Nudges, norms and moral progress

Ivanković, V.; Kudlek, K.

Mind & Society (2025) 24:713–737
https://doi.org/10.1007/s11299-025-00337-w

Nudges, norms and moral progress

Viktor Ivanković[1] · **Karolina Kudlek**[2]

Received: 14 January 2025 / Accepted: 20 August 2025 / Published online: 17 September 2025

Abstract
The compatibility of nudges with moral action and reasoning has become a focal point in philosophical discourse. While some argue that nudges may inhibit genuine moral responsiveness, others advocate their use to promote morally desirable outcomes, particularly in other-regarding contexts. This paper assesses whether other-regarding nudges are compatible with, and could contribute to moral progress, as conceived by moral and political philosophers. We contend that other-regarding nudges can be compatible with moral progress when they facilitate and reinforce a morally desirable social norm or help overturn a morally undesirable social norm, specifically when they (1) induce improvements in moral behavior, while (2) not worsening the quality of our moral reasoning or our moral beliefs as they were prior to the intervention. For nudges to achieve this, they must be easy to resist, they should not obfuscate moral reasons for action, and they should help individuals in saving up limited cognitive bandwidth. Our account of the moral progressiveness of nudging has three key strengths: it aligns with less stringent views on moral worth, complements accounts of moral progress emphasizing the insufficiency of moral reasoning, and mitigates stability concerns by anchoring nudges in desirable social norms. By demonstrating how heuristics-triggering nudges can facilitate moral progress, we offer a framework that reconciles their practical effectiveness with the demands of moral philosophy.

Keywords Nudges · Social norms · Moral progress · Behavioral ethics · Moral reasoning

✉ Karolina Kudlek
k.kudlek@phil.leidenuniv.nl

Viktor Ivanković
viktor.ivankovic@ifzg.hr

[1] Institute of Philosophy, Ulica grada Vukovara 54, Zagreb 10000, Croatia

[2] Institute for Philosophy, Leiden University, P.O. Box 9515, Leiden 2300 RA, Netherlands

1 Introduction

There is growing attention in the philosophical literature to whether nudges – behavioral influences that trigger cognitive heuristics to induce predictable behavior change in targeted individuals (Thaler and Sunstein 2008; Saghai 2013) – are compatible with moral thinking and acting (Niker 2018; Frank 2020; Engelen and Nys 2024). It is somewhat surprising that such accounts have only recently emerged, since some of the earliest and most well-known examples of nudges – like the organ donation opt-out default, the spillage-reducing urinal fly, and techniques promoting charitable donations – are *other-regarding*: their purpose is to benefit persons other than those being influenced and to encourage behavior often deemed morally desirable.[1] Additionally, several authors have argued in favor of employing nudges to achieve other-regarding objectives. Examples include: Jamie Kelly, who advocates nudges for benefitting the worst off in line with the Rawlsian difference principle (2013); Andrés Moles, who suggests using nudges to deter individuals from shirking enforceable duties (2015); and Michiru Nagatsu, who argues that nudging people into a "we-frame" of mind to pursue pro-social goals does not compromise preference coherence or personal autonomy (2015). Nevertheless, concerns persist among some of these authors that other-regarding nudges cannot be properly 'moral', the primary worry being that owing to the less-than-fully rational character of their influence on individuals, they inhibit responsiveness to, or diminish the appreciation of, the reasons required for truly moral action.

To examine this concern further, we draw in what follows on the philosophical literature on moral progress, and assess how other-regarding nudges relate to morally progressive social change.[2] Our aim is not only to show that other-regarding nudges can facilitate morally desirable outcomes, but to test whether they are compatible with moral progress (i.e., whether progress is undermined by them). The notion of being "compatible" with moral progress requires some unpacking. For an instance of social change to count as moral progress, narrowly conceived, it should not only bring about a more desirable outcome compared to a preceding state of affairs, but should include morally permissible means in bringing such an outcome about. If the means were not permissible, the very notion of such change being morally progressive would come into question. We claim that nudges can be one such permissible cause in bringing about morally progressive change. Note also that in vindicating this claim, we will not trouble ourselves with disagreements regarding what progressive moral aims are. The focus of the paper is whether nudges are permissible even if a proposed moral aim is agreed upon by all. We will hence simply assume that there are

[1] By contrast, the existing literature has mostly observed nudges through the broad lens of paternalism. Paternalistic considerations are self-regarding; the intended beneficiary is the person being influenced. For representative works adopting this approach, see, for example, Sunstein (2015), Le Grand and New (2015), Berg (2018) and Engelen and Ivanković (2025).

[2] For simplicity, we consider other-regarding nudges to be 'moral' both when they motivate action backed by universal moral obligations and special moral obligations (e.g., those arising from friendships and kinship relations).

at least some moral aims that are largely agreed upon and considered progressive in a liberal democracy, such as enforceable and non-enforceable moral duties.[3]

For their permissibility to be vindicated, nudges will need to face important challenges. One is the difficulty of stability – that nudges, typically conceived as mild influences that can easily be resisted, might not be able to "induce behavioural changes that should be radical and permanent" (Persson and Savulescu 2012, p. 79). Paradoxically, nudges also face the opposite difficulty of rigidity – that the behavioral patterns that they instill or help reinforce can become too socially rigid and inflexible for them to be considered morally progressive. We do not provide an exhaustive account of how nudges, broadly construed, may be argued to bring about stable and non-rigid morally desirable behavioral change. Instead, we offer a case for heuristics-triggering nudges being compatible with moral progress despite their supposed tension with traditionally conceived moral action.

Other-regarding nudges can be compatible with moral progress, we argue, when they facilitate and reinforce a morally desirable social norm or help overturn a morally undesirable social norm, specifically when (1) they induce improvements in moral behavior, while (2) not making the quality of our moral reasoning or our moral beliefs worse than they were prior to the intervention. To avoid worsening the quality of moral reasoning or moral beliefs, such a nudge should be easy to resist (it should not block off or overburden alternatives), it should not obfuscate moral reasons for action, and it should help individuals in saving up limited cognitive bandwidth. We show that such a case for the moral progressiveness of nudging has three strengths: (1) it aligns with less demanding views of when actions lose moral worth (Mill 1863/1906; Raz 1999); (2) it complements accounts of moral progress pointing to the insufficiency of moral reasoning (McTernan 2014; Tam 2020; Klenk and Sauer 2021; Rehren and Blunden 2024); and (3) it avoids stability concerns by latching onto desirable social norms. If such nudges can be deemed progressive, then this may have important implications for policymakers, but the primary aim of the paper is conceptual. It is to show that, in facilitating a morally desirable state of affairs, moral nudges do not undermine its characterization as morally progressive.

We proceed as follows. In the next section, we illustrate several accounts elaborating the character of nudges that could be considered 'moral', show how they fit into views regarding when actions lose their moral worth, and specify areas of moral action in which they could conceivably be implemented, i.e., those in which morally progressive aims are the least controversial. We close off the section by specifying other ways in which nudges may complement moral progress, but ones that we do not explore here. Then, we provide a detailed account of how and why nudges can be hoped to support morally progressive aims. We specify two kinds of moral nudges – social nudges and situationist nudges – that, in most of the examples provided, could promote moral progress in the ways we specified here. We also show how nudges may help us achieve – both individually and socially – the offloading of our moral tasks. To end the section, we face off our view against the objections of stability and rigidity. Next, we point to recent contributions on the importance of social norms in

[3] We elaborate this at length in Sect. 2.3.

the achievement of moral progress, which moral nudges can reinforce, if norms are desirable, or help overthrow, if they are undesirable. The final section concludes.

2 Moral nudges

Before assessing whether nudges can play a role in morally progressive behavioral shifts and tendencies, we must first attend to whether nudged behavior can ever be considered 'moral' in a meaningful sense. Although other-regarding nudges promote behavior that is typically considered morally desirable, and do so quite effectively (Mertens et al. 2022), some general objections to nudges cast doubts on whether nudged behavior could aspire to truly moral action. Because nudges, or at least their heuristic-triggering variety, are often criticized for bypassing the reasoning of influencees (Mills 2015), diminishing their control over their deliberations (Hausman and Welch 2010), or infantilizing them (Bovens 2009), they are standardly perceived as failing to (explicitly[4]) appeal to moral reasons. An other-regarding nudge can be conceived as having one of two effects: (1) bringing the nudgee's behavior into alignment with their own moral beliefs (either by bypassing reasoning or enhancing motivation to act upon a belief), or (2) steering behavior irrespective of any underlying beliefs. The worry is that such effects might not be sufficient for developing the kind of practical reasoning associated, for instance, with cultivating moral virtue (Niker 2018, p. 153). While the intentions behind other-regarding nudges are often morally minded, they might often at best lead influencees to *conform* with certain moral obligations, rather than genuinely *comply* with them. Mere moral conformity may be incompatible with truly moral action (and, by extension, with moral progress) (Buchanan and Powell 2018, p. 68). We elaborate how these objections have recently been resisted.

