kant (1797/1996: 6:232) allows for (legal) coercion to enforce adherence to the moral law in the case of perfect duties.<sup>11</sup> Even if our fear of being punished for stealing is sufficient to ‘keep us in line’, this does not preclude (most of) us from adhering to the moral law *for moral reasons*. Clearly, for Kant, nonmoral and moral motivations can coexist. When both are at play and people have mixed motives that causally ‘overdetermine’ actions, the presence of the former doesn’t diminish the moral worth of actions. If Kant is right, then nudges that – much like coercive policies – trigger nonmoral motivations, can be perfectly compatible with actions having moral worth. A proper understanding of Kant thus does not lend support for the Incompatibility Claim but provides reasons for denying it.

In sum, the purist reading of Kant is overly strict. If it were right, and only those actions have moral worth that are motivated *exclusively* by one’s good will, then one should continuously seek out or even create difficult circumstances and overcome those by sheer strength of will. Kant admits this makes no sense. Instead, and this is quite the reverse, Kant believes that we should try to eliminate or reduce the strength of motives that inhibit us from acting out of duty.

Having a good will, in this sense, is compatible with having feelings and emotions of various kinds, and even with aiming to cultivate some of them in order to counteract desires and inclinations that tempt us to immorality. (Johnson & Cureton, 2021)

Moral nudges – both Type 1 and Type 2 – can help cultivate the kinds of motives that have such counteracting potential and thus facilitate moral action. The fact that doing the right thing becomes less difficult when nudges trigger non-moral motives doesn’t make it any less morally worthy (Frank 2020).

<sup>11</sup> If such coercion would preempt the moral worth of our actions, then Kant’s ‘philosophy of right’ would thwart his moral project. In line with what we are arguing here, however, there is no need to suppose that it does.

An interesting case in this respect is self-nudging. You can, for example, buy a shower device to help yourself take shorter showers. Or, you can set up an automatic system of monthly payments to a charity because you know you will be too forgetful or weak-willed to make the payments you take to be morally required. When you engage in such “moral offloading” (Frank 2020: 373) and outsource certain mental tasks to external tools, the self-nudges you install here do not undermine your authenticity, autonomy, or agency at all. After all, your own moral reasons are a crucial factor in the entire causal story: they are what made you buy the shower device and set up monthly payments and in that sense they both justify and motivate you doing the right thing.

A similar story can be told on the part of Aristotle. While Kant acknowledges that acting morally *can* be easy, Aristotle goes further and claims that it *should* come easy to the truly virtuous person. In other words, the less difficult it is for you to do the right thing, the more morally worthy you are. Someone who still experiences some detracting inner force or obstacle that they should overcome – and thus has to exhibit what Aristotle calls self-control or ‘enkrateia’ – falls short of the moral ideal of the *phronimos* (Gould 1994: 174). According to Aristotle, teeth-grinding moral heroes are second-best at best. While we expand on this later on, the implication for now is similar to that of our non-purist reading of Kant: nudges that facilitate doing what morality requires, do not necessarily preclude moral worth at all.<sup>12</sup>

## 4.2 Nudges can promote acting for the right reasons

This brings us to our next point. While we have argued so far that nudges can *leave intact* the moral worth of actions performed for the right reasons, we now make the stronger claim that nudges can actually *promote* this. Moral nudges can and may actually be needed to get people to act for those reasons (instead of precluding them from doing so). Here, we focus on scenarios where actions, absent any nudges, would *lack* moral worth. People often do *not* think about their water and energy consumption when showering, or *fail* to register as organ donors or to donate to charities or shop fair trade, which is exactly why moral nudges are needed, designed and implemented.

Often, the problem is not that people are immoral or lack moral principles. Most of us care about others and the environment and do not want to be involved in exploitative practices. But these moral principles often insufficiently motivate our actions. While we have all the right reasons, we often fail to act on them, because we are weak-willed or suffer from ‘ego-depletion’ (Frank 2020: 371), that is, we have a hard time overcoming temptation or lack the mental bandwidth to attend to all the complexities of the many morally laden situations we encounter. When this happens, nudges can help bridge the resulting “intention-behavior gap” (Papies 2017) and provide the necessary push to actually *act on* our own moral reasons, by making this easier or simply ‘more fun’. As Frank (2020: 3741) puts it, they can “help people

<sup>12</sup> Note that truly virtuous people can still deviate from nudges if that were morally better and, for example, stop ahead of the polar bear’s digital demise or pick a fair-trade product even when placed below or above eye level.



behave more consistently with their own deeply held moral convictions”.<sup>13</sup> When we play the polar bear game, donate to charity after seeing some emotional campaign or buy conspicuously labeled fair trade products, the most plausible causal story is similar to what happens in Capraro’s study: these nudges remind us of our moral values and principles and (re)invigorate us to act upon them (see also: Frank 2020: 381). Instead of inhibiting or bypassing moral reasoning, the less-than-fully-rational processes invoked by nudges can actually be key to triggering and encouraging it.

