Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Environmental Sustainability

When is green nudging ethically permissible?

C Tyler DesRoches[1,*], Daniel Fischer[2,#], Julia Silver[5], Philip Arthur[1], Rebecca Livernois[3,†], Timara Crichlow[1,‡], Gil Hersch[3,§], Michiru Nagatsu[4,$] and Joshua K Abbott[1,$]

This review article provides a new perspective on the ethics of green nudging. We advance a new model for assessing the ethical permissibility of green nudges (GNs). On this model, which provides normative guidance for policymakers, a GN is ethically permissible when the intervention is (1) efficacious, (2) cost-effective, and (3) the advantages of the GN (i.e. reducing the environmental harm) are not outweighed by countervailing costs/harms (i.e. for nudgees). While traditional ethical objections to nudges (paternalism, etc.) remain potential normative costs associated with GNs, any such costs must be weighed against the injunction to reduce environmental harm to third parties.

Addresses
[1] Arizona State University, Tempe Campus, United States
[2] Wageningen University & Research, Netherlands
[3] University of Toronto, Canada
[4] University of Helsinki, Helsingin Yliopisto, Finland
[5] Latino Policy and Politics Institute, University of California, Los Angeles, United States

Corresponding author: DesRoches, C Tyler (tyler.desroches@asu.edu)
[*] **ORCID**: 0000-0002-7318-6948
[#] **ORCID**: 0000-0001-5691-0087
[†] **ORCID**: 0000-0001-5842-3506
[‡] **ORCID**: 0000-0002-2124-1562
[§] **ORCID**: 0000-0003-0992-0164
[$] **ORCID**: 0000-0001-6566-0307

Current Opinion in Environmental Sustainability 2022, **60**:101236

This review comes from a themed issue on **Open Issue**

Edited by **Opha Pauline Dube, Victor Galaz** and **William Solecki**

Available online xxxx

https://doi.org/10.1016/j.cosust.2022.101236

Introduction

Most sustainability scientists recognize that human behavior is a key leverage point for transforming socio-ecological systems [1], mitigating anthropogenic climate change [2–4], and fostering biodiversity conservation [5,6]. One promising behavioral policy intervention type among sustainability scientists is 'green nudges' (GNs). Unlike ordinary 'welfarist nudges' (WNs), which aim to make nudgees better-off by their own subjective standards, GNs aim to promote environmentally benign behavior to reduce environmental harm to third parties [7].[7]

While the prospect of GNs has been judged favorably among sustainability scientists, some scholars have raised ethical objections against nudging in general. Nudging can be paternalistic and violate individual autonomy [8–11]. Others have suggested that nudging is disrespectful or insulting [12]. Sustainability scientists who endorse GNs should take these objections seriously.

We argue that even if such objections succeed against WNs, they have less traction against GNs. Why? There is a significant ethical difference between intervening to make people better-off by their own standards and intervening to reduce harm to a third party. Other things being equal, GNs are more easily justified than WNs.

By synthesizing a small and fragmented literature, we provide a new perspective on the ethics of GNs [7,13–15]. We advance a new model — the *Pro Tanto* Model — for assessing the ethical permissibility of GNs. On this model, which is designed to provide high-level normative guidance for policymakers, a GN is ethically permissible when the intervention is efficacious, cost-effective, and has advantages (i.e. reducing the environmental harm to others) that are not outweighed by countervailing costs or harms for nudgees. While traditional ethical objections to WNs present potential normative costs associated with GNs, these costs must be weighed against the injunction to reduce harm to third parties. We conclude by suggesting that, in some cases, GNs might not only be permissible but a moral obligation.

Welfarist nudges versus green nudges

Welfarist nudges

For decades, behavioral scientists have shown that people make mistakes or errors in choice situations, from the perspective of rational choice theory [1,16].

[7] All nudges are aspects of the choice architecture that predictably alter people's behavior without forbidding any options or significantly changing their economic incentives [20].

Individuals are subject to a host of psychological biases and have been shown to exhibit preference reversals, weakness of will, and a lack of self-control in decision-making contexts [16–19]. WNs are interventions that exploit these biases to predictably alter behavior without significantly changing economic incentives (prices) (e.g. changing the default option in retirement plans as opt-in or arranging healthy food in a salient spot in cafeterias). To count as WNs, those who are nudged, called nudgees, must be made better-off by their own subjective standards [20,21]. Nudges are often claimed to be more consistent with libertarian principles than various forms of regulation, such as banning activities or taxation [22] because the former are presumably easy to avoid. Nudges preserve option-freedom.