2.1 Accounts favoring the possibility of moral nudges

Several accounts have argued that nudges can be moral in a meaningful sense. Most recently, Bart Engelen and Thomas Nys (2024) contended that moral nudges can help close, or at least alleviate, the "intention-behaviour gap" (Papies 2017) for moral principles that most individuals endorse, but often fail to enact. Other-regarding nudges can supposedly help people "behave more consistently with their own deeply held moral convictions" (Engelen and Nys 2024, pp. 16–17), although, they believe, not all nudges accomplish this. The organ donation default, for instance, does not prompt in agents "any moral consideration" (Engelen and Nys 2024, p. 18), since the morally desirable outcome results from mere inaction. Some other techniques do not foreclose the opportunity for moral engagement entirely, but may crowd out moral motives with prudential reasoning, like considerations of social status, and in even more problematic cases, techniques may accomplish moral ends by increasing the salience of morally questionable reasons, like exploiting "sexist biases to induce men to hire more women" (Engelen and Nys 2024, p. 11). The morality of such nudges,

[4] For the discussion on whether and in what way heuristics-triggering nudges could in fact appeal to reasons, see Levy (2019–2020) and Douglas (2022).

Engelen and Nys contend, is undoubtedly suspect. However, they argue that a set of other heuristics-triggering nudges can habituate moral action and help individuals reflect on moral reasons after the fact. Despite the usual notion that people learn from their mistakes, Engelen and Nys believe that success can also be instructive: "Setting people up for success" does not preclude the opportunity for moral learning (Engelen and Nys 2024, p. 24).

Lily Eva Frank (2020) defends the possibility of moral nudging in a similar vein. She argues that nudges can help people overcome not only the intention-behavior gap, but also their own "cognitive and affective limitations" in engaging with the moral dimensions of situations (2020, p. 371). Crucially, by making it easier for people to act in accordance with some supposed morally correct course of action, moral nudges allow for "moral offloading" (2020, p. 373), wherein some portion of their reflective moral work is done for them. Frank thus seems to propose a less demanding standard for engaging with moral reasons when agents operate morally; it seems sufficient that they act morally well in the appropriate situations. For her, moral nudges – alongside other moral technologies – obviate the need for "difficult moral deliberation and internal moral struggle," which may at times lead to moral overload and burnout (2020, pp. 373–375). Unburdening agents from such struggle not only lowers their stress and improves their well-being, but also frees up limited cognitive resources for situations where moral action is not facilitated (2020, pp. 377–378). Like Engelen and Nys, Frank argues that moral nudges can habituate moral action as a training aid that allows people to cultivate virtue: if people were to perform virtuous action before being virtuous, but in the hopes of becoming so, they would not be acting impermissibly (2020, p. 380). Finally, Frank resists the Kantian objection that, under the influence of nudges, individuals merely act *in accordance with* duty, but not *from* duty. Moral worth is preserved on Kantian standards, she argues, "as long as one holds a standing commitment to respect for the moral law and duty. Living with the help of moral technologies can be done out of respect for the moral law itself" (2020, p. 382).

Finally, Michiru Nagatsu has mounted a defense of social nudges (2015). While he never explicitly refers to these as 'moral', there are two reasons why his account is relevant in making a case for the possibility of moral nudges. First, the aim of Nagatsu's social nudges is morally minded in the sense described earlier – it is to steer individuals toward promoting or safeguarding public goods, which are typically pursued, at least in part, due to moral motivations. Second, Nagatsu defends social nudges against precisely the kinds of objections that raise doubts about the possibility of 'moral' nudging. Most importantly for our purposes, he addresses the autonomy concern – the worry that a nudge may steer an individual's behavior while bypassing their capacity to appreciate reasons (2015, p. 484). Following Bacharach (2006), Nagatsu claims that the willingness to cooperate in collective action scenarios depends on whether individuals expect others to cooperate as well, and on whether they are facing the decision in an "*I*-frame" ("What should *I* do?") or a "*we*-frame" ("What should *we* do?"). When they anticipate others to do their share and engage the decision in a *we*-frame, individuals are more likely to opt for a pro-social course of action. Social nudges can facilitate this shift in mindset. Nagatsu argues that the "Don't Mess With Texas" anti-littering campaign, which appealed to Texan pride rather than reasons against littering, was one such social nudge. Its success, he

claims, can be explained by the expectation that enough Texans would conform to the appeal, and by the engagement of individuals with the frame of how they, as Texans, should behave; such a nudge thereby engineers a social norm (Nagatsu 2015, pp. 488–489).[5] Does changing the frame undermine the autonomy of decision-makers? According to Nagatsu, it does not. Social nudges engage individuals in practical reasoning through "the right kind of belief-desire psychological process" (2015, p. 489). This kind of practical reasoning, Nagatsu concludes, does not deviate from the kind of reasoning we typically engage in as authors of our own choices, and, hence, does not undermine our personal autonomy (2015).

2.2 Moral worth

The worry about the lack of appreciation of moral reasons associated with nudged behavior stems from the long-standing debate about moral worth. Kant (1785/1998, 4, p. 398) argued that actions not performed *from* duty lack moral worth; the presence of other motives steering the agent in the direction of the duty might obscure what drives the action, even to the acting agent. If nudges undermine the appreciation of moral reasons that give rise to such duties by engaging nudgees with non-moral reasons or failing to engage them with any reasons whatsoever, then nudgees may merely *conform* to moral duties rather than *comply* with them, with moral worth thereby being thwarted. To conform with a reason, we need only act in the direction favored by the moral reason, whereas to comply requires our action to be *guided specifically by* the moral reason, that is, in accordance with *why* the action is favored by that reason. Imagine that Martha is dreaded by the mere sight of doctors, and her friend George can allay her anxiety by tagging along for a check-up at her general practitioner. George would be complying with a moral reason if he were motivated by the fact that his presence would ease Martha's distress, but he would be merely conforming if he was using the opportunity to spend this time reading a novel in the waiting room.[6]

Engelen and Nys' and Frank's accounts seem to push back against the strictness of the Kantian view. First, they point to interpretations of Kant suggesting that it is sufficient for individuals to commit to an overarching maxim and act in accordance with it (Potter 1997, p. 494, in Engelen and Nys 2024, p. 15), or to adopt a moral technology that is compatible with respect for the moral law (Frank 2020, p. 382). Second, Engelen and Nys point to a less demanding account of when moral actions lose moral worth, namely, that of John Stuart Mill. Mill argued that it is permissible to be driven by non-moral motives when performing a duty, unless the duty itself condemns such motives (1863/1906, p. 26). It is sufficient for Mill that individuals intend to perform a morally desirable act without the moral motivation being the sole driver of their behavior; in fact, he contends, the great majority of our actions are driven by

[5] We attend to the relation between other-regarding nudges and social norms later in the paper.

[6] For Aristotle, virtuous action is similarly produced "according to the right reason […] at the right times, with reference to the right objects, towards the right people, with the right motive, and in the right way" (2009, p. 1106b20-22).

non-moral motivation.[7] Engelen and Nys apply this reasoning to contexts like traffic safety, where it seems of little relevance to moral worth whether safe driving 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 […] In domains like this, […] morality doesn't require motivational purity" (Engelen and Nys 2024, p. 8).

Joseph Raz offers another lax view about moral worth. We primarily have reasons, Raz says, to ensure that moral reasons for action are conformed to, be it through compliance (which trivially implies conformity) or through some other motivation. He emphasizes the case of omissions: although we continuously have moral reasons not to kill people, we rarely consciously engage with these reasons, since the thought of killing (hopefully) never occurs to us. Raz believes that the absence of this thought, which implies not vividly engaging with the moral reasons not to kill (and thus merely conforming), is in fact more admirable than vividly engaging with the reasons and deciding not to kill as a result (Raz 1999, pp. 180–182). But it is unclear why we would not apply the same reasoning to certain actions as we do to omissions. In the case of actions for which moral reasons frequently arise or continually lurk in the background, it might be more admirable to conform to them habitually, as though they were part of our second nature, than to re-engage with reasons each time the opportunity for said action arises. We may think that if agents have no reason to doubt certain morally required actions, it could be morally preferable that these are carried out almost instinctively. Moral nudges may help individuals develop such moral habits and instincts. In Sect. 3, we explore how developing these moral habits and instincts can complement the appreciation of reasons in achieving moral progress. Before that, we illustrate the areas in which moral nudges could most plausibly be utilized.