Think of ‘cruelty free’ labels, a prototypical Type 2 nudge. While most consumers are against animal cruelty and want to act accordingly, such moral concerns are typically not on their minds when shopping.<sup>14</sup> The labels, however, saliently prompt relevant moral commitments, at the right time, thus helping people to translate those into action. According to Niker (2018: 158), some nudges can change “how a person sees a situation – perhaps by making relevant reasons more salient, so that they are more easily perceived”. While “automatic-behavioural” or Type 1 nudges only change behavior, according to Niker (2018), these “discernment-developing” or Type 2 nudges help cultivate people’s capacities to see the morally relevant features of their choice options. Salient labels highlight relevant moral features and increase the likelihood of these actually playing a role in the (practical) reasoning processes of consumers.<sup>15</sup>

Even Type 1 nudges, we argue, can promote people’s awareness of and attentiveness and responsiveness to the right moral reasons. As Neil Levy (2019: 289) points out, nudges often “have the function of making considerations salient to us” and are thus addressed to – rather than bypass – our reasoning mechanisms. This applies to both Type 1 and Type 2 nudges. Take Type 1 nudges that smartly frame specific options and, in doing so, highlight specific considerations and direct our attention to them. As such, they make us more aware of and attentive, sensitive and responsive to reasons for acting accordingly (Levy 2019: 292). The less-than-fully-rational processes invoked by nudges – like shocking cigarette packages, cruelty-free labels and messages about other hotel guests reusing towels – do not inhibit but facilitate and encourage our awareness, recognition and consideration of relevant moral reasons.

<sup>13</sup> One can wonder how this paper relates to Frank’s, especially given this two-fold claim of hers: “First, technologies to improve individuals’ moral capacities and behaviors are realizable. Second, such technologies will actually help them get morality right and behave more consistently with whatever the ‘real’ right thing to do turns out to be.” (Frank 2020: 373) The main differences lie in 1) our more specific focus on nudges (whereas Frank focuses on ‘moral technologies’ more broadly) and 2) our willingness to question these claims (whereas Frank uses them as “assumptions” to argue that moral struggle, which can be reduced by moral technologies, is not necessary for moral progress).

<sup>14</sup> Note that Kant allows for this: “Consider the extreme claim that (1) we ought always to be thinking explicitly of *the moral law* and moved by that thought whenever we conform to duty. If acting from duty means being moved at that time by an explicit thought of some formula of the Categorical Imperative, then (1) is an unrealistic demand and arguably we should not even try to fulfill it.” (Cureton & Hill 2018: 5)

<sup>15</sup> Note that Niker (2018: 162) only argues that such nudges have educational value and help *cultivate* virtue. In her view, the resulting virtues “fail to meet the conditions necessary for full virtue”. Our claim is more radical: if nudges help people see the moral reasons relevant in a particular situation and motivate them to act on those, we see no reason to deny that they promote genuine virtue.

The Incompatibility Claim then wrongly assumes that nudges work exclusively via processes that inhibit or bypass our reasoning capacities.

Of course, some people do not care at all about other people, animals, the environment, or third-world farmers. How do moral nudges affect them? First, and most likely, the easy resistibility of nudges leaves amoral or even immoral people unaffected. A lot of people ignore fair trade labels and buy cheaper but exploitative products. Second, nudges might still encourage some of these people to reflect on their commitments. If one faces loud and clear signals that one's decisions raise ethical worries, nudges can make those moral reasons salient and get people – perhaps for the first time – to actually *consider* them. Nudges can thus kickstart a process of moral awareness and reasoning over time, for example when people have more mental bandwidth.

Moral nudges can then succeed in multiple ways. They can *remind* people and *make them (more) aware* of these reasons, they can make those reasons *gain motivational force* and they can *trigger* moral reflection, a first step in *reconsidering* their reasons for action.

This does not imply that all of the nudges listed above *always* work in one of these ways. Nudges can make use of salience (labels) but also of morally irrelevant emotions (cigarette packages), perceptual illusions (road design), cognitive biases (anchors for higher donations), mere laziness (defaults), and so on. While labels that highlight relevant moral considerations may count as genuine moral nudges, other nudges may not. We fully accept that point. It all depends on the causal story at hand and the (psychological) mechanisms at play.

Take organ donation defaults. Switching from opt-in to opt-out can increase donor numbers without triggering any moral considerations. When there is no accompanying campaign to make people aware of this policy and its ethical justifications, these good outcomes do not arise from the right moral reasons. When defaults work through ignorance, laziness, conformism or the status quo bias, they cannot be said to promote genuine moral thinking, feeling or acting. According to Moti Gorin (2018: 239), nudges that work in this way “do not reliably track reasons”. This is also why they arguably constitute manipulation, which Gorin (2018: 237) defines as “a process of influence that deliberately fails to track reason”.