Ethical objections against welfarist nudges

WNs have been subject to two main ethical objections.[8] First, scholars have argued that WNs are paternalistic [8,9]. WNs typically involve an authority, such as a policymaker, intervening in a person's 'choice architecture' to help them make better decisions. Some scholars thus argue that regulating a person's behavior merely for the purpose of benefiting them is unjustified and objectionably paternalistic. Second, WNs may violate individual autonomy [10].[9] Some WNs, particularly those that are difficult to detect, risk abrogating the nudgee's control over their personal evaluations, deliberations, and choices. This has led some scholars to argue that policymakers should strive to preserve the autonomy of nudgees by rationally persuading them rather than circumventing their faculty of rationality.[10] These objections to nudging should give sustainability scientists pause when promoting nudges that aim to bolster a desired behavior, pro-environmental or otherwise.

Green nudges

What are GNs? Broadly construed, GNs involve the use of behavioral insights to tackle environmental problems. In his overview of GNs, Christian Schubert defines them as a subset of "nudges that aim at promoting environmentally benign behavior" [7]. We accept this basic definition but refine it further. *GNs are behavioral interventions that aim to reduce or eliminate environmentally mediated harms to a third party*.[11] For the most part, it will be useful to consider GNs as interventions designed and deployed to reduce negative environmental externalities. We return to this point below.

What are the main differences between GNs and WNs? While GNs aim to reduce environmentally mediated harm ('environmental harm') associated with a negative externality, collective action problem, or a social dilemma, WNs aim to make nudgees better-off by their own subjective standards.[12] Compared with a WN, an efficacious GN primarily benefits a different group than the one intervened on, namely, those who would have been harmed without the GN. Crucially, however, GNs might have welfare effects for nudgees as well, leaving them potentially worse-off, equally well-off, or better-off, even when the policymaker intends only to reduce environmental harm to a third party.[13]

As depicted in Table 1 below, scientists and scholars have identified at least three different types of GNs, including 'signaling a green self-image,' 'following the herd,' and 'green defaults' [7].

GNs designed to maintain a green self-image generally simplify product information on packaging or make certain characteristics more salient based on pre-existing consumer preferences to practice 'green behavior' and thereby maintain a positive self-image.[14] 'Following the herd,' GNs typically stimulate concerns over social status, social norms, or identity competition between nudgees to emulate the green behaviors of others in one's social group. The third type of GN makes default options green.

There are at least two arguments for reducing harm to third parties: one accepted by medical practitioners and some philosophers, and the other favored by environmental economists and policymakers generally. The ancient Greek physician, Hippocrates, argued that medical practitioners have a duty to do no harm. "First,

[8] Two peripheral objections to WNs include being disrespectful and insulting toward nudgees [12].

[9] Others argue that nudges are autonomy-damaging when they cause inconsistent preferences or induce preference change that disturbs the coherence of a person's all-things considered preferences [11,13].

[10] Nudges contrast with 'boosts,' which are behavioral interventions that target an individual's skills and knowledge, the available set of decision tools, or the environment in which decisions are made [23].

[11] Harm as evaluated by the agents themselves.

[12] On our account, GNs are a subset of social nudges (the provision of the public good in question involves an environmental externality problem). Social nudges are interventions that "encourage the voluntary provision of public goods, which are characterized by nonrivalry and nonexcludability (it is impossible or difficult to exclude people from benefiting from the goods)" [13].

[13] Pure GNs should be distinguished from another class of GNs: *hybrid* GNs (HGNs). HGNs generally aim to reduce harm to a third party *and* leave nudgees better-off, simultaneously. For example, people have been nudged toward healthy patterns of consumption (making healthy choices) that, as a consequence, reduce environmental harm *and* make nudgees better-off [14,35]. Because GNs and HGNs are distinct, one should expect that, other things being equal, the ethical permissibility of GNs and HGNs may not always coincide. This short review article focuses on pure GNs.

[14] Merely supplying consumers with information may not constitute a nudge [10]. There is a substantial literature in economics that models ecolabels as information provision for 'credence attributes' of a good apart from any assumptions of behavioral foibles (e.g. [36]).

Table 1

Three types of GNs.

GN type	Example	Representative studies
Signaling a green self-image	**Eco-labeling** — Organizations label their products in ways that highlight the 'green' elements, with the intention to increase visibility and sales of the product based on consumers' desire to buy 'green products.'	[24–26]
Following the herd	**Towel reuse by hotel guests** — using descriptive social norms to encourage guests to reuse their towels because that is what the other guests are doing.	[27–29]
Green defaults	Automatic enrollment in a **green energy** option, subject to opt-out.	[30–34]

do no harm" (*primum non nocere*) is widely known as the 'Hippocratic Oath.'[15] Setting aside potential objections to this duty for the moment, one might suppose that an argument for reducing harm to a third party can be constructed as follows: 1) an activity is harming others; 2) people have a negative duty to do no harm to others; 3) therefore, those engaged in the harmful activity should cease and desist. If sound, this argument provides grounds for individuals to stop activities harming a third party.