2.3 The areas of moral nudging

In this section, we introduce several areas in which moral nudges may steer behavior in seemingly morally desirable directions and take part in moral progress. Note that our aim is to ascertain whether moral nudges can be compatible with moral progress, particularly in light of their supposed tension with standardly conceived moral action; it is not to settle the correct aims of morality. Some might object that the directions of nudges we describe are not clearly morally progressive, or that we lack a moral standard for assessing the validity of moral aims. While this could be true, we think that, for the purposes of this paper, there is a more interesting puzzle concerning whether and how nudges can be part of a morally progressive behavioral shift, regardless of whether the pursued aim is morally controversial or uncontroversial. For this reason, we simply assume, following Persson and Savulescu, that a functioning society

[7] Mill admits, however, that whether such actions are motivated by moral motives may reflect on moral character, which is relevant for utilitarians, insofar as it bears on whether individuals will be disposed to repeat such actions in the future, and thereby maximize morally desirable consequences. Yet some doubts may be raised as to whether Mill's consequentalist evaluation is correct or sufficiently nuanced. McTernan, whose account we attend to later, argues that moral progress may be more easily accomplished via social norms than through the cultivation of moral virtues (2014).

largely agrees on moral aims regarding harm, easy rescue, and reciprocity, and that morality "requires *more* than most of us do" (Persson and Savulescu 2015, pp. 52–53; emphasis in original). Our claim in the paper can be conceived as a conditional: "If X could be agreed upon as a valid moral aim, then nudges facilitating said X would be compatible with the moral desirability of X". Our task, then, concerns the evaluation of moral nudges themselves, not the aims that they are promoting. Below we simply assume that X is occupied by some aim that members of a liberal democratic society are most likely to morally converge upon. Similar approaches have been undertaken by authors endorsing moral bioenhancement, who claim that there are "points of overlapping consensus among competing, reasonable moral perspectives" that should allow us to target traits like reduced empathy, significant prejudice, weak will and the inability to focus on unpleasant realities (DeGrazia 2014, p. 364).[8] Those discussing moral progress have also mostly agreed that some events in history can uncontroversially be characterized as morally desirable directions, like the abolition of slavery, extending suffrage to women, or the decriminalization of homosexuality (Sauer et al. 2021, p. 1). In our attempt to assess the permissibility of moral nudges, we carefully follow in these footsteps.

Consider first that individuals often endorse certain morally permissible courses of action but fail to live up to their own commitments. This gives rise to a profusion of intention-behavior gaps in areas such as environmental behavior [e.g., green energy practices (Pichert and Katsikopoulos 2008) or preserving bird species (Kahneman et al. 1993)], organ donation (Kurtz and Saks 1996) and vegetarianism (Loughnan et al. 2010, p. 156; Schwitzgebel and Rust 2014, p. 97). Although accounts that explain each specific gap tell an elaborate tale, we can usually pin the effects onto three broad factors. First, the current state of our moral psychology might not be up to the task: although we might wish to help those beyond our immediate social circle, we are often unmoved by the plights of "strangers, distant in space or time" (Persson and Savulescu 2019, p. 8), and our compassion often fails to scale with the numbers of suffering people and animals (Persson and Savulescu 2012, p. 30, pp. 62–63; Desvouges et al. 1993).[9] Second, living up to a commitment can be obstructed by choice architectures that are already in place: for instance, T. J. Kasperbauer observes that "[e]nergy consumption is massively influenced by external factors and is thus already subject to nudge-like interventions", such as peer pressure that often draws us away from energy conservation and solar panel adoption (2017, pp. 52, 55). Third, individuals can fail to follow through on their commitments whilst "in the grip of unjust social norms", in which case unjust social expectations deter compliance with moral reasons (Tam 2020, p. 79).[10]

[8] Similarly, Thomas Douglas has claimed that attenuating a strong aversion to other racial groups or the impulse of aggression could be claimed uncontroversially to qualify as moral enhancements on virtually all plausible moral and psychological theories (2008, p. 231).

[9] In the debate on moral progress, Klenk and Sauer (2021, pp. 938–941) highlight the pernicious effects of biases and distortions of this kind on the ability of individuals to exert control over their moral cognition and convert their moral conclusions into genuinely morally progressive behavior.

[10] Motivational internalists and externalists debate over whether it is possible to have genuine moral beliefs and commitments with hardly any implications for the mustering of moral motivation. We set the issue aside here. Svavarsdottir (1999) and Rosati (2016).

Consider a person who believes she should do more for environmental protection or poverty alleviation and is willing to bind herself to a nudge that will help her overcome the above-mentioned obstacles but does not, by hypothesis, appeal explicitly to moral reasons (see, e.g., Douglas 2013, p. 163). Would the nudge undermine her appreciation of moral reasons and make her non-compliant? Unless we require that moral reasons "vividly resonate" at all times whilst the action is being performed, which seems overly demanding, using moral nudges as a crutch appears sufficient for what we will call 'distant compliance'. Binding ourselves to moral nudges, as we would to a self-nudge in our private lives (Lades 2014), would then not seem to undermine our appreciation of moral reasons. We will address some objections to this later.

Another area in which moral nudges might be appropriate is the fulfillment of duties. Some duties are enforceable – governments can impose coercive rules to see them respected and discharged. For liberals, inspired by Mill's proposal (1859/1991, p. 14), one such duty is to not cause others harm or prevent, at minimal cost to oneself, that others suffer harm. Assume, however, that individuals have difficulties discharging an enforceable duty, and that the government is unable to codify effective rules that would coerce people into abiding by it. For instance, assume that traffic, an area with great potential for causing harm to others, is fraught with rules and guidelines that citizens readily acknowledge,[11] but which fail to secure the intended outcome – that they avoid behaviors harming others. It would seem nudges are permissible in getting citizens to conform to these rules. The duty to prevent harm in traffic is akin to the duty of easy rescue. Using moral nudges to prompt people to rescue others at little or no cost to themselves in at least some circumstances would be permissible, since we would be getting them to conform in ways already prescribed by an enforceable moral duty (Clayton and Moles 2018, p. 241). If agents were aware of the cognitive shortcomings preventing them from abiding by such duties, they would have good reasons to accept moral nudges that bear little or no costs to themselves (Persson and Savulescu 2019, p. 7). And even if they did not endorse such nudges, they would hardly have reasonable complaints against them, especially if the absence of the nudge entailed a failure to abide by the duty. This could also be extended to non-enforceable moral duties (Moles 2015, p. 660), like environmental protection, in which case the government fails to produce any set of practicable rules ensuring that they are discharged. If moral nudges could reduce the intention-behavior gap and increase abidance with enforceable and non-enforceable moral duties, then this is where they will likely extend their candidacy to take part in moral progress.

2.4 Roads to progress not taken

Finally, before assessing whether heuristic-triggering nudges could conceivably be part of morally progressive shifts, we set aside three alternative strategies to potentially achieve this. This is not because these strategies are not viable; in fact, all three are independently plausible. Rather, they are set aside because we are trying to assess

[11] However, endorsement might not be relevant for these kinds of cases, since agents are to discharge enforceable duties regardless of whether they endorse them.

the possibility of moral progress in light of the most morally controversial aspect of moral nudges: their ambiguous relationship with moral action. The potential threat of nudges to the appreciation of moral reasons is at the heart of the skepticism regarding their compatibility with moral progress. We explore whether moral progress is possible despite this tension. The three other strategies pursue moral progress in cases where the tension is altogether absent, or where progress is not a consequence of standardly conceived moral action.

The first two strategies rely on the idea that other-regarding nudges could be morally progressive because they, in fact, *do* appeal to moral reasons, and are thus able to improve both the behavior and the reasoning of the influenced individual. The first strategy adopts the broad definition of nudges, according to which any alteration in choice architecture that predictably results in behavior change while preserving the option set counts as a nudge, regardless of whether it bypasses reasoning. On this broad reading, interventions like information-giving, disclosures and reminders count as nudges (Sunstein 2013, p. 42). While such nudges could conceivably be compatible with moral progress, they overlap too closely with standard forms of reasons-giving to be theoretically interesting and challenging. Some other techniques, like Oxfam-style pictures of destitute children that move us to donate, could boost motivation, while plausibly highlighting relevant reasons for moral compliance. These interventions not only seem non-threatening to the appreciation of morally relevant reasons, but may in fact enhance such appreciation, so we set them aside here.[12]

The second strategy follows Neil Levy's proposal that nudges do not bypass reasoning capacities because they do, in fact, provide reasons for action. While these reasons might not be consciously recognized, it is because of them that behavior is altered (Levy 2019). For instance, setting up a default on Levy's account is communicative: "people frame options in ways that highlight particular choices because they take them to be good ones, and their communicative intent is recognized by those who respond to the framing" (2019, p. 290). As with the first strategy, if Levy's proposal holds, the tension between other-regarding nudges and the appreciation of moral reasons would be altogether defused. Thus, in order to attend to the (im)possibility of moral progress in light of the tension, we assume that at least some moral nudges do not appeal to moral reasons, at least not explicitly or primarily.