Three comments are in place here. First, conceding that some nudges – like defaults – in some cases and for some people – when they work exclusively through ignorance and laziness – do not promote genuine moral worth does not imply that moral nudging is an oxymoron. The Incompatibility Claim that this is impossible – for anyone, for any technique, in any case – is simply too strong. Second, some nudges trigger both moral and nonmoral considerations. The message that a majority of hotel guests reuses its towels (Goldstein et al 2008), for example, partly relies on psychological mechanisms with no moral worth – like conformism – but can also make people consider the underlying moral reasons. The same goes for defaults, which can work in different ways and invoke different psychological mechanisms. They can play into people's tendency to avoid cognitive effort, provide a new reference point and they can even be perceived as recommendations (Grüne-Yanoff 2016). Some of these considerations likely count as moral reasons while others do

not. Third, note that even prototypical Type 1 nudges like defaults can in practice play out as Type 2 nudges and prompt moral soul-searching. “If the government wants me to do this”, some citizens may start wondering, “Why is that? Do I agree? Is this perhaps what I should do?” If not for the nudge, some people would not have considered these moral reasons and would not have acted ‘spontaneously’ for the right reasons (which was indeed our assumption in this section).

### 4.3 Objections and responses

Let us consider two objections to our claim that nudges can leave intact and promote moral worth.

The first is the ‘Not Really Nudges’ objection. Moral nudges, as we described them, arguably do not work *qua nudges*. After all, the mechanisms at play here are conscious and reflective: the key causal factors in our examples (salient labels, emotional charity campaigns, shower devices, hotel towel messages) are not the quasi-automatic perceptual processes, cognitive heuristics and emotional responses but the information that is at least implicitly presented and the awareness and reflection triggered by the intervention.

We provide two responses. First, this objection – in our view wrongly – assumes that only Type 1 nudges are ‘real’ nudges. But remember there are also Type 2 nudges, which trigger quasi-automatic System 1 processes that in turn facilitate more reflective System 2 processes. The less-than-fully-rational processes at play – salience, conformism, laziness, et cetera – are what makes these nudges work differently (and typically better) than mere information provision and rational persuasion. Second, even if you would – for some reason – exclude Type 2 interventions from being ‘real’ nudges, our claims about Type 1 nudges still hold: they can leave moral worth intact (remember our examples of the gym and the painted lines) and, in some cases, even promote it (remember how defaults can prompt moral soul-searching).

The second objection is the ‘All Too Easy’ objection. The idea is that morality requires people acting *on their own*, without the help or assistance from (semi-automatic) prods and pushes. Because doing the right thing in non-nudged environments is difficult<sup>16</sup>, willpower is needed to act ‘from duty’ (as opposed to just ‘going along’ with nonmoral motives). When people reduce shower time and buy cruelty-free products *in the absence of nudges*, their actions are clearly motivated by moral motives. If these products are on the bottom shelf, in inconspicuous packaging, then it takes *effort* to find them: getting down on all fours to examine the small print shows true commitment to moral reasons. If nudges make moral action all too easy, they prevent people from showing their *good will*, much like training wheels on a bike prevent people from showing others that they can ride a proper bike.

In response, the fact that some people indeed lose the opportunity to reveal their moral merit to others is mostly an epistemic worry (how can we know whether others are morally motivated?) and not a motivational one (can people remain morally

<sup>16</sup> Nudged environments can make it even more difficult, when the nudges trigger nonmoral considerations, like in-your-face “sale!”-labels on products that are not fair trade or cruelty-free.

motivated?). This objection is about the ability to *showcase* one's moral motivations, not about *having* such motivations, which is our concern here. While it may be hard to spot the difference, people who can ride a bike can also ride a bike with training wheels, without these wheels *doing* anything. Likewise, you can pick fair trade products for the right reasons, even when these are placed at eye level and adorned with salient labels. While such nudges make your moral commitments less conspicuous, they don't diminish them in any way.

But consider this refined version of the 'All Too Easy' objection. By making it too easy, nudges actually do undermine genuine moral motivation, as people no longer face ethically challenging situations or internal moral struggles; experiences which are necessary prompts and catalysts for moral deliberation and reflection (Frank 2020: 373-375).

Three responses are in order. First, facilitating moral action doesn't eliminate the need for moral motivation. Even when nudged, people still need to reach out for cruelty-free products, physically turn off the shower and say 'no' to cigarettes. At least some motivation is needed to make that decision, however easy it becomes. (The exception here are defaults, which can rely on inaction instead of action.)

Second, remember our previous claim: when nudges facilitate doing the right thing, a plausible causal story is one of overdetermination, where both the nudged choice environment and the right moral reasons cause the action at hand, without the former necessarily crowding out the latter. In line with Philip Pettit (1995: 323), one can understand nudges as "standby causes of certain patterns of behaviour: they may be potential causes that would serve to sustain those patterns, did the actual causes fail." Moral motivations can actually drive people, while nudges function like the safety tires besides the racing track, ensuring they do not steer off too much.