A second argument, the policymaker's argument concerning social welfare optimization, runs as follows: 1) there is a negative environmental externality (third-party spillover) associated with the consumption or production of some good or service, X; 2) reducing an externality increases social welfare[16]; 3) other things being equal, policymakers should aim to increase social welfare; 4) therefore, policymakers may intervene in some way to curb an externality and thus bring about the socially optimal quantity of X.

Generally, goods, services, and activities generating negative externalities are overproduced in underregulated markets. For the sake of Pareto efficiency, these activities should be curtailed — ideally to the point where the overall costs to society of an additional increment of the good, service, or activity are just balanced by its benefits [38].[17]

To be clear, economic analysis generally concerns costs rather than harms. While any reduction in welfare constitutes a cost, we consider 'harm' to be a subset of non-negligible costs that require redress [39,40]. Nonetheless, environmental economists tend to ascribe policy relevance to negative externalities in a way that tracks the 'harm principle' of liberal theory, which holds that state interference is limited to preventing harm to third parties [41,42].

On our account, GNs are one type of policy intervention that can reduce environmental harms to third parties. We assume that such interventions, when efficacious, help to move *in the direction of* the social optimum.[18]

The ethics of green nudging: a fragmented literature

The ethics of GNs has been broached by several scholars, but the literature remains fragmented and underdeveloped. Kasperbauer [15] argues that GNs designed to promote sustainable energy consumption and production are ethical, but he reaches this conclusion by responding to the standard objections to WNs. Yet, there is a significant ethical difference between interventions to make someone better-off and interventions to reduce harm to a third party. Analyzing GNs as if they were WNs yields an incomplete picture of the ethical permissibility of GNs.

Without referring to GNs specifically, Sunstein [43] and Guala and Mittone [14] recognize the special ethical considerations that arise with 'market failure nudges' or social nudges in general. Sunstein states, "if the government is trying to reduce a collective action problem that produces high levels of pollution, it does not raise the kinds of ethical concerns that come into play if the government is acting paternalistically. It follows that market failure nudges should not be especially controversial in principle" [43].[19] Similarly, Guala and Mittone [14] argue that, holding other factors constant, if the consequences of behavior are entirely private, then the ethical case for nudging is much weaker than in cases characterized by the motivation to reduce harm. When conduct harms third parties, the justification for nudging is stronger than the justification of nudging for purely

[15] The philosopher Thomas Pogge [37] has argued that in realizing global justice, we have a negative duty to do no harm.

[16] The truth of this premise depends on the details of a specific policy, including the costs of implementing the policy relative to the benefits brought about by the policy. These types of concerns, however, are outside the scope of this short article. The aim here is to outline the standard economic argument for policy interventions to reduce negative environmental externalities.

[17] An outcome is Pareto efficient if there is no possible reallocation of resources that could improve the welfare of one agent without reducing the welfare of another agent.

[18] Other interventions to address a negative environmental externality include voluntary collective action, Coasean bargaining, 'market-based' approaches such as pollution taxes or 'cap and trade' programs, or conventional regulatory approaches [38].

[19] Also, see Ref. [44].

welfarist considerations.[20] For both Sunstein [43] and Guala and Mittone [14], GNs are distinct from WNs and this fact warrants a separate and distinct ethical analysis. As we see it, our analysis below is consistent with Sunstein [43] and Guala and Mittone [14].

Arguably, the most sophisticated attempt to establish a framework for assessing the ethical quality of GNs is due to Schubert [7]. In short, Schubert argues that GNs should (1) respect autonomy, (2) avoid interfering with a person's private ability to self-legislate, and (3) be fair.[21] On this account, (1)–(3) appear to be necessary and sufficient for the ethical permissibility of GNs. However, it remains unclear how the imperative to reduce harm to third parties is to be weighed against (1)–(3), a question that becomes especially palpable when GNs mitigate the risk of catastrophic environmental harms to third parties. Schubert's framework stops short of systematically assessing the ethical permissibility of GNs.

The *Pro Tanto* Model

Our main question is this: under what conditions are GNs ethically permissible? Let us begin with a strong proposal that people have a negative duty to do no harm to others and that some activity is known to be causing a non-negligible environmental harm to a third party, which may include members of present and future generations.[22] The harm could be ordinary (e.g. losing $100) or catastrophic (the expected consequences of unmitigated anthropogenic climate change).[23] Suppose further that when there is an activity causing such harm, an authority, such as a policymaker, has reason to curtail the activity. There is a cheap and efficacious GN available. Is the GN permissible?