The third and final road not taken here is Jeremy Evans's account of moral progress (Evans 2017). On this view, the most reliable indicator of moral progress across cultures is an increase in mean population welfare. Since the most prominent case for government nudging has been grounded in the promotion of welfare, nudges could be regarded morally progressive insofar as they demonstrably advance welfare in a sufficiently consistent manner. We do not contest this view. It is entirely plausible that welfare-promoting, self-regarding nudges may play a significant role in securing the welfare-related conditions in which moral progress takes place. Yet, like the first

[12] Niker (2018) seems to refer to such techniques, which help "discern the [moral] aspects of a situation that are relevant for choice-making", when she examines the capacity of nudges to facilitate virtue cultivation. It would seem permissible to rely on such techniques and allow for virtue cultivation, but some caution about whether such virtues can deliver moral progress should be exercised. McTernan shows, drawing on social psychology, that virtues may have a limited capacity in delivering desirable outcomes (2014). Additionally, these nudges would not take part in moral offloading, which we elaborate on in Sect. 3.3.

two strategies, this approach sidesteps the core tension motivating our argument in this paper.

3 Nudged into moral progress

There are three distinctions in the moral progress debate that are relevant to whether moral nudges are compatible with moral progress: broad vs. narrow, individual vs. collective, and cognitive vs. behavioral moral progress. First, *broad* moral progress concerns changes that are considered improvements from a moral point of view (such as the promotion of welfare or reduction of crime), but, unlike narrow moral progress, do not include the exercise of moral powers (moral reasoning, moral motivation, moral virtues) (Buchanan and Powell 2018, pp. 48–50). Second, moral progress can be understood either as reducible to *individual* moral improvement (Singer 1981) or as a distinctly *social* phenomenon, i.e., an improvement of collective morality that manifests in social institutions (Musschenga and Meynen 2017, p. 5; Macklin 1977, p. 370). Third, we distinguish between progress achieved through improvements in moral reasoning, that is, by understanding moral concepts, values and principles (Bodlović and Kudlek 2025), and progress in affective and behavioral components of morality (Musschenga and Meynen 2017, p. 5; Moody-Adams 1999, pp. 169–170).[13] We are interested in the capacity of moral nudges to be progressive on the narrow conception, to improve morality on both individualist and collectivist understandings, and to improve the behavioral component of morality without undermining the cognitive component. In what follows, we examine this capacity.

3.1 Nudging, moral behavior, and moral reasoning

We proposed at the outset that an other-regarding nudge is compatible with moral progress when (1) it can induce improvements in moral behavior, while (2) not making the quality of our moral reasoning or our moral beliefs worse than they were prior to the intervention. That nudges can induce improvements in morally desirable behavior is evident from their demonstrated effectiveness. However, the improvement need not be optimal. The progressiveness of the nudge is assessed relative to the status quo, and not to some hypothetical state of affairs in which behavior is improved to an even greater degree. The aim of the paper is to make sense of the progressiveness of moral nudges, not to identify the best possible moral nudge.

Condition (2) calls for further clarification. We do not suggest that moral progress can occur without any kind of baseline of moral reasoning or moral beliefs. Nudges would not be progressive if they improved behavior in terms of a morally desirable direction, while moral reasoning was altogether absent, i.e., if agents were completely oblivious to any reasons that would justify moral action. If that were the case, only broad, not narrow, moral progress would obtain. Instead, we assume that individuals and collectives possess some awareness of reasons that are meant to

[13] This corresponds to the conceptual vs. practical distinction often brought up in the moral progress literature.

guide moral behavior, but often fail to do so. A moral nudge compatible with moral progress will thus improve behavior without worsening this awareness, however developed it may be.

Some might argue that a moral nudge would still be progressive if it significantly increased social conformity with a moral reason for action, while only slightly worsening agents' appreciation of said reason. For instance, a nudge might significantly boost citizens' donations to charity, but ever-so-slightly suppress their engagement with reasons for doing so. Although there is some intuitive pull to this suggestion, we resist it here. It is difficult to assess whether improvements (and deteriorations) in moral behavior and in moral reasoning are commensurable, and how to make sense of trade-offs between them. We would face a similar uncertainty if we collectively advanced our moral reasoning, but at such a cost in time or energy that, consequently, our moral behavior (slightly) worsened. Thus, cases where regression in moral reasoning is weighed against progress in moral behavior as a result of a nudge are not clear candidates of nudges being compatible with moral progress.

Some accounts of moral progress suggest that progress in moral reasoning and moral behavior are intimately linked (Hermann 2019), and that both are to be improved if "full progress" is to be achieved (Klenk and Sauer 2021, pp. 942–943). Our proposal is more minimalistic: we only require that a nudge not worsen moral reasoning. But the consequences of a nudge could well satisfy what prominent accounts of moral progress require. We do not rule out the possibility that, by facilitating certain moral behaviors, nudges might indirectly highlight moral reasons for action and foster changes in moral beliefs. Nudgees will often observe *ex post* that the behavior a nudge has prompted highlights reasons to behave similarly in the future. As Engelen and Nys aptly observe, "doing the right thing – for whatever reason – helps us to understand why we should act in this way" (2024, p. 24). As one type of institutional reform, nudges can achieve moral progress in reasoning as institutional features that can transform moral attitudes (Sauer 2019) and shape individual choice and agency (Madva et al. 2023), even if only indirectly and after the fact. This is one possible version of what Sauer et al. call the institutional bypassing of the limits of our moral psychology (2021, p. 6). Note, however, that we are not committed to the view that every moral nudge must improve reasoning in this way. If we had to pick between two moral nudges – one significantly improving behavior while leaving moral reasoning intact, and another only slightly improving behavior but also indirectly improving reasoning – our account does not require favoring the latter.

3.2 Two kinds of moral nudges

There are two kinds of moral nudges that differ in whether and how they fulfill the requirements for moral progress outlined for them. Consider the following two groups of examples:

1) A message in a hotel bathroom pointing out that other guests reuse towels (Nolan et al. 2008).
2) Telling people that their energy consumption exceeds that of their neighbors (Schultz et al. 2007).

3) Telling taxpayers that others have already filed their taxes (Behavioral Insights Team 2012).
4) A robot greeting passengers as they enter a train station (Borenstein and Arkin 2016).

As compared to:

1) Setting up a default vegetarian menu in a cafeteria (Campbell-Arvai et al. 2012).
2) Setting up a double-sided printing default (Egebark and Ekström 2013).
3) Setting up an organ donation default (Beraldo and Karpus 2021).
4) Setting up a default for green energy use (Pichert and Katsikopoulos 2008).
5) Painting lines on the road further apart and then closer together (Thaler and Sunstein 2008, pp. 37–39).
6) Getting people to sign their tax forms before filling them out (Behavioral Insights Team 2012).
7) Placing images of eyes in school hallways and classrooms (Haley and Fessler 2005).
8) Painting images of babies in streets with high crime rates (Berg and Kim 2019).
9) Adjusting lighting in streets with active nightlife (De Kort et al. 2014).

The relevant distinction here is between *social* nudges and *situationist* nudges (McTernan 2014, p. 100). The rationale behind social nudges is to get people to conform with moral ends by appealing to social rather than moral reasons, specifically that others like us are engaging in the same kind of action.[14] Like Nagatsu shows, these nudges signal that others are doing their part and that we should reciprocate. They can help establish or reinforce a social norm simply by pointing out what others are doing and what is expected of us, by explicitly appealing to an existing norm, or by increasing the prevalence of norm-abiding behavior.

Although some may find it questionable to have moral progress achieved on the back of seemingly conformist behavior or peer pressure, Agnes Tam argues that social reasoning of this kind extends beyond mere prudential reasoning, as it immerses agents in a "we-reasoning" that provides several social membership goods, like group cohesion, social trust and fellow-feeling, the normative force of which is not to be underestimated for persons (2020). In the words of Mols et al., "humans are social beings who derive meaning" and "significant value […] from identity-affirming behaviour" (2015, p. 89). Thus, social nudges may reinforce valuable social norms, or loosen the grip of detrimental norms, engaging nudgees in a socially desirable we-reasoning.

[14] There may be some difficulty in teasing apart social reasons from particularistic moral reasons (those arising from our special relationships, such as friendship). It is true that these are sometimes difficult to tell apart. Acting on particularistic moral reasons may often involve a we-frame of reasoning, much like acting on social reasons. But particularistic moral reasons are always backed by universal moral reasons about what we owe to one another when taking part in these special relationships. By contrast, acting on social reasons may entirely lack this component; what matters is merely what we expect from one another as members of a community, not what universal morality requires. We thank an anonymous reviewer for pressing us on this point.

It is worth noting that, while social nudges appeal to social reasons for action – namely, that we should engage in activities that relevant social others are also engaging in – moral reasons for these actions remain transparent. It is typically obvious what the moral reasons behind these nudges are: promoting energy conservation, meeting an enforceable duty like paying taxes, or fostering social etiquette. Insofar as nudgees recognize these reasons lurking in the background, they can adopt them as reasons driving their community's collective moral cause. And, crucially for our case, the transparency of these reasons may significantly bear on the assessment whether moral reasoning is undermined. In short, moral reasons are not crowded out in action here by social reasons. When agents act on a combination of such reasons, their moral reasoning is not undermined, so moral progress can be achieved on our account.