The third response denies that morality can be 'too easy'. If morality is demanding as it is, why make it even harder instead of facilitating it? Doing your duty may be hard but need not be. While moral struggle *can* prompt moral deliberation and reflection, it is not necessary for this. In fact, the opposite may be true. Nudges that facilitate or even automate moral action can free up mental bandwidth that people can use to attend to other (previously overlooked) ethical issues (Frank 2020: 377-378). Setting up automated monthly donations to a charity, for example, frees up time you can spend on investigating which charity actually does most good. Clever designs that nudge healthcare workers to wash their hands without much thought not only prevent infections but also enable workers to provide better care for their patients.

Let us return to Kant whose view of morality is criticized for being overly demanding and requiring Herculean willpower. However, we have argued that doing one's duty will come easier when one's desires are aligned with this instead of posing a "powerful counterweight" (Groundwork 4:405; Kant 1785/1998: 17). In line with our duty to become a moral person, we have a "subsequent duty to establish the conditions that human beings, as psychological and sensible creatures, depend on in order to be motivated to advance the moral world" (Dubink & Van Liedekerke 2020: 383). Given that human nature is made from "crooked timber" (Kant 1784/2009, sixth proposition) and morality often insufficiently motivates us, we

should design external circumstances and choice environments to facilitate, and not hinder people complying with it.

Kant (1797/1996: 6:457) himself gives an example of this, arguing that we have “a duty not to avoid the places where the poor who lack the most basic necessities are to be found but rather to seek them out” in order to “cultivate the compassionate natural feelings in us, and to make use of them”. Being confronted with salient stimuli about the plight of the poor will *remind* us of our duties, trigger feelings of care and decrease our self-love, thereby “removing or diminishing impediments to” the moral law (see also Johnson & Cureton, 2021). It is what charities do when they portray identifiable beneficiaries who are obviously in need. The trick is to select those environments that ensure that our nonmoral and moral and motivations line up, with the former supporting instead of thwarting the latter.

Moral nudges then consist in deliberate (re)designs of choice environments that facilitate doing what morality requires, either by *reminding* people of moral considerations and/or by *lessening* their inclinations to deviate from them. Because the following four scenarios can occur, the Incompatibility Claim is false. Scenarios 1 and 2 feature people who would not have done the right thing, absent moral nudges (i.e. nudges that promote moral worth), while those in 3 and 4 would have done so (i.e. nudges that leave moral worth intact).

1. ‘Mindset nudges’ (Type 2) that induce a moral mindset and bring people’s own moral considerations to the front of their minds (like the nudges Capraro and co-authors discuss).
2. ‘Pointer nudges’ that make people consider specific moral reasons (Type 1 nudges like defaults and Type 2 nudges like labels, graphic pictures and games).
3. ‘Stabilizing nudges’ that keep morally motivated people on track (Type 1 nudges like defaults, automatic enrollment and perceptual illusions).
4. ‘Reminder nudges’ that remind morally motivated people of specific moral reasons and focus people’s attention on those (Type 2 nudges like labels, graphic pictures and games).

That said, nudges – both Type 1 and Type 2 – can also reduce moral worth.

5. ‘Immoral nudges’ that make people do the right thing for the wrong (immoral or nonmoral) reasons (as in the sexist hiring example).
6. ‘Crowding out nudges’ where nonmoral motives replace moral motives (as can happen with opt out defaults for organ donation).

While consequentialists and non-consequentialists disagree on whether scenario 5 counts as moral nudging, they agree that scenario 6 is at odds with morality, both for consequentialist reasons (people are less likely to do the right thing when the force of moral motivations wears thinner) and for non-consequentialist reasons (the right reasons no longer motivate so moral worth decreases).

## 5 When moral nudging is desirable

Having shown that moral nudging is possible, we now want to ask when it is desirable to engage in it. Without aiming or claiming to provide an ‘all-things-considered’ assessment of its desirability, we investigate which kinds of circumstances provide reasons for implementing moral nudges in situations in which people fail to exhibit moral behavior (section 5.1) or are in need of moral education (section 5.2).

### 5.1 Nudges that scaffold morality

As mentioned, most people have moral concerns – they care about the environment and the plight of those in need – but fail to be properly motivated by these at times (due to akrasia, laziness, shortsightedness, biases, limited mental bandwidth, and so on). When people are imperfectly moral (which often holds for all of us, at least some of the time), we not only can but also should ‘scaffold’ morality. As architects of the (physical, digital, social and institutional) choice environments of ourselves, our families and friends and our fellow citizens, we should try to redesign these environments in ways that promote rather than inhibit morality. Improving the design of ‘moral ecologies’<sup>17</sup> then is an indispensable part of enabling ourselves and others to lead good lives. Whenever possible (and opportunities are plenty), we should make the right path easy and the wrong path hard.<sup>18</sup>

This argument has wider repercussions. Speed bumps, for example, quite forcibly prevent people from hurting themselves and others in traffic, but the moral reasons to build them are massive. Or think of anonymous grading, where a simple technology can render it impossible for sexist, racist or implicit biases to cause unfair discrimination. The moral reasons for implementing such stronger policies also hold for weaker instruments – like nudges – that can achieve similar aims.