Below, Figure 1 shows the relevant welfare consequences and permissibility or impermissibility of a GN grounded by a negative duty to do no harm. The minus signs indicate negative net welfare effects, the plus signs indicate positive net welfare effects, '0' indicates no change in net welfare. Because every GN is assumed to be efficacious, they have a positive net welfare effect (harm is reduced) for the relevant third party. However, GNs may simultaneously leave nudgees worse-off (e.g. by violating their autonomy or creating an inconvenience), equally well-off, or better-off (e.g. by unintentionally encouraging welfare-improving decisions). Because we are supposing that there is a negative duty to do *no* harm, GNs that cause any harm are impermissible. The impermissibility of GNs are represented by the cells shaded with *red* in Figure 1. Conversely, harmless GNs eliminate or reduce harm to third parties without reducing anyone's welfare, including that of nudgees. Other things being equal, these GNs are permissible, as represented by the cells shaded with green. Permissible GNs achieve their intended goal, which is to eliminate harm to third parties, while abiding by the negative duty to do no harm. In other words, permissible GNs are permissible by the 'no harm' principle when they create a Pareto improvement [45].

While grounding the permissibility of GNs with a negative duty to do no harm might appear to track our ethical intuitions, this principle will encounter problems in many policy contexts. For example, a negative duty to do no harm is silent in contexts when every GN leaves at least one person worse-off. While it seems reasonable to suppose that a GN that causes a small quantity of harm to a few people is preferable to a GN that causes a large quantity of harm for many, a negative duty to do no harm fails to discriminate between such GNs. Both are harmful and, therefore, impermissible. Second — and relatedly — it seems reasonable to suppose that some GNs should be permissible, even if they cause some harm to some nudgees. For example, there is a case to be made for a GN that causes minimal harm for some while simultaneously blocking a significant harm (catastrophic environmental outcome). Yet, if policymakers ground the permissibility of GNs with a negative duty to do no harm, the intervention would be ruled out *a priori*.

An alternative principle, one that is pragmatic and familiar to social scientists, grounds the ethical permissibility of GNs in *maximizing social utility (or welfare)*. This principle begins by expressing harms in terms of welfare effects and then defines a social welfare function defining a social preference ranking resolving how gains and losses in individual welfare are to be traded off against one another [50]. This ranking enables policymakers to evaluate the trade-off between harms created by a GN and harms that would be imposed on third parties without the GN.[24] Suppose, for example, that a certain GN would violate the right to autonomy for a small segment of the population but would also prevent the violation of a right to life for a significant number of people. While the principle 'do no harm' seems to preclude any action in this case, the social utility principle

[20] Nagatsu [13] analyzes two ethical objections to 'social nudges' (behavioral interventions that aim to facilitate voluntary cooperation in social dilemma situations) and concludes that neither objection is definitive.

[21] Schubert [7] determines fairness according to 1) the 'redistributive impact' of nudges; and 2) nudges' risk of distracting attention away from socio-institutional factors that cause sustainability problems.

[22] For considering third parties beyond the human species, see Ref. [46].

[23] For more on this distinction, see Refs. [47–49].

[24] Our analysis takes into consideration a broad range of noneconomic goods, such as respect and individual autonomy. Violations of autonomy, disrespect, and (objectionable forms of) paternalism count as reasons against the ethical permissibility of GNs.

Figure 1

Prevented Environmental Harms	Net Welfare Effects			
	Unintended			Intended
	Nudged			*Third Party*
Ordinary	+	0	−	+
Catastrophic	+	0	−	+

The consequences of an efficacious GN (*Do No Harm to Others*).

Figure 2

Prevented Environmental Harms	Net Welfare Effects				
	Unintended				Intended
	Nudged				*Third Party*
Ordinary	+	0	− (small)	− (large)	+
Catastrophic	+	0	− (small)	− (large)	+

The consequences of an efficacious GN (*Maximize Social Welfare*).

dictates that a GN is permissible if the harm (cost) created by the policy is outweighed by the harm avoided by the policy.

If policymakers also accept something akin to John Stuart Mill's [41] 'harm principle,' then an activity that harms others provides an authority, such as a government or policymaker, with a *pro tanto* reason to regulate that activity. There may be countervailing reasons against regulating a harmful activity (the regulation might cause harm), but an authority always has a *pro tanto* reason to regulate it. On this account, policymakers have a special *pro tanto* reason to reduce an activity known to be harming a third party. GNs aim to weigh the costs and benefits of the policy to the point where the social benefits of reducing the harmful activity through the GN exceed the social costs.

On the *Pro Tanto* Model, a GN is permissible if and only if the GN is efficacious, cost-effective, and the advantages of the GN (reducing the harm to a socially acceptable level) are not outweighed by countervailing reasons (significant harms for nudgees). This model is compatible with political liberalism, broadly construed, since restricting someone's liberty to prevent them from harming others is generally seen as permissible and justified.