By contrast, the moral reasons in at least some examples of situationist nudges are somewhat less transparent. For instance, citizens may easily fail to notice changes in street lighting altogether, let alone infer that its purpose is to reduce aggression. Nor are they likely to deduce that signing their tax form before filling it out reduces the likelihood of false disclosures. For some situationist nudges, then, the background moral reasons remain opaque, leaving individuals unaware of why they are engaging in certain behaviors. These situationist nudges could thus plausibly worsen the quality of moral reasoning. Moreover, it would be odd to consider defaults as facilitating anyone's progress, individual or collective, at least when they represent cases of inaction in which moral reasons are mostly unappreciated or appreciated less than before.

But these problems do not afflict all cases of situationist nudging. Noticing the change of default, as in the case of the vegetarian menu, can alert nudgees to the underlying moral reasons behind the use of the technique, leaving the quality of moral reasoning roughly unchanged. Or consider that those influenced by the lines-on-the-road technique to reduce their speed will often grasp its purpose once it has taken effect on behavior. Furthermore, whether citizens are aware of the moral reasons behind particular techniques will often depend on how much those reasons are publicly discussed, or how easily they can attain information about them (Ivanković and Engelen 2019; Schmidt 2017). Finally, if these nudges are aligned with and facilitate a desirable social norm, then they increase its prevalence, thereby boosting social signals about acceptable behavior and we-reasons.

Two objections may be raised at this point. First, recall that the areas we have identified as appropriate for nudging to take part in morally progressive change are commitments and enforceable duties. In such cases, individuals will typically be aware of the moral reasons for action, but may struggle to act on them. If nudges could move them to follow through, then individuals will have achieved what we have called a distant compliance. But one worry is that distant compliance might not suffice for maintaining morally unified agency. According to Christine Korsgaard (2009), a morally unified agent must act from principled ends to sustain moral identity and personhood. Building on this idea, Christian Schubert expresses concern that nudges are incompatible with morally unified agency, given their tendency to discourage active choice (2017). So, if choices prompted by nudges are not in the proper sense 'active', then morally unified agency requires vivid, and not just distant compliance.

But this seems much too demanding. Allowing some moral choices to be secured through clever choice designs hardly seems sufficient to constitute a breach in morally

unified agency. If it did, breaches would occur all the time, e.g., whenever individuals fail to live up to their principled ends due to cognitive shortcomings. Moreover, moral nudges are proposed here in a limited range of cases, leaving ample room for individuals to engage in vivid compliance in other domains.

The second objection is that the nature of nudging – namely, that it triggers cognitive heuristics into what is often semi-conscious or entirely unconscious action – is such that we cannot seriously insist that the quality of moral reasoning remains intact, and thus hope to achieve any version of moral compliance. Our response is that nudging leaves compliance *no worse* than it would have otherwise been under the influence of existing cognitive heuristics. Recall that prior to being exposed to a nudge, individuals are already subject to biases that may be detrimental from a moral point of view – planning fallacy, hindsight bias, scope neglect, etc. Nudges do not detract individuals from moral reasoning any further than standardly occurring biases do. If so, the level of engagement with moral reasons under the influence of a nudge (i.e., a distant compliance) is a mere continuation of the prior level of engagement, albeit one that increases the likelihood of moral conformity with desirable moral ends.[15]

3.3 Moral offloading

In Sect. 2.1., we mentioned Frank's proposal that moral nudges, like many other moral technologies, may provide individuals with a form of moral offloading, relieving them of the mental distress and exhaustion associated with moral struggle (2020). We now want to suggest, following Frank, that the achievement of moral offloading through moral nudges is also an important holistic point about the progressiveness of moral nudges in the more general scheme of moral reasoning, both from individualistic and collectivist understandings of moral progress. Consider that vivid compliance with moral reasons often requires the investment of cognitive resources. Some obligations will require such vivid compliance, like tending to close social relationships or reflecting on reasons in cases of moral uncertainty. The obligation to appreciate such reasons will thus be a significant constraint on how we should organize our cognitive resources and, by extension, our mental lives.

Other obligations do not obviously require vivid compliance, and they seem sufficiently settled that performing them "vividly" would hardly lead to value revisions. Nudging the fulfillment of such obligations might save up cognitive resources for more contested cases of moral reasoning and action. By prompting the fulfillment of socially established obligations, nudges could presumably help free up limited cognitive resources for obligations that call for a more vivid appreciation of reasons. For both individuals and collectives, moral nudges can therefore help with the conservation and optimization of our moral energies and, in that sense, our moral reasoning. When we are committed to certain moral acts that do not require sustained vivid compliance, or when we cannot bring ourselves to discharge our enforceable

[15] This consideration also echoes Thaler and Sunstein's claim that choice environments will inevitably influence individuals in some way (2008, pp. 10–11). A more recent restatement of this argument holds that if choice architects can predict the behavioral effects of various available choice environments, it is inevitable for them to pick some option, and should in that case nudge individuals in whichever direction seems most normatively justifiable (Ivanković and Moles 2025).

moral duties, reflecting on moral reasons could amount to cognitive waste; our cognitive energies could and perhaps should be utilized elsewhere. The systematic conservation and optimization of our moral energies seems to us intuitively to manifest moral improvement and progress. As moral agents, we are limited, both individually and collectively, in our capacities to appreciate the moral dimensions of our circumstances. If moral nudges can help us in systematically optimizing the use of these capacities, then that seems a plausible indication of moral progress.

Finally, moral nudges could conserve our cognitive energies not only for laborious moral reflection in controversial cases, but for using these energies on our own personal projects. Requiring people to expend most or all of their cognitive energies on moral controversies, leaving little or none for their personal autonomous pursuits, would verge on demanding moral sainthood. The use of moral nudges thus not only ensures the investment of reflection on moral cases that matter, but also spares individuals from moral preoccupation in a broader sense, one that would hinder them in dedicating themselves to their autonomous pursuits.

3.4 Stability and rigidity

Before showing how our endorsement of moral nudging complements recent accounts emphasizing the insufficiency of moral reasoning and the importance of social norms in moral progress, we face our account with two objections. One suggests that nudges might be unable to produce stable patterns of behavior associated with moral progress, and the other, conversely, that nudges might yield value rigidity and dogmatism – outcomes that would intuitively undermine the claim that they are morally progressive.

Consider the worry of stability first. Several authors note that, due to their stipulated mildness, nudges might be inadequate in resolving the social concerns for which they are designed. Thaler and Sunstein acknowledge early on that for, say, environmental problems, "gentle nudges may appear ridiculously inadequate – a bit like an effort to capture a lion with a mousetrap" (2008, p. 184). Evan Selinger and Kyle Whyte have also noted that nudges at best work like techno-fixes that "cannot solve complex policy problems" (2012, p. 26), and, similarly, John et al. have stated that nudges "will not be enough on their own to combat climate change" (2009, p. 368). Moreover, the stability condition plays an important role in the moral progress debate, where genuine moral progress is often taken to require deliberate, deep, long-lasting, and not easily reversible change (Schinkel and de Ruyter 2017; Rehren and Blunden 2024; Bina et al. unpublished). In short, moral nudges might fail to produce the stable behavioral patterns associated with a morally progressive behavioral shift.

However, these objections merely show the inadequacy of nudges as the *sole* means of stabilizing behavioral patterns. But our analysis of moral nudges that are most likely to be progressive has pointed to the potential synergy between nudges and social norms. The question, then, is not whether nudges can stabilize morally progressive behavior on their own, but whether they can stabilize the norms that perform said function. In other words, progressive moral nudges should be sufficient in enabling people to uphold society's norms in a not-easily-reversible, lasting manner, or in destabilizing unjust norms that inhibit people's responsiveness to moral reasons.

Whether moral nudges will have such effects is ultimately an empirical matter, and a difficult one to assess. While they might be one of the stabilizers of a desirable social norm (or destabilizers of an undesirable one), other factors – like public education, momentous events, or suitable socio-economic conditions – may play significant roles in contributing to the stability of the norm. Exactly how much nudges contribute to stabilizing or destabilizing a behavioral pattern will often be difficult to tell given other influences.

We can offer two remarks in response. First, nudges seem to be doing the brunt of the work in at least some cases, like when they lead to dramatic increases in organ donation (Johnson and Goldstein 2003) or charitable giving (Small et al. 2007). If nudges can bring about such significant behavioral shifts, they could plausibly, in such cases, be sufficient for affecting the signals associated with social norms. Second, the worry about stability points to a broader methodological concern in the debate on moral progress: the factors responsible for some morally desirable, stable behavioral patterns are often difficult to identify with pinpoint precision. Thus, the methodological hardships should not befall moral nudges more severely than other proposed contributors to moral progress.