In sum, moral nudging is likely desirable – on both consequentialist and non-consequentialist grounds – when and because 1) a lot is at stake and doing the right thing matters (for example, when preventing huge and immediate harms), 2) choice architecture is inevitable (choices are always framed and structured in *some* way) and 3) there is no ignoring our increased knowledge about what influences decisions. Choosing to refrain from moral nudging is a decision itself with predictable but undesirable consequences (for example, causing harms to persist or foregoing desirable consequences). Given that the enormous challenges we currently face – like climate change – are due to human behavior, we urgently need to change that

<sup>17</sup> This refers to John Stuart Mill’s notion of “ethology” (Mill 1882), a science devoted to character formation based on empirical insights about the human mind (Ball 2000) and Liz Hurley’s more recent notion of “public ecology” (Hurley 2011). Both refer to the importance of environments that are conducive and supportive instead of impeding or being counterproductive (see also: Niker 2018: 153).

<sup>18</sup> We put aside here the huge question what the right (or wrong) path is and what constitutes a good (or bad) life. One justification for this is that we only focus on the means (the ‘nudge’ aspect) and not the ends (the ‘moral’ aspect) of moral education. Another is that we focus here only on inculcating uncontroversial moral principles, like not harming innocent others and helping those in need through no fault of their own.

behavior. If nudges work (better than alternative interventions), we have a reason to implement them, and we should not refrain from doing so on the grounds that this would involve a loss of moral worth.

## 5.2 Nudges that morally educate people

Imperfect moral human beings not only need scaffolding but also moral education. Given how nudges in educational settings have been discussed elsewhere (see for example: Damgaard & Nielsen 2018), we limit ourselves to a general claim here, namely that nudges are very suitable for moral education. Because we need moral education that is not indoctrinating (see: Croce 2019), in which nudges can play a pivotal role (Engelen et al 2018), moral nudges are valuable educational tools.

Given the long and winding path to full moral development, moral education takes time and effort. Aristotle emphasized (critical) habituation as a key ingredient in a virtuous person's character formation (see also: Niker 2018). In order to become courageous, for example, people need to learn what to fear, what not and why, and need a lot of practice: "we become just by doing just actions, temperate by doing temperate actions, brave by doing brave actions" (Aristotle, *Nicomachean Ethics*, 1103b1-2). One needs to do *what* one should do, before understanding *why* (Kristjansson, 2006: 110). Learning to 'do the right thing', even if not yet 'in the right state', is a first but necessary step in virtue cultivation. As Frank (2020: 380) puts it, such "moral technologies (...) may function as a kind of training aid, helping people acquire virtues thorough habituation".

Perhaps a visual illusion nudges you to slow down before a dangerous intersection or a cleverly designed bottle placed at eye-level pushes you into buying a cruelty free shampoo. Having done the right thing, without any moral considerations, is a first step in developing the habits, dispositions and virtues required by morality. Nudges can help to habituate people into driving carefully and buying cruelty-free products, but also into giving to charity and quitting smoking.

Now, the worry with nudges as educational tools is that they make people rely on external help and thus fail to cultivate the autonomous and reflexive capacities that morality requires (Niker 2018). Two of the four "compelling reasons to reject nudging" that Frank Furedi (2011) lists apply here: "1) It denigrates moral independence" and "2) It erodes our capacity to make judgments of value". Both objections are based on Kant's Enlightenment adagio that people should develop their capacities for independence and judgement needed for "the exercise of moral autonomy" and the practical wisdom to conduct a good life (Furedi 2011). Nudges arguably inhibit this cultivation process: even if they encourage people to do the right thing, they thwart the development of their moral capacities to tell right from wrong themselves.

Three responses are in order here. First, nudges can do much more than quasi-automatically steer us into doing the right thing, into *what* it is we should do. Nudges can also help us understand *why* we should act like this. Remember how Type 2 or "discernment-developing" nudges (Niker 2018) can help people appreciate the

morally relevant aspects of the choices they face and thus support the development of practical reasoning, i.e. of people's capacity to judge what is relevant and required in specific situations.

Second, and this is where we diverge from Niker (2018), even Type 1 nudges can actually help in this respect. Often, doing the right thing – for whatever reason – helps us to understand why we should act in this way. Hitting the brakes, even when done automatically, can help see how dangerous an intersection is and why caution is required. Reducing shower time, even when you only want to win a silly game, can help you reassess the trade-off between your comfort and the environment.

Third, it would be hard to imagine moral education without such assistance and habituating tools. Throwing people in the deep end of the pool is not the best way to teach them how to swim. Setting people up for success often works better than making them fail. While mistakes can provide great opportunities for learning and while there can be value in making occasional mistakes, as nudge critics Waldron (2014) and Furedi (2011) stress, you don't need to make mistakes in order to become a good person. Harming others – in traffic, for example – is in no way necessary to realize that this is bad.