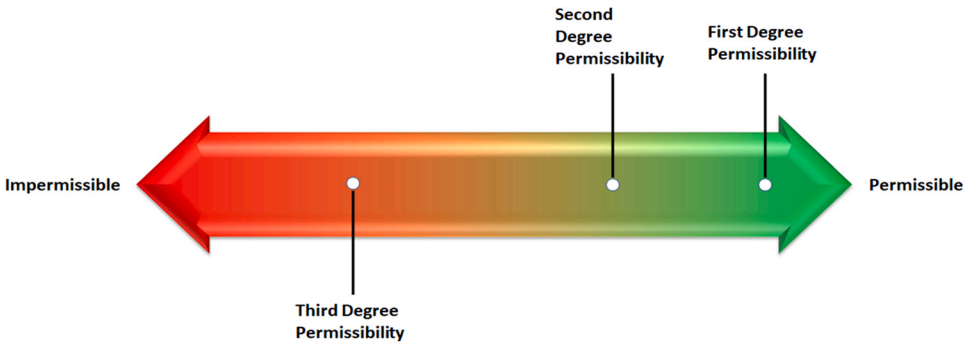
Figure 2 depicts the relevant consequences associated with an efficacious GN that aims to maximize social utility, which includes the prevented (ordinary or catastrophic) harms to a third party and the net welfare effects for nudgees and a third party. Similar to a GN grounded by a negative duty to do no harm, an efficacious GN that helps to move in the direction of the social optimum makes third parties better-off by reducing harm, the intended consequence of the GN. This harm reduction, which is equivalent to a benefit, is represented by the plus signs that populate the cells in the right-hand column of Figure 2. Policymakers should expect that GNs will almost always have unintended welfare consequences for nudgees as well. GNs can make nudgees worse-off, equally well-off, or better-off by their own subjective standards.

The colors in Figure 2 represent the permissibility and impermissibility of GNs. However, rather than assuming a GN is either permissible or impermissible, it may be useful for practical purposes to frame the permissibility of GNs as *a matter of degree*. Figure 3, below, represents degrees of permissibility with different colors along a continuum. From left to right along the continuum are red, orange, yellow, light green, and dark green (these colors map onto the colors in Figures 1 and 2).

Red indicates impermissible GNs: the negative consequences (typically borne by nudgees) associated with the GN exceed some threshold of acceptability. On this account, *bona fide* instances of disrespect or violations of autonomy caused by a GN are treated as noneconomic harms, which provide countervailing reasons to regulate the harmful activity.

Consider an example of an impermissible GN. Suppose that a group of people are engaged in some economic activity characterized by a negative production externality and the relevant harm is non-negligible and ordinary. Further suppose that a GN is the only intervention available to reduce this activity to its socially optimal level but, unfortunately, this intervention violates the autonomy of nudgees in an objectionable way, for example, by bypassing their reasoning capacity and

Figure 3



Green nudging: degrees of ethical permissibility.

treating them disrespectfully. In this case, the harm to nudgees is sufficiently large to make the intervention unjustified and impermissible, despite aiming to reduce harm to a socially optimal level. The countervailing reasons are sufficiently strong to overrule the *pro tanto* reason to regulate the environmentally harmful activity.

Even if GNs reduce harm to third parties, they may simultaneously cause small or large negative net welfare effects for the party being nudged. These possible outcomes are represented by the four cells in Figure 2 that contain minus symbols, each of which represents the negative net welfare effects borne by nudgees. Orange and yellow in Figures 2 and 3 indicate a *third degree of permissibility*, which is a GN characterized by harms, likely borne by nudgees. These GNs are objectionable, but less so than impermissible ones. In Figure 2, the yellow cell situated along the top 'ordinary' row reduces a non-negligible harm to third parties but at a cost to nudgees who are made worse-off in some respect (small net negative welfare effect) by the GN. The yellow cell situated along the bottom 'catastrophic' row reduces what would otherwise be a catastrophic harm to a socially acceptable level, but nudgees are simultaneously made much worse-off (large net negative welfare effect) by the GN.

The shade of light green indicates a GN with a *second degree of permissibility*: GNs cause minor net negative effects for some welfare subjects. Among the four cells in Figure 2 with a minus sign, the light-green cell is the most permissible GN. Why? This GN blocks a catastrophic harm to third parties while causing a relatively small negative net welfare effect for nudgees. When evaluating the justification and ethical permissibility of GNs, dark green is the gold standard. These GNs reduce harm to third parties and without making anyone worse-off. This *first degree of ethical permissibility* is indicated by the dark shade of green in Figures 2 and 3.

Objections and replies

Before concluding, we consider four potential objections to the *Pro Tanto* Model.