An opposite kind of worry is that of rigidity, which arises if nudges do, in fact, successfully produce stable behavioral patterns. What if they reinforced a social norm to the point of rigidity? Would the outcome still count as morally progressive? The concern is that nudges and norms could mutually reinforce one another, thereby increasing pressure on behavior, especially when the nudge explicitly appeals to the now-reinforced norm. As the norm's pressure grows, the nudges appealing to it and reinforcing it might cease to be easily resistible, as they are often conceived.[16] A further concern is that this could undermine public deliberation. Since democratic dissent is sometimes seen as the only means of revising established beliefs in society (Mill 1859), its loss would be deeply regrettable and would likely undermine the claim that the behavioral pattern promoted by the nudges and norms is progressive.

We share these concerns. If nudges were to engineer incredibly strong social norms, perhaps even taboos, the case for their moral progressiveness could well be undermined. But much like the issue of stability concerns the interplay between nudges and norms, so too does the issue of rigidity. It is a worry that befalls not only our account, but any account that emphasizes the role of social norms in moral progress. If social norms were to exert a level of pressure that stifles public deliberation, an account relying on them should address whether moral progress is thereby compromised.

Our brief responses address the more general concern about the pressure of social norms and the potential for nudges to reinforce them. First, while norm rigidity might be regrettable, it is unclear how its badness comes anywhere near the badness of certain harms that could be avoided or alleviated through greater pressure on individual behavior. For example, in cases of harm-related enforceable duties, it is unclear that avoiding rigidity should come at the expense of significantly risking such harms. In

[16] Owen G. Schaefer raises a similar concern about direct moral bioenhancements threatening the freedom to hold dissenting opinions and publicly criticizing the pursuit of some moral ends, leading to a uniformity of opinion between the enhancers and the enhanced (Owen Schaefer 2015, pp. 263–266).

fact, some injustices like slavery lead to such harms that individuals seem justified in opposing them rigidly. Second, it is similarly unclear why we should not accept some risks to public deliberation in pursuit of much greater moral conformity; it is perfectly conceivable that public deliberation elucidates moral reasons for individuals and collectives without prompting them into action. As we show in the next section, the elucidation of these moral reasons is often insufficient for moral progress and may require additional means to prompt action, even at the risk of some norm rigidity. It seems cautious enough to allow some such risks in cases of established enforceable duties and moral ends. Third, to guard against norm rigidity, governments utilizing nudges might also be required to take proactive steps in fostering open public deliberation about the nudged ends, thereby mitigating the chances of rigidity. For instance, John et al. believe people should be encouraged into "participatory budgeting, the mini-publics of citizens' assemblies and juries, and online forums" (2009, p. 365), as venues for reflecting on various moral dilemmas.

4 Nudging social norms

So far, we have argued that other-regarding nudges can facilitate moral progress if they reinforce a morally desirable social norm or help overthrow a morally undesirable social norm. We have shown that many nudges in both the social and situationist categories have the capacity to highlight the kinds of social reasons that are associated with social norms. In many examples, we argued, neither of the two kinds obfuscate the moral reasons that, if the norm is indeed morally desirable, underpin it or operate in its background. Alternatively, these nudges could also contribute to behavioral shifts by loosening the grip of a morally undesirable social norm, i.e., one that is at odds with moral reasons. We have thereby established the connection between moral nudges and morally (un)desirable social norms. But insofar as social norms are the intermediary step toward moral progress, we have yet to establish the connection between social norms and moral progress. We do so by pointing to recent contributions to the moral progress literature that specifically appeal to the importance of social norms in the achievement of moral progress.

Typically, these accounts start from the claim that moral reasoning is insufficient for moral progress (McTernan 2014; Tam 2020; Klenk and Sauer 2021; Rehren and Blunden 2024). We remain largely agnostic on this point. Much like our contention is that moral progress can be achieved if moral behavior is stably improved while moral reasoning is not worsened, we do not rule out the possibility that some notion of moral progress may be achieved if the opposite obtains – if moral reasoning is stably improved while moral behavior remains constant. Nevertheless, since our argument points to a synergy between moral nudges and social norms, then, it complements any account of moral progress that takes social norms to be its building blocks.

We first call back Tam's account (2020) on the importance of "we-reasoning". Tam argues that, historically, moral reasons raised in democratic deliberation were largely insufficient to trigger momentous social shifts due to the obstructive influence of unjust social norms. Acting in accordance with a social norm – be it just or unjust – provides citizens with membership goods like group cohesion and recipro-

cal conformity, which are important for establishing their common identity (2020, pp. 73–75). We might think that societies striving toward moral progress should be able to cast aside social reasons that arise from (unjust) social norms in favor of acting upon moral reasons. But Tam rejects this view. It is not merely psychologically understandable that individuals act on social reasons, but there is also normative value to it. It can be socially destructive to overlook social norms, which provide membership goods "such as fellow feeling, social trust, and group honor" (2020, p. 75); such goods are "fundamental for parochial beings like us" (2020, p. 91). In other words, these membership goods carry normative weight rather than just providing a feasibility-grounded excuse for failing to act as morality requires. Insofar, acting on social reasons is not mere conformism: "The ought of sociality extends beyond the ought of prudence and stops short of the ought of morality" (2020, pp. 82–83). Tam argues that social norms establish joint commitments between individuals that steer their behavior and foster a sense of belonging to shared projects (2020, p. 84). The reasons that arise from these commitments are recognized not as universal, such as those of morality, but as reasons that we associate specifically with *our* community, i.e. we-reasons (2020, p. 87). The road to moral progress on Tam's account is to overcome the conflict between moral and social reasons by embedding the former into the latter, i.e., by making moral reasons part of our common projects, identity, and we-reasoning (2020, p. 91).

The social and situationist nudges that we have portrayed in Sect. 3.2 can be seen as one way to promote moral reasons on the back of social reasons. They thus represent a kind of embedding that Tam has in mind. They are also proposed in areas in which we can expect people to be in the grip of what some might regard as unjust social norms, such as a carnivorous diet, reliance on non-clean energy, ecologically wasteful behavior, a widespread reluctance to donate organs, etc. Where these unjust social norms prevail, moral nudges forge connections between moral and social reasons for action, or increase the prevalence of behaviors that can, over time, create important social signals in establishing more just social norms. Crucially, moral nudges do not undermine social norms in a way that would jeopardize the membership goods Tam associates with them. They merely ease the transition from an unjust norm to a more just one.

McTernan (2014) offers a similar case for the importance of social norms. On her account, they are instrumental not to the achievement of moral progress, but to securing functioning, flourishing, and stable social institutions. Philosophers, she notes, usually want to achieve this through the cultivation of political virtues. But according to research from social psychology, virtues are unlikely to deliver such an outcome. The dominant theories of social psychology – first situationism, and later interactionism – deny that (attempts at) virtue cultivation can guarantee stable behavior across a variety of contexts and situational factors. According to various experiments in psychology, attempts at cultivating virtue typically have little impact on whether individuals help others in a social context, let alone consistently (2014, pp. 89–93). It is simply that "the kinds of traits that the dominant paradigm of social psychology supports are not of the kind that could fulfil the instrumental roles that many liberals assign to virtue" (2014, p. 93). Conversely, social psychology ascribes great causal efficacy to social norms, which is confirmed by behavior in experiments

much like those we have discussed earlier: reusing towels in hotel rooms, lowering energy consumption in response to cues about the consumption of others, filing taxes on time, etc. McTernan concludes that social norms are a significantly more reliable instrument for achieving the flourishing and stability of social institutions than virtues (McTernan 2014, p. 98).

Note that McTernan is skeptical about the use of nudges in reinforcing such norms. This is because she raises doubts about social and situationist nudges belonging to the same kind of things; whereas social nudges, according to McTernan, are merely appeals to social norms of the kind we both describe, situationist nudges rely entirely on situational factors and do not impact social norms (2014, p. 100). Our first disagreement is terminological. While McTernan is right to point out that the term 'nudge' may be overly heterogenous, it now seems to be standard practice to take peer pressure and social norms as some of the psychological effects that nudges may harness. Here we simply stick with that practice. From a normative standpoint, we both seem to endorse these techniques as reliable tools for achieving normative ends. We do have some normative disagreement with regard to situationist nudges. In Sect. 3, we have noted that the use of defaults may often be easily noticeable or highlighted so that moral reasons are not obfuscated. The same holds for social reasons. Noticing a change of default may give us a sense of what others are doing and what they are expecting us to do. Noticing the effects of a nudge *ex post*, such as a speeding technique in traffic, may similarly convey public expectations. Finally, if such nudges are truly effective and increase the prevalence of some kind of behavior, then this will, in time, change relevant social signals.