In sum, the main claims in both subsections (5.1 and 5.2) are closely connected. As with training wheels, fully morally developed people perhaps no longer need the supporting nudges. Ideally, one could say, people gradually gain independence and take control over their own actions and circumstances.<sup>19</sup> But given how difficult it is to be(come) moral, support is likely needed and conducive to (the development of) people's moral capacities.

## 6 Conclusions

Moral nudging – deliberately (re)designing choice environments to facilitate genuinely moral thinking, feeling and acting – is both clearly possible and often desirable. Whenever people fail to act on their moral considerations because of their human (all too human) psychology and their (badly or randomly designed) choice environments, moral nudges can help trigger, ensure, scaffold and improve people's moral behavior and reasoning.

Moral nudging, so we have argued, is not an oxymoron, because morality does not require motivational purity and because nudges can leave intact or even promote people's attentiveness and responsiveness to moral reasons. When designed and implemented well, nudges can facilitate doing the right for the right reasons and, hence, promote moral worth (on both consequentialist and non-consequentialist understandings of this). While nudges can crowd out moral motivations in some

<sup>19</sup> Like Terence Ball (2000: 31), we refer to John Stuart Mill (17 ch. 2, §3) here: "If they could place us under the influence of certain circumstances, we, in like manner, can place ourselves under the influence of other circumstances. We are exactly as capable of making our own character, *if we will*, as others are of making it for us."



cases, as critics rightly stress, this is far from inevitable. We can and should try to smartly design people's moral ecologies, especially when they need help to be or to become true moral agents.

**Acknowledgements** Earlier versions of this paper were presented at the 2021 OZSW Annual Conference at Tilburg University (December 2021) and at the Centre for Philosophy, Politics and Economics of the University of Groningen (January 2022). We thank participants for their valuable questions and comments. We also thank two anonymous reviewers of this journal for their thoughtful comments and suggestions on an earlier version of the paper.

**Conflict of interests** The authors have no relevant financial or non-financial interests to disclose and have no competing interests to declare that are relevant to the content of this article.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Arieli, Dan. 2010. *Predictably Irrational, Revised and Expanded Edition: The Hidden Forces That Shape Our Decisions*. New York: Harper Collins.
- Aristotle (350 B.C.E.). *Nicomachean Ethics*. Translated by W.D. Ross. Available online: <http://classics.mit.edu/Aristotle/nicomachaen.html>.
- Ball, Terence. 2000. The Formation of Character: Mill's "Ethology" Reconsidered. *Polity* 33 (1): 25–48.
- Banerjee, Sanchayan, and Peter John. 2021. Nudge Plus: Incorporating Reflection into Behavioural Public Policy. *Behavioural Public Policy* 5717: 1–16.
- Baron, Marcia. 1995. *Kantian Ethics Almost Without Apology*. London: Cornell University Press.
- Bovens, Luc. 2009. The Ethics of Nudge. In *Preference Change: Approaches from Philosophy, Economics and Psychology*, ed. Till Grüne-Yanoff and Sven Ove Hansson, 207–219. New York: Springer.
- Byerly, Hilary, Andrew Balmford, Paul J. Ferraro, Hammond Wagner, Palchak Courtney, Polasky Elizabeth, Ricketts Stephen, H. Taylor, Aaron J. Schwartz, and Brendan Fisher. 2018. Nudging Pro-environmental Behavior: Evidence and Opportunities. *Frontiers in Ecology and the Environment* 16 (3): 159–168.
- Capraro, Valerio, Glorianna Jagfeld, Rana Klein, Mathijs Mul, and Iris van de Pol. 2019. Increasing Altruistic and Cooperative Behaviour with Simple Moral Nudges. *Scientific Reports* 9: 11880.
- Congiu, Luca, and Ivan Moscati. 2022. A Review of Nudges: Definitions, Justifications, Effectiveness. *Journal of Economic Surveys* 36 (1): 188–213.
- Croce, Michel. 2019. Exemplarism in Moral Education: Problems with Applicability and Indoctrination. *Journal of Moral Education* 48 (3): 291–302.
- Cureton, Adam, and Thomas Hill. 2018. Kant on Virtue: Seeking the Ideal in Human Conditions. In *The Oxford Handbook of Virtue*, ed. Nancy Snow. Oxford: Oxford University Press.
- Damgaard, Mette Trier, and Helena S. Nielsen. 2018. Nudging in Education. *Economics of Education Review* 64: 313–342.
- Dubbink, Wim, and Luc van Liedekerke. 2020. Rethinking the Purity of Moral Motives in Business: Kant Against Moral Purism. *Journal of Business Ethics* 167: 379–393.
- Engelen, Bart. 2019. Nudging and Rationality: What is There to Worry? *Rationality and Society* 31 (2): 204–232.
- Engelen, Bart, and Thomas R.V. Nys. 2020. Nudging and Autonomy: Analyzing and Alleviating the Worries. *Review of Philosophy and Psychology* 11 (1): 137–156.