- **Objection 1: Must Every Environmental Harm to Third Parties be Eliminated?** Environmental harm to third parties is a ubiquitous feature of liberal democratic societies. Policymakers should not be expected to eliminate every such harm. Afterall, this would require continuous interference in our lives, which seems objectionable for people committed to the ideals of liberal democratic society [51]. **Reply**: For the purpose of public policy, it seems reasonable to insist that for any given (expected) environmental harm to be taken seriously by policymakers, the harm should be *non-negligible* or impose a genuine cost on a third party. Moreover, the positive probability of some activity causing environmental harm to a third party should exceed some *de minimis* threshold so that the 'mere possibility' of an activity causing harm does not immediately trigger a GN.
- **Objection 2: GNs Are Not the Best Policy Intervention to Reduce Environmental Harm.** Perhaps there is a more efficacious and cost-effective policy intervention for preventing harm to a third party than any GN on offer. Maybe GNs are too conservative for the deep transformational change required to achieve sustainability. Worse still, GNs might crowd out public support for more effective policy measures such as a carbon tax [2]. Therefore, policymakers should use this alternative non-GN intervention rather than a GN. **Reply**: GNs are but one policy option in the behavioral sustainability scientist's toolkit. The *Pro Tanto* Model merely serves to assess the ethical permissibility of GNs. Non-GN policy interventions might sometimes be the best way to reduce or eliminate harm to third parties. While the primary purpose of this article was *not* to compare alternative policy interventions, GNs appear to have

at least three advantages over equally efficacious policy interventions: they are relatively cheap, non-coercive, and preserve option-freedom. Concerning their efficacy and interactions with other policy measures, one should be aware of crowding out as well as synergies and complementarities between nudges and non-nudge policies. For example, GNs may be used to steer people to accept otherwise unpopular policies such as taxes [2].

- **Objection 3: The *Pro Tanto* Model is too Abstract. Reply:** From the vantage point of sustainability science, two points seem particularly salient for future research. First, procedural questions are salient when deciding on interventions in human behavior. From a deliberative democratic point of view, there is an open question about democratic legitimacy. For example, are citizens mandated to codetermine the conditions under which harms are considered negligible, ordinary, or catastrophic [52]? Second, because sustainability science is widely conceived to be based on public participation, how might GNs with their inherent tendency to bypass peoples' reasoning capacity be implemented without jeopardizing the idea of sustainable development as a reflexive and deliberative democratic process [2]? Answers to such questions are highly relevant to sustainability science but beyond the scope of this review article.
- **Objection 4: The Problem of Collective Harm.** For some sustainability challenges, we, collectively, are harming present and future generations with our greenhouse gas emissions, and yet no individual is causing the harm to any appreciable extent. Afterall, if any individual were to cease and desist from emitting greenhouse gases, climate change would endure [53,54]. Yet, if no individual is causing environmental harm and the aim of GNs is to reduce such harm, then it would appear that the justification for GNs would be significantly diminished, especially if GNs would make nudgees worse-off than they would have been otherwise. **Reply**: First, this is not a special problem for GNs, but can be leveled against all policy interventions that aim to neutralize environmental harm. Second, the problem of collective harm appears to play on the differences between individual and collective action. If the GN in question targets me and others, jointly, then the intervention does nothing to single me out as an individual.

Conclusion

This article aimed to galvanize a mutually enriching conversation between sustainability scientists, philosophers, and ethicists, thus responding to recent calls from sustainability scientists for more engagement with the humanities and behavioral sciences (e.g. [55–57]).

While some sustainability scientists might have supposed that, because GNs are pro-environmental, they are always ethically permissible [58]. The *Pro Tanto* model, by contrast, insists that, whenever the advantages of GNs are outweighed by sufficiently strong countervailing reasons, then such interventions are impermissible. With that said, however, policymakers should still expect that GNs are almost always easier to justify than WNs. Interventions that reduce environmental harm to a third party are significantly different from interventions that merely help people make better decisions for their own well-being.

Finally, it is important to note that the use of a GN to shift intergenerationally harmful behavior toward a 'safe operating space' (or something like it) is almost certainly permissible on the *Pro Tanto* Model [59]. Indeed, if people today are, collectively, causing an unacceptably high probability of catastrophic harm and GNs are part of the most cost-effective policy package to reduce this probability, then GNs might not only be permissible but a moral obligation.

Data Availability

No data were used for the research described in the article.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. • Schill C, Anderies JM, Lindahl T, Folke C, Polasky S, Cárdenas JC, Crépin AS, Janssen MA, Norberg J, Schlüter M: **A more dynamic understanding of human behaviour for the Anthropocene**. *Nat Sustain* 2019, **2**:1075-1082.

This article argues for the expansion of behavioural economics and cognitive psychology by adopting a more dynamic and systemic understanding of human behaviour. The complex adaptive systems approach endorsed by this article allows one to capture behaviour as 'encultrated' and 'enearthed', coevolving with socio-cultural and biophysical contexts.

2. • Hagmann D, Ho EH, Loewenstein G: **Nudging out support for a carbon tax**. *Nat Clim Change* 2019, **9**:484-489.