Finally, on the example of reducing meat consumption, Joshua May and Victor Kumar suggest that findings from psychology can be employed in important ways to improve moral behavior (2023).[17] May and Kumar state that individuals continue eating meat because the practice is so ubiquitous that this "serves as positive social proof that it is acceptable, that nothing is morally amiss", despite their more abstract declarations about, say, the immorality of factory farming (2023, p. 374). Like Tam and McTernan, May and Kumar describe this responsiveness to social cues as a "social form of reasoning" (2023, p. 377). Observing others is an important source of information regarding which rules and principles are considered socially relevant, and which violations might be overlooked, either "because the rule is unnecessary or because punishment is unlikely" (2023, p. 378). One way in which we may be able to send important social signals to others, they argue, is precisely through nudges: a default vegetarian menu signals not merely a moral cue, but a social cue about what others will be eating (2023, p. 380), and how we, as a community ought to behave. Once more, a clear path emerges from nudge to social norm to moral progress.

[17] We should point out that May and Kumar's view does not stem from skepticism about the capacity of moral reasoning. In another paper, they argue that such skepticism is unfounded, but their understanding of moral reasoning there seems to be intertwined with social reasoning: "moral reasoning, particularly when embedded in social networks with mutual trust and respect, is integral to moral progress" (Kumar and May, forthcoming).

5 Conclusion

In this paper, we explored the potential for other-regarding nudges to be part of morally progressive change, despite their supposed tension with traditionally conceived moral thinking and acting. We argued that such nudges can be compatible with moral progress when they enhance moral behavior without degrading the quality of moral reasoning or beliefs. By reinforcing desirable or overturning undesirable social norms, nudges can play a pivotal role in fostering progressive social change. However, for nudges to meet these criteria, we argued, they must be easy to resist and economize cognitive resources without obfuscating moral reasons. We faced our account with objections pertaining to potential instability and excessive rigidity that might result from nudge-induced behavior. Our take is that as long as nudges are carefully designed to promote morally desirable social norms, these concerns can be avoided. We also demonstrated how nudges can effectively signal moral and social expectations, leading to behavioral shifts that support stable institutions and challenge unjust norms. While nudges may not be a silver bullet for overcoming moral and social challenges, their thoughtful design and application can serve as powerful tools for fostering morally progressive change.

Acknowledgements We thank the audiences at Change: International Conference (Belgrade, 2024), OZSW Annual Conference (Eindhoven, 2024), 4TU.Ethics Conference (Enschede, 2024), and the Uehiro Oxford Institute, as well as two anonymous reviewers, for their insightful comments and suggestions on the paper's earlier versions.

Funding This article is an output of the project "Moral Progress: Individual and Collective" supported by the Croatian Science Foundation (Grant No. IP-2022-10-5341). VI acknowledges financial support by the project Ethics and Social Challenges (EDI) at the Institute of Philosophy, reviewed by the Ministry of Science and Education of the Republic of Croatia and financed through the National Recovery and Resilience Plan 2021–2026 of the European Union – NextGenerationEU. KK acknowledges funding from the Starting grant project "Do we owe anything to ourselves?", funded by the Dutch Government (Grant No. 1570 00 3901).

Data availability The manuscript contains no associated data.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

Aristotle (2009) The Nicomachean ethics. Oxford University Press, Oxford

Bacharach M (2006) Beyond individual choice: teams and frames in game theory. Princeton University Press, Princeton

Behavioural Insights Team (2012) Applying behavioural insights to reduce fraud, error and debt. Available at: https://assets.publishing.service.gov.uk/media/5a789ae740f0b62b22cbb536/BIT_FraudErrorDebt_accessible.pdf

Beraldo S, Karpus J (2021) Nudging to donate organs: do what you like or like what we do? Med Health Care Philos 24(3):329–340. https://doi.org/10.1007/s11019-021-10007-6

Berg N (2018) Decentralization mislaid: on new paternalism and skepticism toward experts. Rev Behav Econ 5:361–387. https://doi.org/10.1561/105.00000099

Berg N, Kim J-Y (2019) Economics of the broken window theory. Asian J Law Econ 10(3):1–9. https://doi.org/10.1515/ajle-2019-0026

Bina F, Liberti M, Reichlin M, Songhorian S, Vaccarezza M (Unpublished manuscript) individual moral progress: a virtue-based approach

Bodlović P, Kudlek K (2025) Knowledge versus understanding: what drives moral progress? Ethical Theory Moral Pract 28(1):17–39. https://doi.org/10.1007/s10677-024-10465-w

Borenstein J, Arkin R (2016) Robotic nudges: the ethics of engineering a more socially just human being. Sci Eng Ethics 22(1):31–46. https://doi.org/10.1007/s11948-015-9636-2

Bovens L (2009) The ethics of nudge. In: Grüne-Yanoff T, Hansson S (eds) Preference change: approaches from philosophy, economics and psychology. Springer, Berlin & New York, pp 207–219

Buchanan A, Powell R (2018) The evolution of moral progress: A biocultural theory. Oxford University Press, Oxford

Campbell-Arvai V, Arvai J, Kalof L (2012) Motivating sustainable food choices: the role of nudges, value orientation, and information provision. Environ Behav 46(4):453–475. https://doi.org/10.1177/0013916512469099

Clayton M, Moles A (2018) Neurointerventions, morality, and children. In: Birks D, Douglas T (eds) Treatment for crime: philosophical essays on neurointerventions in criminal justice. Oxford University Press, Oxford, pp 235–251

De Kort Y, Ijsselsteijn et al (2014) De-escalate: Defusing escalating behaviour through the use of interactive light scenarios. In: De Kort, Y, Ijsselsteijn. (eds) Proceedings of experiencing light 2014: International conference on the effects of light on wellbeing, Technische Universiteit Eindhoven, Eindhoven, pp 94–97

DeGrazia D (2014) Moral enhancement, freedom, and what we (should) value in moral behaviour. J Med Ethics 40(6):361–368. https://doi.org/10.1136/medethics-2012-101157

Desvouges W, Johnson F, Dunford R, Boyle K, Hudson S, Wilson K (1993) Measuring natural resource damages with contingent valuation: tests of validity and reliability. In: Hausman J (ed) Contingent valuation: A critical assessment. North-Holland, New York, pp 91–164

Douglas T (2008) Moral enhancement. J Appl Philos 25(3):228–245. https://doi.org/10.1111/j.1468-5930.2008.00412.x

Douglas T (2013) Moral enhancement via direct emotion modulation: a reply to John Harris. Bioet 27(3):160–168. https://doi.org/10.1111/j.1467-8519.2011.01919.x

Douglas T (2022) If nudges treat their targets as rational agents, nonconsensual neurointerventions can too. Eth Theo Mor Prac 25:369–384. https://doi.org/10.1007/s10677-022-10285-w

Egebark J, Ekström M (2013) Can Indifference Make the World Greener? IFN Working Paper 975

Engelen B, Ivanković V (2025) What I really, really want: the role, nature, and value of true preferences in the ethics of nudging. J Appl Philos 42(4):1127–1150. https://doi.org/10.1111/japp.12784

Engelen B, Nys T (2024) Pushed for being better: on the possibility and desirability of moral nudging. J Val Inq. https://doi.org/10.1007/s10790-023-09977-y

Evans J (2017) A working definition of moral progress. Ethic Mor Prac 20(1):75–92. https://doi.org/10.1007/s10677-016-9753-2

Frank LE (2020) What do we have to lose? Offloading through moral technologies: moral struggle and progress. Sci Eng Ethics 26(1):369–385. https://doi.org/10.1007/s11948-019-00099-y

Haley K, Fessler D (2005) Nobody's watching? Subtle cues affect generosity in an anonymous economic game. Evo Hum Behav 26(3):245–256. https://doi.org/10.1016/j.evolhumbehav.2005.01.002