- Engelen, Bart, Alan Thomas, Alfred Archer, and Niels van de Ven. 2018. Exemplars and Nudges: Combining Two Strategies for Moral Education. *Journal of Moral Education* 47 (3): 346–365.
- Entwistle, Tom. 2021. Why Nudge Sometimes Fails: Fatalism and the Problem of Behaviour Change. *Policy & Politics* 49 (1): 87–103.
- Frank, Lily Eva. 2020. What Do We Have to Lose? Offloading Through Moral Technologies: Moral Struggle and Progress. *Science and Engineering Ethics* 26: 369–385.
- Frey, Bruno. 2012. Crowding Out and Crowding In of Intrinsic Preferences. In *Reflexive Governance for Global Public Goods*, ed. Eric Brousseau, Tom Dedeurwaerdere, and Bernd Siebenhüner, 75–83. Cambridge (MA): The MIT Press.
- Furedi, Frank. 2011. Defending Moral Autonomy against an Army of Nudgers. *Spiked*. 20 January 2011. Available online: <http://www.spiked-online.com/newsite/article/self-help/10102>.
- Gigerenzer, Gerd. 2001. The Adaptive Toolbox. In *Bounded Rationality: The Adaptive Toolbox*, ed. Gerd Gigerenzer and Reinhard Selten, 37–50. Cambridge (MA): The MIT Press.
- Glod, William. 2015. How Nudges Often Fail to Treat People According to Their Own Preferences. *Social Theory and Practice* 41 (4): 599–617.
- Goldstein, Noah J., Robert B. Cialdini, and Vlasdas Griskevicius. 2008. A Room with a Viewpoint: Using Social Norms to Motivate Environmental Conservation in Hotels. *Journal of Consumer Research* 35 (3): 472–482.
- Gorin, Moti. 2018. Paternalistic Manipulation. In *The Routledge Handbook of the Philosophy of Paternalism*, ed. Kalle Grill and Jason Hanna, 236–247. London: Routledge.
- Gould, Carol. 1994. A Puzzle about the Possibility of Aristotelian Enkrateia. *Phronesis* 39 (2): 174–186.
- Gråd, Erik, Arvid Erlandsson, and Gustav Tinghög. 2021. Do Nudges Crowd Out Prosocial Behavior? *Behavioural Public Policy* 8: 1–14.
- Grüne-Yanoff, Till. 2016. Why Behavioral Policy Needs Mechanistic Evidence. *Economics & Philosophy* 32 (3): 463–483.
- Hagman, William, David Andersson, Daniel Västfjäll, and Gustav Tinghög. 2015. Public Views on Policies Involving Nudges. *Review of Philosophy and Psychology* 6 (3): 439–453.
- Hansen, Pelle Guldborg. 2016. The Definition of Nudge and Libertarian Paternalism: Does the Hand Fit the Glove? *European Journal of Risk Regulation* 7 (1): 155–174.
- Hansen, Pelle Guldborg, and Andreas Maaløe Jespersen. 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.
- Harris, John. 2013. Moral Progress and Moral Enhancement. *Bioethics* 27: 285–290.
- Hausman, Daniel M., and Brynn Welch. 2010. Debate: To Nudge or Not to Nudge. *Journal of Political Philosophy* 18 (1): 123–136.
- Herman, Barbara. 1981. On the Value of Acting from the Motive of Duty. *The Philosophical Review* 90 (3): 359–382.
- Huang, Pei-Hua. 2018. Moral Enhancement, Self-Governance, and Resistance. *Journal of Medicine and Philosophy* 43 (5): 547–567.
- Hurley, Susan. 2011. The Public Ecology of Responsibility. In *Responsibility and Distributive Justice*, ed. Carl Knight and Zofia Stemplowska, 187–215. Oxford: Oxford University Press.
- John, Peter, Graham Smith, and Gerry Stoker. 2009. Nudge Nudge, Think Think: Two Strategies for Changing Civic Behaviour. *The Political Quarterly* 80: 361–370.
- Johnson, Robert, and Adam Cureton. 2021. Kant's Moral Philosophy. *The Stanford Encyclopedia of Philosophy* (Spring 2021 Edition), Edward N. Zalta (ed.), Available online: <https://plato.stanford.edu/archives/spr2021/entries/kant-moral/>.
- Kahneman, Daniel. 2011. *Thinking, Fast and Slow*. New York: Farrar, Straus and Giroux.
- Kant, Immanuel. 1784/2009. Idea for a Universal History with a Cosmopolitan Aim. Allen Wood (trans.). In *Kant's Idea for a Universal History with a Cosmopolitan Aim*, ed. Amélie Oksenberg Rorty, and James Schmidt, 9–23. Cambridge: Cambridge University Press.
- Kant, Immanuel. 1785/1998. *Groundwork of the Metaphysics of Morals*. Mary J. Gregor (trans.), Cambridge: Cambridge University Press.
- Kant, Immanuel. 1797/1996. *The Metaphysics of Morals*. Mary J. Gregor (trans.), Cambridge: Cambridge University Press.
- Khan, Carrie-Ann Biondi. 2005. Aristotle's Moral Expert: The Phronimos. In *Ethics Expertise: History, Contemporary Perspectives, and Applications*, ed. Lisa Rasmussen, 39–54. Dordrecht: Springer.
- Kristjánsson, Kristján. 2006. Habituated Reason: Aristotle and the 'Paradox of Moral Education.' *Theory and Research in Education* 4 (1): 101–122.