Across six experiments, these authors show that introducing a green energy default nudge diminishes support for a carbon tax. The authors propose that nudges decrease support for substantive policies by providing false hope that problems can be tackled without imposing considerable costs.

3. Steel D, DesRoches CT, Mintz-Woo K: **Climate change and the threat to civilization**. *Proc Natl Acad Sci USA* 2022, **42**:e2210525119.

4. DesRoches CT: **On the concept and conservation of critical natural capital**. *Int Stud Philos Sci* 2019, **3–4**:207-228.

5. •• Nielsen KS, Marteau TM, Bauer JM, *et al*.: **Biodiversity conservation as a promising frontier for behavioural science**. *Nat Hum Behav* 2021, **5**:550-556.
Authors argue that human activities are degrading ecosystems worldwide, posing existential threats for biodiversity and humankind. They outline the core components for building a robust evidence base and suggest priority research questions for behavioural scientists to explore in opening a new frontier of behavioural science for the benefit of nature and human well-being.

6. Selinske MJ, Garrard GE, Gregg EA, Kusmanoff AM, Kidd LR, Cullen MT, Cooper M, Geary WL, Hatty MA, Hames F, Kneebone S, McLeod EM, Ritchie EG, Squires ZE, Thomas J, Willcock MAW, Blair S, Bekessy SA: **Identifying and prioritizing human behaviors that benefit biodiversity**. *Conserv Sci Pract* 2020, **2**:e249.

7. Schubert C: **Green nudges: do they work? Are they ethical?** *Ecol Econ* 2017, **132**:329-342.

8. Sugden R: **The Community of Advantage: a Behavioral Economist's Defence of the Market**. Oxford University Press; 2018.

9. Zizzo MJ, Whitman G: **Escaping Paternalism: Rationality, Behavioral Economics and Public Policy**. Cambridge University Press; 2019.

10. Hausman DM, Welch B: **Debate: to nudge or not to nudge**. *J Political Philos* 2010, **18**:123-136.

11. Bovens L: **The ethics of nudge**. In Preference Change: Approaches from Philosophy, Economics and Psychology. Edited by Grüne-Yanoff T, Hansson SO. Springer; 2009:207-219.

12. Wilkinson T: **Nudging and manipulation**. *Political Stud* 2013, **61**:341-355.

13. Nagatsu M: **Social nudges: their mechanisms and justification**. *Rev Philos Psychol* 2015, **6**:481-494.

14. Guala F, Mittone L: **A political justification of nudging**. *Rev Philos Psychol* 2015, **6**:385-395.

15. Kasperbauer TJ: **The permissibility of nudging for sustainable energy consumption**. *Energy Policy* 2017, **111**:52-57.

16. Thaler RH, Sunstein CR: **Nudge: Improving Decisions about Health, Wealth, and Happiness**. Yale University Press; 2008.

17. Kahneman D, Knetsch JL, Thaler RH: **Experimental tests of the endowment effect and the Coase theorem**. *J Political Econ* 1990, **98**:1325-1348.

18. Schkade DA, Kahneman D: **Does living in California make people happy? A focusing illusion in judgements of life satisfaction**. *Psychol Sci* 1998, **9**:340-346.

19. Tversky A, Kahneman D: **The framing of decisions and the psychology of choice**. *Science* 1981, **211**:453-458.

20. Thaler RH, Sunstein CR: **Nudge: Improving Decisions about Health, Wealth, and Happiness**. Yale University Press; 2008.

21. Tyler DesRoches C: **Value commitment, resolute choice, and the normative foundations of behavioural welfare economics**. *J Appl Philos* 2020, **37**:562-577.

22. Sunstein CR, Thaler RH: **Libertarian paternalism**. *Am Econ Rev* 2003, **93**:175-179.

23. Grüne-Yanoff T, Hertwig R: **Nudge versus boost: how coherent are policy and theory?** *Minds Mach: J Artif Intell Philos Cogn Sci* 2016, **26**:149-183.

24. Felonneau ML, Becker M: **Pro-environmental attitudes and behavior: revealing perceived social desirability**. *Int J Soc Psychol* 2008, **21**:25-53.

25. Ginsberg JM, Bloom PN: **Choosing the right green marketing strategy**. *MIT Sloan Manag Rev* 2004, **46**:79-84.

26. Kahsay GA, Samahita M: **Pay-what-you-want pricing schemes: a self-image perspective**. *J Behav Exp Financ* 2015, **7**:17-28.

27. Levitt SD, List JA: **What do laboratory experiments measuring social preferences reveal about the real world?** *J Econ Perspect* 2007, **21**:153-174.

28. Goldstein NJ, Cialdini RB, Griskevicius V: **A room with a viewpoint: using social norms to motivate environmental conservation in hotels**. *J Consum Res* 2008, **35**:472-482.

29. Ölander F, Thøgersen J: **Informing versus nudging in environmental policy**. *J Consum Policy* 2014, **37**:341-356.