Hausman D, Welch B (2010) Debate: to nudge or not to nudge. J Polit Philos 18(1):123–136. https://doi.org/10.1111/j.1467-9760.2009.00351.x
Hermann J (2019) The dynamics of moral progress. Rat 32(4):300–311. https://doi.org/10.1111/rati.12232
Ivanković V, Engelen B (2019) Nudging, transparency, and watchfulness. Soc Theo Prac 45(1):43–73. https://doi.org/10.5840/soctheorpract20191751
Ivanković V, Moles A (forthcoming) The inevitability argument for choice architecture and the evidence-based view. Beh Pub Pol. https://doi.org/10.1017/bpp.2025.1
John P, Smith G, Stoker G (2009) Nudge nudge, think think: two strategies for changing civic behaviour. Pol Quar 80(3):361–370. https://doi.org/10.1111/j.1467-923X.2009.02001.x
Johnson E, Goldstein D (2003) Do defaults save lives? Sci 302(5649):1338–1339. https://doi.org/10.1126/science.1091721
Kahneman D, Ritov I, Jacowitz K, Grant P (1993) Stated willingness to pay for public goods: a psychological perspective. Psychol Sci 4(5):310–315. https://doi.org/10.1111/j.1467-9280.1993.tb00570.x
Kant I (1785) /1998 Groundwork of the Metaphysics of Morals. Cambridge University Press, Cambridge
Kasperbauer T (2017) The permissibility of nudging for sustainable energy consumption. En Pol 111:52–57. https://doi.org/10.1016/j.enpol.2017.09.015
Kelly J (2013) Libertarian paternalism, utilitarianism, and justice. In: Coons C, Weber M (eds) Paternalism: theory and practice. Cambridge University Press, New York, pp 216–230
Klenk M, Sauer H (2021) Moral judgement and moral progress: the problem of cognitive control. Phil Psych 34(7):938–961. https://doi.org/10.1080/09515089.2021.1931670
Korsgaard C (2009) Self-Constitution: agency, identity, and integrity. Oxford University Press, Oxford
Kumar V, May J (forthcoming), Copp D, Rosati C (eds) The Oxford handbook of metaethics. Oxford University Press, Oxford
Kurtz S, Saks M (1996) The transplant paradox: overwhelming public support for organ donation vs. under supply of organs: the Iowa organ procurement study. J Corp Law 21(4):767–806
Lades L (2014) Impulsive consumption and reflexive thought: nudging ethical consumer behavior. J Econ Psychol 41:114–128. https://doi.org/10.1016/j.joep.2013.01.003
Le Grand J, New B (2015) Government paternalism: Nanny state or helpful friend. Princeton University Press, Princeton and Oxford
Levy N (2019-2020) Nudge, nudge, wink, wink: nudging is giving reasons. Ergo 6(10). https://doi.org/10.3998/ergo.12405314.0006.010
Loughnan S, Haslam N, Bastian B (2010) The role of meat consumption in the denial of moral status and mind to meat animals. Appetite 55(1):156–159
Macklin R (1977) Moral progress. Eth 87(4):370–382
Madva A, Kelly D, Brownstein M (2023) Change the people or change the policy? On the moral education of antiracists. Eth Theo Mor Prac 27:91–110. https://doi.org/10.1007/s10677-023-10363-7
May J, Kumar V (2023) Harnessing moral psychology to reduce meat consumption. J Am Philos Assoc 9(2):367–387. https://doi.org/10.1017/apa.2022.2
McTernan E (2014) How to make citizens behave: social psychology, liberal virtues, and social norms. J Pol Phil 22(1):84–104. https://doi.org/10.1111/jopp.12015
Mertens S, Herberz M, Hahnel U, Brosch T (2022) The effectiveness of nudging: a meta-analysis of choice architecture interventions across behavioral domains. Proc Natl Acad Sci USA 119:e2107346118. https://doi.org/10.1073/pnas.2107346118
Mill JS (1859) /1991 On Liberty. In: Gray J (ed) On Liberty and Other Essays, Oxford University Press, Oxford, pp 5–131
Mill JS (1863) /1906 Utilitarianism. Reprinted from the Thirteenth London Edition. The University of Chicago Press, Chicago. Available online: https://www.utilitarianism.com/utilitarianism-jsmill.pdf
Mills C (2015) The heteronomy of choice architecture. Rev Philos Psychol 6(3):495–509. https://doi.org/10.1007/s13164-015-0242-7
Moles A (2015) Nudging for liberals. Soc Theo Prac 41(4):644–667. https://doi.org/10.5840/soctheorpract201541435
Mols F, Haslam A, Jetten J, Steffens N (2015) Why a nudge is not enough: a social identity critique of governance by stealth. Eur J Pol Res 54(1):81–98. https://doi.org/10.1111/1475-6765.12073
Moody-Adams MM (1999) The idea of moral progress. Metaphil 30(3):168–185. https://doi.org/10.1111/1467-9973.00120
Musschenga A, Meynen G (2017) Moral progress: an introduction. Eth Theo Mor Prac 20(1):3–15. https://doi.org/10.1007/s10677-017-9782-5

Nagatsu M (2015) Social nudges: their mechanisms and justification. Rev Philos Psychol 6(3):481–494. https://doi.org/10.1007/s13164-015-0245-4
Niker F (2018) Policy-led virtue cultivation: can we nudge citizens towards developing virtues? In: Harrison T, Walker D (eds) The theory and practice of virtue education. Routledge, London, pp 153–167
Nolan J, Schultz P, Cialdini R, Goldstein N, Griskevicius V (2008) Normative social influence is underdetected. Pers Soc Psychol Bull 34(7):913–923. https://doi.org/10.1177/0146167208316691
Papies E (2017) Situating interventions to bridge the intention-behaviour gap: a framework for recruiting nonconscious processes for behaviour change. Soc Personal Psychol Compass 11(7):e12323. https://doi.org/10.1111/spc3.12323
Persson I, Savulescu J (2012) Unfit for the future: the need for moral enhancement. Oxford University Press, Oxford
Persson I, Savulescu J (2015) The art of misunderstanding moral bioenhancement: two cases. Camb Q Healthc Ethics 24(1):48–57. https://doi.org/10.1017/S0963180114000292
Persson I, Savulescu J (2019) The duty to be morally enhanced. Top 38(1):7–14. https://doi.org/10.1007/s11245-017-9475-7
Pichert D, Katsikopoulos K (2008) Green defaults: information presentation and pro-environmental behaviour. J Environ Psychol 28(1):63–73. https://doi.org/10.1016/j.jenvp.2007.09.004
Potter N (1997) Supererogation and overdetermination in Kant's ethics: analysis and interpretation at their best in Baron. Jahr Rec Und Eth 5:489–496
Raz J (1999) Practical reasons and norms, 2nd edn. Oxford University Press, Oxford
Rehren P, Blunden C (2024) Let's not get ahead of ourselves: we have no idea if moral reasoning causes moral progress. Phil Exp 27(3):351–369. https://doi.org/10.1080/13869795.2024.2363876
Rosati C (2016) Moral Motivation. In: Zalta E (ed) The Stanford Encyclopedia of Philosophy. Available at: https://plato.stanford.edu/entries/moral-motivation/
Saghai Y (2013) Salvaging the concept of nudge. J Med Ethics 39(8):487–493. https://doi.org/10.1136/medethics-2012-100727
Sauer H (2019) Butchering benevolence moral progress beyond the expanding circle. Eth Theo Mor Prac 22(1):153–167. https://doi.org/10.1007/s10677-019-09983-9
Sauer H, Blunden C, Eriksen C, Rehren P (2021) Moral progress: recent developments. Philos Compass 16(10):e12769. https://doi.org/10.1111/phc3.12769
Schaefer G (2015) Direct vs. indirect moral enhancement. Ken Ins Eth J 25(3):261–289. https://doi.org/10.1353/ken.2015.0016
Schinkel A, de Ruyter D (2017) Individual moral development and moral progress. Eth Theo Mor Prac 20(1):121–136. https://doi.org/10.1007/s10677-016-9741-6
Schmidt A (2017) The power to nudge. Am Polit Sci Rev 111(2):404–417. https://doi.org/10.1017/S0003055417000028
Schubert C (2017) Green nudges: do they work? Are they ethical? Eco Econ 132(23):329–342. https://doi.org/10.1016/j.ecolecon.2016.11.009
Schultz P, Nolan J, Cialdini R, Goldstein N, Griskevicius V (2007) The constructive, destructive, and reconstructive power of social norms. Psychol Sci 18(5):429–434. https://doi.org/10.1111/j.1467-9280.2007.01917.x
Schwitzgebel E, Rust J (2014) The moral behavior of ethics professors: relationships among self-reported behavior, expressed normative attitude, and directly observed behavior. Phil Psychol 27(3):293–327. https://doi.org/10.1080/09515089.2012.727135
Selinger E, Whyte K (2012) Nudging cannot solve complex policy problems. Eur J Risk Regul 3(1):26–31. https://doi.org/10.1017/S1867299X0000177X
Singer P (1981) The expanding circle: ethics, evolution, and moral progress. Princeton University Press, Princeton and Oxford
Small D, Loewenstein G, Slovic P (2007) Sympathy and callousness: the impact of deliberative thought on donations to identifiable and statistical victims. Org Behav Hum Decis Process 102(2):143–153. https://doi.org/10.1016/j.obhdp.2006.01.005
Sunstein C (2013) Simpler: the future of government. Simon & Schuster, New York
Sunstein C (2015) Why nudge? The politics of libertarian paternalism. Yale University Press, New Haven & London
Svavarsdottir S (1999) Moral cognitivism and motivation. Phil Rev 108(2):161–219
Tam A (2020) Why moral reasoning is insufficient for moral progress. J Polit Philos 28(1):73–96. https://doi.org/10.1111/jopp.12187

Thaler R, Sunstein C (2008) Nudge: improving decisions about health, wealth, and happiness. Yale University Press, New Haven