- Levy, Neil. 2019. Nudge, Nudge, Wink, Wink: Nudging is Giving Reasons. *Ergo: an Open Access Journal of Philosophy* 6 (10): 281–302.
- Markovits, Julia. 2010. Acting for the Right Reasons. *The Philosophical Review* 119 (2): 201–224.
- Mill, John Stuart. 1863/1906. *Utilitarianism*. Reprinted from the Thirteenth London Edition. Chicago: The University of Chicago Press. Available online: <https://www.utilitarianism.com/utilitarianism-jsmill.pdf>.
- Mill, John Stuart. 1882. *A System Of Logic, Ratiocinative And Inductive*. New York: Harper & Brothers. Available online: <https://www.gutenberg.org/files/27942/27942-h/27942-h.htm>.
- Mills, Chris. 2015. The Heteronomy of Choice Architecture. *Review of Philosophy and Psychology* 6: 495–509.
- Moseley, Alice. 2020. Nudging in Public Policy. In *Oxford Research Encyclopedia of Politics*, ed. William R. Thompson. Oxford: Oxford University Press.
- Niker, Fay. 2018. Policy-Led Virtue Cultivation: Can We Nudge Citizens Towards Developing Virtues? In *The Theory and Practice of Virtue Education*, ed. Tom Harrison and David I. Walker, 153–167. London: Routledge.
- O’Meara, KerryAnn, Dawn Culpepper, and Lindsey L. Templeton. 2020. Nudging Toward Diversity: Applying Behavioral Design to Faculty Hiring. *Review of Educational Research* 90 (3): 311–348.
- Papies, Esther K. 2017. Situating Interventions to Bridge the Intention-Behaviour Gap: A Framework for Recruiting Nonconscious Processes for Behaviour Change. *Social and Personality Psychology Compass* 11 (7): e12323.
- Persson, Ingmar, and Julian Savulescu. 2012. *Unfit for the Future: The Need for Moral Enhancement*. Oxford: Oxford University Press.
- Pettit, Philip. 1995. The Virtual Reality of “Homo Economicus.” *The Monist* 78 (3): 308–329.
- Potter, Nelson. 1997. Supererogation and Overdetermination in Kant’s Ethics: Analysis and Interpretation at their Best in Baron. *Jahrbuch for Recht Und Ethik* 5: 489–496.
- Rebonato, Riccardo. 2012. *Taking Liberties: A Critical Examination of Libertarian Paternalism*. London: Palgrave Macmillan.
- Ridge, Michael. 2002. Mill’s Intentions and Motives. *Utilitas* 14 (1): 54–70.
- Ruehle, Rebecca, Bart Engelen, and Alfred Archer. 2020. Nudging Charitable Giving: What (If Anything) Is Wrong With It? *Nonprofit and Voluntary Sector Quarterly* 50 (2): 353–371.
- Schmidt, Andreas. 2019. Getting Real on Rationality — Behavioral Science, Nudging, and Public Policy. *Ethics* 129 (4): 511–543.
- Sheehan, Kim Bartel, and Joonghwa Lee. 2014. What’s Cruel About Cruelty Free: An Exploration of Consumers, Moral Heuristics, and Public Policy. *Journal of Animal Ethics* 4 (2): 1–15.
- Singer, Peter. 2015. *The Most Good You Can Do: How Effective Altruism Is Changing Ideas About Living Ethically*. New Haven: Yale University Press.
- Staake, Thomas, Verena Tiefenbeck, Samuel Schöb, and Anna Kupfer. 2016. *Effects of Real-Time Feedback on Hot Water Use*. Bits to Energy Lab, ETH Zurich and University of Bamberg. Available online: <https://www.amphiro.com/en/how-it-works/studies>.
- Sunstein, Cass R. 2016. *The Ethics of Influence: Government in the Age of Behavioral Science*. Cambridge: Cambridge University Press.
- Thaler, Richard H., and Cass R. Sunstein. 2008. *Nudge: Improving Decisions about Health, Wealth, and Happiness*. New Haven: Yale University Press.
- Thi Nguyen, C. 2020. *Games: Agency As Art*. Oxford: Oxford University Press.
- Waldron, J. 2014. It’s All for Your Own Good. *The New York Review*. October 9, 2014. Available online: <https://www.nybooks.com/articles/2014/10/09/cass-sunstein-its-all-your-own-good/>.
- White, Mark D. 2013. *The Manipulation of Choice: Ethics and Libertarian Paternalism*. New York: Palgrave Macmillan.

**Publisher’s Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.