30. Sunstein C: **Green defaults can combat climate change**. *Nat Hum Behav* 2021, **5**:548-549.

31. Liebe U, Gewinner J, Diekmann A: **Large and persistent effects of green energy defaults in the household and business sectors**. *Nat Hum Behav* 2021, **5**:576-585.

32. Vetter M, Kutzner F: **Nudge me if you can-how defaults and attitude strength interact to change behavior**. *Compr Results Soc Psychol* 2016, **1**:8-34.

33. • Kaiser M, Bernauer M, Sunstein CR, Reisch LA: **The power of green defaults: the impact of regional variation of opt-out tariffs on green energy demand in Germany**. *Ecol Econ* 2020, **174**:106685.
In Germany, green energy defaults are effective among those concerned with climate change. This finding, based on real-world rather than experimental evidence, attests to the power of automatic enrollment in addressing environmental problems in Germany.

34. Sunstein CR, Reisch LA: **Automatically green: behavioral economics and environmental protection**. *Harv Environ Law Rev* 2014, **38**:127.

35. Garnett EM, Balmford A, Sandbrook C, Pilling MA, Marteau TM: **Impact of increasing vegetarian availability on meal selection and sales in cafeterias**. *PNAS* 2019, **116**:20923-20929.

36. Mason CF: **The economics of eco-labeling**. *Int Rev Environ Resour Econ* 2013, **6**:341-372.

37. Pogge T: **Real world justice**. *J Ethics* 2005, **9**:29-53.

38. Keohane NO, Olmstead SM: **Markets and the Environment**. Island Press; 2016.

39. Brink DO: **Mill's deliberative utilitarianism**. *Philos Public Aff* 1992, **21**:67-103.

40. Fuchs AE: **Mill's theory of morally correct action**. In The Blackwell Guide to Mill's Utilitarianism. Edited by West HR. Blackwell Publishing; 2006.

41. Mill JS: **On Liberty**. John W. Parker and Son; 1859.

42. Satz D: **Why Some Things Should Not be For Sale**. Oxford University Press; 2010.

43. Sunstein CR: **The ethics of nudging**. *Aust Nurs Midwifery J* 2016, **24**:413-450.

44. Hands DW: Libertarian paternalism: making rational fools. *Unpublished manuscript*; 2021.

45. Varian HR: **Microeconomic Analysis**. Norton; 1992.

46. Budolfson M, Spears D: **Public policy, consequentialism, the environment, and nonhuman animals**. In The Oxford Handbook of Consequentialism. Edited by Portmore DW. Oxford University Press; 2020.

47. Bartha P, DesRoches CT: **Modeling the precautionary principle with lexical utilities**. *Synthese* 2021, **199**:8701-8740.

48. Christiansen A: **Rationality, expected utility theory and the precautionary principle**. *Ethics Policy Environ* 2019, **22**:3-20.

49. Steel D: **Philosophy and the Precautionary Principle**. Cambridge University Press; 2014.

50. Adler MD: **Measuring Social Welfare**. Oxford University Press; 2019.

51. Nozick R: **Anarchy, State and Utopia**. Basic Books; 1974.

52. Parkinson J: **Deliberating in the Real World: Problems of Legitimacy in Deliberative Democracy**. Oxford University Press; 2006.

53. Nefsky J: **Collective harm and the inefficacy problem**. *Philos Compass* 2019, **4**:e12587.

54. In Philosophy and Climate Change. Edited by Budolfson M, McPherson T, Plunkett D. Oxford University Press; 2021.

55. Hulme M: **Meet the humanities**. *Nat Clim Change* 2011, **1**:177-179.

56. Jetzkowitz J, van Koppen CK, Lidskog R, Ott K, Voget-Kleschin L, Wong CML: **The significance of meaning. Why IPBES needs the social sciences and humanities**. *Innov: Eur J Soc Sci Res* 2018, **31**:38-60.

57. •• Laplane L, Mantovani P, Adolphs R, Chang H, Mantovani A, McFall-Ngai M, Rovelli C, Sober E, Pradeu T: **Opinion: why science needs philosophy**. *Proc Natl Acad Sci USA* 2019, **116**:3948-3952.
Authors argue that philosophy can have an important and productive impact on science. They illustrate their point with three examples taken from various fields of the contemporary life sciences.

58. Gsottbauer E, van den Bergh JCJM: **Environmental policy theory given bounded rationality and other-regarding preferences**. *Environ Resour Econ* 2011, **49**:263-304.

59. Rockström J, Steffen W, Noone K, Persson Å, Chapin FS III, Lambin E, Lenton TM, Scheffer M, Folke C, Schellnhuber H, *et al*.: **Planetary boundaries: exploring the safe operating space for humanity**. *Ecol Soc* 2009, **14**:32.